

FOVEATION FOR 3D VISUALIZATION AND STEREO IMAGING
Arzu Çöltekin



TEKNILLINEN KORKEAKOULU
TEKNISKA HÖGSKOLAN
HELSINKI UNIVERSITY OF TECHNOLOGY

FOVEATION FOR 3D VISUALIZATION AND STEREO IMAGING

Arzu Çöltekin

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Department of Surveying, Helsinki University of Technology for public examination and debate in Auditorium M1 in Helsinki University of Technology (Espoo, Finland) on the 3rd of February, 2006, at 12 o'clock noon.

Helsinki University of Technology
Department of Surveying
Institute of Photogrammetry and Remote Sensing

Teknillinen korkeakoulu
Maanmittausosasto
Fotogrametrian ja kaukokartoituksen laboratorio

Distribution:

Helsinki University of Technology
Institute of Photogrammetry and Remote Sensing
P.O.Box 1200
FIN-02015 TKK
Tel. +358 9 451 3901
Fax: +358 9 451 3945
E-mail: publications@foto.hut.fi

©Arzu Çöltekin

This work may be freely distributed and copied so long as the original source is acknowledged.

Image on the front cover

© (GNU Free Documentation License) Arzu Çöltekin

ISBN 951-22-8016-7 (printed)

ISBN 951-22-8017-5 (PDF, located at <http://lib.hut.fi/Diss/2006/isbn9512280175/>)

ISSN 1796-0711

ABSTRACT

Even though computer vision and digital photogrammetry share a number of goals, techniques, and methods, the potential for cooperation between these fields is not fully exploited. In attempt to help bridging the two, this work brings a well-known computer vision and image processing technique called foveation and introduces it to photogrammetry, creating a hybrid application. The results may be beneficial for both fields, plus the general stereo imaging community, and virtual reality applications.

Foveation is a biologically motivated image compression method that is often used for transmitting videos and images over networks. It is possible to view foveation as an area of interest management method as well as a compression technique. While the most common foveation applications are in 2D there are a number of binocular approaches as well.

For this research, the current state of the art in the literature on level of detail, human visual system, stereoscopic perception, stereoscopic displays, 2D and 3D foveation, and digital photogrammetry were reviewed. After the review, a stereo-foveation model was constructed and an implementation was realized to demonstrate a proof of concept. The conceptual approach is treated as generic, while the implementation was conducted under certain limitations, which are documented in the relevant context.

A stand-alone program called *Foveaglyph* is created in the implementation process. *Foveaglyph* takes a stereo pair as input and uses an image matching algorithm to find the parallax values. It then calculates the 3D coordinates for each pixel from the geometric relationships between the object and the camera configuration or via a parallax function. Once 3D coordinates are obtained, a 3D image pyramid is created. Then, using a distance dependent level of detail function, spherical volume rings with varying resolutions throughout the 3D space are created. The user determines the area of interest. The result of the application is a user controlled, highly compressed non-uniform 3D anaglyph image. 2D foveation is also provided as an option.

This type of development in a photogrammetric visualization unit is beneficial for system performance. The research is particularly relevant for large displays and head mounted displays. Although, the implementation, because it is done for a single user, would possibly be best suited to a head mounted display (HMD) application.

The resulting stereo-foveated image can be loaded moderately faster than the uniform original. Therefore, the program can potentially be adapted to an active vision system and manage the scene as the user glances around, given that an eye tracker determines where exactly the eyes accommodate. This exploration may also be extended to robotics and other robot vision applications. Additionally, it can also be used for attention management and the viewer can be directed to the object(s) of interest the demonstrator would like to present (e.g. in 3D cinema).

Based on the literature, we also believe this approach should help resolve several problems associated with stereoscopic displays such as the accommodation convergence problem and diplopia. While the available literature provides some empirical evidence to support the usability and benefits of stereo foveation, further tests are needed. User surveys related to the human factors in using stereo foveated images, such as its possible contribution to prevent user discomfort and virtual simulator sickness (VSS) in virtual environments, are left as future work.

PREFACE

”Excuse me, what was the word again?”

This must be the sentence that I uttered when my supervisor Prof. Henrik Haggrén mentioned the word “foveation” about three years ago when I was telling him one –yet another- of my new research ideas about level of detail management. Foveation immediately intrigued me – because it had a hint of non-technological science in it, and because I obey the universal law that says “anything that is not *your* work is extremely interesting”.

The discussion on foveation that afternoon has given me the basis to build a feasible and original frame of research, finally leading me on a straight path from my long wandering interests. The stereotypical PhD is of course constituted precisely of long wandering research interests, by the way, maybe more so in Finland. I believe the approach is based on another universal rule that says “if it does not kill you, it will make you stronger”. What I mean is, if you are a fellow PhD student reading this, hang on dear, there is light at the end of the tunnel! I think I can see it now. Where are my shades?

Now it is time to mention those noble ones who have held the torch for me until I got to this point. I will, most naturally, start with Prof. Haggrén, my supervisor, who is a true scientist and has always inspired me with his sincere interest in research. I cannot thank him enough for the long, insightful late afternoon discussions. There could not be a better motivator than his constant faith in me.

While I was doing this work, I divided my time between two labs: The Institute of Photogrammetry and Remote Sensing (IPRS), and The Institute of Cartography and Geoinformatics (ICG). I did the research in the first, and I taught and dealt with administrative things in the second.

Therefore next person I would like to thank is Prof. Kirsi Virrantaus, my patron saint (I wish I could use a smiley!). Prof. Virrantaus was my employer in ICG, from where I have received my regular full-time salary and nearly all of my expenses for worldwide conference trips. She has been a great example of an open minded, global thinking academic. I am deeply indebted for her support and the generous opportunities she has provided for me to gain experience in international academic life.

I am grateful to Prof. Christian Heipke who kindly agreed to be in the thesis evaluation committee to pre-examine my work. I feel most privileged to have his valuable and sound scientific input in my work. Dr. Martin Reddy was my second pre-examiner, to whom I am thankful beyond words. Dr. Reddy has a gift of being firmly scientific and professional, yet perfectly supportive, encouraging and gentle. He has not only pre-examined my thesis and given substantial and stimulating scientific advice, he actually proofread it carefully. I modified and rephrased many sentences in this thesis based on his extremely useful, sophisticated language tips.

Well, I am not done with my professors!

An extraordinary person who had a great positive influence on me when I was taking my baby steps in academia while I was still in Istanbul is Prof. Ayhan Alkış, whom I consider myself lucky to have met and worked with. Another great photogrammetrist (and a most welcoming host) I had the chance to work with is Dr. Sabry El-Hakim, of Visual Information Technology Group in National Research Center, Ottawa, Canada, where I stayed about two months during my PhD studies.

In addition to my regular teaching assistant salary, I received financial support from Finnish Cultural Foundation for three years, for which I am thankful.

Among my previous co-workers, I am grateful to a number of them for various reasons, but I would particularly like to acknowledge the efforts of Füsün Şanlı and İbrahim Çetin: they helped me with a significant amount of paperwork back at home when I was abroad.

My co-workers in Geomatics unit of HUT have provided me with a nurturing environment to do research and teaching, and a pleasant environment to have long coffee breaks.

Petteri Pöntinen was always available for my photogrammetry questions even when the question was posed from the other side of the ocean. Katri Koistinen has always been helpful and made me feel welcome from the beginning. Miltiadis Daniil has helped me to orient myself in this new working environment (and outside the work, shamelessly claimed that lahana dolması was Greek!). Ulla Pyysalo has kept me up to date with life in Finland and life in the department at our coffee breaks.

In “the other lab”, my friend Jukka Krisp has provided the social glue for us, keeping the lunch and coffee times with German punctuality but everything else in great flexibility. Paula Ahonen never rejected my untimely calls about work and my questions about the final stages of a PhD. And Riikka Henriksson has been irreplaceable, like-minded company in late work evenings.

David Brown, and Stefan Nesbitt have checked my English. Not only have they hunted down my fuzzy long sentences and the hiding "the"s, they did this at very short notice. Trevor Joyce has been the last-minute language supervisor. I am truly grateful for their generous and friendly help.

Only half a page left. I have to hurry up and prove that during this work, I also had a life outside the cave. Pardon me, I mean the office.

I will always remember the silly fun we had as a team of four: Ole Jensen, Freya Jensen (then it was Johnson!), Flamine Alary and myself. Through Ole and Freya I also met and enjoyed stimulating conversations (and yummy food!) with some great people including Titia van Zuijen and Karen Johanne Pallasen.

Furthermore, Jan von Pfaler, Timo Alakoski, Min Gong, Pierre-Olivier Pineau, Michael Ross, Ali Nadir Arslan, Cumhuriyet Erkut, Johnny Skåning, İlke Şenol, Wolfgang Ludwig, Nida Şen and most recently Can Ersen Fırat have made my stay in Finland interesting and rich. Can Bican and I had countless interesting talks including some computer vision topics; he will clearly remember the times when I asked “what is disparity, anyway?”. And thanks to all of my old and not so old friends who were mostly placed in physically remote locations, but happily bugged me with emails and messages day and night. You know who you are.

None of what I have achieved in life would be possible without the support of my parents. I thank them for everything. Their rich library poisoned my mind irreversibly at an early age, planting analytical thinking, presenting global adventures, which resulted in a passion to go and see the world and to scratch the surface to understand. And last, because his impact is most likely the biggest, is my brother, who has helped me in so many ways including tapping me on my shoulder when I was grumbling about it all and also the computer science help for this thesis. I do not believe I have the words to express my gratitude, therefore I will continue to give him useless little gifts and dark chocolate.

Thanks to you all. Sağ olun. Kiitos kaikille.

Arzu Çöltekin, Espoo, January 2006.

Contents

ABSTRACT	I
PREFACE	III
LIST OF ABBREVIATIONS	XI
LIST OF FIGURES.....	XII
LIST OF TABLES	XIV
LIST OF SYMBOLS	XV
CHAPTER 1. INTRODUCTION	1
1.1. Overview: The Big Picture	1
1.2. Motivation	2
1.3. Cross-disciplinary Nature of the Work	6
1.4. Relevance of This Work to Photogrammetry	8
1.5. Main Points of Scientific Contribution.....	8
1.6. About This Thesis	9
1.7. Typographic Conventions.....	10
1.8. Summary	10
CHAPTER 2. BACKGROUND - STATE OF THE ART	11
2.1. The Human Visual System.....	11
2.1.1. The Structure of the Eye.....	12
2.1.1.1. Blind Spot.....	13
2.1.1.2. Pinhole Camera Model	14
2.1.2. Visual Acuity (Resolution)	15
2.1.2.1. Types of Acuity	16
2.1.2.2. Discussion: Superacuity or Hyperacuity?	17
2.1.3. Contrast Sensitivity	17
2.1.4 Foveal Vision	17

2.1.5. Motion Sensitivity	20
2.1.6. Summary	20
2.2. Depth Perception.....	21
2.2.1. Depth Cues	22
2.2.2. Stereoscopic Perception	23
2.2.2.1. Horopter.....	26
2.2.2.2. Panum’s Fusional Area and Diplopia.....	28
2.2.2.3. How Far Can We See Stereoscopically?	30
2.2.2.4. Cyclopean Eye.....	31
2.2.3. Limited Depth of Field	32
2.2.4. Summary	34
2.3. Stereoscopic Viewing Techniques.....	35
2.3.1. Time Multiplexed Displays (TMDs).....	37
2.3.2. Head Mounted Displays	38
2.3.3. Autostereoscopic Displays.....	38
2.3.4. Retinal Projection Displays	39
2.3.5. Crossed Eye and Parallel Viewing	39
2.3.6. Anaglyphs.....	39
2.3.7. How Much Can We See on a Stereoscopic Display?	40
2.3.7.1. Panum’s Fusional Area and Stereoscopic Displays	40
2.3.7.2. Brain Pixels and Optimal Display	42
2.3.8. Problems with Stereoscopic Displays and Suggested Solutions	43
2.3.8.1. Frame Cancellation.....	44
2.3.8.2. Accommodation Convergence Conflict	44
2.3.8.2.1. Cyclopean Scale	45
2.3.8.2.2. Large Screens	46
2.3.8.3. Diplopia and Its Possible Solutions.....	46
2.3.9. Summary	47
2.4. Level of Detail.....	48
2.4.1. Culling	49
2.4.2. Perceptually Motivated LOD Techniques.....	49
2.4.2.1. Distance LOD	50

2.4.2.2. Size LOD	50
2.4.2.3. Priority LOD.....	51
2.4.2.4. Hysteresis	51
2.4.2.5. Eccentricity LOD.....	51
2.4.2.6. Velocity LOD	52
2.4.2.7. Depth of field LOD	52
2.4.3. Summary	52
2.5. Foveation	53
2.5.1. What is Foveation?	53
2.5.2. Active Vision	54
2.5.3. Common Foveation Methods, Models and Examples	55
2.5.3.1. Log Polar Mapping and Foveation	56
2.5.3.2. Foveation Techniques.....	58
2.5.3.3. Eccentricity LOD.....	60
2.5.3.4. Depth Aware Foveation?	62
2.5.3.4.1. Compression of Stereoscopic Image Pairs	63
2.5.3.4.2. Focus/Foveation	63
2.5.4. Are Depth of Field Simulation and 3D Foveation the Same Thing?	65
2.5.5. Examples of Foveation and Depth of Field Rendering	65
2.5.6. DOF Simulations	67
2.5.7. Foveation and Photogrammetry.....	68
2.5.8. User Studies on Foveation: <i>Perceptually Lossless?</i>	69
2.5.9. Summary	70
2.6. Correspondence and Reconstruction.....	71
2.6.1. The Correspondence Problem (Image Matching)	71
2.6.2. The Reconstruction Problem.....	72
2.6.2.1. Defining the Geometry: The camera model	72
2.6.2.2. Normal Case of Stereography.....	73
2.6.2.3. Epipolar Geometry for Normal Case of Stereo	76
2.6.3. Summary	77
CHAPTER 3. DEVELOPMENT AND IMPLEMENTATION.....	78
3.1. Development.....	78
3.1.1. The Dream Algorithm	78
3.1.2. Limitations for This Implementation.....	79

3.1.3. What is the Implementation for – Questions Before Coding	79
3.1.4. Planning the Implementation	80
3.1.4.1. Input	80
3.1.4.2. 3D Information	80
3.1.4.3. Display Method	80
3.1.4.4. Compression Approach	80
3.1.4.5. Eye Tracking or Not	81
3.1.4.6. Foveated Image Composition	81
3.1.4.7. Evaluation	81
3.2. Foveaglyph: The Implementation	81
3.2.1. The Processes	83
3.2.2. Explanation of the Tasks	84
3.2.2.1. Image Acquisition and Camera Setup	84
3.2.2.2. Camera Calibration	85
3.2.2.3. Camera Information for the Test Images	87
3.2.2.4. Creation of the Anaglyph	87
3.2.2.5. Image Matching and Disparity Map Calculation	87
3.2.2.6. Depth Discontinuities by Pixel-to-Pixel Stereo	88
3.2.2.7. Building the Foveation Pyramid	91
3.2.2.8. 2D Foveation	93
3.2.2.9. 3D Foveation	95
3.2.2.10. LOD Function	98
3.3. Shortcomings	99
3.4. Summary	100
CHAPTER 4. RESULTS	101
4.1 Things That Affect the Results	101
4.1.1. What Affects the Compression Rates	101
4.1.1.1. The Wandering POI	102
4.1.1.2. The Scene Content	102
4.1.2. What Affects the Performance of Image Matching	103
4.2. A Method to Measure the Compression Rates: Effective Pixel Count	104
4.3 Results	105
4.3.1. Performance	105
4.3.1.1. Why Emphasize the Image Matching?	106
4.3.1.2. On the Performances of Foveation and Image Matching	106
4.3.1.3. Performance Test	106
4.3.2. Foveation Results	107
4.3.2.1. 2D versus 3D	109
4.3.2.2. 2D versus 3D: Compression Rates	111
4.4. Evaluation of the Results	114
4.5. Summary	115

CHAPTER 5: DISCUSSION	116
5.5.1. Geovisualization.....	116
5.5.2. Visual Attention Management and Progressive Image Loading	118
5.5.3. Stereo Foveation an Alternative to Stereo JPEG Compression?	118
5.5.4. WWW use for 3D foveation: VRML and QuicktimeVR.....	119
5.5.5. Summary	119
CHAPTER 6. CONCLUSIONS	120
6.1. General Remarks.....	120
6.2. Perceptual Issues	121
6.3. Future Work	122
6.3.1. Usability	122
6.3.2. Specific Photogrammetric Tasks	123
6.3.3. Comparison of Alternative Foveation Techniques.....	123
REFERENCES.....	124
APPENDICES	142
Appendix 1: An explanation of Arc Minutes	142
Appendix 2: Glossary	143
Appendix 3: Index of Test Images	148
Appendix 4: The GUI Menus of <i>Foveaglyph</i>	149
Appendix 5: Snellen Eye Chart	151
Appendix 6: Radial Eye Chart	152

List of Abbreviations

2D: Two-dimensional	GTK+: GIMP Toolkit +
3D: Three-dimensional	H3D: A stereo format commonly using TGA and JPG
AI: Artificial Intelligence	HMD: Head Mounted Display
BSF: Binocular Sensory Fusion	HVS: Human Visual System
BMP: Bitmapped Image Format	ICT: Information and Communication Technologies
BMS: Stereo BMP	IPD: Inter Pupillary Distance
CAD: Computer Aided Design	JPEG or JPG: Joint Photography Experts Group
CAVE: Collaborative Virtual Environment	LCD: Liquid Crystal Display
CG: Computer Graphics	LOD: Level of Detail
CPU: Central Processing Unit	MAR: Minimum Angle of Resolution
CRT: Cathode Ray Tube	MSE: Mean Square Error
CV: Computer Vision	NYU: New York University
DOF: Depth of Field (not Depth of Focus)	P2p or p2p: Pixel-to-pixel stereo (stereo matching program)
DCTDP: Disparity Compensated Transform-Domain Predictive Coding	POI: Point of Interest
DP: Digital Photogrammetry	PNG: Portable Network Graphics
DPW: Digital Photogrammetric Workstation	PNS: Stereo PNG
DTM: Digital Terrain Model	RAW: This is not an abbreviation, but used like one as an image format
FOV: Field of View	RS: Remote Sensing
GIF: Graphics Interchange Format	SID: The Society of Information Display
GIS: Geographical Information System(s), also for Stereo GIF.	TGA: True Vision Graphics Display Adapter)
GPS: Global Positioning System	TIFF: Tagged Image File Format
GIMP: GNU Image Manipulation Program	TMD: Time Multiplexed Display
GUI: Graphical User Interface	QTVR: Quicktime Virtual Reality
GNU: GNU is a recursive acronym for “GNU's Not UNIX”	VSS: Virtual Simulator Sickness
GPL: General Public License	VPE: Virtual Planetary Exploratorium
GPU: Graphics Processing Unit	VR: Virtual Reality
GPGPU: General Purpose Computation on GPUs	VRD: Virtual Retinal Display
	WWW: World Wide Web

List of Figures

Figure 1: Concepts relating to stereoscopic vision and technology.	2
Figure 2: Illustration of the “stereodrome” by Henrik Haggrén	6
Figure 3: Across fields, there is a common interest in 3D modeling.	7
Figure 4: The organization of the thesis.	9
Figure 5: A plan view of the brain.	12
Figure 6: A cross-section of the eye.	13
Figure 7: The lens focuses a small, inverted picture of the objects onto the retina.....	14
Figure 8: An acuity graph showing how the resolution changes	15
Figure 9: Stereo acuity.	16
Figure 10: Rod and cone distribution across the fovea.	18
Figure 11: Tangential section through the human fovea.	19
Figure 12: The human visual field for a person gazing ahead.	19
Figure 13: Disparity and stereopsis.	24
Figure 14: Kepler’s projection theory.	25
Figure 15: The drawings of Charles Wheatstone.....	26
Figure 16: Vieth-Muller Circle is a theoretical horopter.	27
Figure 17: Horopter.	27
Figure 18: Panum’s fusional area in relation to a simple stereo display.....	28
Figure 19: The horopter, Panum’s area and the zone of stereopsis.....	29
Figure 20: Limits of stereo vision.....	30
Figure 21: A basic illustration of Cyclopean View.	31
Figure 22: The methods of viewing stereoscopic images and graphics.....	37
Figure 23: Cross-eye technique versus parallel viewing technique.....	39
Figure 24: Anaglyph viewing.	39
Figure 25: Human FOV and current display technology.....	41
Figure 26: Brain pixels in and their relationship to the display size.....	43
Figure 27: Ware’s cyclopean scale illustration.	45
Figure 28: The famous “Stanford Bunny” demonstrates the polygon count.	48
Figure 29: Culling techniques.....	49
Figure 30: How the distance makes a difference in our perception.	50
Figure 31: Mesh simplification is not visible if it is proportional to the object size.....	51
Figure 32: An illustration of foveation to demonstrate its principle idea.....	54
Figure 33: a) Cartesian Plane b) Log-polar plane.	57
Figure 34: Cartesian, log-polar and foveated maps of the same picture	57
Figure 35: Shows the gradually changing pixel size throughout the image.	59
Figure 36: The relationship between the POI, the distance to the point and the eye. ...	59
Figure 37: Panum’s fusional area and the positions of ϕ and ϕ_0 angles.	62
Figure 38: Foveation/Focus compression scheme.....	64
Figure 39: Visual results from Reddy’s Percept.	66
Figure 40: Visual results from UTEXAS’s Foveator.....	66
Figure 41: Visual results from Chang et al.’s online Java based foveation demo.	67
Figure 42: Comparison of normal case versus convergent camera configurations.	72
Figure 43: Bird’s eye cross-section view of xz plane for the normal case of stereo.....	74
Figure 44: The perspective projection	75
Figure 45: Image planes are coplanar and epipolar lines are parallel.....	76
Figure 46: Foveaglyph’s command line options.....	82
Figure 47: A screenshot of Foveaglyph’s graphical user interface.....	82
Figure 48: Schematic description of the application.	83

<i>Figure 49: Stereo picture acquisition equipment.</i>	85
<i>Figure 50: Distortions and the effects of affinity.</i>	86
<i>Figure 51: Pixel-to-pixel stereo: disparity maps and the depth discontinuity maps.</i>	89
<i>Figure 52: Image matching CPU timings</i>	90
<i>Figure 53: The image pyramid of 4 levels.</i>	91
<i>Figure 54: The image pyramid of 8 levels.</i>	92
<i>Figure 55: Illustration of the level switching in 2D foveation.</i>	93
<i>Figure 56: Illustration of how the steps are calculated 2D foveation</i>	95
<i>Figure 57: Level switching and calculation of steps for 3D foveation.</i>	97
<i>Figure 58: How the disparity and distance relationship works.</i>	98
<i>Figure 59 The pull-down menu for optional LOD functions.</i>	99
<i>Figure 60: The options menu showing the user inputs for foveation settings.</i>	101
<i>Figure 61: An illustration of changing resolution in the space variant image.</i>	102
<i>Figure 62: The information from Table 5 in a graph.</i>	104
<i>Figure 63: Foveation and p2p time comparison as a graph.</i>	107
<i>Figure 64: A 2D foveated image. 8 levels of detail, POI is the center (mid-point).</i>	108
<i>Figure 65: A 3D foveated image. POI and LOD values are the same as in 2D.</i>	108
<i>Figure 66: The results as the POI changes in the gymball image.</i>	110
<i>Figure 67: The graph obtained from various POIs in the gymball image.</i>	112
<i>Figure 68: The original left input image and the approximate locations of the POIs.</i>	113
<i>Figure 69: The results in Table 8 as a graph for an easier visual comparison.</i>	114
<i>Figure 70: A GIS showing the roads layer in focus.</i>	117

List of Tables

<i>Table 1: Depth of field values calculated based on the viewing distance.</i>	33
<i>Table 2: Advantages and disadvantages of active and passive stereo viewing glasses.</i>	38
<i>Table 3: CPU times for image matching in relation to the image size.</i>	89
<i>Table 4: Notation used in this chapter.</i>	94
<i>Table 5: The different user-given maximum disparity values, and CPU times.</i>	103
<i>Table 6: Foveation p2p times in seconds, these are the CPU times.</i>	107
<i>Table 7: Gymball image test results for 2D and 3D foveation.</i>	111
<i>Table 8: 2D and 3D foveation results.</i>	113

List of Symbols

These symbols appear in Chapter 3 and Chapter 4 and Section 2.6 where the contribution is original. The symbols encountered in the rest of Chapter 2 are treated locally as they are taken from the literature and are not modified.

- POI***: Point of interest
- P***: When walking on an image matrix starting at POI, each visited pixel.
- B***: Base or inter-ocular distance between the two cameras.
- c***: Camera constant, which is the focal length after calibration.
- P_x***: Horizontal parallax.
- d***: Euclidian distance between two points.
- d_{max}***: Maximum distance in the current work space.
- D***: A threshold that determines the LOD switch.
- lod***: Of the (0 to L-1) levels in the image pyramid, the index number of each level.
i.e. 0th level is the best resolution, and would be expressed as $lod = 0$.
- L***: Maximum number of levels.
- S***: The scale ratio.
- P_{lod}***: The number of pixels taken from lod^{th} image in the pyramid.

CHAPTER 1. INTRODUCTION

1.1. Overview: The Big Picture

I'm not crazy about reality, but it's still the only place to get a decent meal.
--Groucho Marx

Imagine an environment where nothing is real, yet it all feels real. When the system senses a demand or takes an order, it transmits signals via the varying physical means simulating the world addressing all of your senses: sight, sound, smell, taste and touch. Such a possibility has captured the imagination of many in films, novels and games. Not only because it fascinates us and because people are willing to spend millions of euros in amusement, but also because it would be highly useful for communication and design tasks. It would also be highly useful for navigating, discovering, understanding, treating problems and learning effectively from each other.

The bigger picture of this work relates to this image, but of course the work itself is in a smaller corner, stereoscopic three-dimensional (3D) visualization.

Photogrammetry is a field that deals with techniques and processes that allow us to do accurate measurements from photographs and digital imagery. The output can be orthophotos or two-dimensional (2D) or 3D models. In contemporary photogrammetry the main task is drawing accurate 3D graphics.

This work, in its essence, tries to show that if we understand human vision and manage our 3D graphics based on that understanding, we would gain considerably from it. Humans are not always better designed compared to machines. Cameras, for instance, can see better detail than eyes. As stated in Schenk 1999, “the quality of the retinal image is far inferior to that of any disposable camera” (Schenk, 1999), but humans are, in general still much superior than computers or any machine that we have ever made in information processing when performing tasks such as visualization.

While working within this frame of inspiration, we are focusing on stereoscopic vision and technology. Stereoscopy is an instrument for reconstructing scenes in 3D as well as a directly immersive medium when it is used for visualization such as in 3D cinema, photography and other arts. It can be considered a form of virtual or augmented reality for its immersive quality. Even though it does not necessarily include sensory feedback or interaction¹, there are applications of it, which have both qualities integrated. It can,

¹ There are many definitions of virtual reality. Scherman and Craig in their book “Understanding Virtual Reality” listed interactivity as one of the four main criteria for a system to be considered virtual reality. Others are “a virtual world”, “immersion” and “sensory feedback” (responding to user input). While the sensory feedback is provided by the system as a result of tracking,

not only be used for macro objects, in planetary mapping, terrain modeling and such, but it also has been used in micro environments such as in medical applications for diagnosis of deformation in bones, tissues, microscopic entities or, when generated real time over varying imaging systems and platforms, in surgery.

Still narrowing the scope of the thesis, we are introducing a level of detail management method called foveation to close range photogrammetric applications and present its potential in our field with an implementation. A graphical representation of the relevant concepts is presented below.

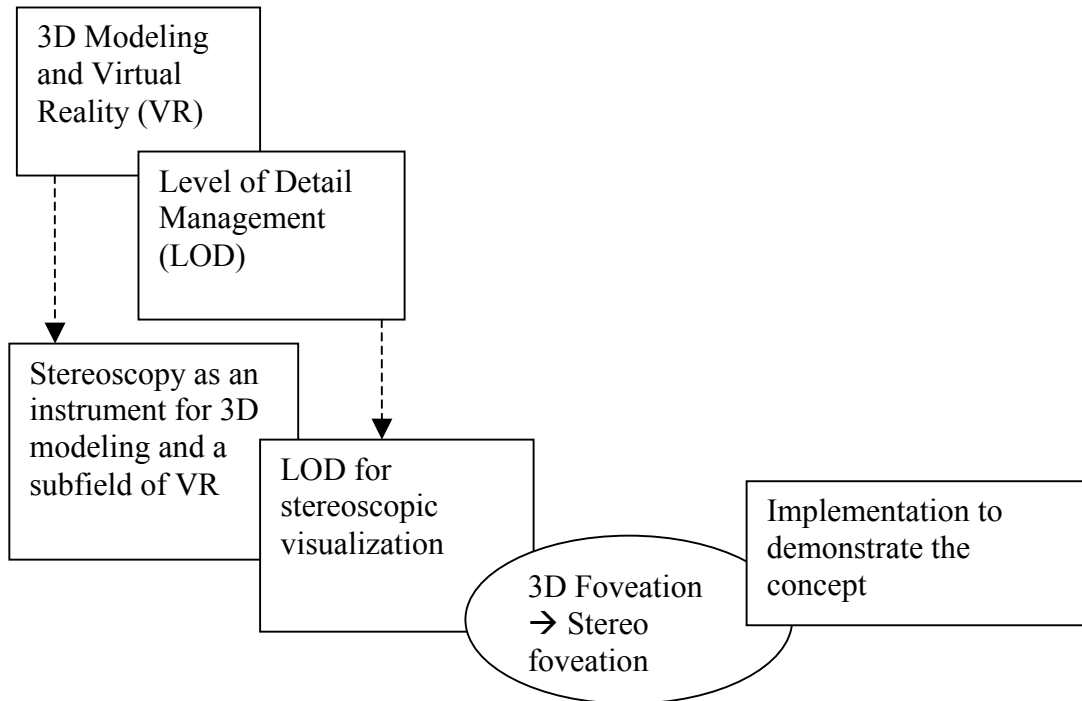


Figure 1: Concepts relating to stereoscopic vision and technology organized within the framework of this research.

1.2. Motivation

In a number of fields including digital photogrammetry (DP), computer vision (CV), virtual reality (VR) and geographic information systems (GIS), the size of the data and the demand for complex operations continue to compete with Moore's law² for computer hardware development. In other words, the processing power doubles every

interactivity occurs when user provides an intentional request from the system. (Scherman and Craig, 2003)

² The term **Moore's law** was used by the press and has been since adopted by the scientific communities. It expresses an observation by Gordon Moore in his paper (Moore, 1965) in 1965. According to his observation, the number of transistors per integrated circuit followed an exponential growth curve and he predicted that this trend would continue. This was only four years after the first planar integrated circuit was invented. The prediction has proved roughly correct until today and is predicted to continue to be so for another two decades (Intel Research, 2005).

year or every second year, but the demand for it grows in an equal fashion. This means we are forced to seek innovative software solutions rather than relying on the latest computer hardware.

In addition, disposing of the previous generations of computational power right after the better, faster and newer ones enter the market is often not economically feasible. Yet further, the best processing power is not readily available to a large part of the world for financial reasons. The term digital divide³ is used to express this concept.

Another development that makes this work relevant is the recent technological trend and a big consumer market for mobile embedded devices such as mobile phones, Global Positioning System (GPS) devices and hand held computers. These devices have limited resources and a limited size available for their displays. Yet there is potential for 3D applications for maps and navigation or games in such devices and Sharp Electronics has already introduced an auto stereoscopic display to the market in one of its laptop computers and soon after in a mobile phone model, SH251iS (Sharp, 2004; Stereoscopy, 2004).

“Following the successful commercialization of Sharp 3D technology developed here at SLE, the first Sharp 3D product was the SH251iS mobile phone, featuring a 2D/3D switchable TFT display at 176x220 resolution. This product sold over 1.5 million units in the first 6 months of sales, more than all previous 3D products (ever) combined.

The second Sharp 3D product was the SH505i mobile phone, featuring a 2D/3D switchable Continuous Grain Silicon display at 240x320 resolution.” (Sharp3D 2005)

The smaller field of view (FOV) in such mobile devices might seem to be a disadvantage as the area is already small. Therefore, a yet smaller area of interest from that might seem not plausible at first. But in real time image and video transmissions, foveation still holds relevance, as the bandwidth is more limited in mobile devices.

Sanghoon et al. demonstrated the benefits of using foveation for video streams in noisy wireless low bandwidth applications, which is today’s norm for common mobile devices. In their words “the results clearly underline the significant potential of foveated video communication protocols for wireless multimedia applications.” (Sanghoon et al., 2005).

In addition, while 2D foveation is most relevant in large field of view screens, 3D foveation alone is relevant in very narrow fields of view, which makes it interesting for smaller displays running 3D applications.

³ The digital divide is a social/political issue referring to the socio-economic gap between communities that have access to computers and the Internet and those who do not. The term also refers to gaps that exist between groups regarding their ability to use ICTs (Information and Communications Technologies) effectively, due to differing levels of literacy and technical skills, as well as the gap between those groups that have access to quality, useful digital content and those that do not. The term became popular among concerned parties, such as scholars, policy makers, and advocacy groups, in the late 1990s. (Wikipedia, 2005).

In the light of these factual and semi-political arguments, the scientific, commercial and individual computational needs of the world require us to consider the most efficient ways to manage available resources. The visualization of data, particularly for 3D and stereoscopic environments, therefore needs to be done in a calculated manner with adaptation of compression techniques and other data management approaches. In the opening of the textbook “Level of Detail for 3D Graphics”, Luebke et al. state that,

“Level of Detail (LOD) is as relevant today as ever, for despite tremendous strides in graphics hardware; the tension between the fidelity and speed continues to haunt us” (Luebke et al., 2003).

In photogrammetry and remote sensing projects, higher resolution imagery is as a means to achieve higher accuracy and gather more information about the scene. Depending on the project resolution requirements change, but in almost all cases, it is desirable to obtain and store highest resolution possible. This naturally leads to performance problems.

In the world of very large datasets, being able to transmit this data over a network is another problem to solve. This argument naturally includes the Internet. It is also meaningful for environments like the CAVE⁴s and panoramic, stereoscopic or other types of display systems with multiple projectors.

When dealing with images and 3D worlds, a number of compression and data management methods have been applied and it continues to be a field of research with many innovative new approaches.

Many of the most interesting are biologically inspired. Visual attention is a major topic in psychology, neurobiology and computational aspects of vision (Yamamoto et al., 1996). If we consider Artificial Intelligence (AI) as an umbrella to fields such as robotics, virtual reality and perhaps also photogrammetry, it is possible to say that all things ever done under that umbrella had some direct or indirect inspiration from biological fundamentals. Looking at how humans deal with the world has been a guide to configuring computers to do so. Computers seem to be destined to tackle the tasks that humans do, eventually, in some cases with an even better accuracy and efficiency. This includes vision, hence visualization.

In terms of vision, robotics and camera researchers try to build “eyes” that can see just the way the human eyes do. In terms of visualization, the researchers are keen on rebuilding the scenes and pictures in a similar way that the eyes process the real world. In an artificial world, things do not need to have their full existence at all times. They

⁴ CAVE is said to stand for “Collaborative Virtual Environment”, “Computer Assisted Virtual Environment” as well as “Computer Automatic Virtual Environment” - A virtual reality system that uses projectors to display images on three or four walls and the floor (answers.com, 2005). It also has been suggested to be a recursive acronym (Cave Automatic Virtual Environment) and a reference to “The Simile of the Cave” found in Plato’s Republic, in which the philosopher explores the ideas of perception, reality, and illusion. Plato used the analogy of a person facing the back of a cave alive with shadows that are his/her only basis for ideas of what real objects are. (Cruz-Neira et al., 1993).

are not made of hard material but instead are softcopy objects, easily editable to varying resolutions.

If we adopt the philosophical position that biological vision is the upper limit of a feasible mechanism for sensory information processing, then we have a clear motivation to produce biologically motivated solutions in the context of machine vision (Yamamoto et al., 1996).

Human eyes work with varying resolutions and they -in cooperation with the brain- are very effective in processing visual input. When the eyes look at the world around them, they actually find a point of interest, focus (accommodate) on that, and the rest of the world is actually fuzzier as one goes further and further away from the point of interest, in all directions. When the point of interest changes the whole reconstruction happens again, sharp in the middle, not so sharp towards the periphery. The periphery is good for navigating and comprehending the whole, but the brain does not need to have it **all** in focus at once.

So, eyes have a *depth of field* (DOF) and they **foveate**. These two features are at the core of our conceptual model, and will be explained in detail in later chapters (see Section 2.2.3. and Section 2.5.).

Why could this not be done with images and scenes that are computer generated or digitally formed? Indeed, the question was asked earlier too, and the answer is yes, it can be done, even though it is rather complicated to do so.

Trying to understand how the human eyes work and using “natural compression” in implementing a level of detail management ought to have some advantages. That is where the second half of the motivation for this work enters.

Additional motivating factors are the fact that DOF simulation is not exploited as much as it could be, even though it addresses certain usability issues. Ware and others make more than one reference to address this as a problem (Ware, 2000), and that there is some evidence in literature that DOF simulation might be helpful in reducing *diplopia* (double vision), which creates discomfort in using stereoscopic displays (Linde, 2003 and 2004).

Diplopia occurs when the disparity is greater than can be fused by the human brain. There is only a certain area where binocular fusion is possible (discussed in Section 2.2.2). Diplopia is one of the sensory conflicts when stereo displays are used, while the other one is the *accommodation convergence* conflict. While the conflict is explained in Section 2.3.8.3, a brief definition of these two terms and their relationship is quoted from Hung as follows:

Accommodation, or focusing, is driven by blur of the target image on the retina, whereas vergence, or rotation of the two eyes in opposite directions, is driven by the disparity of the images on the retinas (Hung, 1997).

Simulating depth of field is said to help with both of these problems (diplopia and accommodation convergence conflict) as well as providing a more natural visual experience.

In addition to these scientific motivations, another inspiration comes from research and experiments with stereoscopic displays in the institute where this research was conducted.

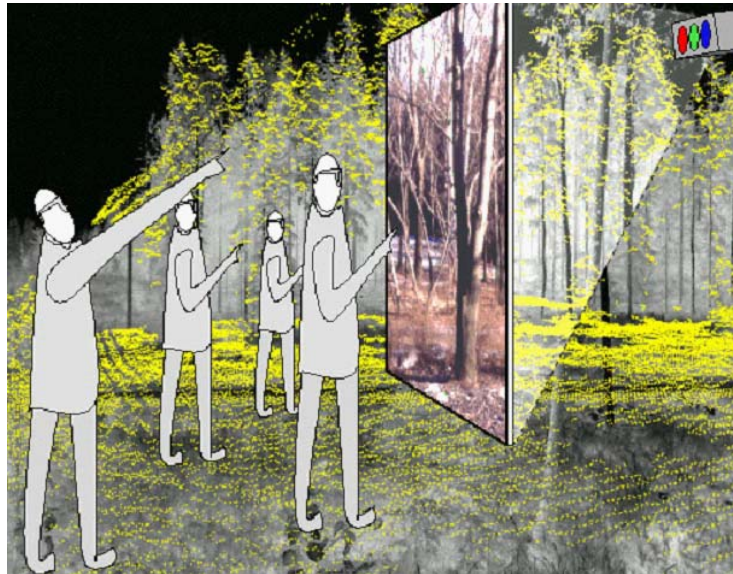


Figure 2: Illustration of the “stereodrome” by Henrik Haggrén, Helsinki University of Technology. Reprinted from Haggrén 2005 with permission.

In a large display with a wide FOV that can visualize things at their real size (1:1 scale) the area of interest is obviously smaller than the whole screen. See section 2.3.7 for a numerical analysis of how much of the screen we can actively utilize. A seamless cut on the excess data where the user can still have the peripheral information in lower resolutions to navigate seemed a most appropriate and attractive way to add intelligent image management to our system.

Large FOV displays, such as the one in Figure 2 and Head Mounted Displays (HMD)s, assist in spatial awareness, telepresence and remote manipulation scenarios (Linde, 2003). On the negative side, a wide FOV is reported to exacerbate the effects of virtual simulator sickness (VSS), (Kolasinski, 1995) and rarely occurs when using displays with lower than 60° FOV. In fact, Simulator sickness is particularly associated with displays with a wide FOV (Skerjanc et al., 1997).

In addition to the optimization of the display use, simulating depth of field is suggested to increase human performance by helping to reduce the negative side effects of stereoscopic displays.

1.3. Cross-disciplinary Nature of the Work

Measuring 3D information is an old occupation for photogrammetry, though it is worth noting that today 3D modeling is cross disciplinary.

The models of all natural or human made objects are essential in a Virtual Environment (VE), therefore, be it topographic or non-topographic, the byproducts of photogrammetry are used in such environments.

Another connection between virtual reality (VR) and photogrammetry is the similarity in the immersive quality of the stereoscopic visualization and VR. Stereo visualizations are rarely used only for entertainment in photogrammetry, but their immersive quality becomes a tool in education.

As for CV, there are a number of overlapping areas of interest, the correspondence and reconstruction problems being at the center. Digital photogrammetry research makes use of nearly all that the computer vision and computer graphics (CG) fields have to offer. Computer vision researchers have also employed photogrammetric methods.

Among other things, CG addresses the needs of computer aided design (CAD) systems. Almost all photogrammetric systems have CAD support and almost all CAD systems have some means to take input from photogrammetry. Therefore it is fair to say that CAD is closely related to photogrammetry even though the former is often for modeling ideas into products and the latter for converting real objects into abstract models so that they can be measured.

At this point it has to be mentioned that while half of photogrammetric products are raster, as in digital orthophotos and other projects making use of stereo visualizations, the other half will be vector, such as the 3D coordinates and vector models of the scene. While stereo imaging can be used as a base to produce vector maps, in itself it can be the final product for certain visualization tasks.

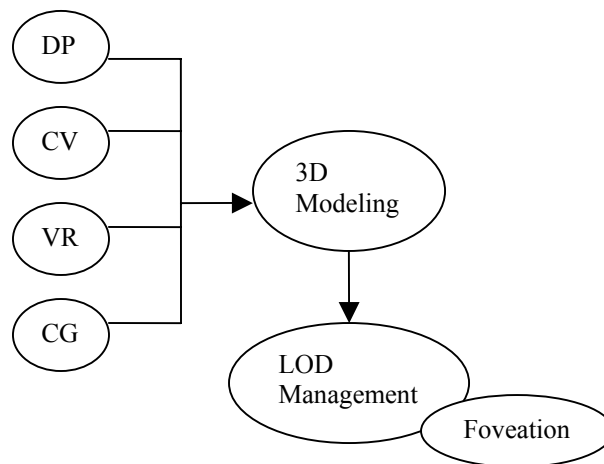


Figure 3: Across fields, there is a common interest in 3D modeling. An important issue in 3D modeling is the level of detail management. In the figure, acronyms are DP: Digital Photogrammetry, CV: Computer Vision, VR: Virtual Reality, CG: Computer Graphics, LOD: Level of Detail.

The above mentioned fields also cooperate in tasks such as interpreting remotely collected data like planetary mapping, planning a new product, visualizing microscopic entities, or represent abstract concepts.

These fields certainly have more than one common problem, although in this study, we are dealing primarily with just one of them, level of detail management.

1.4. Relevance of This Work to Photogrammetry

Photogrammetry traditionally has dealt with topographic data. Today, often in connection with Remote Sensing (RS) and GIS, it continues to do so. In addition to its traditional involvement in terrain modeling, close-range applications of photogrammetry such as architectural, terrestrial, medical, microscopic, x-ray and moiré have always been active and in demand.

This work relates more to the close-range applications of photogrammetry since stereoscopic vision is more relevant when the objects are closer.

When observing a mountain range at a distance of 30 km stereo vision contributes almost nothing to our understanding of the spatial shape. However, if we create a stereo pair of images with the viewpoint separated by 5 km we will obtain a useful enhanced "hyper stereo" image (Ware et al., 1998).

An obvious extension to the close range photogrammetry is into the realm of virtual reality and animation (Fryer, 1996; Atkinson, 1996). The level of detail and area of interest management techniques are highly active topics of research in virtual reality and animation fields (see Reddy, 1997; Luebke et al., 2003). The exploration of these techniques by photogrammetrists can make a valuable impact and contribution to the photogrammetry field.

Stereo image acquisition from the air and ground and stereo image processing to recover 3D information are central interests to photogrammetry, therefore research on stereo image foveation should clearly be of interest to the field.

Also, the more digital cameras that are employed in photogrammetric applications the more overlapping research interests that the photogrammetry and computer vision communities share. As Cooper et al. stated: "One of the results of the increasing use of digital cameras for photogrammetry is the transfer of machine vision algorithms and concepts into photogrammetric process" (Cooper et al., 1996). In this research, this is exactly what we are doing, bringing machine vision concepts and algorithms to photogrammetry.

1.5. Main Points of Scientific Contribution

The contribution to the existing knowledge made by this research may be summarized as follows:

- This thesis makes a novel connection between disparate research areas, namely photogrammetry, computer vision, computer graphics and the human visual system.
- It combines foveation with depth of field simulation and creates stereo foveation for 3D visualization tasks on stereoscopic displays.

- It demonstrates that level of detail management for stereoscopic images can be successfully realized by stereo foveation.
- It implements an independent application of stereo foveation as a proof of concept.

In doing these, a detailed overview of the state of the art and background information in the human visual system, depth perception, stereoscopic displays, level of detail, foveation, and the correspondence and reconstruction problem is presented.

A potential contribution to the research on eyestrain problems caused by accommodation convergence conflict and diplopia associated with stereoscopic displays is discussed. This research also offers a discussion on how foveation could be useful for visual attention management and geoinformatics related tasks.

1.6. About This Thesis

This work is organized in five main chapters as presented in the following figure:

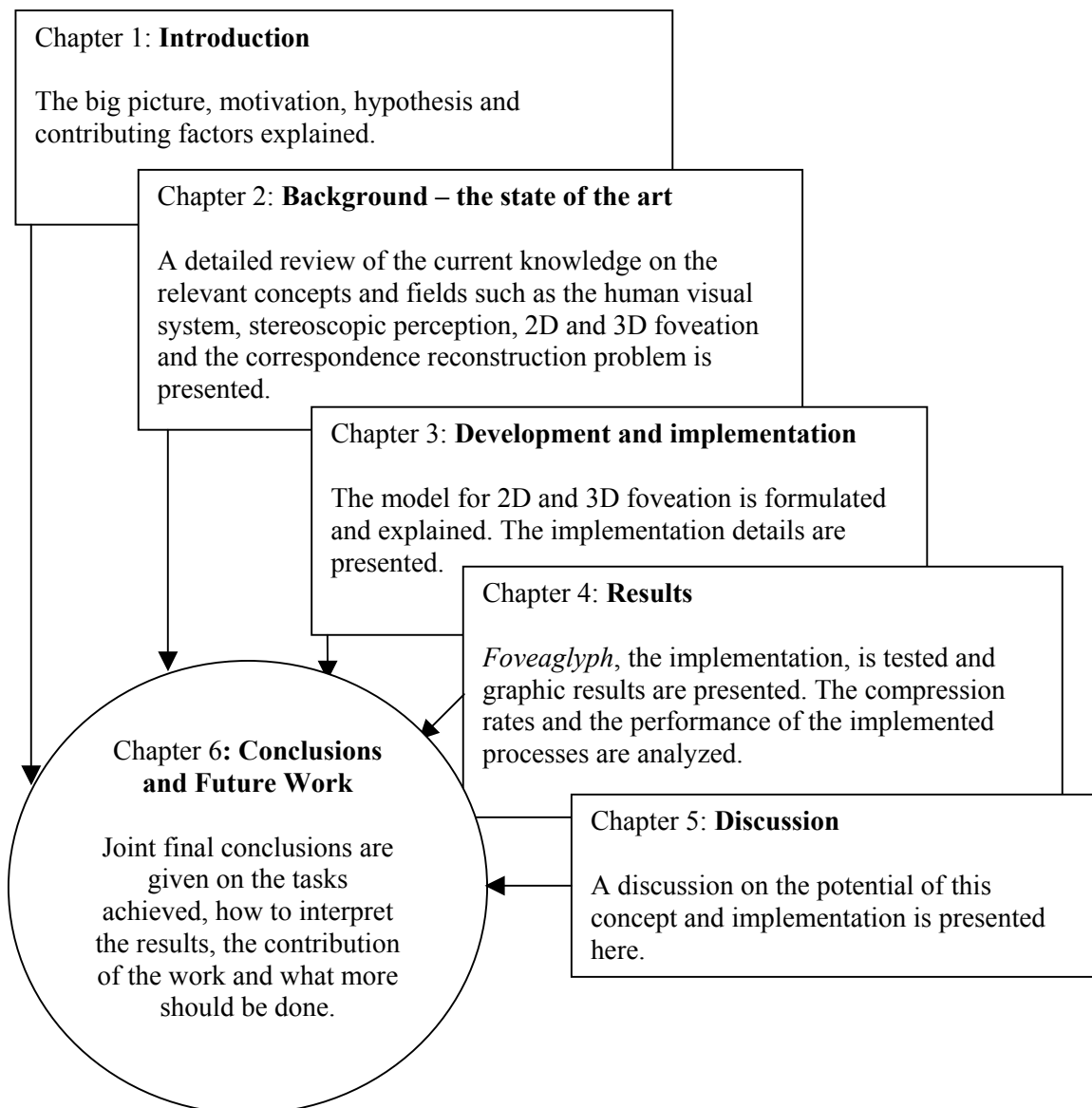


Figure 4: The organization of the thesis.

After every chapter, a summary is given. This is done for each sub-chapter in Chapter 2, as there are literature reviews on 6 different topics. After the final chapter, references, a glossary of terms and appendices can be found.

1.7. Typographic Conventions

Modified from Reddy (1997), the following typographic conventions are used in this thesis:

- The default paragraph font is Times New Roman, 12 points.
- Times new roman 10 points is used for figure and table captions, and for the text within a figure or a table. These sentences are centered when appropriate.
 - Indent forward is used for portions of text that are reprinted as is from another source. If the sentence is integrated to the text, the quotation marks are used. The reference is normally stated at the beginning or at the end of the quotation.
- *Italics* format is used for the quotes in the beginning of the chapters and when a new term is introduced for the first time. A glossary containing many of these terms is provided as an appendix.
- To emphasize a word or a portion of a sentence, **boldface** font is used.

1.8. Summary

In this chapter, the vision that inspired this work was described and the big picture was drawn. We also indicated where in the big picture the actual scope of this thesis lies. The motivating factors for conducting this research were analyzed, and a brief account was given considering where in the field of photogrammetry this research finds its relevance. The main points of contribution are listed, the thesis structure is explained, and typographic conventions are given.

CHAPTER 2. BACKGROUND - STATE OF THE ART

Fact are facts, Watson, and after all you are only a general practitioner with very little experience and mediocre qualifications.
-- *Sherlock Holmes (Arthur Conan Doyle)*

2.1. The Human Visual System

Ferwerda pointed out that, “an understanding of early visual processing is currently driving the development of perceptually based algorithms that are improving both the efficiency and the effectiveness of graphics methods” (Ferwerda, 2001).

In this chapter, we will present the current knowledge on the eye and human vision. A general overview of the human visual system is presented, visual acuity is introduced and available information is reviewed on foveal vision. Stereoscopic perception is reviewed separately.

Humans and other animals interpret the reflected light that they receive from the environment. It is a complex biological, optical (physical) and psychological process.

The visible light constitutes a very small part of the electromagnetic spectrum. Some animals, such as snakes can see infrared while certain insects can see in the ultraviolet. Humans can perceive light only in the range of 400 to 700 nanometers. At wavelengths shorter than 400 nm are ultraviolet light and X-rays. At wavelengths longer than 700 nm are infrared light, microwaves and radio waves (Ware, 2000).

The visible light is reflected by the objects in the environment and received by the eyes, then transferred through visual pathways and interpreted by the brain’s visual cortex.

Once this had been thought to occur differently, as we learn from Lenny Lipton: “In his projection theory in 1611, Kepler imagined that mental rays travel outward from the eyes, in straight lines, a concept starting with the ancient Pythagoreans” (Lipton, 1982, derived from Kaufman, 1974). Kepler’s projection theory of stereopsis will be presented further in the following section (Section 2.2.2) where stereoscopic perception is explained.

Vision is the most powerful of the senses, and it is by far the most neurologically demanding, with over 70% of all sensory receptors in the human nervous system dedicated to its functioning (Marieb, 2000). Visual intelligence occupies almost half of your brain’s cortex (Hoffman, 2000), 70% of all receptors, 40+% of cortex and 4 billion neurons are dedicated to vision and we can see much more than we can mentally image (Ware, 2000).

The human visual system is capable of discriminating between different intensities and

color, adapting to changing illumination and color, adapting locally within an image – it can enhance the edges and can sense and perceive depth.

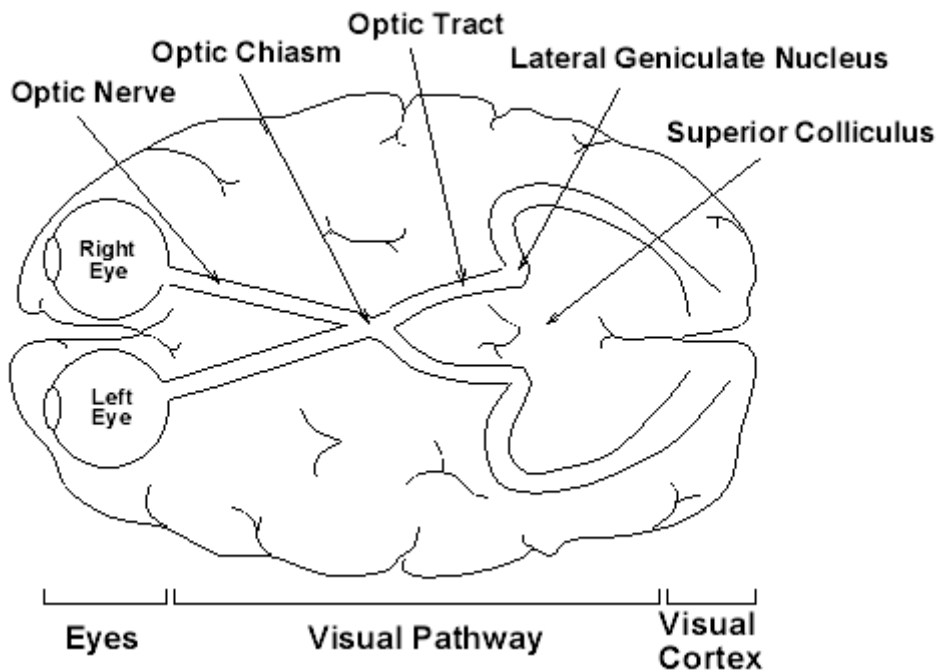


Figure 5: Reprinted from Reddy (1997) with permission, this image shows a plan view of the brain generalized to mark eyes, visual pathway and the visual cortex. The visual cortex is also referred to as the striate cortex, *Area 17*, and *V1*.

Being the most complex of the senses, it is also one of the most exploited, yet we still do not have a full comprehension and a numerical model. In fact, as stated by Schenk, the lack of a detailed understanding of vision is the reason why it is so difficult to program a computer to analyze and understand images (Schenk, 1999).

2.1.1. The Structure of the Eye

The spherical shape of the eye body (known as the eye-globe) is typically 22mm from posterior to anterior node (Kolb et al., 2001). Its diameter is reported as being approximately 20-21mm.

Focal length of an average eye is reported as ranging from 16mm up to 50mm in different documents. 50mm is totally absurd (Clark, 2005). Most commonly it is cited as being ~16mm and when focused at infinity ~22.4mm (Yasayan, 1996; Schenk, 1999; Allen, 2003 and Clark, 2005).

The following is reported for a "standard European adult" (via Clark, 2005):

- Object focal length of the eye = 16.7 mm (often rounded to 17mm)
- Image focal length of the eye = 22.3 mm (often rounded to 22 mm)

The object focal length is for rays coming **out of the eye**. But for an image on the retina, the image focal length is what one wants. (Clark, 2005)

A cross-section of the eye can be seen in the following figure (Figure 6):

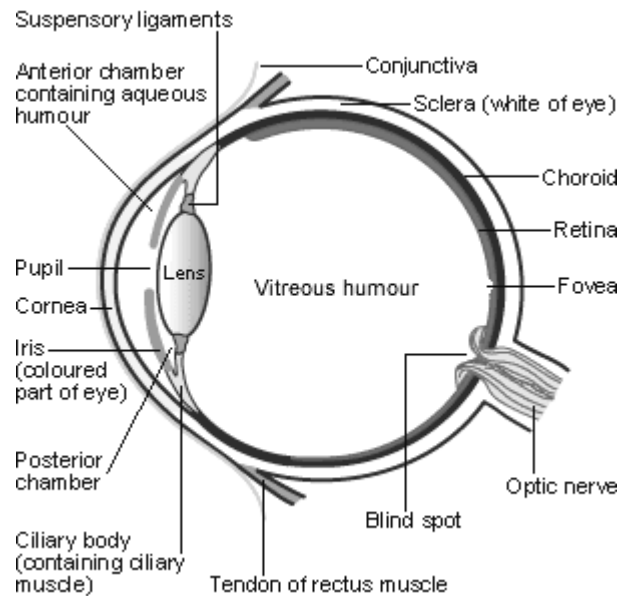


Figure 6: A cross-section of the eye.

While the fovea is the region where vision is the sharpest, the iris determines the amount of light that enters the eye. The large eye muscles enable eye movements. The *blind spot* is caused by the absence of receptors where the retinal arteries enter the eyeball. The two principal optical elements are the lens and the cornea. (Ware, 2000)

Foveal vision will be explained in detail in Section 2.1.4.

2.1.1.1. Blind Spot

The blind spot, as illustrated in Figure 6 is an area on the retina, where all the axons of the retinal *ganglion cells* meet to form the optic nerve. The blind spot is 5-7 degrees wide and is located at approximately 17 degrees of eccentricity⁵ (Reddy, 2001). Even though we compensate for the blind spot by having two eyes in normal vision, it may be interesting to look into the subject a little more carefully when the images are presented one by one to each eye in split screen displays such as in HMDs.

Luebke et al. considered the question as follows:

[...] There are no photoreceptors in this region, so we cannot detect any light that falls on the blind spot. Furthermore, the angular size of the blind spot is quite large, about 5 degrees (Andrews and Campbell, 1991). This therefore raises the question: could we reduce the detail of objects that fall onto a user's blind spot?

⁵ In Reddy's words eccentricity is "a measure of the extent to which a stimulus lies in a subject's peripheral vision, measured in units of degrees of arc from the fovea".

Unfortunately, the answer is “no” under normal stereoscopic vision. The blind spots for both eyes are in different regions of our visual hemisphere. Therefore, any part of the scene within one eye’s blind spot will be visible to the other eye.

For applications that render a separate image for each eye –such as virtual reality using head-mounted displays or stereo video projection- reducing the detail is conceptually possible, but still seems hardly worthwhile. For example, when using discrete LOD, an entire object would have to be projected onto the blind spot before we could degrade its detail or remove it from the scene. (Luebke et al., 2003)

It could also be interesting to consider blind spot in terms of stereoscopic perception, as the regular retinal disparity information should be missing in those bands. We leave this question for further future consideration.

2.1.1.2. Pinhole Camera Model

The human eye optics work similarly to a pinhole camera model:

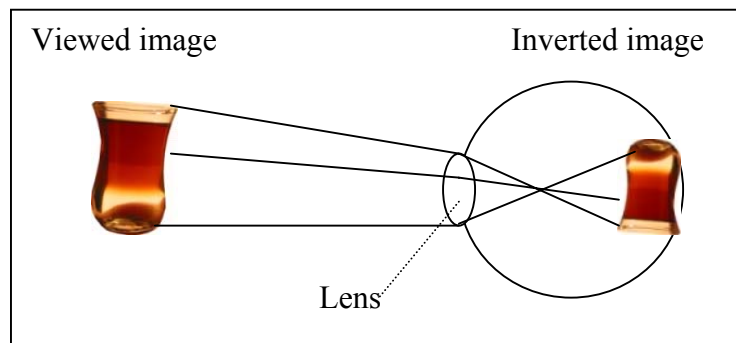


Figure 7: The lens focuses a small, inverted picture of the objects onto the retina.

This basic optical similarity between the pinhole camera model and the human visual system should not be taken to be very straightforward as there are a number of other perceptual issues involved in the way the human eyes function. In fact, as stated in Luebke et al., researchers have observed that 3D graphics systems should be based more on how our human visual system works than how a pinhole camera works:

A number of perceptual factors that can affect the amount of detail we can perceive under different circumstances are typically ignored in computer graphics. For example, we can perceive less detail for objects in our peripheral vision, or for objects moving rapidly across our gaze. (Luebke et al., 2003)

While the statement “perceptual factors are typically ignored in computer graphics” may be true particularly for the commercial portion of the computer graphics community, in the computer vision, robotics and stereoscopic displays research, there is considerable interest in modeling visual perception. Academic interest on the subject also covers a number of ways to model visual acuity to provide only the desired level of detail in the presented scene.

2.1.2. Visual Acuity (Resolution)

The eye's ability to recognize fine details is usually called visual acuity. In its simplest form, acuity is expressed as a measure of resolution (Schenk, 1999). This is important in display technologies because they give us an idea of the ultimate limits on the information densities that we can perceive (Ware, 2000).

Visual acuity is often measured as minimum angle of resolution (MAR) in units of minutes of arc. This has been determined experimentally by measuring a subject's ability to resolve two stimuli presented at different eccentricities (Linde, 2003). See Appendix 1 for a conversion of arc minutes to metric.

According to Reddy, visual acuity is a measure of the smallest detail that a person can resolve. This is only a measure of size and does not take into consideration the contrast of a target. Visual acuity is therefore normally assessed under optimal illumination conditions, e.g. black letters on a white background under bright lighting (Reddy, 1997).

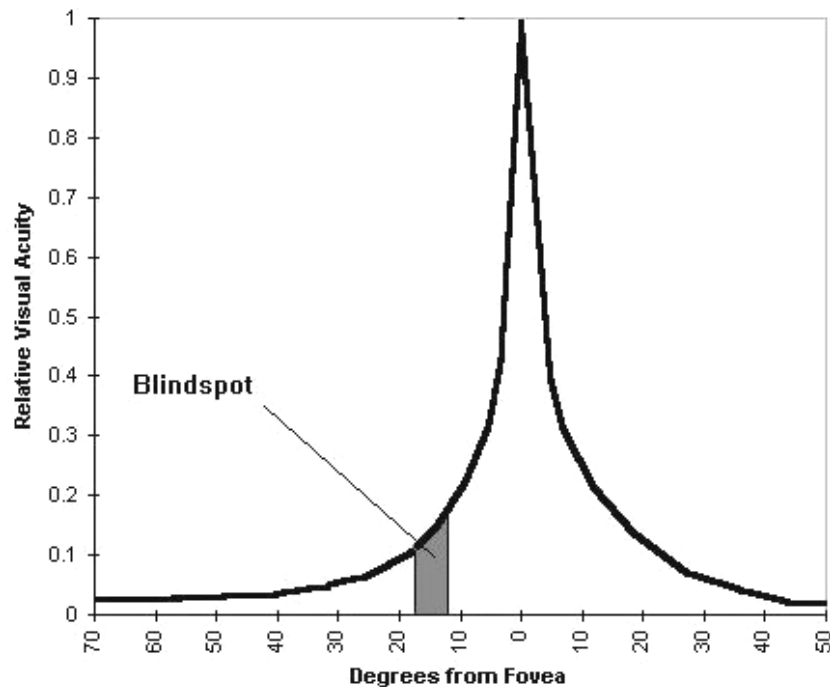


Figure 8: An acuity graph showing how the resolution changes from the center of fovea to the periphery. This graph is attributed to Coren and is also known as “Coren’s acuity graph” (See Blanke et al., 2002 and Coren et al., 1999)

Visual acuity is defined as $1/a$ where a is the response in x/arc-minute. The problem is that various researchers have defined x to be different things. However, when the different definitions are normalized to the same thing, the results agree. When we define x to be a line pair, as is normally done in modern optics, the $1/a$ value is 1.7 under good lighting conditions (Clark, 2005)

According to Clark, this was first determined by Konig in 1897 and was encountered in publications of Hecht (1931) and Pirenne (1967). Continuing to quote:

[...] The acuity of 1.7 corresponds to 0.59 arc minute per line pair. Thus, one

needs two pixels per line pair, and that means pixel spacing of 0.3 arc minute. Clark then asks the question “how many megapixels equivalent does the eye have”? Noting the fact that normally the eye’s field of view is 180 degrees, he calculates the brain pixels as 576 megapixels and concludes that it would require a “large format camera to record this kind of image detail”.

Similarly Colin Ware talks about brain pixels. As Ware states, brain pixel is basically another word for retinal ganglion cell receptive fields (Ware, 2004). Ware demonstrates that as the display size increases, the stimulated brain pixels in the *parafoveal* areas decrease. This leads to a conclusion that the bigger the display is, the more area is “lost”. See Section 2.3.7.2 for a more detailed explanation of this concept.

2.1.2.1. Types of Acuity

There are a number of different types of visual acuities. Ware lists a number of different types of acuities. The following is taken from the book as is (Ware, 2000):

Point acuity (1 minute of arc): The ability to resolve two distinct point targets.

Grating acuity (1-2 minutes of arc): The ability to distinguish a pattern of bright and dark bars from a uniform gray patch.

Letter acuity (5 minutes of arc): The ability to resolve letters. The Snellen eye chart (see Appendix 5 for a graphical illustration) is a standard way of measuring this ability. 20/20 vision means that a 5-minute letter target can be seen 90% of the time.

Stereo acuity (10 seconds of arc): The ability to resolve objects in depth. The acuity is measured as the difference between the two angles (a and b) for a just detectable depth difference.

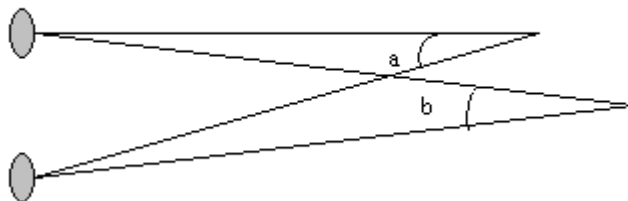


Figure 9: Stereo acuity (adopted and redrawn from Ware, 2000 by permission).

Vernier acuity (10 seconds of arc): The ability to see if two line segments are collinear.

While these are listed as “basic acuities” in the aforementioned book (Ware, 2000), the author also lists vernier acuity as a *superacuity*. Ware’s definition of superacuity is as follows:

A superacuity is the ability to perceive visual properties of the world to a greater precision than could be achieved based on a simple receptor model. Superacuities can be achieved only because postreceptor mechanisms are capable of integrating the input from many receptors to obtain better than single-receptor resolution (Ware, 2000).

2.1.2.2. Discussion: Superacuity or Hyperacuity?

In a large portion of other relevant literature, grating, vernier and stereo acuities are referred to as hyperacuities instead of superacuities. In several sources, vernier acuity is in fact used for defining *hyperacuity*.

The word hyperacuity can be found in modern English dictionaries, e.g. dictionary.com gives us “greater than normal acuteness especially of a sense; specifically: visual acuity that is better than twenty-twenty (see Section 2.2.3 to make sense of 20/20).

Sometimes, to determine which term is more commonly used in a language the author of this thesis, as a child of the Internet age, runs a search engine and looks at the number of results for each. We are aware of the fact that while this approach works and is in the limits of common sense, it is not a scientific statistical method. Still, when we realized the terms superacuity and hyperacuity are essentially referring the same concept, in an online scientific search engine (i.e. Google Scholar) we applied this method to determine which one to use. The term hyperacuity returns 1410 results where superacuity returns only 8 results. Judging from the context of these 8 publications and a number of others, we conclude that the two concepts are actually referring to the same phenomenon. Maybe standardization of this term is needed.

Among these we are most interested in stereo acuity.

2.1.3. Contrast Sensitivity

Another description of the human eye’s ability to resolve detail is given based on contrast sensitivity. According to Reddy (1997) the most common experimental device for accurately measuring a subject’s visual acuity is the contrast grating. This is a palette with a sinusoidal pattern on it consisting of bars switching the luminance values (contrast). To measure the contrast sensitivity, spatial frequency must be known. Spatial frequency is a measure of the spacing between bars, defined in units of contrast cycles per degree of *visual field* (c/deg).

Contrast sensitivity at threshold vision is analogous to grating acuity.

2.1.4 Foveal Vision

In the eye, there is a small yellow spot on the retina known as macula, and in the center of the macula lies a *cone* free area, densely filled with *rods*, called fovea centralis, often shortly referred as fovea. Fovea is the unit that controls the visual acuity and it does not fully develop before 4 years of age (Hendrickson et al., 1984).

The fovea is located at 11.8° or 3.4 mm temporal to the optic disk edge, and the cross diameter of the central fovea -from foveal rim to foveal rim- is reported as being 1.2-1.5 mm (Kolb et al., 2001; Polyak, 1941). At the fovea, only cones are present (no rods) with an approximate spacing of 2-3 μm (Schenk, 1999).

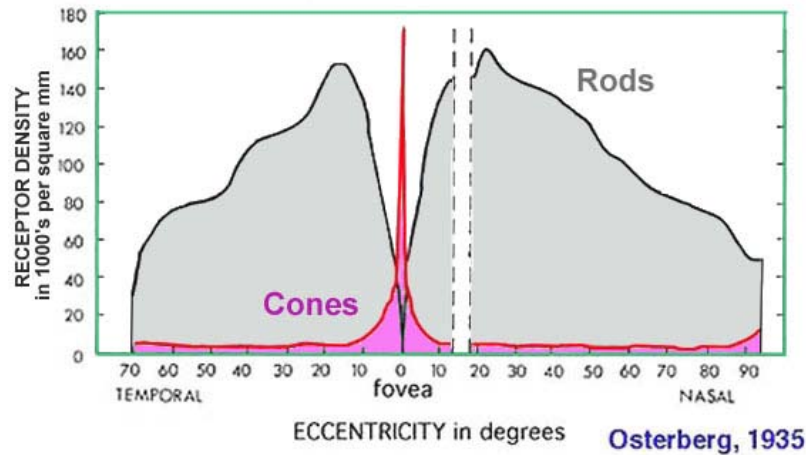


Figure 10: Rod and cone distribution across the fovea by Osterberg 1935. Cones peak in the fovea while rods are dense in *parafovea*.

Cones and rods are two types of photoreceptor cells in the retina. While the total number of cones in the fovea is reported as approximately 200,000 (17,500 cones/degree²), the rod free area is 1° thus there are 17,500 cones in the central rod-free fovea. The density in the center of fovea (50 x 50 μm) is estimated at between 96,900 and 161,900/mm² by several different researchers (Osterberg, 1935; Ahnelt et al., 1987; Curcio et al., 1987).

Cones capture high resolution and color information while rods record lower resolution and monochromatic information. Thus as we progress towards the periphery from the central foveal area, the quality of vision will decline both in terms of resolution and color.

While the retina can resolve detail of around 0.5 min of arc, there are 130 million *photoreceptors* / 1 million ganglion cells. In peripheral vision, the highest sensitivity to spatial detail is at the fovea, the central 4 to 5 degrees of vision, and from fovea towards periphery a 35-fold reduction is reported (Reddy, 1997; Nakayama, 1990).

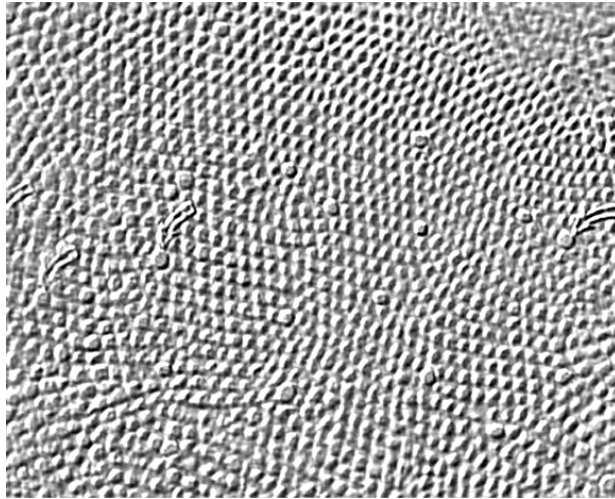


Figure 11: Tangential section through the human fovea. Larger cones (marked by arrows) are blue cones. The hexagonal shape of the cones optimizes the packing density (Schenk, 1999).

Current understanding is that the 6 to 7 million cones can be divided into "red" cones (64%), "green" cones (32%), and "blue" cones (2%) based on measured response curves. They provide the eye's color sensitivity. The green and red cones are concentrated in the fovea centralis. The "blue" cones have the highest sensitivity and are mostly found outside the fovea, leading to some distinctions in the eye's blue perception (Nave, 2005).

While these numbers document the facts for each eye, because we are interested in stereoscopic vision, the following graphic showing the binocular visual field is relevant to this section:

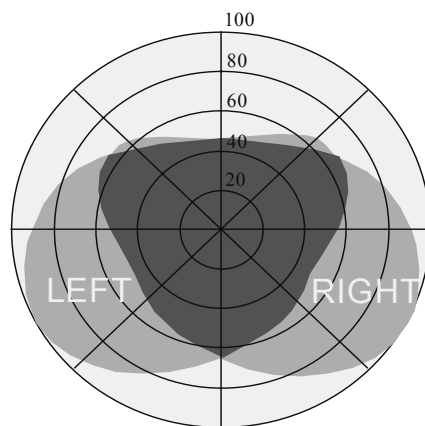


Figure 12: The human visual field for a person gazing ahead. The darker gray area shows the region of binocular overlap. Reprinted from Ware (2000) by permission.

2.1.5. Motion Sensitivity

Motion is another parameter that has significance for understanding the human visual perception. The eye is less sensitive to detail moving across retina and fast moving objects become “blurred” (Reddy, 1997).

Peripheral vision, while recording a lower resolution image, offers an excellent ability to detect movement. This is because it has a wide range of illumination levels (Scott et al., 2005). Being able to detect the motion in high sensitivity in the peripheral vision is a valuable tool for survival for humans.

2.1.6. Summary

This section gave a thorough overview of current knowledge on human visual system with reference to its modeling when it was necessary. The structure and basic optical numerics about the human eye is provided and foveal vision is described in detail. Visual acuity and its types as well as models were also covered.

2.2. Depth perception

*“eyes twinned make the world deep”
In the poem “Saws” by Trevor Joyce*

This section is an introduction to depth cues and it focuses on stereoscopic depth. The term, cues, has been utilized to formalize the specification of stimulus conditions for space perception (Carr, 1935, via Ostnes et al., 2004).

Depth perception is achieved through both monocular and binocular cues. Disparity only conveys depth to a distance of 25m (Ware, 1995)⁶, and only where the binocular viewpoints overlap, that is 120° horizontally, and 135° vertically. In addition to that, conjugate points do not present excessive disparity, i.e. object distances greater than 10cm. (Linde, 2003)

Several fields of computer science, robotics, and virtual reality have exploited aspects of stereoscopic perception for recovering depth information from the real world using photographs. Photogrammetry itself is a result of biological understanding and a technical application of stereovision.

Humans are predators, and like other predators their eyes are in the front of their heads, as discussed in the quoted text from Lipton 1982 below:

For stereopsis to be possible, both eyes must be able to converge on an object so that the image can be fused into a single three-dimensional view. Eyes placed on opposite sides of the head simply cannot accomplish this. They do have a distinct advantage over eyes that can converge: excellent coverage of the visual field, a nearly panoramic view of the world. (The advantage of such a view is obvious. For example, a grazing animal can constantly be on the lookout for predators without having to move its head.) Binocular vision, on the other hand, restricts the field of view to the direction faced. (Lipton, 1982)

Having two eyes located to the front is what gives us stereopsis, which is a valuable depth cue. It is not the only one, and we know that monoscopic depth cues are sufficient for a person to maintain his/her life normally, but it does provide us extra valuable information and it is very relevant to this thesis. We will therefore give a summary of depth cues and focus on binocular (or stereoscopic) vision in the following sections.

Pfautz summarized the following on depth perception from the relevant literature (Pfautz, 2002):

- The ability to fuse two images into a single image is described in terms of the horopter, a circle in space defined by points that fall onto corresponding points

⁶ This distance is reported variously because of differing parameters and interpretation. Ware (1995, 1998, 2000) and Linde (2003, 2004) report 25m or 30m, while it is reported as 300m to 670m in other books and research papers. Part of the reason also is that while taking the absolute limit at zero disparity, what is useful depends on interpretation. See Section 2.2.2.3 How Far Can We See Stereoscopically for an explanation.

on the two retinae. Points that lie on the horopter will be fused into a single image (Graham, 1951).

- Panum⁷'s area is the range in front and behind the horopter where single images can be seen (Buser & Imbert, 1992). This range is mainly a function of the viewing distance and is important in the design of stereoscopic display systems.
- Binocular vision can only occur where the fields of view of the two eyes overlap. The horizontal binocular visual field is about 120° out of a possible 200°.
- Stereopsis plays an important role in fine discrimination of objects in the near- and mid-fields but has a diminished role for objects more than ten meters from the viewpoint (Nagata, 1993).
- Stereo vision is more useful for relative depth comparisons than absolute judgments (Gillam, 1995).

The references in the above bulleted list are all via Jonathan D. Pfautz, as the list is taken as is from his report (Pfautz, 2002).

2.2.1. Depth Cues

As mentioned, having two forward facing eyes creates stereopsis and this is a major *depth cue*, though we need to remember:

“Stereoscopic disparity is only one of the many depth cues that the brain uses to analyze 3D space, and it is by no means the most useful one. In fact as many as 20% of the population may be stereo-blind, yet they function perfectly well and in fact are often unaware that they have the disability” (Ware, 2000)

Depth cues are often considered under two main categories: oculomotor and monocular cues.

Oculomotor cues are based on our ability to sense the position of our eyes and the tension in our eye muscles (Goldstein, 2002). These are created by *convergence* and *accommodation*. While convergence refers to the coordinated inwards movements of the eyes, accommodation is a term used for describing the eye's behavior when focusing on the object of interest (also see the glossary for more in-depth definitions). These two normally work together, and in stereoscopic displays their separation is reported to be a cause for eyestrain. Several problems are associated to stereoscopic displays in relation

⁷ The man who gave his name to this area, Peter Ludvig Panum was a Danish medical scientist who lived between 1820 and 1885. In 1858 Panum published a monography on "Physiologische Untersuchungen über das Sehen mit zwei Augen". He proposed the concept of corresponding circles of perception instead of the absolute identity of corresponding points on the retina. This was met with opposition particularly by A. W. Volkmann of Halle who tried to explain all stereoscopic phenomena psychologically (Arch. Ophthalmol., 1859). Both authors defended their views with numerous experiments. Panum's results in "Über einheitliche Verschmelzung verschiedenartiger Netzhautindrücke beim Sehen mit zwei Augen" (Arch. Anatomie, 1861) are still valid today (Piper, 1999).

to the viewer's comfort, including the accommodation convergence problem. Seeking solutions for these problems is an active area of research as will be presented in later chapters.

Of the oculomotor cues, convergence is essentially binocular, because it cannot happen without two eyes, but accommodation will also work monocularly.

Monocular cues are many, including *depth of focus* (accommodation), occlusion, pictorial cues, relative height, relative size, cast shadows, familiar size, atmospheric (aerial) perspective, linear perspective, and texture gradient and shape-from-shading (See Goldstein, 2002 and Ware, 2000 for detailed explanations of these terms).

Motion *parallax* is another depth cue but it is a special case and sometimes categorized separately as a "monocular dynamic" cue while the others are listed as static.

2.2.2. Stereoscopic Perception

"Stereoscopy is the science and art that deals with the use of binocular vision for the observation of overlapping photographs or other perspective views and the method by which such views are produced. Essentially most of us with "normal" eyesight have stereoscopic vision i.e. the ability to see and appreciate depth of field through the perception of parallax." (RSCC, 2005)

The human eyes are horizontally separated by an inter-pupillary distance (IPD). This is sometimes called inter-ocular separation. Because of this separation, the retinal images in the left and right eyes are slightly different from each other. The IPD, also referred to as eye base in photogrammetric literature, is reported to range from 50 to 76mm in adults (Robinett, 1999), though it is commonly generalized to an average of 65mm.

The closer the stimulus object is to the observer's eyes, the greater the difference between the retinal images acquired. The differences between two retinal images are referred to as disparities, and it is through analysis of image disparity that depth may be estimated. (Linde, 2003)

The role of stereoscopic disparity in depth perception should not be overrated. As mentioned earlier, there are many other depth cues. However, there are certain advantages that need to be mentioned.

3D coordinates from stereo pairs gives us valuable information. There is also some justification for plain visualization. As a visual effect it clearly fascinates the majority of people when they see a 3D picture (Holliman, 2005).

"[...] Stereoscopic displays can provide a compelling sense of a three-dimensional virtual space, and for certain tasks, they can be extremely useful (Ware, 2000)"

Holliman, 2005 has listed the following as the benefits of stereoscopic vision:

- Relative depth judgment. The spatial relationship of objects in depth from the viewer can be judged directly using binocular vision.
- Spatial localization. The brain is able to concentrate on objects placed at a certain depth and ignore those at other depths using binocular vision.
- Breaking camouflage. The ability to pick out camouflaged objects in a scene is probably one of the key evolutionary reasons for having binocular vision (Richards, 1970)
- Surface material perception. For example, lustre (Helmholtz, 2000), sparkling gems and glittering metals are in part seen as such because of the different specular reflections detected by the left and right eyes.
- Judgment of surface curvature. Evidence suggests that curved surfaces can be interpreted more effectively with binocular vision.

To further justify the contribution of stereoscopic vision, Campbell and Green (1965) have found that binocular viewing improves the visual acuity in general by 7% as compared to the monocular viewing. They also found a $\sqrt{2}$ improvement in contrast sensitivity (Ware, 2000).

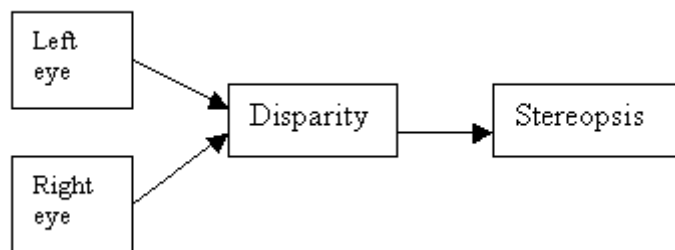


Figure 13: Disparity and stereopsis (adopted from Goldstein 2002). Stereopsis is the sense of depth resulting from the information provided by binocular disparity.

Kepler also did research into the curious fact that we have two eyes and how two inputs turn into a single image (he had more than one problem with his sight which might have motivated some of his interest in vision related topics). Quoting from Lenny Lipton (1982):

Kaufman holds that Kepler's construct is actually isomorphic with the modern projection theory of stereopsis. That is, it is structurally identical and produces the same explanations of phenomena despite the fact that modern scientists no longer believe in mental rays of light originating from the eyes, a concept, by the way, that did not begin with Kepler but originated with the ancient Greek philosophers, the Pythagoreans.

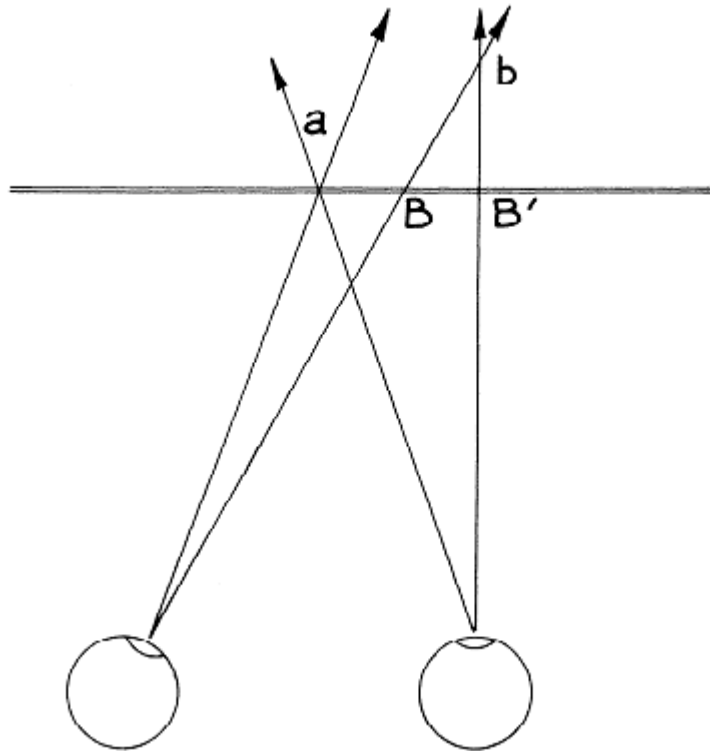


Figure 14: Kepler's projection theory. Mental rays cross at point a, leading to a perception of a single image. Object at point b will then be seen double. The straight line containing points B and B' is the horopter of Aguilonius. Reprinted from (Lipton, 1982).

While Kepler's theory of projection goes as far back as 1611, the first practical stereoscopic drawings are credited to Charles Wheatstone in 1838, some years before the first stereo photographs.

Sir Charles Wheatstone was an English physicist and inventor whose work was instrumental in the development of the telegraph in Great Britain. His work in acoustics won him (1834) a professorship of experimental physics at King's College, London, where his pioneering experiments in electricity included measuring the speed of electricity, devising an improved dynamo, and inventing two new devices to measure and regulate electrical resistance and current: the Rheostat and the Wheatstone bridge named after Wheatstone, as he was the first to put it to extensive and significant use (Katz, 2005).

Wheatstone also invented a stereoscope in 1833 and have published a paper in 1838 with the title "Contributions to the Physiology of Vision- Part the First: On some remarkable, and hitherto unobserved, Phenomena of Binocular Vision", where he explained that doubleness of vision, caused by retinal disparity, actually produced the depth sensation stereopsis (Lipton, 1982).

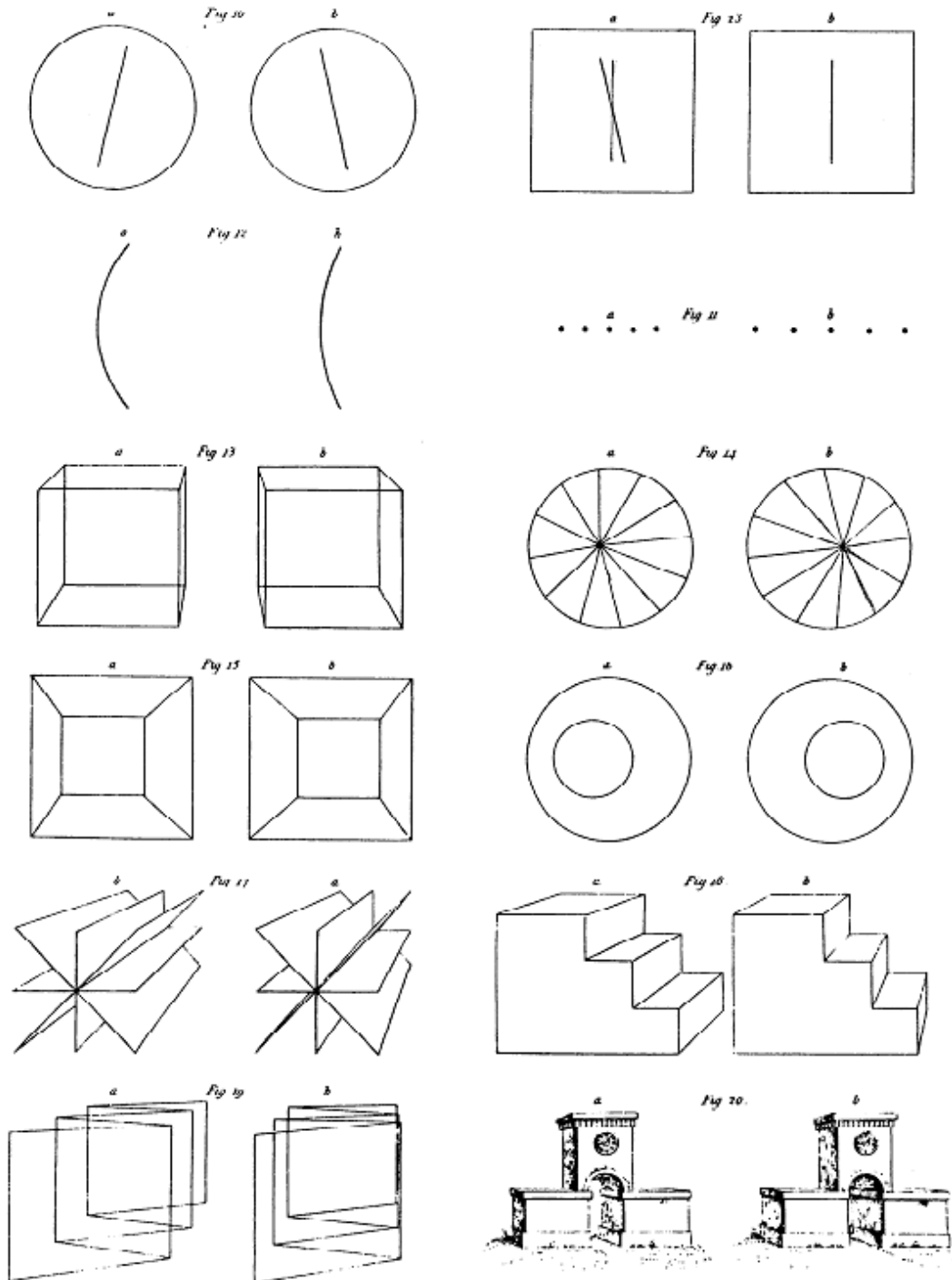


Figure 15: The drawings of Charles Wheatstone as presented in Lipton, 1982.

2.2.2.1. Horopter

Also referred to as Vieth-Muller’s horopter, or Vieth-Muller circle, this is a concept related to disparity. When we focus on an object with two eyes, the eyes are positioned so that the projected images on two retinas correspond. This is the “zero disparity” position.

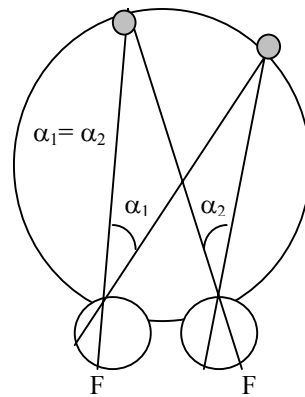


Figure 16: Vieth-Muller Circle is a theoretical horopter. The circle represents the theoretical locus of points in space that stimulates corresponding retinal points. Image redrawn and caption modified from Webvision 2005, Vieth Muller.

The horopter is the imaginary 3D surface that extends from the focused object to include all other points at which the images fall onto corresponding places in both eyes. (Huk, 1999)

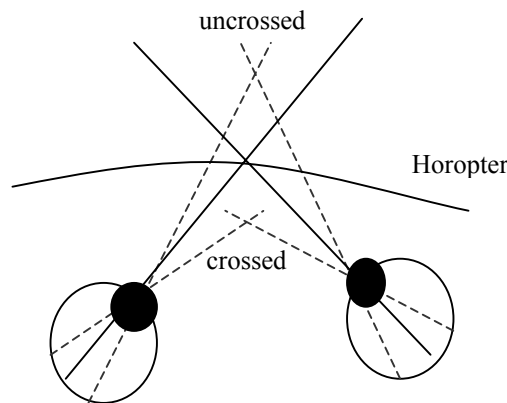


Figure 17: Horopter. What is marked as “crossed” in this picture is referred to as “negative parallax” in stereo displays, while the horopter is the display’s surface and that behind the screen labeled “uncrossed” is “positive parallax”.

Uncrossed disparity occurs when an object is farther away than the horopter. It requires the viewer to 'uncross' the eyes to fixate on it. Crossed disparity, meanwhile, occurs when an object is closer than the horopter. It requires the viewer to 'cross' the eyes to fixate on it. (Huk, 1999)

Accurately speaking, there exists an area around the horopter where fusion is perceived. This area is called Panum’s fusional area (Ohshima et al., 1996).

2.2.2.2. Panum's Fusional Area and Diplopia

Panum's fusional area (also known as Panum's space or Panum's area) is where binocular vision occurs. In consistence with the ganglion cell receptive fields, this space narrows at the fixation point and expands in the periphery. Its boundaries are within a short distance on either side of the horopter. The area that corresponds to this region on the retina is Panum's fusional area.

It does not have a fixed size:

“Panum's area extends **approximately** ± 600 arc second (10 arc minutes) on either side of the horopter, it does not have a fixed size, but varies depending on stimulus conditions. It is larger for big, moving objects, but is narrower for detailed and stationary objects. Objects far from the horopter, that is, objects that are outside of Panum's space (Panum's area), cause very large disparities on the retinas, and they cannot be fused. They are seen in diplopia.”(Salmon, 2005)

Panum's fusional area is frequently referred to in relevant literature. It is a model that defines the boundaries of binocular fusion by investigating the greatest amount of horizontal disparity. If the disparity is greater than a certain value, fusion will not happen. When this is the case, diplopia (double vision) occurs. Panum's fusional area is the 3D area without diplopia occurrence.

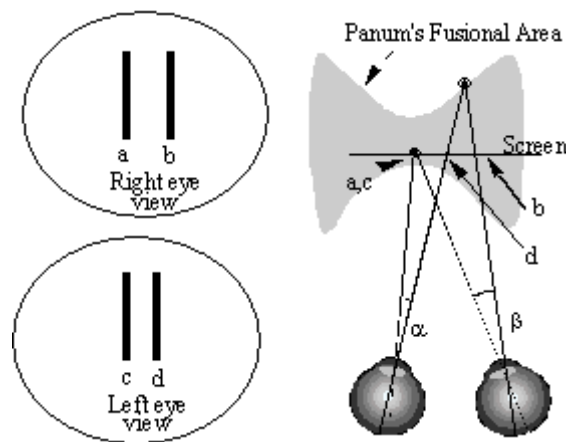


Figure 18: Panum's fusional area in relation to a simple stereo display. Image reprinted from Colin Ware by permission (Ware, 1998 and 2000). While angular (retinal) disparity is α - β ; screen disparity (parallax) is $(c-d)-(a-b)$.

Agreeing with the previous statement that its size is not fixed, the boundaries of Panum's fusional area are not crisp:

This classical notion was considered to place an absolute limit on the amount of disparity between primitives for fusion to occur regardless of the presence of neighboring primitives. Recently studies have shown a need for the reformulation of the Panum idea. It now seems that neighboring primitives do have an effect on the fusion chances and so the idea of an absolute limit seems

unlikely. With a great many of the accepted research works accepting the Panum notion as fundamental, there now seems a need to re-evaluate their results.

Therefore, the fusional limit is not absolute but corresponds to what has been called a **disparity gradient**. The disparity gradient uses primitive size and proximity in deciding whether fusion is possible or not. Experiments on human subjects showed that this limit was approximately 1. (CVOnline, 2005).

“Panum's fusion area in adults is approximately 10 arc minutes... If the size of Panum's fusion area in infants is proportional to peak contrast sensitivity, as it is in adults, the extent of Panum's fusion area in 9- to 10- week olds should be 200 arc minutes” (Aslin, 1993).

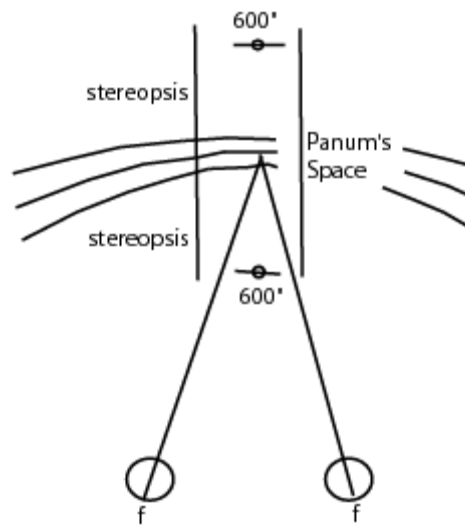


Figure 19: Relationship between the horopter, Panum’s area and the zone of stereopsis. Reprinted from Salmon, 2005 with permission. Suggested 600 arc seconds (10 arc minutes) on either side of the horopter are marked.

Ware, citing from Patterson and Martin, 1992, reports that this area has remarkably little depth. At the fovea the maximum disparity before the fusion breaks down is 1/10 degree and at a 6-degree eccentricity the limit is 1/3 degree (Patterson, et al., 1992 via Ware, 2000).

However, the size of Panum's fusional area is highly dependent on a number of visual display parameters such as exposure duration to the images and the size of the targets. Moving targets can be fused at greater disparities. Depth judgments can be made outside of the fusion area, although these are less accurate (Ware, 1998; Ware, 2000).

What these numbers imply in terms of stereo displays is studied in Section 2.3.

2.2.2.3. How Far Can We See Stereoscopically?

If the disparity is too small, it will not be sufficient to elicit a sense of stereoscopic depth (Salmon, 2005). The disparity is too small when the object is too far away. In Section 2.1.2.1 we introduced “stereo acuity” which defines the limits for this.

Stereo acuity is reported as being 10” (seconds of arc: see Appendix 1 for an explanation of this unit) in Ware, 2000. In other sources, this value is reported as varying between 1.8” to 20”. This number varies due to the individual differences in people (Holliman, 2005). 20” is suggested as a working limit by Diner et al.1993.

The stereo acuity is the smallest perceptible change in angular disparity. An illustration can be seen in Figure 20 below, where the difference between a and c gives the stereo acuity. Holliman reports the following:

“[...] A person with a stereo acuity of 20” and an eye separation of 65mm will be able to perceive depth differences between small objects of just 0.84mm at a distance of 750mm from the eyes.”

It is also possible to calculate a geometric value for the furthest possible range of stereo vision which occurs when the vergence angle between the two visual axes is equal to or less than the stereo acuity (Holliman, 2005).

This value is calculated as 670m with a stereo acuity of 20” and an eye separation of 65mm.

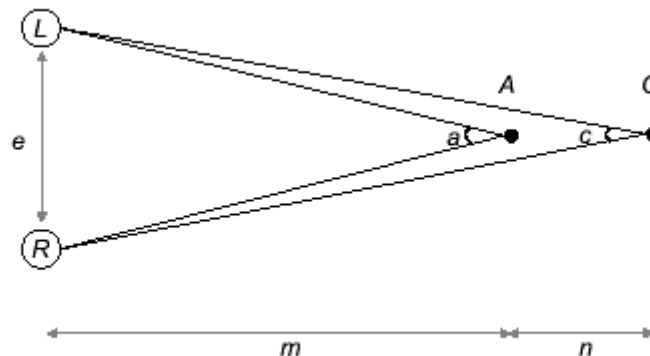


Figure 20: Holliman calculates $n = 0,84\text{mm}$ when $m = 750\text{mm}$ from the eye, and maximum $m = 670\text{m}$ when IPD (base) is $e = 65\text{mm}$, and the stereo acuity $(a-c) = 20''$. See Holliman, 2005 for the formulae used for this calculation. Reprinted from Holliman, 2005 by permission.

In Figure 20, points such as C, at a distance of 670m or more from the observer, can not be distinguished from point A in terms of distance using binocular vision alone. Just before this limit is reached the smallest distinguishable depth difference between points will have increased to over 300m and it is clear only gross differences in depth will be perceived at the furthest limits of stereoscopic perception (Holliman, 2005).

While giving a definite distance limit for stereoscopic vision depends on all kinds of parameters, with the above formula it stops at 670m for a 20" stereo acuity. The interpretation of what is **useful** however, depends on the task at hand.

In Ware's interpretation, which is accepted by Linde as well, further away than 25-30m (see Section 2.2.) we do not receive useful stereoscopic vision (Ware, 2000 and Linde, 2003). It is also important to remember that Ware takes the stereo acuity as 10" instead of 20" (see Section 2.1.2.1.). Though this interpretation does not depend on the stereo acuity and eye base as the generic formula Holliman uses, but is an interpretation of the retinal disparity.

2.2.2.4. Cyclopean Eye

Also known as "the center eye"; this is used in modeling stereoscopic vision.

As the two eyes obtain two potentially quite different images, one of the eyes is dominant, and its viewpoint masks that of the other eye to create a clean image, which is perceived to originate from midway between the two eyes, known as the cyclopean viewpoint (Yeh, 1993). Biologically this process is known as binocular sensory fusion (BSF). (Linde, 2003)

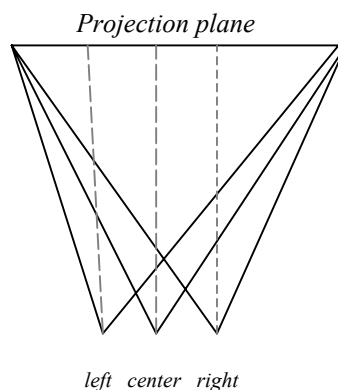


Figure 21: A basic illustration of Cyclopean View.

As we see the world as one image and not two, a single eye can represent binocular vision. This is called the *cyclopean eye*. The cyclopean eye is an imaginary eye situated midway between the two eyes. Using the cyclopean eye, crossed and uncrossed diplopia can be explored (Webvision, 2005, space perception)

Ware et al. have suggested a "*Cyclopean Scale*" which is based on cyclopean view concept. It is suggested to be helpful with problems associated with stereoscopic displays. More on the Cyclopean Scale can be found in Section 2.3.8.2.1.

2.2.3. Limited Depth of Field

A relevant concept to depth perception, which also appears in nearly all photography and camera related texts, is depth of field, abbreviated as DOF. The same abbreviation is also used to express depth of focus, which is different from depth of field⁸ even though the two terms are used interchangeably even by professionals.

Throughout this thesis, the abbreviation DOF stands for depth of field and **not** *depth of focus*.

DOF defines the range where vision is the sharpest. It is not dependent on stereovision. A single eye also has a depth of field. Once the eye accommodates on an object, there is a certain range where the image depth is the sharpest. Accommodation numerics for the human eye are briefly as follows:

The lens cannot accommodate for an object closer than 10cm, and the stimulus will be blurred. Beyond 6m, the lens is entirely flat (fully accommodated). Convergence only occurs for stimuli closer than 10m. (Linde, 2003)

The 6-meter limit is also used in eye examinations that use the Snellen chart. When the Snellen chart is used for measuring visual acuity, the chart is placed 6 meters away from the person. The symbols in the 4th line from the bottom of this chart (as can be seen in Appendix 5) are designed so that a “normal” eye can recognize those at a 6m distance. When this is the case, the person is said to have 6/6 vision (or 20/20 in feet). If a person has 6/30 vision, s/he would have to get as close to the chart as close as 1.8 meters to be able to read the same line. It is also possible that an individual’s visual acuity is better than normal.

For human eyes, as the size of the pupil changes, DOF also changes. Assuming a 3mm pupil diameter and that the eye focuses at infinity, objects between about 3m and infinity are in focus and this corresponds to about 1/3 *diopter* (Ware, 2004). If the eye is then focused to distance d_0 , then the objects within

$$\frac{3d_0}{d_0 + 3m} < \frac{d}{m} < \frac{3d_0}{d_0 - 3m} \quad \text{(Equation 1)}$$

appear in focus.

⁸ Depth of Focus is also called “focus spread” and it differs from Depth of Field in that it describes the distance over which light is focused at the **camera’s sensor**, as opposed to how much of the subject is in focus (McHugh, 2005).

For some common viewing distances, DOF values are calculated based on this formula as follows (Puolamäki, 2004):

Viewing distance d_0 (m)	$\Delta \frac{1}{f}$	Near (m)	Far (m)
0.5	2.0 m^{-1}	0.43	0.6
1.0	1.0 m^{-1}	0.75	1.5
2.0	0.5 m^{-1}	1.2	6.0
3.0	0.33 m^{-1}	1.5	∞

Table 1: Depth of field values calculated based on the viewing distance (Puolamäki, 2004). Colin Ware has the same results with the same formulae except he refers to the calculated distance as “depth of focus” (Ware, 2004). We chose to differentiate between depth of focus and depth of field as explained throughout this chapter and in this case the distance refers to the depth of field.

In the human eye, the lens shape and hence the focal length/refractive power changes during the fixation of objects at various distances, as accommodation is applied. The DOF will change accordingly; being very small for fixations on close objects when the lens is curved, and large/infinite for distant stimuli, where the lens is far flatter (Linde, 2003).

Lenny Lipton suggests that the *depth range* and the depth of field are two different concepts even though it is easy to see them used interchangeably or in confusing manners (Lipton, 1992). His explanations are geared more towards display technology than the biological aspects, but nonetheless it is worth noting at this point:

“Depth of field applies to planar as well as stereoscopic photography and is used to establish limits for acceptable focus in front of and behind the plane in space on which the lens is focused. Thus we can draw a comparison between near and far limits of focus.

I would like to state that depth of field and depth range are entirely distinct physical entities and have no relationship to each other.

It may be obvious, but it should be said: The lenses of a stereo-camera before used for one distance and converged on another. If the limits of depth of field are exceeded, the image will simply be out of focus. Exceeding the limits of depth of field does not result in pain or discomfort. If the depth range criteria are exceeded, homologous image points go beyond allowable standards for fusion, and eyestrain does result.

Depth of field effectively applies to vision with one eye, even if the image is being seen with two. Depth range applies to both eyes working as a team while viewing a stereoscopic effigy” (Lipton, 1982).

Even though DOF is a monocular concept in essence, simulating DOF for each eye is a technique that is suggested to help with problems associated with stereoscopic displays. Some examples of DOF Simulation can be seen in the later Section 2.5.6.

2.2.4. Summary

We introduced human depth perception and gave definitions of the relevant terms and concepts. The focus was on stereoscopic perception, which is central to this work. Several models of stereoscopic perception such as the horopter, Panum's fusional area and the cyclopean eye were covered in this section. Limited depth of field in terms of human vision was also reviewed.

2.3. Stereoscopic Viewing Techniques

We have previously documented the available knowledge on how humans perceive depth and in particular, stereo. Now the current state of how we display depth, particularly stereoscopic depth will be reviewed. How the human brain processes received information from the two eyes and how a display system produces a 3D visualization based on the input from two cameras are principally similar processes, but in practice there are important differences between them.

A perspective display format can utilize most of the depth cues except for binocular disparity, which can only be provided by using a stereoscopic display (Ostnes et al., 2004).

In computer graphics, the improvements in speed, resolution, and economy make interactive stereo an important capability (McAllister, 2002). According to Robinett, one of the desirable qualities for a virtual reality display is that it has a wide field of view and be stereoscopic (Robinett, 1999).

Utilizing the human stereoscopic vision capability of fusing two retinal images into one image, the stereoscopic display generates a powerful additional depth cue based on stereopsis (Ostnes, 2004).

There are several other advantages to stereoscopic displays over monocular ones. Ian van der Linde gives an overview on these advantages in the following two paragraphs:

“Stereoscopic displays provide the user with a compelling sense of presence as they are able to reproduce a greater number of perceptual depth cues than standard monocular displays.

It has been demonstrated by Drascic and Milgram that stereo displays are beneficial for spatial manipulation tasks, by measuring the accuracy by which test subjects were able to place a 3D pointer on a target (Drascic and Milgram, 1991). An increase in accuracy was achieved with a 3D display over the equivalent 2D representation. For a variety of manipulative and observational tasks, the stereoscopic display can support a higher level of spatial accuracy, and hence improve dexterity (Hubona et al., 1999). For communication, and a variety of other applications, the stereoscopic display is generally more appealing than two dimensional representations (Pastoor, 1995)” (Linde, 2003).

As summarized in Ostnes et al., there are other studies providing empirical proof that stereoscopic viewing is superior to its alternatives for several tasks:

“A comparison of 3D perspective⁹ and 3D stereoscopic displays in a simulated tracking task has been presented by Kim et al (1987). The stereoscopic display

⁹ A 3D perspective display can be achieved by projecting an object onto the view (projection) plane and then mapping the view plane onto the display screen. There are two methods to generate the perspective projection: the viewpoint-transformation and the object transformation (Kim et al., 1993 via Ostnes et al., 2004)

resulted in lower tracking error over all visual conditions. However, the perspective display with appropriate visual perspective parameters (i.e. optimal viewing angles in both the vertical and horizontal plane) and visual enhancement depth cues (such as vertical reference lines) resulted in equivalent performance as compared with the stereoscopic display.

Yeh (1992) investigated spatial judgments (relative depth and altitude) with monoscopic and stereoscopic presentation of perspective displays. The results showed that the presence of binocular disparity in the stereoscopic view improved the spatial judgment. In another study, McLean et al. (1994) compared a 3D perspective video display (one camera view without visual enhancement) with a stereoscopic video display for a peg in a hollow task. The results showed that the stereoscopic video was superior to the 3D perspective video. Yeh (1992) discussed the problem associated with perceptual distortions in perspective projection resulting from the enhancement cues.

The benefit associated with using stereoscopic displays was further reported by Barfield and Rosenberg (1995). Their experiment showed that the stereoscopic display was superior to the perspective display (monoscopic) in judging the relative elevation. However, the judgments of relative azimuth angle were not improved by the use of the stereoscopic display” (Ostnes et al., 2004).

Viewing a stereo pair does not always involve a display in the conventional sense of the word. When viewing stereo pairs, a mechanism is required so that the left eye sees only the left eye view and the right eye sees only the right eye view. Systems are established to provide this to happen simultaneously, or in a time sequence. There are many mechanisms, which have been proposed to accomplish this (McAllister, 2002).

The different stereoscopic viewing methods can initially be categorized by two major means; time parallel and time multiplexed methods. Time parallel methods send the two images at the same time, while time multiplexed (also referred to as time sequential or field sequential) methods send them in a sequence. A list of most common stereoscopic viewing and display techniques can be seen in Figure 22.

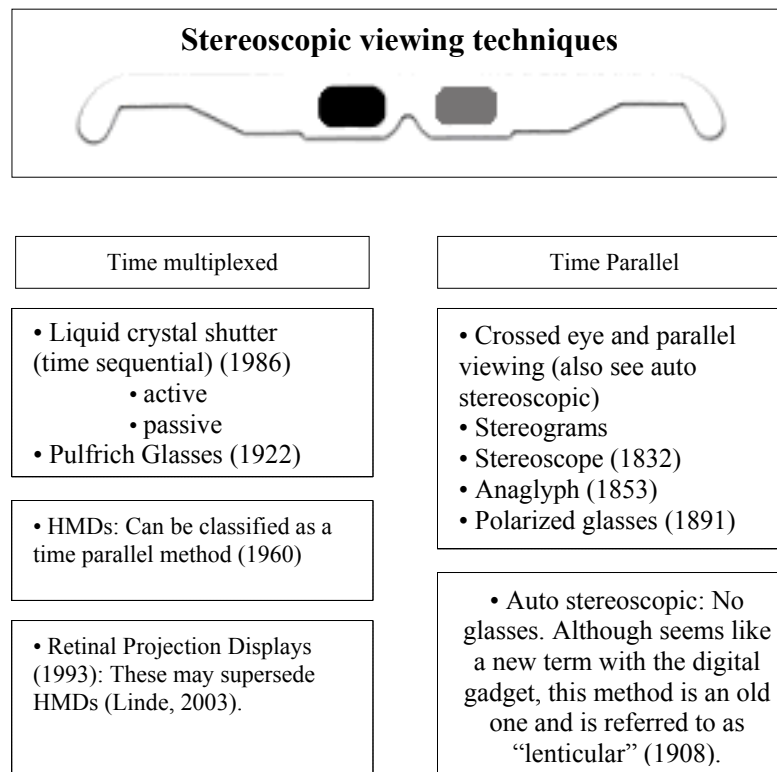


Figure 22: The methods of viewing stereoscopic images and graphics. The illustrated glasses on the top are Pulfrich Glasses. These glasses have a dark lens and a clear lens. Observed by Carl Pulfrich around 1922, the Pulfrich effect shows us that the brain perceives the image through the dark glass slightly later than the clear glass. Also known as Pulfrich method, approximately 10% difference in the shade is sufficient for it to work. Like Kepler, Carl Pulfrich was stereo-blind.

2.3.1. Time Multiplexed Displays (TMDs)

The left and right images are alternated on a single screen and the user wears shutter glasses to occlude the proper image in synchronization with the display. Crosstalk¹⁰ might occur if the synchronization fails (Linde, 2003).

Time multiplexed techniques are sub-classified as active or passive systems.

In a passive system the glasses do not control anything; they only receive the polarized light and interpret it. A polarizing shutter is attached or integrated into the display. In an active system, the glasses do the polarization, and an electronic pulse causes the lens to “open” or admit light from the display device. When no electronic pulse is present, the lens is opaque, blocking the eye from seeing the display device. (See McAllister, 2002 for more).

¹⁰ Stereo *crosstalk* occurs when a portion of one eye view is visible in the other eye. In this case the image can appear blurred or a second or double image appears in regions of the scene being viewed creating a phenomenon called *ghosting*. Cross talk can create difficulty in fusing Left/Right views. When using the same display surface to project both eye views, cross talk can be a problem. When stereo displays are evaluated, the cross talk issue should be addressed (McAllister, 2002).

<i>Active</i>		<i>Passive</i>	
Advantage	Disadvantage	Advantage	Disadvantage
. The display device does not have to polarize the light . Efficiency is higher	. The glasses must be synchronized to the refresh rate of the display device	. Permits multiple viewers . Permits larger FOV	. The display device must produce the polarized image . The screen must be coated with vapor deposited aluminum (silver screen) . Efficiency (transmission) is poor, images appear dark . Initial costs are higher but as it allows for multiple users, this is okay.

Table 2: Advantages and disadvantages of active and passive stereo viewing glasses summarized into a table from McAllister, 2002.

2.3.2. Head Mounted Displays

Also known as “goggle stereoscopes”, this technology was invented to place a human inside computer generated graphic simulations (Sutherland 1968 via Ottoson 2001).

Typically HMDs contain two adjacent Liquid Crystal Display (LCD) screens, which project the correct viewpoint exclusively, and unlike TMDs, concurrently, to the appropriate eye. *Window violation* does not occur in HMDs (Linde, 2003).

Window violation is the fact that the depth illusion breaks down at the edges of the screen if the displayed objects are occluded by the screen boundaries when using regular stereoscopic displays in a fixed position.

2.3.3. Autostereoscopic Displays

Recently adopted by mobile devices such as laptop computers and telephones, this is a most compelling type of stereoscopic display because it does not “require head mounted unit, shutter glasses, or other intrusive equipment to be worn by the user” and “they may be viewed by many users at the same time” (Sharp 2004, Sanghoon et al., 2005). Three types of autostereoscopic displays described in detail in (Halle, 1997 via Linde, 2003) are:

Re-imaging Displays are those using static lenses and mirrors.

Volumetric Displays fill a volume of space, using a revolving or oscillating mirror or flat screen with image content changing in correlation with position.

Parallax Displays have display surfaces capable of emitting light in different directions. Two groups of sites emit light, visible exclusively to each eye.

2.3.4. Retinal Projection Displays

Retinal Projection Displays project the image directly onto the retina of the user. An example of this is known as the Virtual Retinal Display (VRD) (Tidwell et al., 1995). They are reported to have a number of technical, practical and financial advantages compared to HMDs, it is suggested that they might supersede HMDs (Linde, 2003).

2.3.5. Crossed Eye and Parallel Viewing

A stereo pair of images can be viewed with bare eyes with some training. This technique is also referred to as “free viewing”.

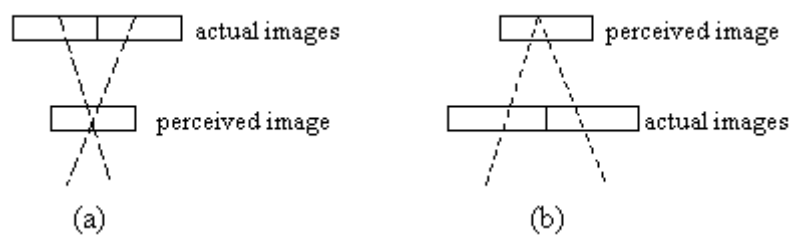


Figure 23: Cross-eye technique (a) versus parallel viewing technique (b) to free view a stereo pair.

In parallel (uncrossed) viewing the left eye image is to the left of the right eye image. In transverse or cross viewing, they are reversed and crossing the eyes to form an image in the center is required. Some people can do both types of viewing, some only one, some neither. (McAllister, 2002)

2.3.6. Anaglyphs

An anaglyph is a method of viewing stereoscopic images using colored spectacles as illustrated in Figure 24. Louis Ducos du Hauron patented the method in 1891, but W. Rollmann in 1853 and J.C. D’Almeida in 1858 had demonstrated similar methods previously. (Gemshein et al., 1969, via Dubois, 2001).

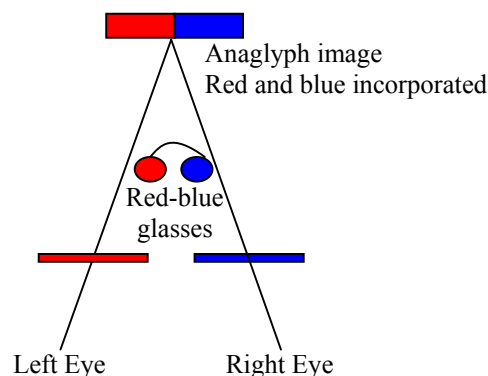


Figure 24: Anaglyph viewing. Figure adopted and redrawn from Yasayan, 1996.

The word anaglyph is a composed of the Greek words “again” and “sculpture”. In the classic method, used for monochrome stereo images, the left view in blue (or green) is

superimposed on the same image with the right view in red. When viewed through spectacles of corresponding colors but reversed, the three-dimensional effect is perceived (Dubois, 2001).

3D movies were often made for anaglyph viewing, which requires the user to wear glasses with red and green (or blue/cyan) lenses or filters. Both images are presented on a screen simultaneously; hence, it is a time-parallel method. Many observers suffered headaches and nausea when leaving the theater, which gave 3D, and stereo in particular, a bad reputation.

A phenomenon called ghosting or cross talk was a significant problem. Colors were not adjusted correctly and the filters did not completely eliminate the opposite-eye view, so the left eye saw not only its image but sometimes part of the right-eye image as well, and vice versa. Other problems included poor registration of the left and right eye images causing vertical parallax, and projectors being out of synch. (McAllister, 2002)

[...] The first anaglyph consisted of a stereogram made by printing a picture in red, and another picture in green was printed directly over it. This was viewed through a pair of “spectacles” containing one red and one green lens (McKay, 1953)

The colors chosen are complementary, “in additive color mixing, any pair of complementary colors will produce white, [...] in subtractive mixing, two complementaries will produce black” (McKay, 1953).

It would seem, therefore, that if a mixture, for example, of orange-red and blue-green will produce white, the separate stimulation of the two eyes by red and green simultaneously should produce a sensation of white through stereo fusion. If this is true, then a red picture and a green one presented separately to the two eyes should produce an achromatic image (neutral or white). If the red image is blanked out by a red filter and the green by a green filter it should produce the necessary stereo differentiation (McKay, 1953).

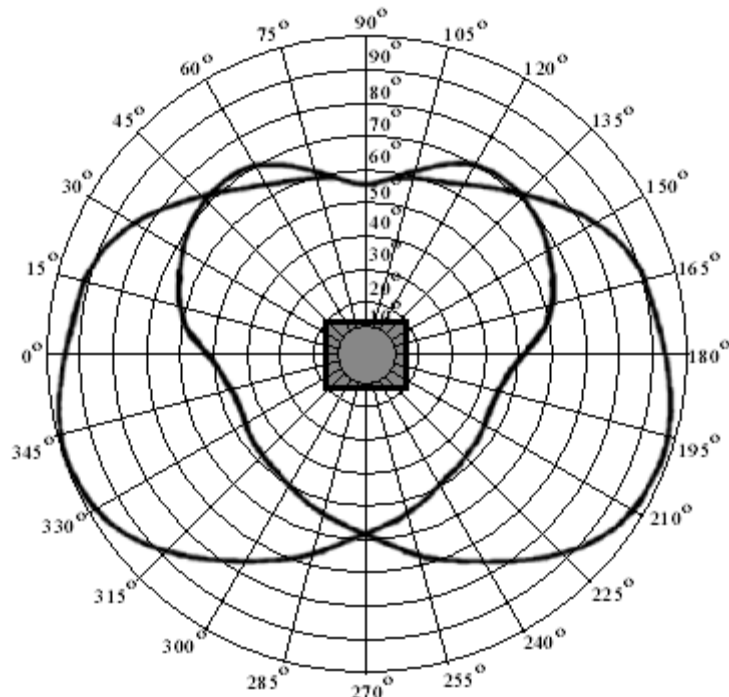
Based on this, red and blue make a better pair, because in the color scale, they are further away from each other than red and green.

In this thesis, foveation is intended for anaglyph glasses, even though technically it is possible to adapt the method/application to any of these display systems.

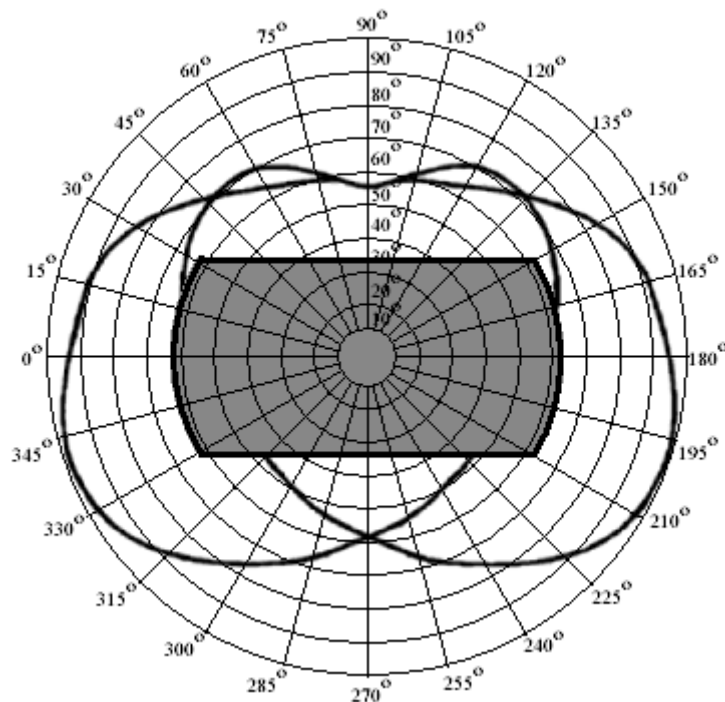
2.3.7. How Much Can We See on a Stereoscopic Display?

2.3.7.1. Panum's Fusional Area and Stereoscopic Displays

Objects within Panum's area result in small disparities, which are fusible. Objects outside Panum's area result in large disparities, which are not fusible, producing double images. Factors affecting the extent of the area include stimulus size, spatial frequency, eccentricity, and temporal modulation of disparity information (Patterson and Martin, 1992).



(a) FOV of 17" CRT at 20cm



(b) FOV of typical stereoscopic HMD

Figure 25: Human FOV and current display technology. The heavy black lines show the left and right eye FOV. The grey box in image (a) represents the relative size of a typical 17" Cathode Ray Tube monitor (CRT) viewed from 50cm. The grey box image (b) shows the FOV for a typical stereoscopic HMD. Reprinted from Pfautz, 2002, with permission.

The disparity limit for fusion increases as the stimulus size increases (e.g. large disparity can be fused with large stimuli), and decreases as the spatial frequency increases. The disparity limit increases with eccentricity (i.e. degrees away from the fovea). The fovea is the most light-sensitive area near the centre of the retina. This is the focal point of the retina, and vision is optimal in this part. The disparity limit also increases as the temporal frequencies of modulation decrease. These factors must be carefully manipulated in order to improve the binocular fusion when designing stereoscopic displays. (Ostnes et al., 2004)

Ware, citing Patterson and Martin (1992), reports that Panum's fusional area has remarkably little depth. At the fovea the maximum disparity before the fusion breaks down is 1/10 degree and in a 6-degree eccentricity the limit is 1/3 degree.

It is worthwhile to consider what these numbers imply for monitor-based stereo displays. A screen 30 pixels/cm, viewed at 57 cm, will have 30 pixels per degree of *visual angle*. The 1/10-degree limit on the visual angle before diplopia occurs translates into about three pixels of *screen disparity*. This means that we can only display three whole-pixel-depth steps before diplopia occur, either in front or behind the screen. It also means that in the worst case, it will only be possible to view a virtual image that extends in depth a fraction of a centimeter from the screen (assuming an object on the screen is fixated). However it is important to emphasize that this is a worst-case scenario. It is likely that anti-aliased images will allow better-than-pixel resolution [...] (Ware, 2000)

Ware relates the human visual system and a display in terms of size of visual acuity using a concept called brain pixels.

2.3.7.2. Brain Pixels and Optimal Display

Brain pixel is another expression for retinal ganglion cell receptive fields. The field size is reported as $0.006(e+1.0)$ (Drasdo, 1977, via Ware, 2004). The size of the smallest distinct characters are expressed with the function $0.046e$ (Anstis, 1974, via Ware, 2004) using an eye chart developed by Stuart Anstis. This chart shows the variation in visual acuity quite vividly (Ware, 2004) as can be see in Appendix 6. In both cases e is eccentricity from fovea measured in degrees of visual angle.

Ware analyzes the efficiency of displays by looking at how many brain pixels are stimulated as a display increases in size. Display efficiency gives the percentage of screen pixels that uniquely influence the visual system (Ware, 2004). This leads to a general conclusion that, as the display size or FOV get larger, a higher percent of the peripheral vision is in fact not perceived by the brain. He states that, "one way to increase the visual efficiency of a display is to have more than one resolution".

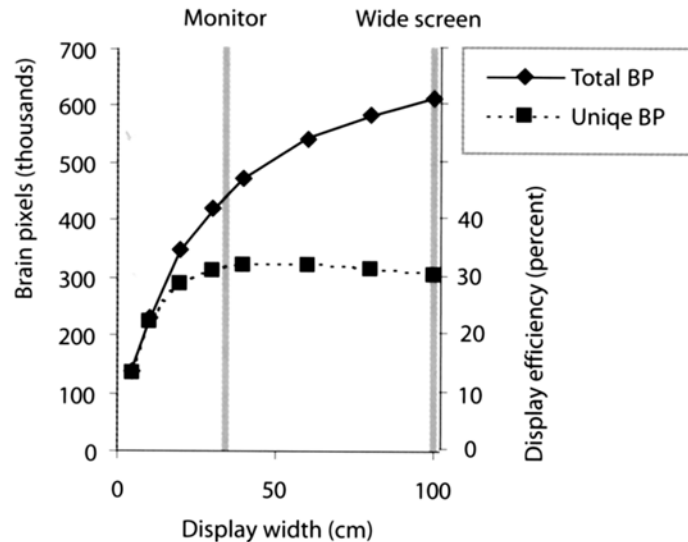


Figure 26: Colin Ware’s presentation of brain pixels in and their relationship to the display size. Reprinted from Ware, 2004 with permission.

There are two types of inefficiency that occur when we view flat displays. [...] At the fovea there are many brain pixels for each screen pixel. To have higher resolution screens would definitely help foveal vision. However, off to the side, the situation is reversed; there are many more screen pixels than brain pixels. We are, in a sense, wasting information, because the brain cannot appreciate the detail and we could easily get away with fewer pixels (Ware, 2004).

These findings and interpretations agree with our research motivation and approach.

2.3.8. Problems with Stereoscopic Displays and Suggested Solutions

It is common for the users of 3D visualization systems with stereoscopic display capabilities to disable stereo viewing once the novelty has worn off, and view the data using a monocular perspective. There are a number of reasons that stereoscopic displays are disliked (Ware, 2000).

These problems with screen-based stereo displays, according to Colin Ware are:

- First, if disparities are too large the result is seeing double (diplopia). The area in which the images can be fused is called Panum's fusional area and this is remarkably small in the worst case.
- A second problem is that objects more than 30 meters away¹¹ have images on the retina that are so similar the brain cannot obtain any useful disparity information.

¹¹ See Section 2.2.2.3 How Far Can We See Stereoscopically for a discussion on the various other numbers reported on this and why.

- A third problem is called vergence focus conflict - this has to do with the coupling of the focusing mechanism in the eye with the mechanism that makes the eyes converge when we see objects at different distances. (Ware, 2005)

2.3.8.1. Frame Cancellation

Ware also lists *frame cancellation*, along with vergence-focus conflict (accommodation convergence problem) and distant objects as problems with stereoscopic displays.

Frame cancellation is a case typical to smaller displays and negative parallaxes. The edge of the screen appears to occlude the virtual object and occlusion overrides the stereo depth information, and the depth effect collapses (Ware 2005).

The vergence-focus problem is the same as the accommodation convergence problem, which is analyzed in the following section.

2.3.8.2. Accommodation Convergence Conflict

In stereo displays, the fact that the eyes accommodate on the display surface but converge on the 3D point behind or in front of that surface creates a conflict. This is also known as vergence-focus conflict. The two processes occur in parallel in the case of human biology, though they are not hard wired to be together. This means we can process them separately, and we do in stereo viewing; unfortunately it results in a large number of people feeling uncomfortable.

In natural vision, accommodation and convergence are covariant but the correspondence of accommodation and convergence is not maintained by artificial stereo display (Linde, 2003). The failure to correctly present focus information coupled with vergence may cause a form of eyestrain (Wann et al., 1995; Mon-Williams and Wann, 1998; via Ware, 2000).

In several sources, a suggested solution to this problem is to artificially simulate the depth of field. For instance Luebke et al. have reported, regarding accommodation convergence conflict, that Ohshima et al.'s depth of field LOD may have some merit in this regard, since we can fuse blurred images more readily than sharp ones (Luebke et al., 2003).

But “accommodation does not occur in stereo displays (because the screen is at a fixed distance), so the DOF effect is absent” (Linde, 2003). Ware also states that “unfortunately, in present day computer graphics systems, particularly those that allow for real-time interaction, depth of field¹² is never simulated” (Ware, 2000).

While this statement is pointing the right direction that there is little awareness in this field, photorealistic computer graphic systems, such as RenderMan implementations

¹² Ware uses the term “Depth of Focus” for what we call “Depth of Field”. See section X2.2.3 for an explanation.

(e.g. prman, or BMRT), so use the depth of field simulation (Ward, 2005), therefore the word “never” is perhaps too strong.

There have been few empirical studies to validate the theory that simulating depth of field could help the problem of accommodation convergence conflict.

Blohm et al. suggest that the known failure of stereoscopic displays to adequately represent accommodation and convergence may be largely eradicated by artificially blurring regions, that are not inside the DOF region (Blohm et al., 1997) according to their distance from the depth of the fixation point to simulate the thin lens effect, which is present in natural vision (Linde, 2003)

Based on another user study, it is suggested that, “it may be preferable, for comfort, to position stereoscopic images in front of the screen rather than behind it. Given that, in general, the accommodative response to a near target is less than the accommodation stimulus, the associated extra vergence-accommodation input should also act beneficially to improve retinal image clarity” (Howart, 1996).

[...] Allowing the user's sensed convergence angle to control a depth of field simulation for stereoscopically displayed objects could reduce viewing discomfort in many applications. This will be true for stereoscopic displays that are otherwise perfectly calibrated, in that the extreme depth of field used in most systems allows extreme horizontal disparities to be presented that result in diplopia. When the stereoscopic images of close-range objects are blurred, however, those same extreme horizontal disparities produce less eyestrain and actually appear quite natural (Martens et al., 1996).

2.3.8.2.1. Cyclopean Scale

Ware reports one approach to solving these problems associated with stereoscopic displays. The basic idea is to scale the whole scene about the midpoint between the viewer's two eyes, hence cyclopean, until the near point lies just behind the screen (Ware, 2004). See Section 2.2.2.4 for an explanation of the Cyclopean Eye concept.

Cyclopean scale brings far objects closer where stereo depth becomes available, reduces the vergence-focus problem and, since everything is moved to behind the frame, there is no frame cancellation effect (Ware, 2004).

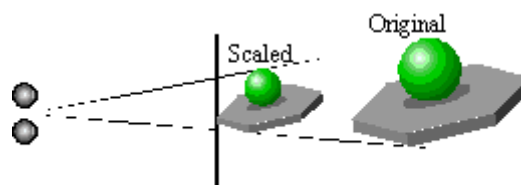


Figure 27: Ware's cyclopean scale illustration. Reprinted with permission from Ware, 2004.

The cyclopean scale method, although useful, does not remove all the possible effects that result in diplopia (Ware, 2004). At this point Ware suggests virtual eye separation

as a solution. Virtual eye separation works similarly as *hyperstereo* to control the range of disparities, but in both ways – therefore the term covers both *hyperstereo* and *hypostereo*.

2.3.8.2.2. Large Screens

The accommodation/convergence problem is exacerbated with small screens, only a foot or two –approximately 30 to 60cm- across, viewed at close distances (30-45cm to 1.5m). This is characteristic of electro-stereoscopic workstation displays. Large-screen displays, viewed at greater distances, may be perceived with less effort. There is evidence to indicate that the breakdown of accommodation and convergence is less severe in this case. (Inoue et al., 1990 via Stereographics 1997)

It is also important to note that for any foveation technique, an increase in display size significantly improves the compression attained (Geisler and Perry, 1998; Geisler and Perry, 1999), since the size of the foveal area remains constant, irrespective of the total angle subtended by the viewing screen. (Linde, 2003)

Their findings are greatly encouraging for our work since in this research we are simulating depth of field, which appears to have good potential for solving accommodation-conflict problems, and building the concept around large screen panoramic displays which should reduce the amount of eyestrain and increase the gains coming from foveation.

2.3.8.3. Diplopia and Its Possible Solutions

According to Ware, double-imaging problems tend to be much worse in stereoscopic computer displays than in normal viewing of the 3D environment. One of the principal reasons for this is that in the real world, objects farther away than the one being fixated are out of focus in the retina (Ware, 2000).

Similarly, Pastoor states that, in naturally viewed scenes, large stereoscopic disparities only exist where the retinal image is blurred (outside the DOF) - objects, which would otherwise produce diplopia, are thereby suppressed (Pastoor 1995). In stereoscopic displays, presenting sharp stimuli with large disparities causes pronounced diplopia and virtual simulator sickness (Linde, 2003).

Since we can fuse blurred images more easily than sharply focused images, this reduces diplopia problems in the real world. In addition, focus is linked to attention and foveal fixation. Double images of non-attended peripheral objects generally will not be noticed. Unfortunately, in present day computer graphics systems, particularly those that allow for real-time interaction, depth of focus¹³ is never simulated. All parts of the computer graphics image are therefore equally in focus, even though some parts of the image may have large disparities. Thus double images that occur in stereoscopic computer graphics displays are very obtrusive (Ware, 2000).

¹³ Colin Ware uses the term Depth of Focus where we use the term Depth of Field. See section 2.2.3 for an explanation.

Several other research papers make similar statements (e.g. Luebke et al., 2003; Linde, 2004; Mulder 2000; Blohm et al. 1997; Reddy, 1997; Martens et al., 1996). The fact that depth of field simulation is considered to be a solution to diplopia is a motivating factor for our research.

It is also suggested that *virtual eye separation* reduces diplopia and expands Panum's fusional area (Ware, 2004).

2.3.9. Summary

In this section, we have introduced stereoscopic viewing techniques, the relationship between what humans can perceive and what stereoscopic displays can present in terms of size, and the problems associated with stereoscopic viewing with suggested solutions. The major problems currently associated with stereoscopic displays and the solutions suggested by experts in the field indicate that our research may make a potentially valuable contribution in this area.

2.4. Level of Detail

First attributed to J.H. Clark (1976), LOD is a relatively old concept in computer years.

“Recognizing the redundancy of using many polygons to render an object covering only a few pixels, Clark described a hierarchical scene graph structure that incorporated not only LOD but other now common techniques such as view frustum culling” (Luebke et al., 2003, also see Clark, 1976).

Today, LOD is a simplification technique known and applied in 3D graphics extensively. In many complex CAD environments and virtual models, it is possible to meet several forms of LOD implemented to gain computational power.

LOD can be discrete, continuous, or view-dependent. In most vector graphics, if it is implemented, it can be seen as a form of mesh simplification. In raster graphics, the term LOD per se is not used very often, even though the concept exists. In this area, the forms of LOD are to be found as non-uniform image representations, such as foveation and depth of field simulation.

There are different techniques for selecting a specific level of detail. These techniques attempt to trade fidelity for performance by removing detail in an object when it becomes imperceptible. This gives the benefits of improvements in system responsiveness without the corresponding cost of detail loss (Constantinescu, 2001).

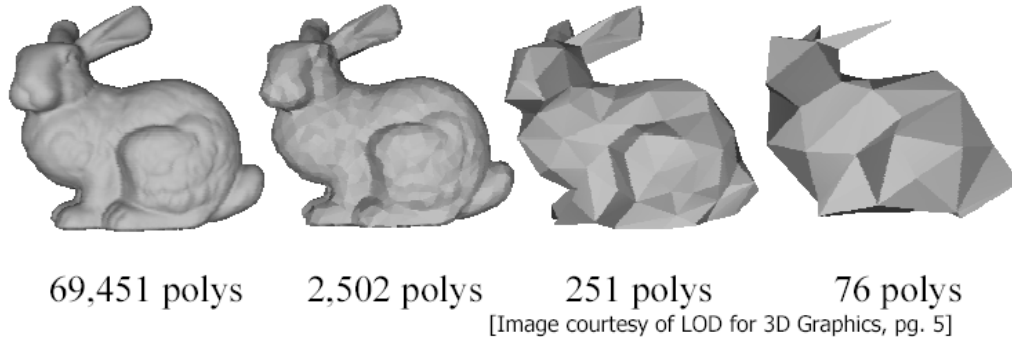


Figure 28: The famous “Stanford Bunny” demonstrates the polygon count and its primary effect. In Figure 30 and Figure 31 it will be shown that when the distance or size changes, the degradation in the resolution is not a problem. The bunny was first produced by Turk et al., 1994.

Two of the main categories of LOD selection are summarized below. This categorization is modified from Constantinescu, 2001, where 4 categories are mentioned separating the first item into two, and adding “another method with the main goal of maintaining a constant high frame rate, regardless of the complexity of the model” as the fourth item.

- Removing the details that do not need to or can not be rendered (i.e. different culling techniques)
- Removing details that can not be perceived by a human viewer (i.e. methods based on eccentricity, depth of field, velocity)

The work in this thesis is of the second category. Foveation employs the rules of stereo perception for a depth of field simulation. The following sections in this chapter however give a brief introduction to general level of detail methods and concepts. In Chapter 2.5 foveation techniques are explained in detail.

2.4.1. Culling

Within the LOD techniques, visibility culling (or view frustum culling) is one used historically, which is not considered to be perceptually motivated. In visibility culling the task is to determine the parts of the scene that are outside of the view frustum at the moment of viewing and to not represent the data in these areas at all.

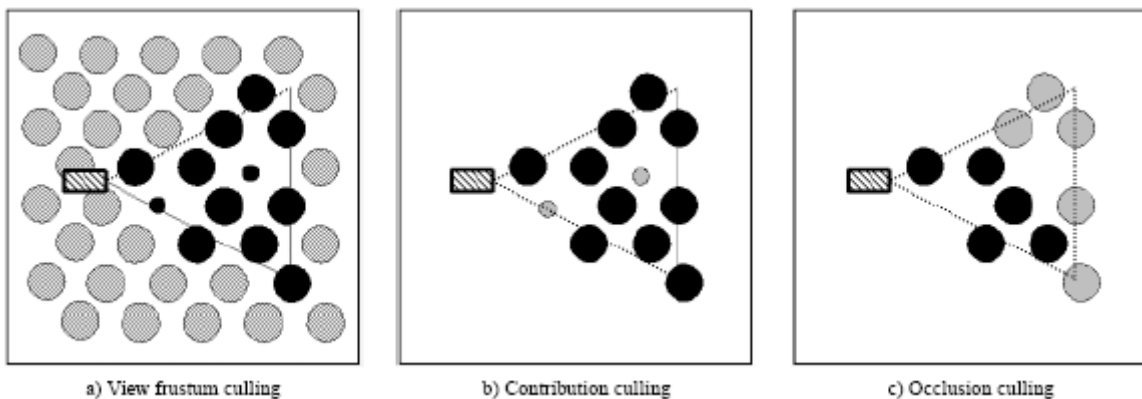


Figure 29: Culling techniques (Cabral, 1997). In a, the objects outside the view frustum will be discarded, in b a threshold defines the visual contribution of the object and even if it is visible, objects that do not meet the threshold is removed, in c, objects that are occluded by other objects therefore not visible will be found and processed accordingly. Figure reprinted from Cabral 1997.

View frustum culling eliminates primitives outside the field of view, and is recalculated frame by frame (Luebke et al., 2003). Occlusion culling, on the other hand, eliminates the objects that are occluded by other objects.

2.4.2. Perceptually Motivated LOD Techniques

The following are listed as LOD selection factors (compiled from Reddy, 1997; Constantinescu, 2001; Luebke et al., 2003):

- Distance
- Size
- Priority
- Hysteresis
- Environments Conditions
- Perceptual factors
 - o Eccentricity
 - o Velocity
 - o Depth of field

Size LOD and Distance LOD also rely on human perception, therefore can be, and sometimes are considered to be perceptual factors. However, they are the most common LOD applications and as such they often appear separately (Coltekin, ICC 2005). We will include a brief definition of the two and the listed perceptual factors. See the referred publications for further information on the others.

2.4.2.1. Distance LOD

The term level of detail is often synonymous with distance LOD (because this is the most prolific use of LOD). A relatively simple and commonly used method, distance LOD takes the Euclidean distance between the viewpoint and a predefined point inside the object into account. The theory behind this is that as the distance between the viewer and the object of interest grows, fewer details are visible. Hence it is possible to use less detail for objects that are at a greater distance than a defined threshold without affecting the fidelity (Reddy, 1997).

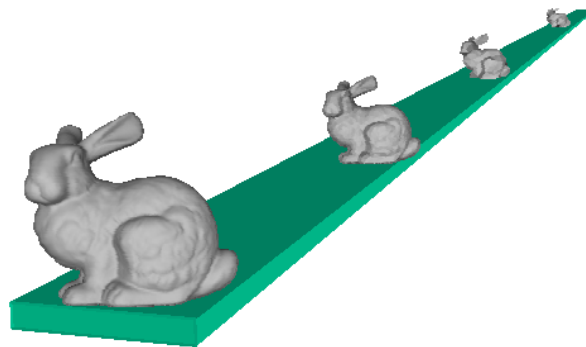


Figure 30: This figure demonstrates how the distance makes a difference in our perception. See Figure 28 for the polygon count in each of the presented bunnies.

2.4.2.2. Size LOD

This is a method similar to distance LOD in some ways, because objects that are further away are also perceptually smaller. But even if they share their Z value (equal distance to the viewer), smaller objects can be drawn with fewer polygons.

“Distance-based criteria measure the distance from viewpoint to object in world space. Alternatively, the system can use a screen space criterion. Since objects get smaller as they move further away, size-based LOD techniques use the projected screen coverage of an object, and switch between LOD based on a series of size thresholds rather than a series of distances.” (Luebke et al., 2003)

The method is computationally somewhat more expensive, though it also has advantages over distance LOD e.g. size LOD is invariant to the scale of objects and to the screen resolution (Luebke et al., 2003).

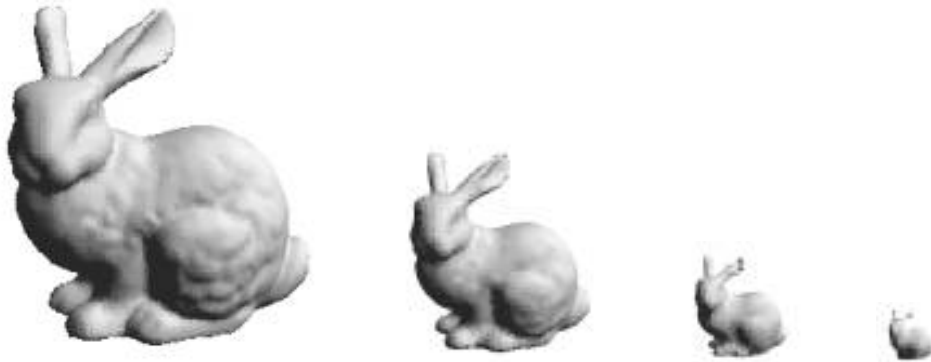


Figure 31: A smaller object does not need to be drawn with as many polygons as a bigger object. In this figure, the bunnies demonstrated in Figure 30 have been scaled down. This is to demonstrate that the effect of mesh simplification is not visible if it is proportional to the object size.

2.4.2.3. Priority LOD

Priority LOD is based on the context of the scene. While modeling, the programmer or modeler will take the context into account and decide certain objects to be of higher interest. LOD selection then is applied giving those objects a higher priority in terms of their resolution (LOD) “meaning that they will retain more detail for longer, or they may in fact never be displayed in low detail”¹⁴.

2.4.2.4. Hysteresis

“Hysteresis is simply a lag introduced into the LOD transitions so that objects switch to lower LOD slightly further away from the threshold distance, and switch to a higher LOD at a slightly closer distance” (Luebke et al., 2003).

2.4.2.5. Eccentricity LOD

This term corresponds to “2D foveation” in the world of computer (vector) graphics. In fact, the most common use of the word foveation is for 2D images; therefore it is possible to say that this term corresponds to foveation. Eccentricity LOD is defined as follows in the relevant literature:

An object’s LOD is based on its angular distance from the center of the user’s gaze, simplifying objects in the user’s peripheral vision more aggressively than objects under direct scrutiny (Luebke et al., 2003).

¹⁴ The sentence in quotation marks is suggested by Dr. Martin Reddy for clearing the concept a little further in terms of what priority LOD means for resolution.

Eccentricity LOD is when an object's representation is selected based upon the degree to which it exists in the visual periphery. Without a suitable eye tracking system, it is generally assumed that the user will be looking towards the centre of the display, and so objects are degraded in relation to their displacement from this point. (Reddy, 1997)

For mathematical models of eccentricity LOD, see Section 2.5.3.3.

The following chapter on foveation will present approaches from image processing and computer vision literature.

2.4.2.6. Velocity LOD

Velocity LOD is relevant in environments where motion is present. In this research the focus is on a stereo image pair where there are no moving objects, therefore only a brief definition is provided as follows:

An object's LOD is based on its velocity across the user's visual field, simplifying objects moving quickly across the user's gaze more aggressively than slow-moving or still objects. (Luebke et al., 2003)

2.4.2.7. Depth of field LOD

For this thesis, the Depth of Field LOD concept is as relevant as the Eccentricity LOD, because it implements a combination of the two.

Depth of Field LOD, as the name implies, is about reducing detail according to depth. Defining the focused volume of interest gives an opportunity to reduce the detail outside this area. As described in Section 2.2.2, human binocular vision is possible within the Panum's fusional area. Outside this area, the fusion of the two images perceived by left and right eyes is not effectively possible. A review of depth of field simulation is presented in Sections 2.5.4 and 2.5.6.

2.4.3. Summary

In this section we introduced several LOD concepts, mostly extracted from computer graphics literature. The concepts are adaptable as principles of 3D perception and the representations hold true both in 3D vector models and in stereoscopic 3D.

There are also some fundamental differences as human-made models offer more flexibility compared to photographs of natural objects or environments. These differences are studied as they become relevant throughout the text and in the discussion (Chapter 5) at the end, where the potential of foveation for maps and GIS is reviewed.

2.5. Foveation

In general, the term LOD is used by the computer graphics community, whereas practitioners in computer vision and image processing use the term foveation. Even though the link between the two is obvious, the literature on these topics does not extensively refer to one another. We view foveation as an LOD management technique.

The main purpose of foveation is providing compression to aid performance in the storage, computation and transfer of large visual datasets. These visual datasets can be images, videos, or 3D models. Other than providing compression, foveation is perceived as a smart LOD management system for such tasks.

The smartness of it lies in the fact that it conforms to the human visual system's principles (Coltekin, ICC 2005), which is reported to have an ameliorating effect on the otherwise uncomfortable side effects of stereo viewing as explained in the previous chapters (Sections 2.1, 2.2 and 2.3).

It is also a process that allows the viewer to take advantage of the full resolution in the area of interest, and for this thesis, more precisely the volume of interest. Having full resolution in the volume of interest means being able to see maximum available detail for the space defined. With regular compression methods where the perception is not modeled, the full detail is never available to the user.

Photogrammetrists would rather not use any compression at all, but this is simply not realistic for most cases considering the large amounts of data that needs to be handled.

In this chapter, we are going to give an overview of the foveation techniques described in the literature, covering both the 2D and 3D approaches.

2.5.1. What is Foveation?

Foveation is a biologically inspired computer vision method. It takes advantage of the finding that animal visual systems have solved the problem of limited resources by allocating more processing power to central than peripheral vision (Tan et al., 2003).

The term comes from the word **fovea**, the part in the eye that controls human spatial vision (Section 2.1 covers the literature review on of human visual system). It is a technique applied mostly in image processing and in robot eyes. Foveation reduces the level of detail gradually; working from a central area outwards in line with how the human eye processes perceived detail, therefore foveation is a compression method.

Foveation is also a *space variant* level of detail control system. *Space variant* means that the resolution of the image or the model varies throughout the spatial domain. This happens according to a pattern or a mathematical model. In fact the term space variant expresses our 3D level of detail control in this thesis quite well.



Figure 32: An illustration of foveation to demonstrate its principle idea. 6 levels of detail are visible.

The human eye is reported to have 20° of central foveal region. The periphery is divided into near and middle areas, which extends 30° out and the far periphery is around 100° (35° nasal and 56° temporal) (Min, 1994, referred via Linde, 2003). Linde calculates the highest acuity area as only $1/1000^{\text{th}}$ of the human field of view, which corresponds to the central 1.5° (Rao et al., 1997, referred via Linde, 2003). All of the foveal area corresponds to only 1.3% of the human field of view.

An approximation of how the human eye's sensitivity to detail varies across the retina can be found in Nakayama 1990 via Reddy, 1997:

“This reduction in visual acuity across the retina is significant, with around a 35-fold difference existing between the fovea and the periphery (Nakayama, 1990)” (Reddy, 1997).

Foveated images have been exploited in computer vision, especially in the context of active vision. But they are also useful in visualization, although this aspect is less well explored (Chang et al., 1997a).

2.5.2. Active Vision

Foveation is a popular concept in active vision applications. Active vision is a term coined for a process where the camera optics and configuration are actively controlled in order to simplify the remaining tasks in computer and robot vision. Characteristics of an active vision system include continuous operation, real time processing, and control of this real time processing, e.g. by managing a region of interest.

The active vision paradigm was espoused by Ballard as a way to overcome the computational complexity of reconstructing a scene from a single image (Kortenkamp et al., 1998). However, the term active vision is also often associated with stereo real

time systems, although there may in fact be more than two cameras in some active vision systems. The basics might be considered as follows:

An active vision system is one that is able to interact with its environment by altering its viewpoint rather than passively observing it, and by operating on sequences of images rather than on a single frame. Moreover, since its cameras can move, the range of the visual scene is not restricted to that of the static view. Active Vision is close, in principle, to the biological systems that inspired it and so it seems intuitively acceptable that as a visual sensor (especially augmented with color) it is perfectly suited to human/robot interaction and autonomous robot navigation in human environments. (RSL, 2005)

The term should not be confused with active sensing image systems, which are ones that have their own source of radiation rather than relying on ambient levels (passive imaging systems). Active Vision refers not to sensing technology but to strategies for observation (Blake et al., (eds.) 1992).

Typical to photogrammetric and some computer vision tasks is the reconstruction of 3D coordinates in a camera coordinate system. According to Crowley this creates difficulties for active vision systems (Crowley, 2005):

Real time response requires limiting the amount of information processed. In vision this means that you cannot look everywhere at the same time. Applying this idea to stereo vision leads to a system in which reconstruction is limited to the region of a scene around a fixation point. Such fixation is achieved by controlling motors for vergence and focus.

Classic stereo correspondence reconstructs the scene in a reference frame based on the stereo cameras. Such an approach poses two major difficulties for active vision. Unfortunately, actively changing the vergence angle and focus modifies the camera parameters, making stereo reconstruction by classic techniques impossible. Active vision systems avoid reconstruction whenever possible. Indeed many visual control tasks such driving a car or grasping an object can be performed by serving directly from measurements made in the image (Espiau et al., 1992). When 3D reconstruction is necessary, active vision systems exploit geometric invariant relations to reconstruct the scene using its own "intrinsic" coordinates (Crowley et al., 1993).

In our implementation *Foveaglyph*, the stereo foveation can be applied with two options. One solution is applied when the camera parameters are known, hence the reconstruction problem is solved and another when they are unknown and the depth is estimated based on disparity information alone. This means the implementation is also eminently suitable for active vision systems.

2.5.3. Common Foveation Methods, Models and Examples

Foveation is encountered in two distinct mediums, applied in hardware, or firmware that is included in the hardware, such as cameras and robot heads, and in software to manage data handling and bandwidth.

The hardware or firmware applications are mostly in active vision tasks. In these systems foveation is done as the images are being acquired. For robot navigation and camera tasks alike, foveation offers faster operation, recognition and processing. The software category relates to visualization, where foveation is applied to a scene that needs to be presented to an audience that has varying interest groups (e.g. virtual reality).

VR scenes are visually very complex and achieving something as close to human vision as possible is relevant both in terms of perceived “realism” (e.g. by applying a 3D foveation to a scene the virtual world is supposed to look more like the real world where visual input is space variant because of the way the eyes work) and in terms of managing large datasets. Another area is perhaps live video transmission, where directing the attention of the audience to desired parts of the scene is important as is dealing with the limited availability of bandwidth.

In all cases, by sending a smaller area in the highest resolution, it is possible to achieve a higher frame rate. While it is quite clearly advantageous for visualization-only tasks, in ones where the scene needs to be reconstructed, this may be questionable. We argue that it is also beneficial for these cases, as in the “working area” we can afford to keep more detail in comparison to overall compression when it is not feasible to keep the raw image as is.

2.5.3.1. Log Polar Mapping and Foveation

Also known as logmap, log-polar mapping is a commonly used method in active foveation systems. Some researchers (e.g. Bernardino, 1999 and 2002, Chang, 1998, others) argue that this coordinate system has certain advantages as it far better matches the “retinal” organization of the eye than the regular Cartesian coordinate system. The following is from Chang et al., 1997:

The complex logmap is a model consistent with empirical data on the mapping from primate retina to the visual cortex (Schwartz et al., 1977). [...] Perhaps the most striking fact is that the data density in such images grows logarithmically with the diameter of the visual field (as opposed to quadratically in the case of uniform images). Such low density images have been exploited in applications such as video phones (Wallace et al., 1992).

The system, as the name implies, is a polar coordinate system based on logarithmic graduation. It is designed, to represent the photoreceptor density as it changes in the eye.

The log-polar transformation is a conformal mapping from the points on the Cartesian plane $x = (x, y)$ to points in the log-polar plane $z = (\xi, \eta)$ (Bernardino et al., 1999):

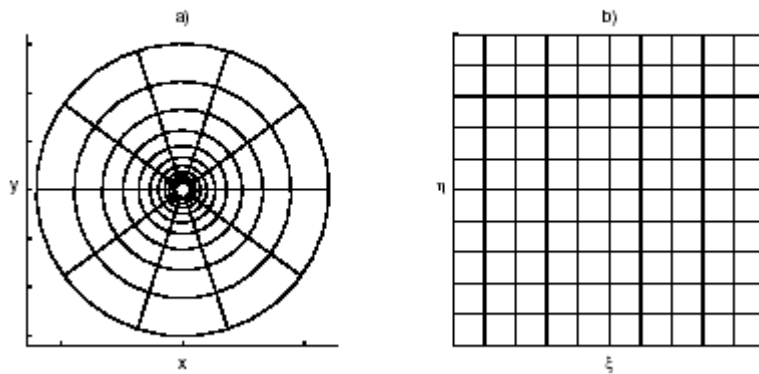


Figure 33: a) Cartesian Plane b) Log-polar plane. Reprinted with permission from Bernardino et al., 1999)

The mapping is described by:

$$\xi = \log \sqrt{x^2 + y^2} \quad \text{(Equation 2)}$$

$$\eta = \arctan\left(\frac{y}{x}\right) \quad \text{(Equation 3)}$$

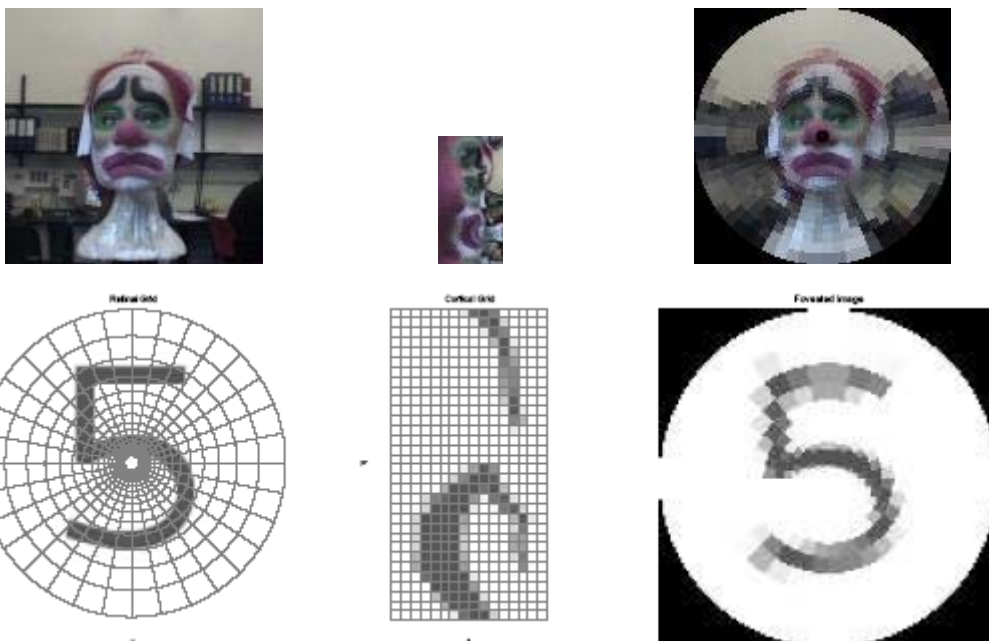


Figure 34: In this figures the Cartesian, log-polar and foveated (retinal) maps of the same picture can be seen, respectively. Images reprinted from Alexandre Bernardino with permission. See Bernardino et al., 2002.

“The main motivation for using a retina-like data reduction technique for robotic vision is that the resulting images are much smaller than the original camera image” (Bolduc et al., 1998). Talbot and Marshall confirm the hypothesis that log-polar mapping is a well-defined representation of visual field in the cortex in a paper written in 1941. A later

study by Schwartz in 1977 further agrees with it (Chang et al., 2000 and via Chang et al., 2000 also see Talbot and Marshall, 1941; Schwartz, 1977).

A number of research papers have reported successful utilization of log-polar sampling for tasks relating to robot vision (Kuniyoshi et al., 1995; Bernardino et al., 2002; Peters et al., 1996; Yamamoto et al., 1996; Panerai et al., 2000).

It is stated that “for engineering purposes it is not necessary to adhere to the logmap” (Basu et al., 1993 via Chang et al., 2000). We have encountered the logmap approach in robot vision applications while the visualization literature does not show this method being used extensively.

2.5.3.2. Foveation Techniques

Common approaches to achieving foveation include foveated lenses (see Rougeaux et al., 1996, via Boyling, 2000), dedicated hardware (see Bolduc et al., 1997, via Boyling, 2000) and software based image resampling (Boyling et al., 2000).

Software based image foveation techniques may introduce blur to the image surface as a function of distance from the fixation coordinate derived from an eye-tracker or other pointing device (Linde, 2004). Some standard methods to achieve this include the following (adopted from Linde, 2004):

- . The application of a low-pass convolution mask that is variable in scale depending on distance from the fixation coordinates,
- . Application of a bank of filters with different cut-off frequencies (Lee et al., 1998), by the multiple application of a fixed size convolution mask,
- . More commonly by the generation and subsequent pixel selection from a low-pass image pyramid (Perry et al., 2000).

Increasing the pixel size gradually away from the point of interest towards the periphery is an early and common approach as well. This approach results in aliasing and blocking effects in the low-resolution areas. Low-pass filtering solves this problem (UTEXAS, 2005; Perry et al., 2002).

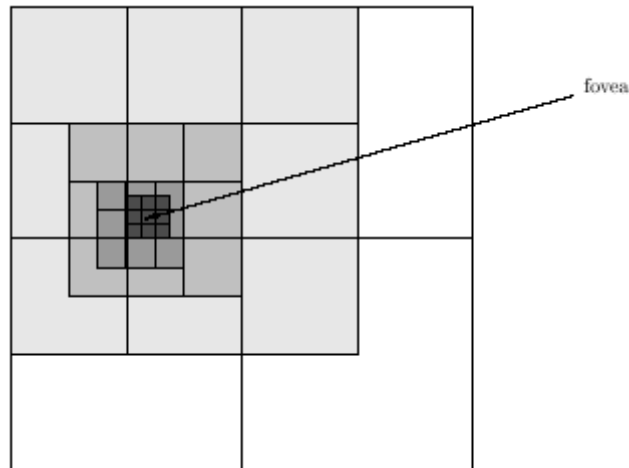


Figure 35: Shows the gradually changing pixel size throughout the image. This is similar to foveal resolution change. In fovea, the number of receptors decreases as the distance from the fovea increases. Figure reprinted from Chang et al., 1997a with permission.

Instead of building a discrete foveation pyramid, it is possible to use continuous foveation. Continuous foveation applies a different low-pass filter at each distance from the gaze point. Perry et al. report that it brings minimal improvement in the foveated image quality and it is computationally complex and intensive. This makes continuous foveation unsuitable for real time applications with current methods. Perry et al. declare that the foveation pyramid method is simple and fast, “Overall the foveation pyramid method is the best method available.” (UTEXAS, 2005; Perry et al., 2002)

Use of super-sampling, Gaussian filters and wavelet compression techniques are common as methods used to “blur” or down-sample the image when creating the foveated image pyramid.

Using the empirical model for the normalized maximum detectable frequency, f_c from the Geisler, 1998, Sheikh et al., 2001 report an ideal model for 2D foveation is as follows:

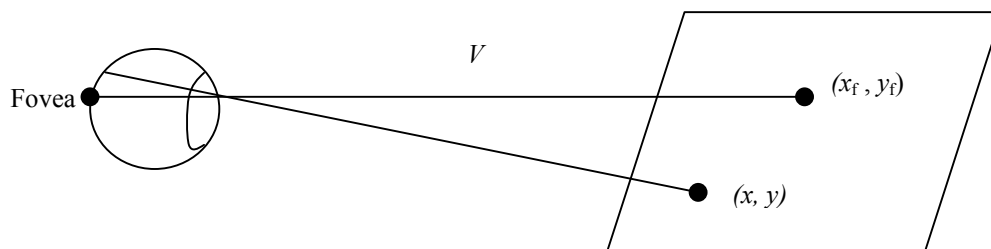


Figure 36: The relationship between the fixation point, which is also assumed to be the point of interest (POI), the distance to the point and the eye. Figure redrawn from Sheikh et al., 2001.

$$f_c(x, y, x_f, y_f, V) = \frac{1}{1 + K \tan^{-1} \left[\frac{(x - x_f)^2 + (y - y_f)^2}{V} \right]} \quad \text{(Equation 4)}$$

Here (x_f, y_f) are the coordinates of the fixation point, V is the viewing distance from the image (see Figure 36), and $K = 13.75$. All distance and coordinate measurements are normalized to the physical dimensions of the pixels on a viewing screen. Thus the ideal foveation of an image would consist of locally bandlimiting the image at coordinates (x, y) to $f_c(x, y)$. (Sheikh et al., 2001)

They further comment on the model saying: “the computational complexity of ideal foveation is enormous. For practical implementations for video coding, faster alternatives must be considered”.

Based on this model, Sheikh et al. eventually suggest another model taking the contrast sensitivity into account (Sheikh et al., 2003).

2.5.3.3. Eccentricity LOD

As summarized in Reddy, 1997 and Luebke et al., 2003, two models of eccentricity LOD documented in the literature are as follows:

$$Interest = \gamma_{static} / distance \quad \text{(Hitchner et al., 1993)} \quad \text{(Equation 5)}$$

This formula was used by Hitchner and McGreevy and was developed and tested for the NASA Ames Virtual Planetary Exploration (VPE). *Interest* represents a measure of importance of the object to the user and *distance* is the user’s gaze measured in 2D screen coordinates, while γ_{static} is an arbitrary scaling factor.

An alternative model by Ohshima et al., 1996 is as follows:

$$f(\theta) = \begin{cases} 1 & (0 \leq \theta \leq \alpha) \\ \exp\left(-\frac{\theta - \alpha}{c_1}\right) & (\alpha < \theta) \end{cases} \quad \text{(Equation 6)}$$

Designed for a head-tracked desktop system, this model represents the decline of visual acuity with eccentricity using the exponential relationship. θ is the angular distance between the center of the object to the user’s gaze fixation, α is the angle from the center of the object to the edge nearest the user’s gaze, and c_1 is an arbitrary scaling factor (assigned 6.2 degrees by the authors).

Reddy develops an equation for threshold spatial frequency taking velocity and eccentricity into account as a model of visual acuity (see Sections 2.4.2.5. and 2.4.2.6

explanations of Eccentricity LOD and Velocity LOD) (Reddy, 1997; Reddy, 1998; Luebke et al., 2003):

H represents spatial frequency (c/deg), v represents angular velocity (deg/s) and E represents eccentricity (deg):

$$H(v, E) = G(v) \times M(E) \text{ c/deg} \quad \text{(Equation 7)}$$

Where;

$$G(v) = \begin{cases} 60.0, & \text{when } v \leq 0.825 \\ -27.78 \log_{10}(v) + 57.69, & \text{when } v > 0.825 \end{cases} \quad \text{(Equation 8)}$$

$$M(E) = \begin{cases} 1.0, & \text{when } E \leq 5.79 \\ 7.49 / (0.3E + 1)^2, & \text{when } E > 5.79 \end{cases} \quad \text{(Equation 9)}$$

The author states this about the formula: “It is worth noting that we have now solved Ohshima et al. (1996)’s dilemma by showing that the product of the velocity and eccentricity scaling factors should be taken; $G(v)$ and $M(E)$ in our model, respectively.” (Reddy, 1998)

While this is presented as a model of human visual acuity for velocity and eccentricity, because “the angular resolution of a computer display limits the size of detail which users can experience” (Reddy, 1998) the model is further elaborated by taking the display into account as follows:

If a highest displayable spatial frequency is ξ , depending upon the output device’s field of view and pixel resolution:

$$\xi = \max\left(\frac{\text{horizPixels}}{2 \times \text{horizFOV}}, \frac{\text{vertPixels}}{2 \times \text{vertFOV}}\right) \text{ c/deg} \quad \text{(Equation 10)}$$

where *horizPixels* and *vertPixels* are the horizontal and vertical pixel resolutions of the display, and *horizFOV* and *vertFOV* are the horizontal and vertical angular resolutions of the display (in degrees). The final visual acuity model is given as:

$$\min(H(v;E); \xi) \quad \text{(Equation 11)}$$

All formulae and further explanation can be found in Reddy, 1997, Reddy, 1998, and Luebke et al., 2003. Reddy’s code implementing this model is called Percept, and it is explained in the later Section 2.5.5.

2.5.3.4. Depth Aware Foveation?

3D foveation for visualization and virtual reality tasks is not exploited as much as 2D. Determining the depth we can perceive, it would be a natural step to reduce the detail outside this area. This should bring further compression, a more natural visual result, and it is also said to help in problems associated to stereoscopic displays such as the eyestrain caused by diplopia in the regions where the parallax is too big and vergence/focus conflict (accommodation/convergence problem) (Linde, 2004; Brooker et al., 2001; Ware, 2000; Blohm et al., 1997) see Section 2.3.8.3 for an explanation of these problems.

“One of the most difficult problems with screen based VR systems is the lack of focus effects. The fixed focal distance of the virtual screen causes the vergence focus conflict that was discussed earlier. There is no effective technical solution to this problem at present.” (Ware, 2000)

The models of depth of field based on the Panum’s fusional area (as introduced in Section 2.2.2.2) are closely related to the mentioned “lack of focus effects”. One model we could refer to is from Ohshima et al., 1996:

$$h(\Delta\varphi) = \begin{cases} 1 & (0 \leq \Delta\varphi \leq b) \\ \exp\left(-\frac{\Delta\varphi - b}{c_3}\right) & (b < \Delta\varphi) \end{cases} \quad \text{(Equation 12)}$$

$\Delta\varphi = \varphi - \varphi_0$ where φ_0 is the angle of convergence for the fixation point, φ is the angle toward the object, b is the threshold width of the fusional area (assigned the value 0 degrees) and c_3 is a scaling parameter (assigned as 0.62 degrees) (Reddy, 1997; Luebke et al., 2003; Ohshima et al., 1996).

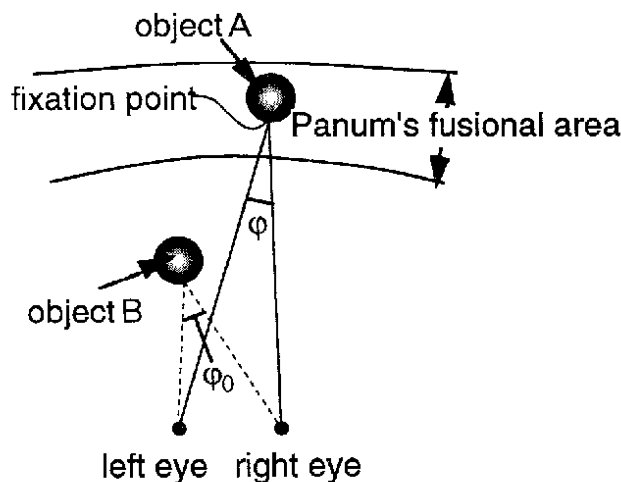


Figure 37: Panum’s fusional area and the positions of φ and φ_0 angles.

Image reprinted from Ohshima et al., 1996.

To read more about Panum’s fusional area, see Section 2.2.2.2.

This is one approach to modeling Panum's fusional area and with further consideration it can be integrated into a stereo foveation model as a measure of depth of field.

2.5.3.4.1. Compression of Stereoscopic Image Pairs

Stereoscopic image compression was generally approached so that one image was considered dominant and the other image was used in low resolution (Fok, 2002).

Two other methods are discussed in Perkins 1992: "disparity-compensated transform-domain predictive coding (DCTDP coding) and mixed-resolution coding. DCTDP coding seeks to minimize the mean-square error (MSE) between the original stereo pair and the compressed stereo pair. Mixed-resolution coding is a perceptually justified technique that is suitable when a human will view the compressed stereo pair. Mixed-resolution coding does not attempt to minimize the mean-square error between the original stereo pair and the compressed stereo pair."

Here the term "mixed resolution coding" should not be confused with space-variant image coding. It is used for describing what has become the most common approach in the stereo image coding process:

In situations where the end user of a compressed stereo pair is a human, it is possible to exploit the way in which the eyebrain processes a stereo pair to achieve compression. This is accomplished by introducing data-saving yet subjectively in situations where the end acceptable distortions. This subsection presents a perceptually justified technique for compressing stereo pairs called mixed-resolution coding. The compression is achieved by presenting one eye with a low-resolution picture and the other eye with a high-resolution picture; the eyebrain can easily fuse such stereo pairs and perceive depth in them. Furthermore, the final percept appears more similar to the high-resolution picture than the low-resolution picture. (Perkins, 1992).

Dinstein et al. (1988) found that using a low quality image for one eye has almost no loss in perceived quality or depth perception (Fok, 2002).

Foveating a stereo image pair has basically not been practiced. One single example is introduced next.

2.5.3.4.2. Focus/Foveation

The closest work in the literature to our work is by Ian van der Linde (Linde, 2004). Linde implements "focus/foveation" with eye tracking enabled HMDs in mind.

Linde implements a DOF simulation on top of a 2D foveation utilizing *image z-buffer* for a synthetic VE, requiring a binocular eye tracker to determine the screen-incident coordinate for each eye. He works with JPEG images and utilizes an image histogram of the z-buffer to determine the spatial distribution of objects within the z-buffer. He summarizes his model for segmenting the z plane as follows (Linde, 2004):

To illustrate the segmentation of the image in the z-plane, we could consider the histogram of the z-buffer. This histogram would show the distribution of objects across the available depth range. Assuming a fixation (i, j) on a single screen, we can query the z-buffer at this coordinate $Z(i, j)$ to determine the fixation depth Fz .

If we create a simple DOF region with only two levels (either inside or outside the focal plane), we could imagine this as a segmentation of image using the z-buffer histogram at two points equidistant from Fz by d depth units (therefore $2d$ gives size of the focused plane), shown in Fig. 2 as $Fz-d$ and $Fz+d$. In the figure, all pixels with corresponding z-buffer elements with value from $Fz-d \rightarrow Fz+d$ would be un-degraded, and all others blurred. To create gradual change, pixels may be blurred according to their absolute distance from Fz . (Linde, 2004)

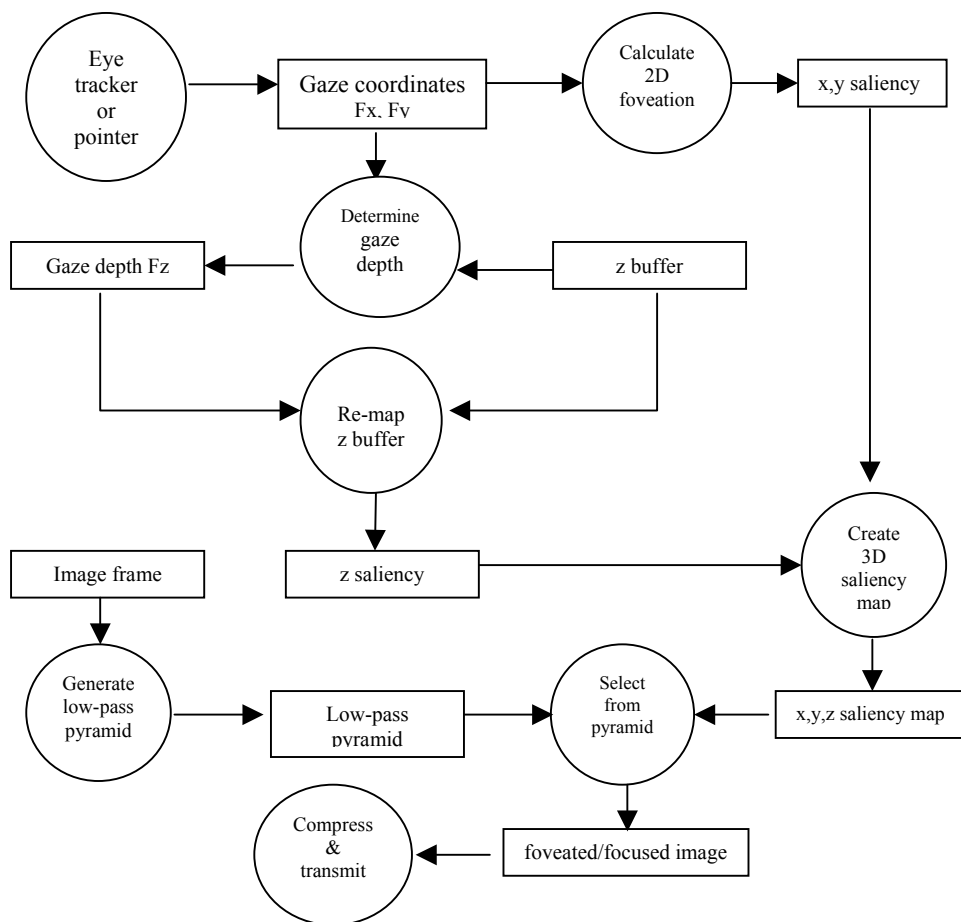


Figure 38: Foveation/Focus compression scheme. Reproduced from Linde, 2004 with permission.

Linde's work is conceptually very similar to our model. However, there are some important differences in the approaches. The major difference is that we are using stereo images instead of synthetic models and we calculate the Z value for each pixel. Other less important differences are that our work is generic while Linde's work is specifically optimized for HMDs and Linde's approach assumes binocular eye tracker while we utilize interactive selection of the point of interest.

There are two other works by Basu et al., 2002 and Schermann et al., 2000 dealing with Java and Java 3D based online 3D foveated visualizations.

Schermann et al., 2000 detail a Java application to transmit medical images on the World Wide Web (WWW). Their work handles both 2D and 3D images but 3D data is treated as a sequence of 2D images, where each 2D image has to be individually requested by the system (Schermann et al., 2000).

Basu et al.'s work is also meant for 3D Web visualizations, particularly textured 3D environments. What they do is “mapping 3D point of interest to 2D texture image, compute fovea and pass to server”. A 3D image is obtained by mapping 2D *texels* (Basu et al., 2002). The foveation is applied to the 2D textures, in a 3D setting.

2.5.4. Are Depth of Field Simulation and 3D Foveation the Same Thing?

Even though the word depth implies stereo, as discussed earlier in Section 2.2.3, depth of field is a monocular concept that depends on accommodation alone. While DOF is monocular, stereoscopic vision provides the ability to see and appreciate the DOF through the perception of parallax (Estes et al., 2005).

However, when referred to as “depth of field simulation” (an expression used in several papers including Ware, 2000; Mulder et al., 2000; Krivanek et al., 2003; Luebke et al., 2003; Linde, 2003 and Linde, 2004) in the context of stereo displays or 3D vector models, it conceptually expresses the same idea as 3D foveation.

2.5.5. Examples of Foveation and Depth of Field Rendering

Several research groups and individual researchers have successfully implemented foveation. The most common approach is to create a foveation pyramid, by scaling the image down in “steps” using a preferred downsampling method and then forming the foveated image by composing a space variant image. Some of these implementations are briefly introduced here.

Percept: An application and its source code in C for a level of detail implementation taking eccentricity into account is published by Reddy under the General Public License (GPL) (Reddy, 1997). The application is called Percept. It is a 2D foveation application. The author uses the term “Eccentricity LOD”. There is no difference between the two concepts, except perhaps the convention – which is that the computer graphics communities dealing with vector graphics talk about types of LOD while the image processing communities have developed it as an area of interest management technique for images and videos and use the term foveation.

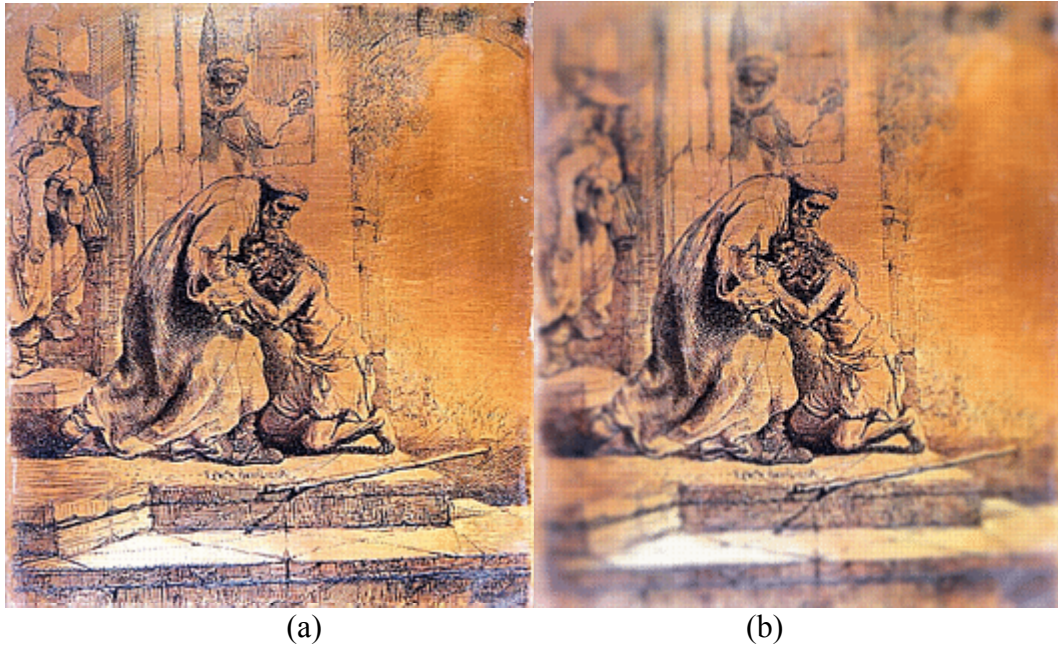


Figure 39: Visual results from Percept. (a) original image of Rembrandt's "Return of the Prodigal Son" (1636). (b) effect of running percept on this image assuming a field of view of 120 x 135 deg and a constant velocity of 50 deg/s. Reprinted from (Reddy, 1997) with permission.

Foveator: An application called “Foveator” that performs 2D image foveation is also distributed freely by UTEXAS 2002 (relevant publications are Kortum et al., 1996; Geisler and Perry 1999; Perry et al., 2002). Foveator allows some options to be configured by the user. It runs fast and the results are visually smooth. It is compiled only for Microsoft Windows and the source code is not published. Some visual results from Foveator can be seen below:

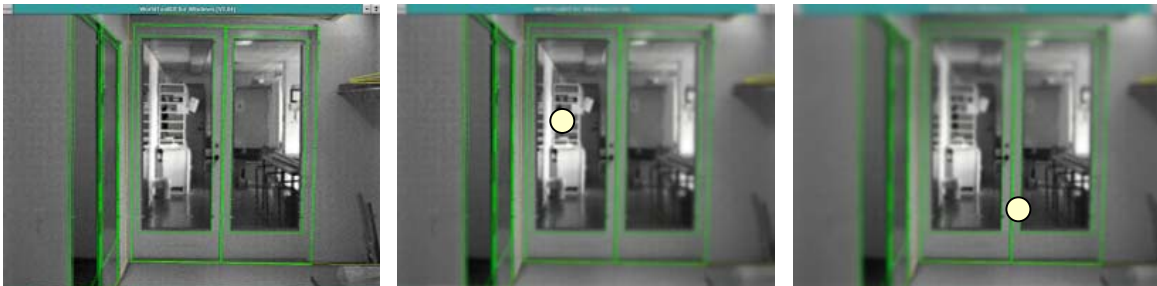


Figure 40: Visual results produced using the Foveator from UTEXAS. The first image is the original. The other two images are foveated and approximate point of interest is marked with a circle.

NYU Demo: Another example where a demo is available is the work of the Active Visualization¹⁵ group at New York University (NYU). The related work is published in Chang et al., 1997a; Chang et al., 1998 and Chang et al., 2000. Their method is based on wavelets and the work is intended for **thinwire** (low bandwidth network) Visualization. A Java based demo of their work is available on the group’s website, which is included in the citation of Chang et al., 1997a.

¹⁵ This term seems to be made by deriving from two other terms, which should not confuse the reader: The terms “active vision” and “visualization” are often used for two separate vision tasks, as described earlier. Bringing the two terms together, they label the multi-resolution space variant visualization as active visualization.

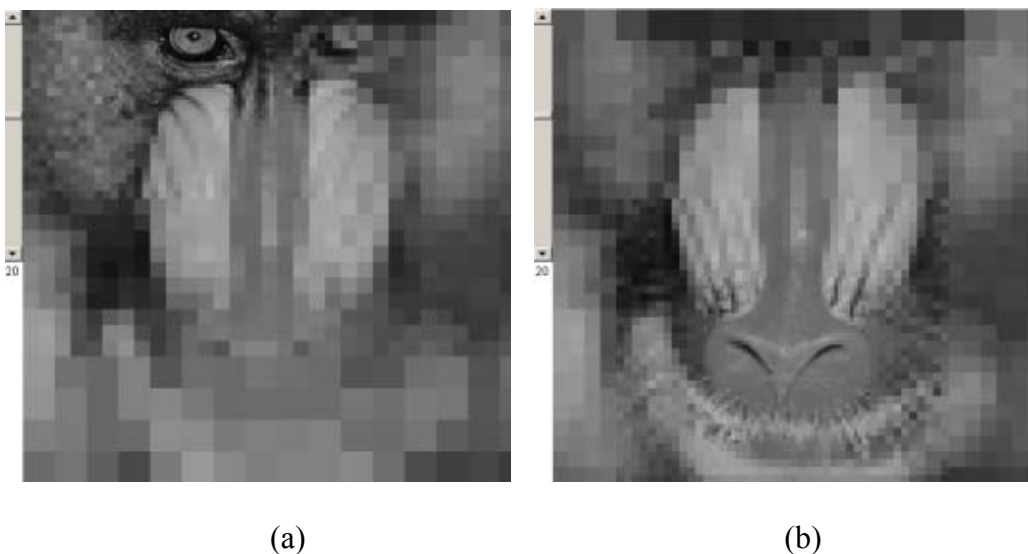


Figure 41: Mandrill image foveated using Chang et al.'s online Java based foveation demo. In the first image the POI is the left eye and in the second image is it the tip of the nose.

For Figure 41, the default radius for the smallest circle is set to 20, based on this transmitted data for (a) was 10404 pixels and for (b) 13783 pixels. When the circle is set for its highest possible value –which is 49-, (a) will become 37080 and (b) 44400. The demo page states that the change occurs roughly quadratically with the radius. The transformation in the demo is done using Haar wavelets.

NYU's Active Visualization group also presented research on foveation and online geovisualization (for Geographic Information Systems). See Section 5.5.1 for a discussion on geovisualization.

2.5.6. DOF Simulations

Categorically there are two kinds of depth of field rendering; post-process filtering (post-filtering) methods or multi-pass rendering (multi-pass algorithms) methods (Mulder et al., 2000, Krivanek et al., 2003). Details of these methods and further references can be found in Mulder et al., 2000 and Krivanek et al., 2003.

Mulder et al., presented an implementation of depth of field simulation in 2000:

The algorithm described here is based on two techniques: a high resolution and accurate technique for the center of attention, and a low accuracy high speed approximation for the remaining part of the scene. (Mulder et al., 2000)

They model the visual space according to a thin lens system (for a single view) and they calculate the *circle of confusion*. The center of attention is a truncated cone lying in the view frustum. Then they apply a technique that gives high resolution results in the center of attention and another that gives low resolution in the periphery. The work was not designed or tested for stereoscopic vision or stereoscopic displays particularly in mind.

Similarly, Krivanek et al. also presented a depth of field rendering.

The basic idea of our algorithm is to blur the individual splats by convolving them with a Gaussian low-pass filter instead of blurring the image itself. It essentially means that each splat is enlarged proportionally to the amount of depth blur appertaining to its depth. To accelerate DOF rendering, we use coarser LOD for blurred surfaces (Krivanek et al., 2003).

Piranda et al.'s work is also related to this as they take the optical depth of field definition into account and continue with calculating the circle of confusion.

Our method consists in placing a set of blurred images before and behind the focus plane. The areas that may cause eyestrain, because of the shifting between two stereoscopic images, can become blurred and as a consequence reduce their interest for the user. (Piranda et al., 2005)

The closest to our study is the foveation research for stereoscopic displays presented by Ian van der Linde (Linde, 2003; Linde, 2004), this is introduced in the Section 2.5.3.4.

2.5.7. Foveation and Photogrammetry

Even though a number of photogrammetric operations can be performed monoscopically, an essential component of digital photogrammetric workstations (DPWs) is the stereoscopic viewing system (Schenk, 1999). 2D foveation is relevant to visualizing monoscopic results, and all systems using binocular/stereo vision to reconstruct the scene in 3D are essentially sharing a number of tasks with photogrammetry. Therefore, foveation applied in such systems is also directly relevant to photogrammetry. There are only a few research papers connecting these two fields.

Boyling et al. have reported a “fast foveated stereo matcher” and a “foveated vision for space variant scene reconstruction” (Boyling et al., 2000 and Boyling et al., 2004) in which they use photogrammetric concepts, however both are meant for active vision systems instead of visualization.

Klarquist et al., 1998 present an approach where they note that “an active foveated image sampling and processing strategy is shown to greatly simplify the problem of establishing correspondence” (Klarquist et al., 1998). By adjusting the camera geometry in an active vision system and integrating foveation in their system, they create a multi-resolution depth map based on a vergent active stereo system. Their work does not mention photogrammetry.

Similarly Boyling et al., utilize an active foveation system to do stereo matching. They “[...] describe the framework and implementation of a foveated stereo-matcher. The application of foveation to a multi-resolution matching algorithm allows data reduction without drastically affecting the quality of the disparity map produced” (Boyling et al., 2000).

In a later work they report an active vision system that integrates photogrammetry:

“As part of a research project exploring space-variant approaches to computer vision, we wished to construct an active vision system capable of building a 3D model of a subject or its environment, based upon multiple observations captured from a binocular robot head, and using photogrammetry to recover 3D measurements” (Boyling et al., 2004).

While these are for active vision systems rather than visualization, there are not many other research papers considering photogrammetry and foveation in the same context. For photogrammetric visualization tasks however, it is evident that foveation would provide acceleration.

2.5.8. User Studies on Foveation: *Perceptually Lossless?*

The compression methods in image processing are roughly classified as *lossy* and *lossless*. Foveation is a lossy compression method, but it is considered as being a *perceptually lossless* method. This is to say that once the area of interest (or volume of interest in case of 3D) is determined, this region may be left with no information loss and the foveated image does not appear differently to a human observer.

There are studies showing that users cannot detect the difference or detect little difference between an original image and a foveated image. In a successful real-time foveation application, this should indeed be the case.

After several experiments Watson et al. have concluded that:

“Results indicate that peripheral LOD degradation is a useful compromise. [...] Two medium resolutions, peripherally degraded displays did not significantly differ from the undegraded, fine resolution display. [...] The fact that these results were achieved without eye tracking is particularly interesting, and suggests that eye tracking may be of limited importance in HMDs when the low LOD periphery is not extremely large.” (Watson et al., 1996)

Watson et al., compare their earlier findings and add that, “the effectiveness of peripheral LOD degradation is highly task dependent.”

Kortum and Geisler have described psychological experiments in which subjects reported little perceptual difference between foveated and uniform images (Kortum et al., 1996).

Parkhurst and Niebur perform a test particularly with standard desktop systems in mind with two different types (velocity and “gaze contingent” by which they mean eccentricity) of level of detail. They are looking at the speed of the task performance.

[...] While these techniques have been previously examined in the context of high-performance rendering systems, it is not clear whether the benefits will necessarily overcome the behavioral costs associated with a reduced LOD on ordinary desktop systems. To answer this question, two perceptually adaptive rendering techniques, one velocity-dependent and one gaze-contingent, were implemented in the Unreal™ rendering engine on a standard desktop computer

and monitor. These techniques were evaluated in separate experiments where participants were required to perform a virtual search for a target object among distractor objects in a perceptually rendered virtual home interior using a mouse to rotate the viewport (Parkhurst et al., 2004).

As the two methods were evaluated separately, here we will include their findings about the eccentricity LOD, which they performed with a system including an eye tracker.

They found that the reaction times to detect the target increased in the periphery, whereas reaction times to localize a target decreased.

[...] Using a medium degree of LOD reduction resulted in a decrease of overall reaction time, i.e., detection plus localization time. These results indicate that perceptually adaptive LOD reduction techniques can be effectively used even on desktop systems.

[...] The LOD reduction may not significantly harm the perceptual quality of the display.

Overall, the behavioral costs associated with perceptually adaptive LOD techniques can be offset by the behavioral performance gains on desktop systems. However, we show that the nature of the task is important in determining the exact cost-benefit trade-off (Parkhurst et al., 2004).

Though researchers have endeavored to eliminate noticeable perceptual artifacts present with earlier foveation methods, little psycho-visual testing has been undertaken to measure the affect of foveation on complex interactive tasks (Linde, 2004).

Perceptually lossless compression is possible with foveated imaging. In general, the greater the resolution of the original image, the greater the compression factor that can be obtained with foveated imaging, while maintaining a perceptually lossless image. A 1024 x 768 image can typically be compressed by a factor of 3-5 without visible loss, when the foveation region is centered on the direction of gaze (UTEXAS, 2005).

Researchers agree that there is a need for further user studies. Nonetheless, current findings indicate that, globally, reduced LOD in peripheral areas does not hinder the use of models or images.

2.5.9. Summary

In this section, we have introduced foveation in detail with its formal definitions, techniques and examples. Related research and terminology on foveation, depth of field simulation, active vision and active visualization were reviewed. Also reviewed were user studies on how a foveated image or model is perceived by a human viewer.

2.6. Correspondence and Reconstruction

In order to be able to implement the foveation in 3D, we need to recover the depth information. Depending on the dataset, information regarding the 3rd dimension can be recovered using several methods. For synthetic VEs, depth data may be accessed from the image z-buffer, requiring a binocular eye tracker to determine the screen-incident coordinate for each eye (Linde, 2004).

For a stereo image pair, it is possible to study the convergence angles of the eyes using binocular eye tracking to determine the crossover point, and hence the depth of the object under fixation (Brooker et al., 2001). If the stereo pair is acquired in a parallel camera configuration (normal case of stereo) or the resulting stereo pair is transformed to the normal case, it is possible to recover the depth information by image matching techniques.

In our work we assume the image acquisition would be done with a parallel camera configuration or that the images would be already converted to parallel case. This is a common approach in photogrammetric applications. If base (inter-pupillary distance) and the camera's interior orientation is known, utilizing the known elements and geometric relationships it is possible to recover the 3D information in the "metric" camera coordinate system which can then be geo-referenced.

It is important to note that while we chose to capture the pictures in parallel camera configuration as a shortcut, this is not a requirement. Orienting and normalizing images that are captured in general case is a readily available alternative and in fact it may give better results because the ideal, 100% parallel configuration is nearly impossible to achieve.

This chapter will introduce the correspondence problem and will give detailed information on the reconstruction process that is typical to photogrammetry and is integrated into our implementation *Foveaglyph*. Even though most of what this section presents can be considered textbook information, we consider it worth including for potential readers of the thesis from different fields.

2.6.1. The Correspondence Problem (Image Matching)

The correspondence problem is defined by the following question: "Given an image point x in the first image, how does this constrain the position of the corresponding point x' in the second image? (Pollefeys, 2004)"

In photogrammetric literature the term "matching" is more often used than "correspondence", citing from Heipke "the matching problem is also referred to as the correspondence problem" (Heipke, 1996). The following paragraph from Heipke 1996 explains where in photogrammetry and related fields the correspondence problem needs to be solved:

In photogrammetry and remote sensing, image matching is employed for relative orientation, point transfer in aerial triangulation, scene registration and Digital Terrain Model (DTM) generation. Also, the reconstruction of the interior

orientation falls within the category of image matching, since the model of a fiducial is usually represented as a gray value image. Image matching (correspondence problem) is inherently an ill posed problem and additional assumptions and constraints have to be introduced to make it well posed (Heipke, 1996).

More information about the correspondence problem and the method employed in this thesis is given in Sections 3.2.2.5 and 3.2.2.6 while explaining the selected method.

2.6.2. The Reconstruction Problem

According to Owens (1997) the reconstruction problem is, given two images formed in the retinal planes P and P' , and the two corresponding points m and m' , computing the 3-D coordinates of M relative to some global reference frame.

The following section focuses on the reconstruction problem and in particular the normal case of stereo setup that is employed in this thesis.

2.6.2.1. Defining the Geometry: The camera model

The geometry of stereography includes the relationships between the camera position, the relative positions of image planes to one another and the position of objects' projection on the image plane. In some literature, the image plane might be referred to as the "retinal plane" taking the analogy from the human visual system.

The camera(s) can be positioned in three different ways when taking stereoscopic photographs: general case (arbitrary camera configuration), convergent (toed-in camera configuration), or normal case (parallel camera configuration or stereo configuration).

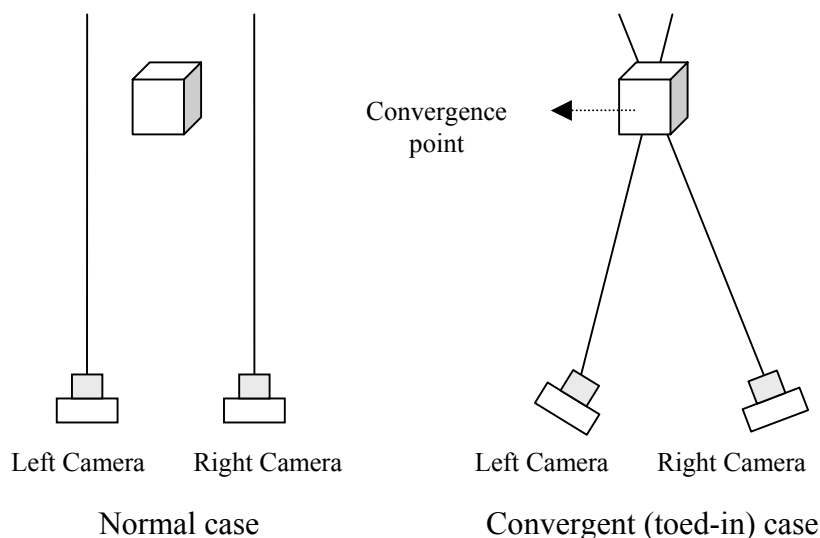


Figure 42: A simplified figure for comparison of normal case versus convergent (toed-in) camera configurations. The general case requires that there is overlap between the captured images, the camera(s) can be basically located anywhere as far as this condition is provided.

The general case is when the cameras see some overlapping area, but their own position and orientation are arbitrary. This is possibly the most common case and relatively more complex to solve.

In convergent (toed-in) systems the cameras are oriented such that their optical axes intersect at a point in space. The angle of the cameras defines a surface in space for which the disparity is zero. This zero disparity surface is often referred as “horopter” (Section 2.2.2.1 has more on horopter). Objects farther than this surface have disparity greater than zero, and objects that are closer, have disparity less than zero (Jain et al., 1995)

Because in many photogrammetric cases a transformation from the general or convergent case to the normal case (original to normalized) is sought after to make the work more convenient (see Schenk, 1999), the following section will introduce the normal case of stereography in detail.

2.6.2.2. Normal Case of Stereography

What is called **normal case** in photogrammetric literature is referred as **parallel camera configuration** in most computer vision and machine vision texts. Throughout the literature, it is also possible to see this case called canonical stereo configuration, ideal case, or stereo configuration.

This configuration refers to a setting where two identical cameras will have parallel optical axes and a coincident image plane. Their *epipolar geometry* here is the simplest.

Note that such a configuration is impossible to achieve physically. However, a stereo rectification algorithm can be used to warp the images to remove the effects of differing internal and external camera geometries. After rectification, the epipolar lines are parallel to the image rows and the epipoles are on the line at infinity (Vincent, 2005).

Several real-time stereo systems have been built around the parallel-camera configuration to minimize computational complexity (Schreer et al., 2001). The old stereo cameras were designed using this principle and "the parallel camera configuration is used in preference to the toed-in (converged) camera configuration" (Grinberg et al., 1994).

While the primary advantage is “the simple and direct formula in extracting depth, the primary problem associated with a stereo arrangement of parallel camera locations is the limited overlap between the fields of views of all the cameras. The percentage of overlap increases with depth” (Kang et al., 1994).

Here we represent object coordinates as x , y and z . The origin of a right-handed (x ; y ; z) coordinate system is located at the projection centre of the left camera. There is no vertical parallax.

$$P_y = y_l - y_r = 0 \quad \text{(Equation 13)}$$

This means that the corresponding features in the left and right images lie in the same horizontal scan line (Jokinen, 1994).

Notice that the image plane is considered to be located between the projection centers and the object. It is customary to assume the image plane is in front of the center of projection. (Jain et al., 1995).

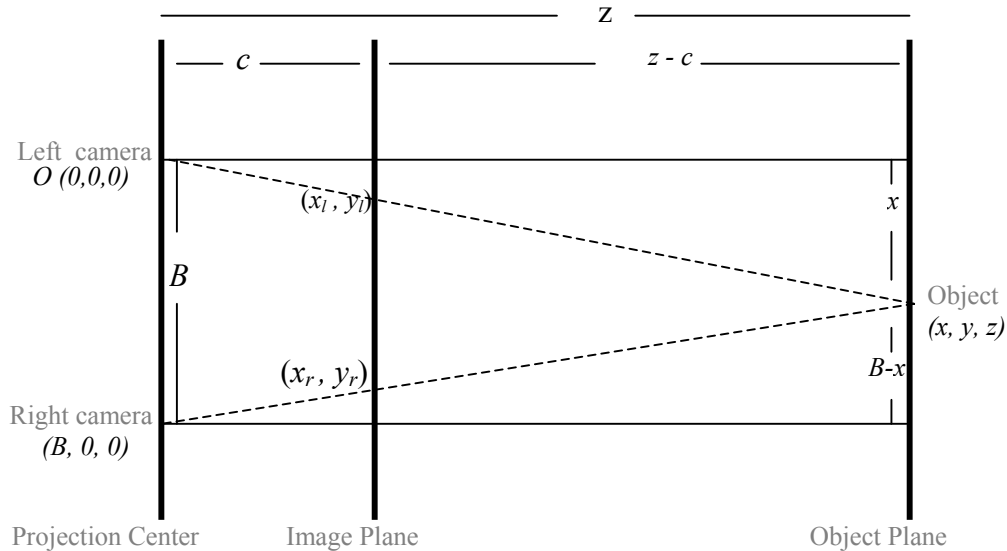


Figure 43: This figure shows a bird's eye cross-section view of xz plane for the normal case of stereo. Origin of the coordinate system is at the center of the left camera and is shown as $O(0, 0, 0)$. Note that image plane in an actual camera is at a distance c (value for calibrated focal length f) behind the center of projection and the projected image is inverted. It is customary to avoid this inversion by assuming that the image plane is in front of the center of projection (Jain et al., 1995). This illustration follows this convention.

Here solving the relationships between the similar triangles we get to the z coordinate of the object taking the parallax¹⁶ in use as follows:

- B : Base or inter-ocular distance between the two cameras,
- c : Camera constant, which is the focal length after calibration,
- P_x : Horizontal parallax

Under perspective projection a point (x, y, z) from the object surface projects onto the left and right images, respectively at: (adopted from Jokinen, 1994)

$$(x_l, y_l) = \left(-\frac{cx}{z}, -\frac{cy}{z} \right) \text{ and } (x_r, y_r) = \left(-\frac{c(x-B)}{z}, -\frac{cy}{z} \right) \quad \text{(Equation 14)}$$

¹⁶ Parallax and disparity in stereo vision related material, often expresses the same phenomena. There are some authors who differentiate the two by associating the word disparity to "retinal disparity" and the word parallax to "screen disparity", but this is not commonly practiced. In this thesis, as there are references to all kinds of literature, the two words are used interchangeably for retinal or screen disparities.

Horizontal disparity is defined as:

$$P_x = x_l - x_r = -\frac{cB}{z} \quad \text{(Equation 15)}$$

The origin of the coordinate system is located in the center of the left camera; object coordinates (x, y, z) are obtained using the left image coordinates by:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{B}{P_x} \begin{bmatrix} x_l \\ y_l \\ c \end{bmatrix} \quad \text{(Equation 16)}$$

$\frac{B}{P_x}$, which equals $\frac{z}{c}$, is commonly denoted as M^{17} . This is the scale factor.

2.6.2.2.1. The scale factor explained

Geometric illustration and a proof of how the scale factor works can be seen in 44 below and the explanation that follows.

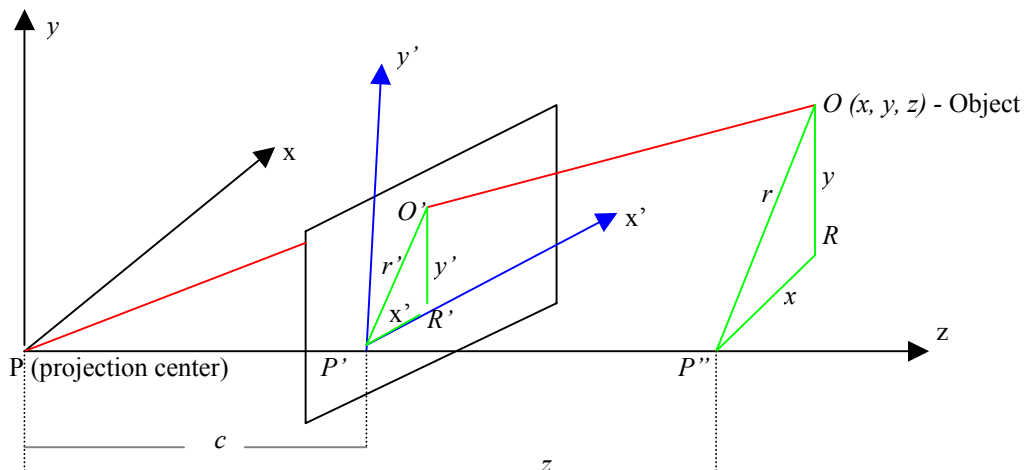


Figure 44: The perspective projection showing the line of sight to demonstrate the relationships between the object point and projected point. Adapted and redrawn from Jain et al., 1995.

The similarity between the two triangles $PO'P'$ and POP'' gives us the following:

$$\frac{c}{z} = \frac{r'}{r} \quad \text{(Equation 17)}$$

¹⁷ The notation M is possibly from the German word “Maßstab”, which means scale. In English it is sometimes said to be for “magnification”.

P'O'R' and P''OR are also similar, therefore:

$$\frac{x'}{x} = \frac{y'}{y} = \frac{r'}{r} \quad \text{(Equation 18)}$$

As a result of the common solution of these:

$$x = \frac{z}{c} x' \quad \text{and} \quad y = \frac{z}{c} y' \quad \text{(Equation 19)}$$

$\frac{z}{c}$ is then denoted as M.

2.6.2.3. Epipolar Geometry for Normal Case of Stereo

Epipolar geometry shows the fundamental relationship between the perspective cameras.

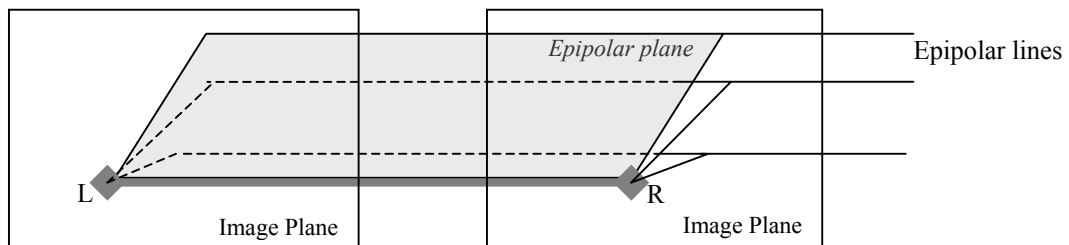


Figure 45: Notice the image planes are coplanar and epipolar lines are parallel. L and R mark the left and right viewpoints (e.g. camera positions) respectively.

Extracting the z value from a stereo pair relies on finding the parallax value. This value is calculated based on knowing the corresponding points (conjugate points) on a given pair. This task is called *image matching* and still challenges researchers with difficulties in achieving precision. The normal case provides another advantage at this point, as the epipolar lines are parallel to each other.

In many cases, pictures that are not taken with a parallel camera configuration are rectified (wrapped) to simplify the epipolar geometry (Pollefeys, 2004) to the normal case. The matching process takes advantage of the geometry of the normal case. After calculating the relative orientation, the positions of the epipolar lines are known. It is then called *stereo matching*.

The image matching method utilized in this thesis is explained in Sections 3.2.2.5 and 3.2.2.6.

2.6.3. Summary

This section has covered the background on an essential stereoscopic imaging concept: correspondence and reconstruction problem. For the reconstruction problem, the mathematical basics that were used in calculating the 3D coordinates in our implementation were provided.

CHAPTER 3. DEVELOPMENT AND IMPLEMENTATION

Action is eloquence.
--William Shakespeare (1564-1616)

The statement “if the human can not sense it, do not display it” (Robinett, 1999) points to a good principle for efficient data management for computer vision and visualization tasks as the presented literature review in the previous chapters also supported. For this thesis, a human visual system (HVS) aware LOD is the principal development. As foveation is such a technique, this research investigates the potential use of foveation for stereoscopic visualizations.

In seeking that, up to this point, we introduced concepts, approaches and methods based on HVS-aware LOD as a principle, discussed what the human eyes see and how stereoscopic displays work. We also looked at how other people have done level of detail management based on the human visual system, paying particular attention to foveation techniques.

By now, we are familiar with the underlying concepts and we know that this thesis claims foveation is a useful approach for managing the data for stereoscopic and other 3D visualizations.

In this chapter, we will present the proof of concept: an implementation of foveation for stereo imaging. It will present a model for stereo foveation and demonstrate how the concept works. The implementation is not optimized for or limited to a certain case – i.e. we did not have any particular visualization data in mind such as a medical, aerial, terrain or 3D cinema. It is readily applicable to any general stereo pair, easily extendible to any 3D vector data set and also suitable for photogrammetric use.

3.1. Development

3.1.1. The Dream Algorithm

If we had total medical comprehension of how the human eyes and the human brain work, an ideal foveation model would take all aspects of visual acuity into account: size/distance, eccentricity, velocity, depth of field. It would also gain additional compression by compressing one of the images more than the other, because this is not believed to have an effect on the stereo perception, as explained in Section 2.5.3.4.1. This model would readily apply to all cases such as 3D vector models, stereo motion pictures, animations or still stereo pairs.

3.1.2. Limitations for This Implementation

The 3D foveation conceptually is valid for all kinds of 3D applications, formats and cases. The prototype implementation however is done for a basic setting of a still stereo pair and a single user, which can be operated and evaluated on a standard computer and a standard display. It combines two of the most relevant visual acuity features for a still stereo pair, the eccentricity and the depth of field.

A still stereo pair and varying points of interest is a typical setting for photogrammetric visualization tasks. The viewer often would study the 3D scene for interpretation, selection of control points and eventually for vectorization.

The practical limitations such as the lack of easy access to a CAVE equipped with trackers at the time of the research also motivated a more simplified and platform-independent approach. The limits do not, however, imply any conceptual or theoretical restrictions, e.g. the extension of this work to a real-time visualization system equipped with trackers would only require solving some practical problems.

3.1.3. What is the Implementation for – Questions Before Coding

When developing a software implementation, it is important to ask the right questions. The answers to those questions then define the frame of the implementation and limits of the work. The first question is, why?

As mentioned in the opening paragraphs of this chapter, the implementation is done as a proof of concept. Let us summarize the concept first.

This research aims at demonstrating that 3D foveation is a novel, little exploited concept that may be useful for disciplines such as stereoscopic displays, virtual reality, computer vision, photogrammetry and geovisualization. While the previous chapters on the state of the art provided the interdisciplinary aspects and demonstrated that the topic was little exploited, its practical usefulness is two fold:

- 3D foveation should help the computer performance by providing compression.
- 3D foveation should help the viewer (human) performance by simulating the depth of field.

Within the frame of this research, the human performance issue is a strong motivator, however it is not our central interest at this point. Therefore it is considered future work and treated based on the literature in the field. Testing how foveation helps the computer performance, on the other hand, is demonstrated by an implementation: *Foveaglyph*. The implementation can be seen as a test bed to measure the computational benefits of 3D foveation. The results chapter (Chapter 4) provides evidence to support our thesis that 3D foveation is a useful method for level of detail management in fields that utilize 3D information.

3.1.4. Planning the Implementation

A number of questions must be answered before the programming could take place. Some in-depth technical considerations are as follows:

3.1.4.1. Input

If we are going to work with 3D, there may be a range of different formats of 3D data sets – so what kind of input should the program take? In principle, it would be easier to work with 3D vector data, as the 3D coordinates could be readily available. But photogrammetry extensively deals with stereo image pairs. Stereo visualization is also a cross-disciplinary research area, with wide public interest through entertainment medium and active academic and commercial applications.

We therefore decided that the program should take stereo image pairs as input. These images should not have any significant vertical parallax. To avoid potential vertical parallax, when possible, in the image acquisition step, a strict normal case of stereo camera configuration should be observed. Also, it is desirable that the geometric and radiometric errors are removed.

3.1.4.2. 3D Information

If the 3D coordinates are not readily available, how should we calculate the 3D coordinates? An image matching approach is needed. Since image matching is a very complex process, and the task is essentially a prototype, if available, utilizing an existing functioning image matching approach will save time.

3.1.4.3. Display Method

If we are to work with stereoscopic data, there are a number of stereoscopic viewing methods and hardware to choose from. What should be the stereo viewing method? Stereo viewing methods vary, and even though time parallel and time multiplexed techniques have different approaches, the underlying perceptual concept of binocular fusion is the same. If a program is developed for one method, it is easily extendible to others when needed. Therefore technically it does not make a difference.

From a practical point of view, we will use anaglyphs because they are common, easy to implement, hardware independent and glasses can be obtained or made cheaply.

3.1.4.4. Compression Approach

Ideally a method that gives the best compression and consumes the least computer resources should be chosen. Among its alternatives, an image pyramid approach is more common, not too complex for a prototype implementation and has proven efficiency.

3.1.4.5. Eye Tracking or Not

Foveation heavily depends on knowing where the point of interest is. How to determine the point of interest? Ideally a 3D eye tracker would tell us where the user is looking and we would assume that is the point of interest. In a prototype which does not have to run real time, it is sufficient that the user selects a point of interest interactively using a pointer.

3.1.4.6. Foveated Image Composition

Once the program knows the point of interest, it should reconstruct the foveated image around it. How to compose the non-uniform scene? To answer this question, we need to define a geometric model and establish the related parameters. These would be utilized to create segments of the scene with varying resolutions around the point of interest. These segments are areas for 2D and volumes for 3D. A distance metric is flexible and easy to implement. A constraint that the core volume should not be smaller than a certain size e.g. Panum's area may be forced.

3.1.4.7. Evaluation

After the non-uniform 3D image is formed, we need a healthy measure of telling how much compression the program can provide. This may be provided by an actual count of the pixels in relation to the way the image pyramid was built.

After these preliminary considerations, we can now introduce the implementation.

3.2. *Foveaglyph*: The Implementation

*"Everything should be made as simple as possible, but not simpler."
--Albert Einstein*

Foveaglyph is our implementation that takes a stereo image pair as input, and provides three different foveation options. The output of the application is a foveated anaglyph image. As it suggests, the name of the program was derived from the two words, fovea and anaglyph.

The application was developed and tested on Linux, however, it can be easily ported to other platforms. It is completely written in C, and makes use of GTK+ and gdk-pixbuf libraries¹⁸ for graphical user interface (GUI) and image operations respectively.

Foveaglyph allows the performance of various tests for stereo foveation, either from the command line, or using a GUI. The GUI allows better graphic interaction while the command line interface makes it possible to run batch tests. Command line options can

¹⁸ GTK+ is an acronym for GIMP (GNU Image Manipulation Program) Toolkit +, which is a toolkit for creating graphical user interfaces. The gdk-pixbuf library provides facilities for image handling. It is available as a standalone library as well as shipped with GTK + 2.

be seen in Figure 46 and a screenshot of the main GUI can be seen in Figure 47. All GUI menus are shown in Appendix 4.

```

arzu@hebe:~/foveaglyph$ ./foveaglyph
Usage: foveaglyph [options] left_image right_image
-h, --help                this help message
-o, --output=FILE         write output to FILE (anaglyph)
-m, --max-disparity=MAX   maximum level of disparity (200)
-p, --point-of-interest=Y,X Set the point of interest for foveation
-l, --level=N             Construct N levels of image pyramid
-c, --camera-constant=N   Camera constant
-B, --base-distance=N     Base distance between stereo cameras (mm)
-2, --two-dimensional     Foveate only in 2D, do not use disparity.
-3, --three-dimensional   Foveate 3D (default),
-d, --3d-disparity        Foveate 3D, only using disparity info,
-r, --write-rcfile        Write calculated coordinates to output_rc
-D, --debug-level         Debug level.

```

Figure 46: *Foveaglyph*'s command line options.

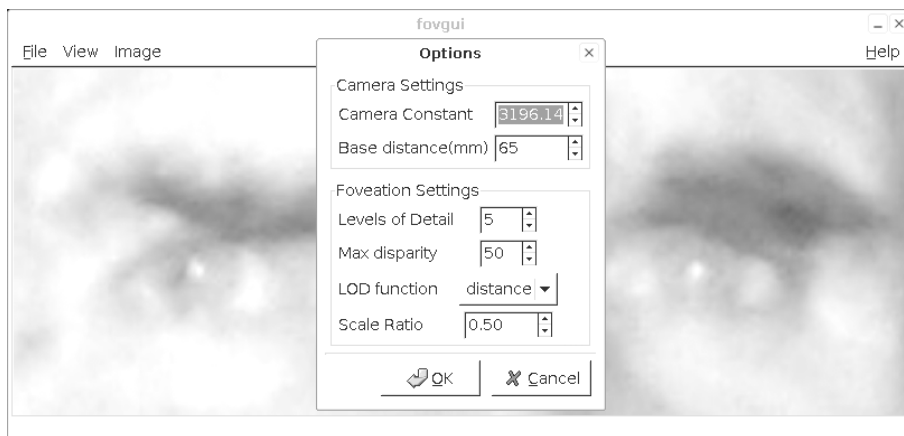


Figure 47: A screenshot of *Foveaglyph*'s graphical user interface. An appendix containing more screenshots is available to demonstrate several operations through *Foveaglyph* menus (see Appendix 4).

An overview of the program can be summarized as follows:

During the intended use of *Foveaglyph*, the system takes a stereo pair of images as input and creates an anaglyph image. The user is able to specify the coordinates of the point of interest when the program is run from the command line. If GUI is used, the anaglyph image is displayed and the user can point-and-click (i.e. using a mouse) at the point of interest. Based on this user interaction, the program presents a foveated version of the stereo image on screen. With the command line interface the image is saved as a file instead of being displayed instantly. See Figure 48 for a schematic description.

3.2.1. The Processes

There are several processes involved in running *Foveaglyph*. First, there are two pre-processes to prepare the image if we have a photogrammetric task. These are:

- Camera calibration
- Corrections of the input images: removal of lens distortions and affinity.

Within *Foveaglyph* the following are performed:

- The stereo image is created.
- The disparity map is calculated. This part is done using the code written by Birchfield et al. and integrated in *Foveaglyph*.
- The 3D coordinates in the object space are calculated according to the normal case geometry.
- The foveation pyramid, an array of downscaled images, is created from the stereo image.
- The foveated image is reconstructed based on the parameters specified by the user or the program defaults and utilizing a LOD function.

Even though its main point is to present the computational gain from 3D foveation, 2D foveation is an option in *Foveaglyph* as well. Not only is 2D foveation an efficient method for single images and would find comfortable use in geovisualization, but also comparing 2D and 3D results for the same point of interest will further justify our claim that it is worth doing 3D foveation. A schematic explanation of the application showing the processes is presented in Figure 48.

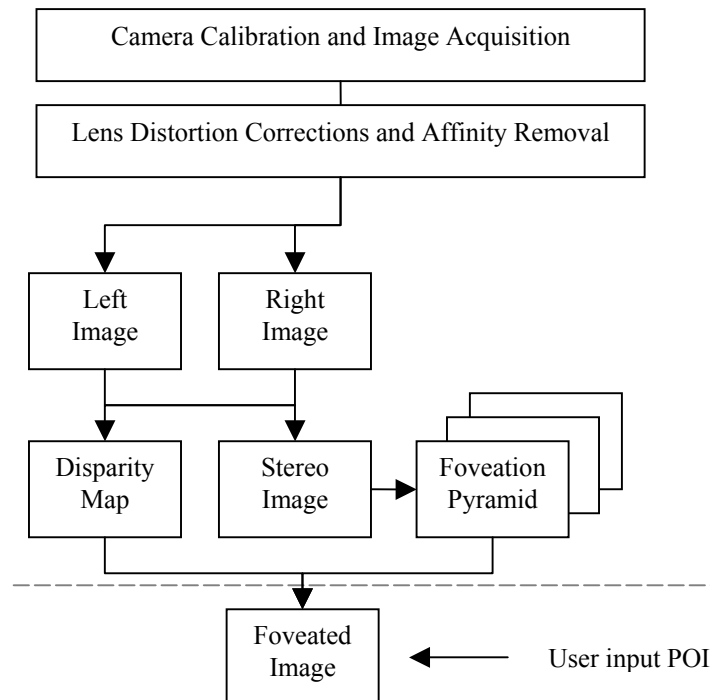


Figure 48: Schematic description of the application. Camera calibration and the corrections of lens distortions and affinity are two very typical steps in photogrammetric tasks.

The input image pair is used for constructing the stereo image and calculating the disparity map. The stereo image (anaglyph) is also used for building a predefined number of downscaled images, which form the foveation pyramid.

The operations listed above the dotted line in Figure 48 are to be pre-computed whenever image pairs are available. This is chiefly because computation of the disparity map and creation of image pyramid may take a long time especially for large images. Computing them in advance is the typical practice and would allow real-time use of 3D foveation. Chapter 4 - Results will demonstrate the numerical values.

After the “dotted line” is the actual foveation process. The foveated image is formed by picking the corresponding pixel from the appropriate member of the foveation pyramid. This is determined by the LOD function. See Sections 3.2.2.8, 3.2.2.9 and 3.2.2.10 for an explanation of *Foveaglyph*'s LOD function.

3.2.2. Explanation of the Tasks

In the previous section, we listed all the tasks performed before and during foveation process using *Foveaglyph*. A more detailed explanation of the tasks performed by *Foveaglyph* is given in the following sections.

3.2.2.1. Image Acquisition and Camera Setup

Obtaining best results in calculating the 3D coordinates from an image pair requires careful planning in the image acquisition stage. Section 2.6 explains the reasons and geometric considerations on the subject. To capture test images, equipment designed for this purpose was utilized.

The camera was set up in normal case (see Section 2.6.2.2 for an explanation of normal case of stereography) by using a rig built in the institute. The rig has a slider and allows the camera to be moved horizontally without permitting vertical motion. It is built to be used with a theodolite tripod, which allows leveling, and therefore offers a high degree of precision in vertical alignment. The emphasis on the vertical fixedness is to ensure better results in the stereo matching as vertical parallax will potentially confuse the process.



Figure 49: A tripod with a slider designed to take sequential stereoscopic photographs in the same vertical plane with adjustable horizontal shift. The camera can move left and right and a ruler allows the base to be adjusted.

If this is a photogrammetric task, typically before the image acquisition, the camera is calibrated. Also, before the image matching the lens distortions and affinity are removed. For our test images, these two processes are realized by using Petteri Pöntinen's software (Pöntinen, 2004). These programs were developed in-house for Helsinki University of Technology's Institute of Photogrammetry and Remote Sensing.

3.2.2.2. Camera Calibration

Recovering 3D structure from images becomes a simpler problem when the images are taken with calibrated cameras (Debevec, 1996). Calibration may have several objectives (Ziemann and El-Hakim, 1982 via Fryer, 1996):

- Evaluation of the performance of a lens,
- Evaluation of the stability of a lens,
- Determination of the optical and geometric parameters of a lens,
- Determination of the optical and geometric parameters of a lens-camera system,
- Determination of the optical and geometric parameters of an imaging data acquisition system.

In a more general categorization, we can think of camera calibration as having three aspects: geometric calibration, image quality evaluation, and in some cases, radiometric calibration (Mikhail et al., 2001). We are concerned with the geometric calibration for this implementation, which means determining the interior orientation parameters.

Interior orientation is the term employed by photogrammetrists to describe the internal geometric configuration of a camera and lens system. Photogrammetrists must know, or be able to compensate for, what happens to the bundle of rays coming from the object

and passing through the lens of their imaging device (Fryer, 1996). Geometric calibration establishes the interior orientation parameters of the camera (Mikhail et al., 2001).

The interior orientation parameters are the location of the principal point, the focal length (camera constant), and the radial and tangential distortion (Mikhail et al., 2001).

The starting point for building a functional model for close range photogrammetry is the central perspective projection (Cooper and Robson, 1996). Following is Leymarie's extract from Cooper and Robson:

The central perspective projection model is only an idealization (and simplification) of the actual optical geometry commonly found in cameras. Camera calibration is concerned with identifying how much the geometry of image formation differs in a real camera. One major difference is found in the optical distortions due to lens.

Radial lens distortion causes variations in angular magnification with angle of incidence. It is usually expressed as a polynomial function of the radial distance from the point of symmetry (usually coinciding with the principal point).

Tangential lens distortion is the displacement of a point in the image caused by misalignment of the component of the lens. The displacement is usually described by 2 polynomials for displacements in x and y (Leymarie, 2000).

In addition to the lens distortions, the differences in length and width of the pixels in the image storage caused by synchronization can be taken into account by an affinity factor (Godding, 2002). Applying an affine transformation to a uniformly distorted image can correct for a range of perspective distortions by transforming the measurements from the ideal coordinates to those actually used (Fisher et al., 2003).

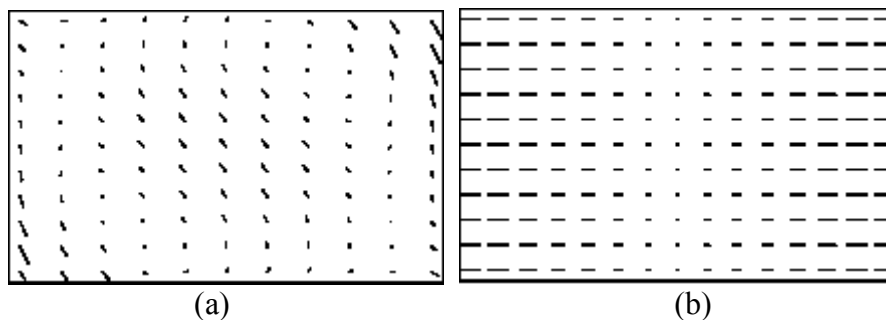


Figure 50: Radially symmetrical and tangential distortions (a) and the effects of affinity (b). Reprinted from Godding 2002 by permission.

The calibration procedures of analogue and digital cameras are similar, with only minor modifications in techniques required (Fryer, 1996).

3.2.2.3. Camera Information for the Test Images

After the generic information about camera calibration, here we provide a little more specific information from the camera calibration that was done for the test images. This information does not particularly imply any critical information as such but in case the experiments are to be repeated, it is to serve as documentation of what was being done.

A Nikon D100 was used for capturing most of the acquired test images. The camera has the following parameters when the focus is set to infinity, and the aperture (f stop) is set to 5.6. After calibration the camera has the following intrinsic (interior) parameters in pixels:

- Camera constant: 3196.144
- Principle point coordinates: 1568.932387, 1065.599147

Most of the images used in the tests in this study were acquired using this camera and have been pre-processed to remove the lens distortions and affinity before they were put into the remaining processes of *Foveaglyph*. While these steps are part of the routine preprocesses in photogrammetric projects, they are not required for visualization-only tasks. *Foveaglyph* can run its main processes with or without the camera information.

3.2.2.4. Creation of the Anaglyph

A more general introduction on how anaglyphs work can be found in Section 2.3.6.

Within *Foveaglyph*, the anaglyph is created simply by taking luminance (gray) values of left and right images, and using them as Red (R) and Blue (B) channels in the resulting image. The green (G) channel is taken from the right image.

Since we are only using the luminance values, the input left/right images are converted to gray scale using the formula

$$Y = (6969 * R + 23434 * G + 2365 * B)/32768 \quad \text{(Equation 20)}$$

Note that the asterisk is used to indicate multiplication in Equation 20. This is an approximation to the ITU recommendation (ITU-R, 1990). The GUI interface allows adjusting the anaglyph by shifting channels to remove the possible excess parallax.

3.2.2.5. Image Matching and Disparity Map Calculation

As expressed earlier, particularly in Section 2.6, image matching is an essential, yet a challenging and complex task. Developing a stereo matching method is not in the scope of this thesis therefore an existing and functioning method is adopted and integrated into *Foveaglyph*.

The integrated image matching algorithm and the C source code for it are by Stan Birchfield and Carlo Tomasi (Birchfield et al., 1998). Named as Depth Discontinuities by pixel-to-pixel stereo and often referred to as p2p, an acronym for pixel-to-pixel, the algorithm and the C code are distributed on a web site published by Stan Birchfield (Birchfield et al., 2003) and obtained from there with the author's permission. The algorithm is explained in several publications including Stanford University Technical Report STAN-CS-TR-96-1573, July 1996 (Birchfield et al., 1996)

The gray scale values of left and right images or R and B channels of the anaglyph are passed to the pixel-to-pixel stereo matching code as input. The output is a matrix of disparity value for each pixel. Adjustable parameters such as maximum disparity can be given either from the command line, or using the GUI.

Since this is the most time consuming step in the process, the application can load the disparity map from a previously saved image file.

3.2.2.6. Depth Discontinuities by Pixel-to-Pixel Stereo

Pixel to pixel stereo (p2p) uses *dynamic programming* to match scan lines individually. Pixels in one image are explicitly matched with pixels in the other image, while occluded pixels remain unmatched. A cost function tries to minimize dissimilarity of the pixel intensities and the number and length of the occlusions (Birchfield et al., 2003).

The code further enhances the result by taking into account that occluded objects and occluding objects have a known topological relationship. That is, the assumption that depth discontinuities are accompanied by intensity variation necessarily implies that an occlusion in the left (right) scanline must lie immediately to the left (right) of an intensity variation (Birchfield et al., 2003). This helps the problem that on the untextured surfaces the locations of the changes in disparity are determined largely by noise.

Also a dissimilarity measure is used that is insensitive to the problems resulting from image resampling.

The post-processing involves overwriting unreliable disparity information with reliable information obtained from neighboring scan lines. The basic idea is to propagate reliable disparities into regions where the disparity is unreliable, where reliability is determined by the number of contiguous pixels in the column agreeing on their disparity. Propagation stops when an intensity variation is reached (Birchfield et al., 2003)

The algorithm works without any extra information on the stereo settings of the image pair. Figure 51 presents visual samples from p2p.

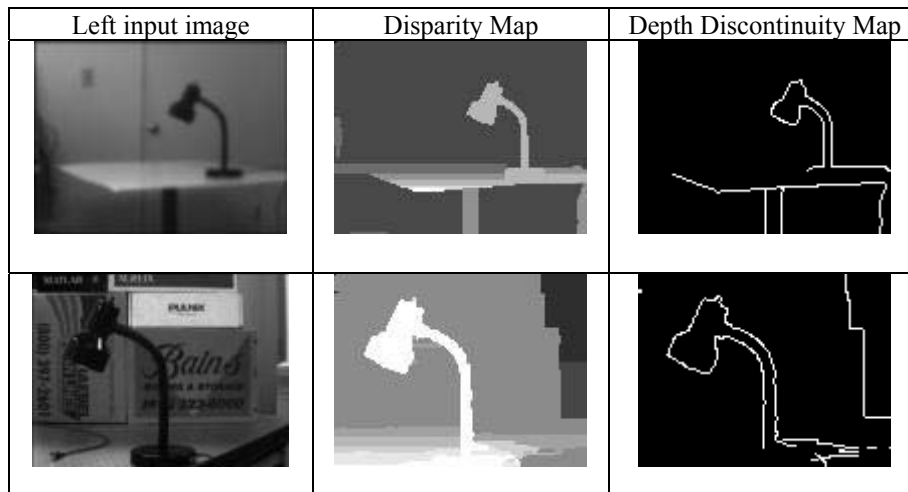


Figure 51: Pixel-to-pixel stereo, visual demonstration of the calculated disparity maps and the depth discontinuity maps. Images are reprinted by permission from Stan Birchfield.

As can be seen in Figure 51, pixel-to-pixel stereo can also calculate depth discontinuity maps, “which are defined as those pixels that border a change of at least two disparity levels” (Birchfield et al., 2003). As observed, the depth discontinuity maps look like the results of edge detection algorithms. Even though it is a default output for p2p, within the frame of *Foveaglyph* we have not utilized this aspect of the program.

As seen in the following figure, Figure 52, p2p’s results are linear to the size of the image in terms of the time the process takes. Table 3 lists the times it takes the complete matching process on a 1.70GHz Intel Pentium4 Central Processing Unit (CPU), running Linux.

<i>Image</i>	<i>Resolution</i>	<i>Pixels processed/s</i>	<i>CPU time (s)</i>
birchf-clorox	630x480	20248.2	14.9
calib field1 (down scaled)	784x521	12970.7	31.5
calib field2 (down scaled)	1569x1042	11794.8	138.6
gymball	3008x2000	14149.9	425.2
Furniture	3008x2000	12677.4	474.5
calib field3 (original)	3137x2084	11812.5	543.9

Table 3: CPU times for image matching in relation to the image size.

The choice of p2p for doing the image matching was mostly based on practical reasons; the code was freely available, it integrated well with the application, it gave visually satisfactory and fast results in a reliable manner.

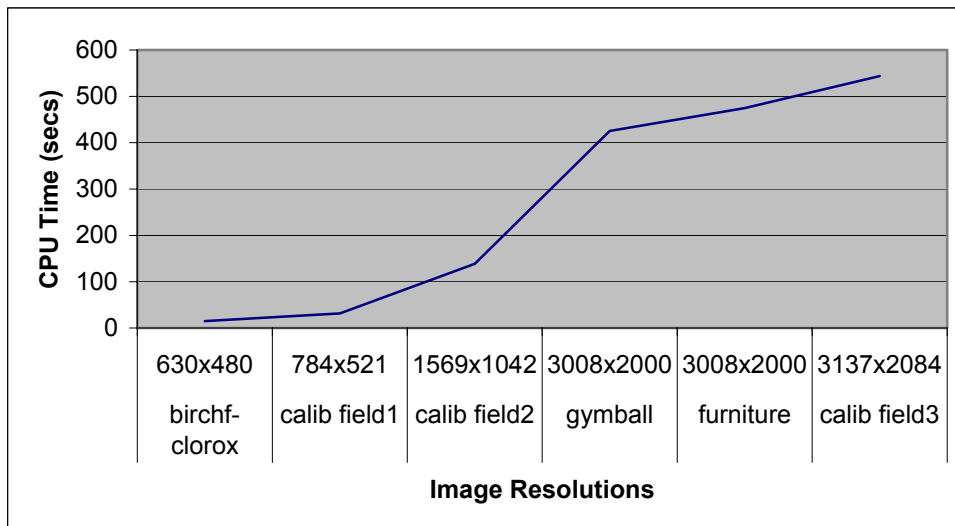


Figure 52: Image matching CPU timings in seconds for 6 images of different resolutions. The CPU time is expressed in seconds.

Problems: Because the program scans lines to match each pixels, even though it checks the neighboring pixels for a confirmation and looks at the line neighbors for further clean-up, as in all stereo matchers, vertical parallax would obstruct the efficiency of the matching.

The program applies a restriction of at least two levels in disparity levels to avoid making wrong judgments on slanted surfaces. Problems also come from this restriction, however, this problem of thresholding is inherent in the task (Birchfield et al., 2003).

Another obstruction typical to matching tasks as well as to p2p's accuracy is that the matching might be difficult if the scene does not have textures or patterns. This occurs on surfaces like walls and windows where there are specular reflections and a lack of texture. If the pixel values are nearly uniform, the matching based on the pixel values is bound to give less successful results.

A comparison of stereo matching algorithms including p2p can be found in Scharstein and Szeliski's publication (Scharstein et al., 2002 and 2003). The updated results are available online on their web site at Middlebury College Stereo Vision Research Page (Middlebury 2005). The web site is designed with the participation of other researchers, and the evaluation of 40 (as of December2005) different stereo correspondence algorithms are presented in comparison to each other with constant inputs and parameters.

Among these presented in Middlebury Stereo Vision Research Page, the ranking of p2p is not the best (ranked 15th at best among 40 algorithms). The main motivation for us to use p2p was that it was available and we were familiar with this algorithm. Also we were pleased with the results that we have got in our experiments. Using a more efficient algorithm for a future implementation similar to *Foveaglyph* may give more optimal results.

3.2.2.7. Building the Foveation Pyramid

Building an image pyramid is an essential process for our foveation algorithm, as we will need the various resolutions throughout the image space when building the final foveated image. Figure 53 is an illustration of the image pyramid.

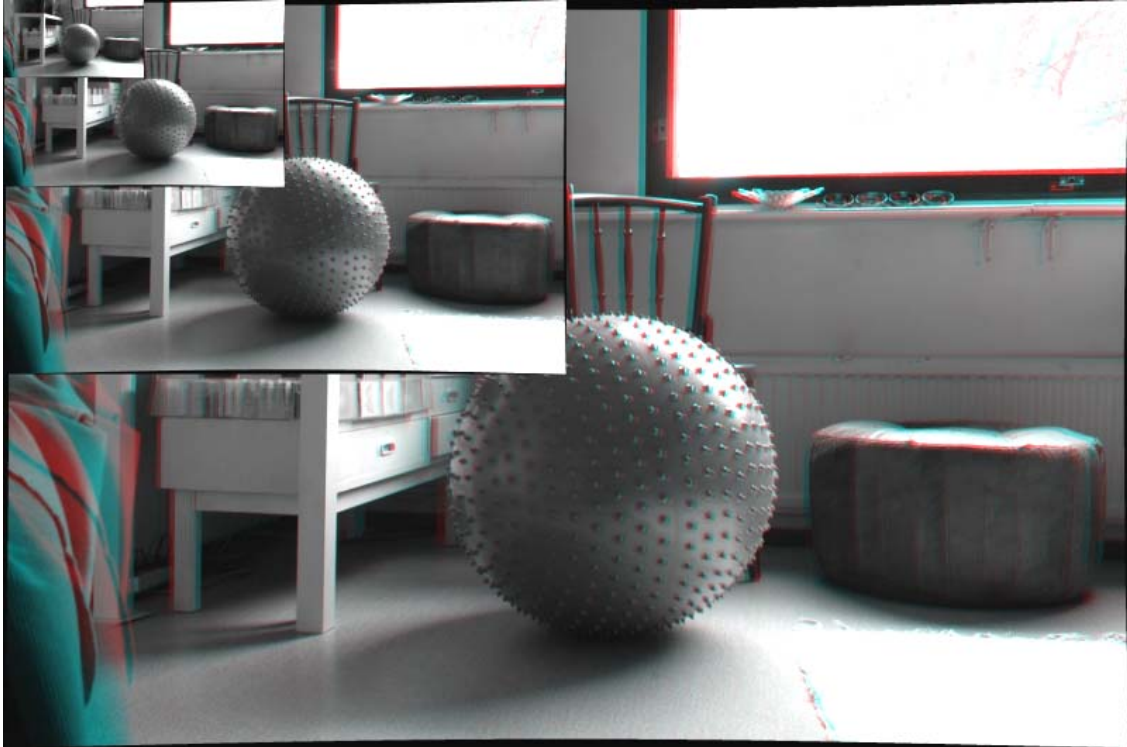


Figure 53: The image pyramid of 4 levels when scale ratio is 0.5. Scale ratio determines the size of the next image in the pyramid.

Our default image pyramid is simply an array of scaled down versions of the anaglyph image. If 2D foveation is requested, then the user-specified single image is scaled down in the same way. The scale ratio determines the size of the next image in the pyramid. By default, the application scales down the next image in the pyramid to half, which amounts to quarter of the area of the previous one. This way, 4 pixels are averaged to create 1 pixel in the new level. Scale ratio, as well as number of levels in the pyramid, is adjustable.

During the implementation of *Foveaglyph*, a proper low-pass filtering was not applied. Should this implementation be developed further, into a more complex program which would serve more than a mere proof of concept, a proper low pass filtering should be applied.

Figure 54 gives more detailed information on the image pyramid formation and how it interacts with foveation.

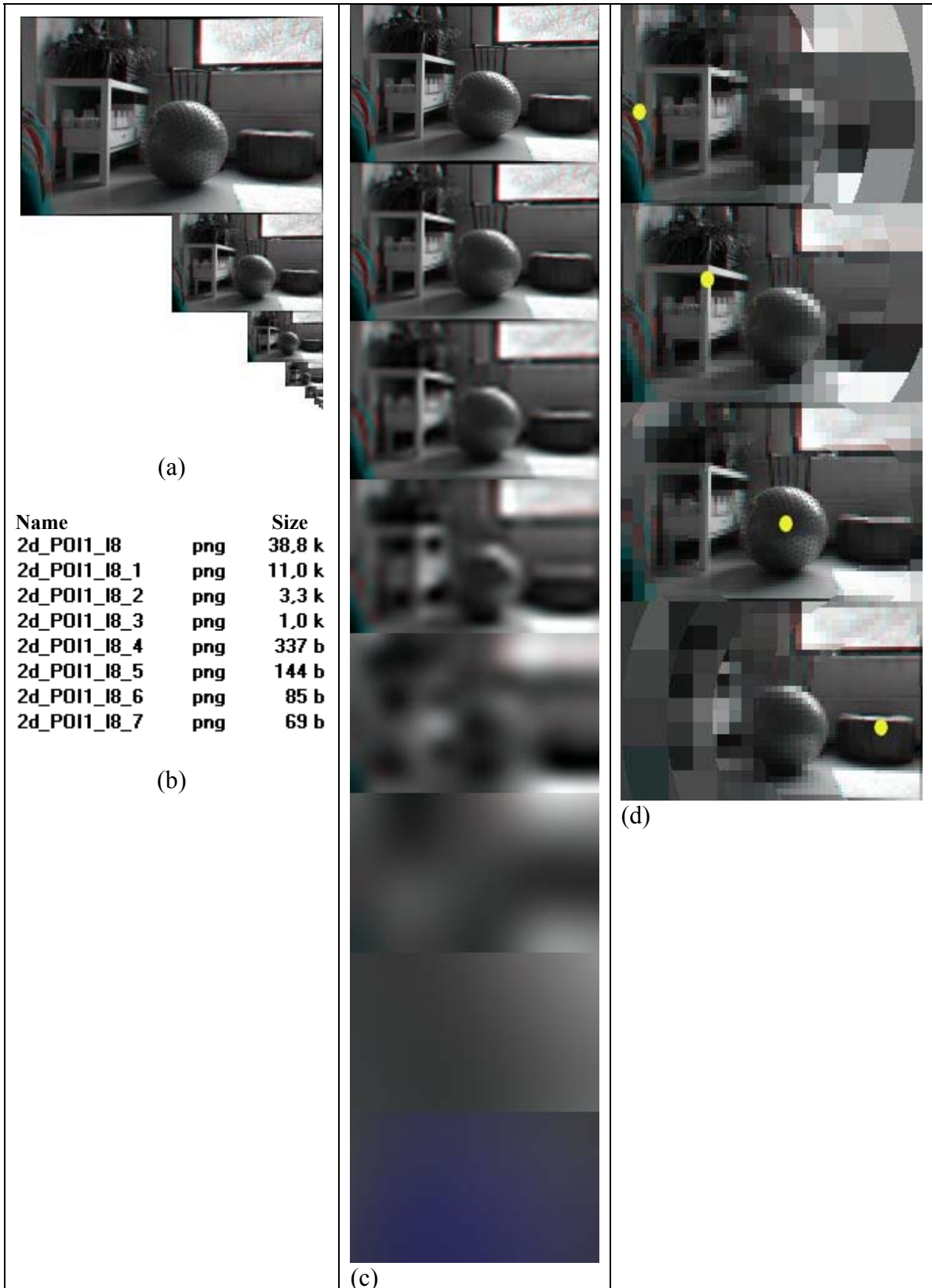


Figure 54: (a) The image pyramid of 8 levels when scale ratio is 0.5. (b) The names and the sizes of each level. (c) An illustration of the visual representations of each level as reflects to the end result. (d) The end results with 4 POIs. POIs are marked with circles and foveation rings are visible. These results are for 2D.

3.2.2.8. 2D Foveation

The foveation model for this implementation is fairly simplistic in terms of perceptual accuracy. The model is based on a step function, which is easily implemented and allows for efficient computations, and the results obtained represent the concept. While a simpler model has been adopted for implementation purposes, the extrapolation to a more perceptually accurate one is trivial.

The foveated image is created based on input from the user. The user specifies the coordinates of POI using either command line options or using the GUI by a mouse click.

Once the POI is specified, the space is segmented into co-centric circular regions using two parameters: image dimensions (maximum distance) and the desired number of level of details.

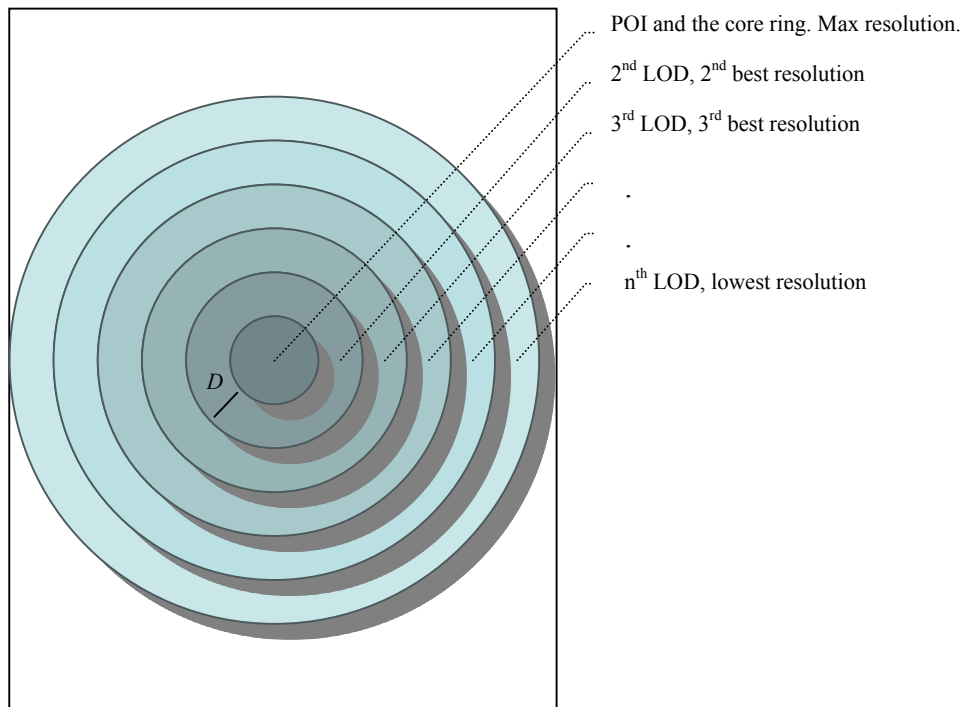


Figure 55: In 2D foveation the input is a single rectangular image. The level switching occurs at the threshold value D that forms the radius of the co-centric circles becoming $2D$ ¹⁹ from the POI in the next level when the scale ratio is set to the default 0.5. 3D foveation works exactly in the same way; only instead of circles, volumes (spheres) are formed around the POI with a radius of D in the core volume.

For each pixel in the foveated image, the pixel's LOD is determined based on its distance from the POI. The LOD will decrease as the distance from POI increases.

As we will meet some formulae after this point the notation used in those formulae is provided here as a chart:

¹⁹ Not to be confused with 2D as an acronym for two-dimensional. Here it means two times D which is the radius of the core circle or volumetric shape.

<i>d</i> :	<i>Euclidian distance between two points</i>
<i>d_{max}</i> :	<i>Maximum distance in the current work space</i>
<i>D</i> :	<i>A threshold that determines the LOD switch</i>
<i>lod</i> :	<i>When printed in lowercase italic letters, of the (0 to L-1) levels in the image pyramid, the index number of each level (i.e. 0th level is the best resolution, and would be expressed as lod = 0).</i>
<i>L</i> :	<i>Maximum number of levels</i> ²⁰ .

Table 4: Notation used in this chapter.

An important point to note is that level of detail values in the program are in reverse order, i.e. value 0 indicates the best quality LOD, and the quality decreases as the “*lod*” increases.

This lowercase italic *lod* is used to refer to the variable in the program; such that if there are 9 levels, and we want to talk about the 7th member in the pyramid, *lod* is 6 (as it is a zero indexed array), demonstrating that the image quality decreases when *lod* is increased. The capital letter LOD is used for expressing the general concept of “level of detail”.

By default, *Foveaglyph* uses a step function, which switches the LOD at a threshold *D*. This threshold is determined based on the maximum possible distance in the image setup, and the number of levels available in the image pyramid. If formulated, *D* is as follows:

$$D = \frac{d_{\max}}{L} \quad \text{(Equation 21)}$$

Where *d_{max}* is maximum possible distance in the working space (e.g. diagonal distance in a 2D image) and *L* is the number of available levels in the image pyramid. The above choice tries to maximize the use of all the levels in the pyramid. If needed, alternative ways to decide the switching threshold *D* could easily be added into the program.

With the above setup, all the pixels closer to the POI than distance *D* are viewed as best quality, the pixels with distances between *D* and *2D* are viewed with the second best quality level and so forth. This practically creates a number of concentric circles for 2D foveation as illustrated in Figure 55 and Figure 56.

²⁰ The maximum number of levels in an image depends on the size of the image, as our function takes image diagonal distance into account when creating the pyramid. User can give this value as an input, and if the user input is smaller than the maximum possible *L* program uses the user input. When *L* is lower than the user input, the maximum possible number of levels obtained from image size is used.

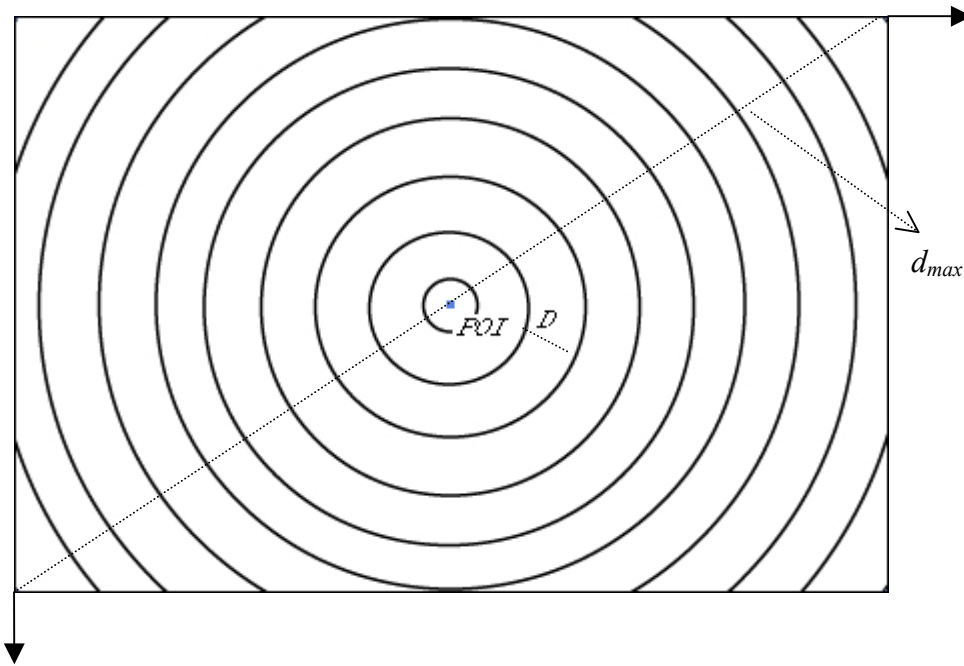


Figure 56: For a 2D image the maximum distance (d_{max}) is the diagonal distance between two opposite corners. The switching threshold value D equals to the radius of the core circle and stays constant for the next levels. It is calculated based on the image dimensions and the maximum or user specified number of LODs.

For example, if an input image has the dimensions of 200x200 pixels, d_{max} would be 282.9 pixels and with a scale factor 0.5, there would be a maximum of 8 levels in the pyramid. The smallest member of the image pyramid cannot be smaller than 1x1. This sets the threshold value D at about 35 and it means that the foveated image will have resolution change (LOD switch) at every 35 pixels starting from the POI.

Note that all of the levels of the image pyramid will only be used when the POI is in one of the corners of the image. When the POI is in the center of the image, only half of the available levels of the pyramid will be used.

In summary, *Foveaglyph* uses the distance between two pixels as a measure of distance.

3.2.2.9. 3D Foveation

In *Foveaglyph*, geometrically, the same distance metric described for 2D can be applied in 3D space (i.e. it is also a simplistic approach only to demonstrate the concept). The 3D coordinates are not so readily available as in 2D, but at this stage the image matching is done and the disparity map, as well as the 3D object coordinates based on camera setup are calculated.

There are two different modes of 3D foveation that can be chosen within *Foveaglyph*. The first approach is calculating the metric coordinates based on stereo camera configuration used while taking the photographs. The second method is using disparity information to decide a relative depth value.

The first approach is more appropriate when the camera configuration is available and more realistic and measurable results are desirable. The second approach is useful when the camera setup information is not available and it gives satisfactory results for visualization purposes. The two are very similar in logic.

For 3D foveation with metric values, the coordinates are calculated by making use of the stereo camera configuration supplied by the user (see Section 2.6.2.2). The user can provide the base distance B , and camera constant c through command line settings or via the options menu of the GUI.

Once the user specifies the point of interest, the distance d is calculated between the POI and each pixel visited. This calculation is based on the metric coordinates (X, Y, Z) which were obtained using the camera information, as opposed to the image based coordinates (x, y, z) which are used if we do not have the camera information and derived as an approximation from the images alone.

At this point, the LOD function uses the same threshold approach for LOD switching, which depends on maximum measurable distance in the calculated camera coordinate system.

Because binocular overlap loses its function after a certain distance, the depth information is lost for points further away from a certain value. We define this 3D working volume as where we have disparity, hence the depth information available. Therefore the near plane is determined by the maximum disparity while the far plane is where the parallax (disparity) is zero. The disparity we take into account here is what comes with the input image. Screen disparity might change as the image is projected to a larger display and if it is scaled up or down by zooming, but the program will take the absolute values from the input image matrices.

A demonstration of the working space can be seen in the following figure (Figure 57).

When the point of interest is denoted as POI , and each pixel visited is denoted as lowercase italic p (note that the capital italic P is for parallax); the general Euclidean distance formula for 3D is as follows:

$$d = \sqrt{(x_p - x_{POI})^2 + (y_p - y_{POI})^2 + (z_p - z_{POI})^2} \quad \text{(Equation 22)}$$

When calculating the metric coordinates, the x and y values are replaced by their corresponding equivalents depending in B , c and P_x :

$$d = \sqrt{\left(\frac{Bx_p}{P_x} - \frac{Bx_{POI}}{P_{x_{POI}}}\right)^2 + \left(\frac{By_p}{P_x} - \frac{By_{POI}}{P_{x_{POI}}}\right)^2 + \left(\frac{Bc}{P_{x_p}} - \frac{Bc}{P_{x_{POI}}}\right)^2} \quad \text{(Equation 23)}$$

The threshold value D is as in 2D foveation. The core sphere has a radius of D and all spherical volume rings are equidistant from each other's bounding surfaces with D :

$$D = \frac{d_{\max}}{L} \quad (\text{Equation 24})$$

All resolution degradation should happen in the regions that correspond to parafoveal perception and not to the foveal perception. To provide that, the core volumetric shape can be restricted so that it would not be smaller than what is the smallest perceivable binocular fusion area by HVS at the current distance of viewing.

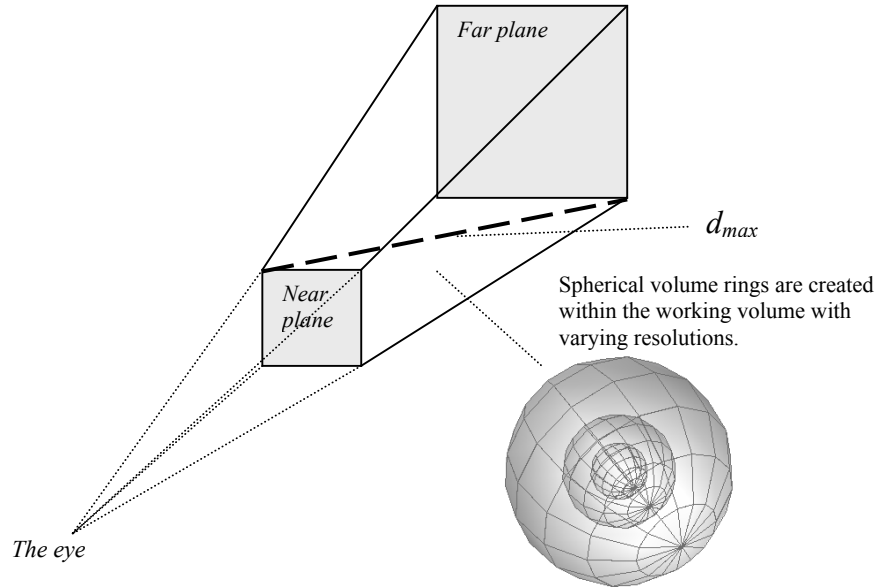


Figure 57: A typical view frustum e.g. used in displays is a truncated pyramid. Our working volume is also a truncated pyramid and d_{\max} here is the diagonal of this pyramid. Near plane and far plane are determined by the maximum and minimum disparity values. Varying resolution volume rings (co-centric spheres) are constructed inside this pyramid which is the bounding box.

As we remember from the previous section, L is the maximum number of levels and the maximum distance in the working area, d_{\max} is calculated as:

$$d_{\max} = \sqrt{X_{\max}^2 + Y_{\max}^2 + Z_{\max}^2} \quad (\text{Equation 25})$$

This distance is between the origin $(0,0,0)$ and $(X_{\max}, Y_{\max}, Z_{\max})$.

As can be seen in Figure 57, in the above equation (Equation 25), $(0, 0, 0)$ would correspond to one of the corners of the far plane and $X_{\max}, Y_{\max}, Z_{\max}$ to one of the corners in the near plane. In 2D foveation it is simply the two corners of the image.

As mentioned earlier, the near and far planes are decided based on the maximum and minimum disparity. Any point on the near and far planes will have the same disparity value and therefore the same Z coordinate. Taking advantage of this, we use opposing corners on near and far planes which gives the diagonal of the truncated pyramid. This is the maximum possible distance in our 3D working space.

If there is no camera information: 3D foveation without camera information uses the same approach. However, in the second approach we do not calculate the coordinates in camera coordinate system. Instead, disparity information is used for finding a z value. The program uses a value W that corresponds to the depth of the image, replacing Bc in the $Z=Bc/P_x$ formula. By default *Foveaglyph* assigns W as the height of the input image. This is an arbitrary choice to limit the depth, but W is a parameter in *Foveaglyph* and the user can specify a different depth value.

In partitioning the depth space, the disparity values are used as indicators of changing depth. But the fact that the disparity of the closer objects are bigger, defines finer depth intervals for near viewing space (see Figure 58 below).

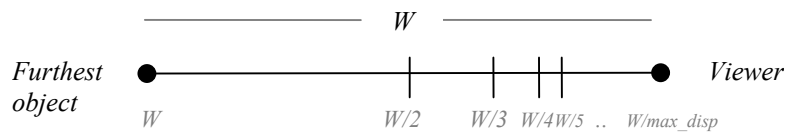


Figure 58: The intervals are defined for determining a disparity dependent z value approximation when it is not possible to calculate the real world coordinates. This is a documentation of how the disparity and distance relationship works.

The z value, calculated using this model (without the base and camera constant) is used in combination with image x, y coordinates as part of the equation to calculate the distance in pixels.

3.2.2.10. LOD Function

The LOD function for *Foveaglyph* is a function that makes use of the information produced by the program described in the previous steps, namely the calculated 3D coordinates and the foveation pyramid. The point of interest (POI) is specified by the user and determines the source of each pixel in the final foveated image. The source for each pixel will be an image in the pyramid. The LOD function is a distance metric between POI and another pixel in the image.

While doing 2D foveation, *Foveaglyph* calculates the distance between the POI and all other the pixels in the image to determine the LOD. During 3D foveation, LOD is still determined by the distance from the POI, but this time depth information from the disparity map is also used.

Foveaglyph, as a prototype, currently uses only one LOD model. For possible future research focusing on comparison of different LOD techniques, alternative LOD functions can be easily added to the program.

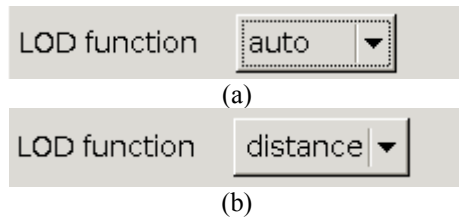


Figure 59 The pull-down menu for optional LOD functions; (a) shows the “auto” selected and (b) shows the “distance” selected. It is possible to extend the program by adding alternative LOD functions.

Currently a pixel’s LOD is determined based on its distance to the POI. Therefore the default LOD function is called distance within *Foveaglyph*. However, it should not be confused with “distance LOD” as a term which decides the LOD of the object taking its distance from the viewer into account. See Section 2.4.2.1 for more on distance LOD.

Regardless of the mode of foveation in use, *lod* for each pixel is calculated using the following function:

$$lod = \frac{dL}{d_{max}} \quad \text{(Equation 26)}$$

This is a step function (e.g. *lod* is an integer) where *d* is the distance between the POI and the pixel to be determined, *lod* is the pixel’s LOD (resolution), *L* is maximum number of levels that are possible and d_{max} is the maximum distance in the workspace.

This causes a linear decrease of quality in steps towards the periphery. The scale factor determining the downscaling ratio used between subsequent images in the image pyramid determines the reduction of quality in each step.

3.3. Shortcomings

The implementation was not done to meet the demands of a certain task. It is rather a proof of concept to demonstrate that foveation can be useful for generic stereoscopic and photogrammetric visualizations and serve as a test bed for further future investigations.

As stated by Geisler and Perry, image foveation techniques are most elegant and optimal where the pointing device used is an eye-tracker (Geisler and Perry, 1999; Geisler and Perry, 1998 via Linde, 2003), and conceptually, the model should best suit when there is an eye tracking system with well functioning binocular tracking capabilities. This definition fits best to HMDs, where the default viewer is only one person and the field of view is wide.

We did not have easy access to an HMD for our tests, nor did we have an eye tracking system running with our stereoscopic display. However, the work is still relevant when there is no eye tracking. Also implementing and testing the way we did is platform independent, which is an advantage because any other interested party can adopt it to their systems.

Our current implementation is aimed mainly at a 1:1 scale screen. 1:1 scale displays are similar to HMDs as they are both wide FOV displays. 2D foveation is particularly ideal for large FOVs. This is because when the FOV is large, the brain neglects more of the parafoveal vision. Our 3D foveation is built upon 2D, so we always foveate in 2D. It is possible to foveate **only** along the Z-axis and therefore simulate the DOF and leave the parafoveal vision alone by not foveating in 2D at all, if the display had a very small FOV.

Also, as it would in HMDs, the model used in this thesis limits the case to a single viewer looking from a single viewpoint. However, limiting the case to one user is not ideal for large screens that are meant for groups viewing a scene. The way we approach this problem is that for a prototype implementation, we are convinced that the frame of implementation conveys sufficient evidence to create interest in the field, which then should lead to a better planned and more laborious programming of foveation for 3D visualization.

3.4. Summary

In this chapter we have described the underlying thinking, models and algorithms used in the implementation. We are using a distance function to segment our working space, be it in 2D or 3D and building an image pyramid based on downscaling. We decide when to switch the LOD and select the appropriate level of detail for each pixel by looking at where they lie in the segmented geometry. The result is then reconstructed utilizing this knowledge and is a much smaller space variant 2D or 3D image.

CHAPTER 4. RESULTS

“The whole of science is nothing more than a refinement of everyday thinking.”
--Albert Einstein

In this chapter, results obtained from *Foveaglyph* and an evaluation of the processes are presented. An explanation of factors that have an effect on the results is offered and a method developed to measure the compression rates is documented.

Tests are performed to demonstrate that foveation can be useful for 3D visualization tasks. Performance evaluation is mostly about *Foveaglyph*'s and p2p's CPU times based on the task at hand. The graphic results will illustrate to the reader what the program does, and how the foveation changes in 2D and 3D based on the point of interest. Compression evaluation explains what these results mean in terms of compression, and how it varies depending on the criteria given by the user.

4.1 Things That Affect the Results

This section provides a brief account on parameters that have an effect on the results. These remarks should prepare the reader for interpreting the results properly.

4.1.1. What Affects the Compression Rates

The compression provided by *Foveaglyph* is not at a fixed rate. This is because of several variables starting with the user-driven parameters in the program. Some of these parameters can be seen in Figure 60 below.

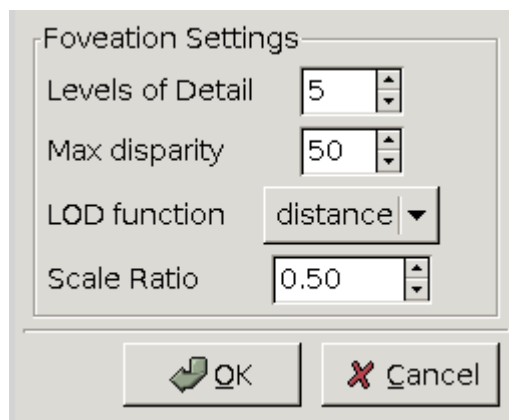


Figure 60: The options menu showing the user inputs for foveation settings. Changing values for Levels of Detail, Maximum Disparity and Scale Ratio have an effect on the results.

These user-given parameters are not alone in having an effect on the compression.

If you imagine that the non-uniform image at the end is made of rings, the compression ring for each ring is different. Because of that, the location of the POI and the image content may have a considerable effect on the foveated output. Figure 61 illustrates how POI location has an effect on the total resolution for images that are the same size in the beginning. The effects of two factors, the location of POI and the scene content, are further explained below.

4.1.1.1. The Wandering POI

If the POI is located towards the periphery rather than the center, the highest resolution areas occupy less of the image and the resulting image has fewer pixels. Hence the resulting image size will be smaller.

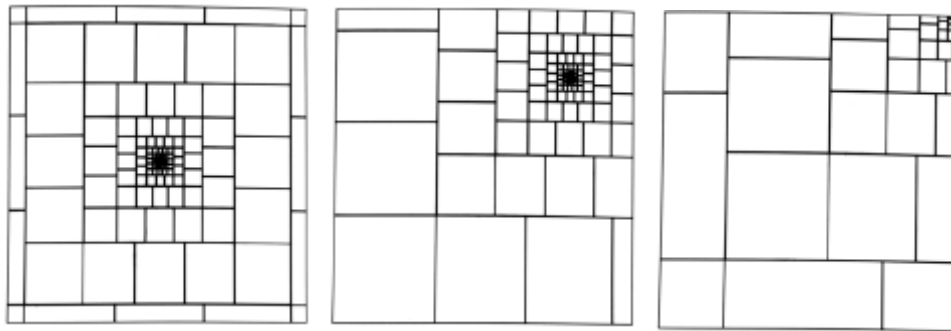


Figure 61: An illustration of changing resolution in the space variant image as the location of POI changes. POI is in the centre of the dark spot. It is easier to judge why the final image size depends on the location of POI from looking at these illustrations. 3D space works similarly. Reprinted from Kortum et al., 1996 with permission.

If two identical-sized images are put into the process, and the POI is located on the exact same location, the resulting image will have the same size in 2D foveation. In cases of 3D foveation however, it is not sufficient that the input image pairs are identical in size. The content of the scene in combination of the POI location will change the results. See below for an explanation.

4.1.1.2. The Scene Content

This relates to 3D foveation.

Let us remember that the Z-coordinate of the POI being in the centre, a sphere defines the core volume of interest. Around this core volume of interest, a number of other *volume strata* are created to form the non-uniform, 3D foveated image.

If there are many objects around the core volume of interest and in the “volume rings” close to it, more data from the highest or second highest resolution images are written in the resulting image. If the core volume and the closest volume rings are not very populated with objects, then there are less data to be recorded on the foveated image, hence the compression is much more efficient.

4.1.2. What Affects the Performance of Image Matching

A numerical evaluation of p2p’s performance was provided in Section 3.2.2.6. In this section we will explain a user specified parameter, which has an effect on p2p’s overall performance.

In p2p, and therefore in *Foveaglyph*, it is possible for the user to specify a maximum disparity value. This is an arbitrary or rather, an educated choice, and most likely not the same as the real maximum disparity calculated from the image matrices.

Typically the program runs a search in the image matrices of the stereo pair for each individual pixel and once it matches the pixel, calculates its disparity. By allowing the user to set a maximum disparity value, p2p provides an option for limiting the search, i.e. it stops searching for disparities when it reaches the provided value. This is to reduce the CPU time invested.

This is a risk taking approach, but it works at the end. The risk is that, in case the user provided value is smaller than the actual maximum disparity, the regions with bigger disparity values will not be included in the disparity map. In other words, the program will skip some areas potentially with existing disparity if they are set off-limit by the user. In the opposite case, when the given maximum disparity value is bigger than what the image pair would in fact have, the program is forced to scan those areas and some processing time will be wasted.

At the end it works, because for an experienced stereography operator, most of the time it is possible to have an accurate enough estimation of what could be the maximum possible disparity and provide the program with something just above that.

The user-given maximum disparity will have an effect on the performance as presented in Table 5 and Figure 62 below.

<i>Maximum Disparity(pixels)</i>	<i>CPU Time(secs)</i>
25	109.9
50	216.3
100	465.6

Table 5: The different user-given maximum disparity values, and CPU times.

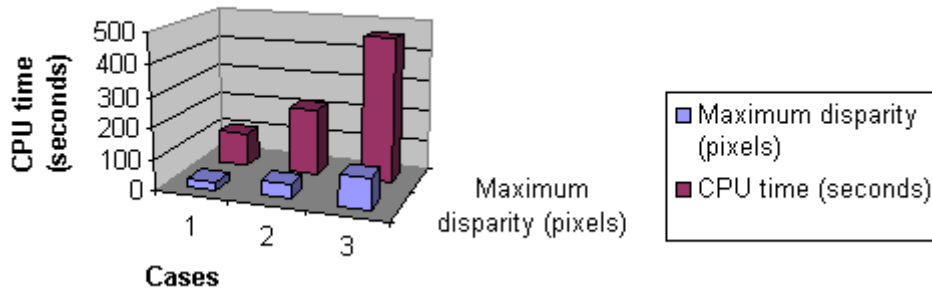


Figure 62: The information from Table 5 in a graph. 3 different maximum disparity settings.

If matching accuracy is important, maximum disparity should be set high. If time is more important and accuracy is of secondary importance, setting a lower maximum disparity will do the job quicker. In most visualization tasks, the image matching for a stereo pair is done only once and it is a pre-process. This is the case in our implementation. Therefore it does not pose a critical problem for our implementation.

Given the current state of general purpose computer hardware stereo matching cannot be carried out in real-time, while foveation can.

4.2. A Method to Measure the Compression Rates: Effective Pixel Count

Before presenting the results from *Foveaglyph*, here we introduce a method to provide a precise measure of compression.

To be able to talk about the compression, we need to be able to compare the size of the input image and the output image. The image size can be measured in a number of different ways. In this work a method was developed called **effective pixel count**. This is for determining the total compression gain in the output images.

Effective pixel count is defined as follows:

$$EPC = \sum_{lod=0}^L S^{lod} \times P_{lod} \quad (\text{Equation 27})$$

P_{lod} in the above formula is:

$$P_{lod} = \sum_{y=0}^{height-1} k \quad \text{where } k = \begin{cases} 1 & \text{if } lod_{xy} = lod \\ 0 & \text{otherwise} \end{cases} \quad (\text{Equation 28})$$

Where:

L is the number of levels in the image pyramid,

S is the scale ratio (default 0.5),

P_{lod} is the sum of the number of pixels taken from lod^{th} image in the pyramid.

lod : Of the (0 to L-1) levels in the image pyramid, the index number of each level i.e. 0th level is the best resolution, and would be expressed as $lod = 0$.

lod_{xy} = The level of detail for the pixel located in x,y coordinates calculated by the lod function.

While for the sake of clarity we have included these two formulas here, a more concise formula was used in Foveaglyph. It is as follows, which includes both of the formulas above:

$$EPC = \sum_{y=0}^{height-1} \sum_{x=0}^{width-1} S^{lod_{xy}} \quad (\text{Equation 29})$$

Typically, as explained in Section 3.2.2.8, the 0th element of the image pyramid has the highest resolution while the next image in the row is downscaled by S – e.g. if S is 0.5 the next image is 1/4th the size of the previous one. The scale ratio is a user-defined parameter and can be changed.

The effective pixel count is a measure for determining which pixel comes from which level, and the total number of each. This is our approach to calculating the resulting number of pixels in a foveated image. It would also be possible to use the image size in kilobytes to compare the results. This would require keeping the compression settings constant as Linde did with JPEG (Linde, 2004), but it is more meaningful to count the effective pixels when we do not control the compression algorithm itself.

Foveaglyph's GUI version currently supports several formats through GTK (gdk-pixbuf) libraries, including common formats such as JPEG, PNG, TIFF and RAW formats. The command line default is PNG (Portable Network Graphics) format.

4.3 Results

The tests were performed to see how much time the main processes will take and how much compression we gain from 2D and 3D foveation. The first is analyzed in the following subsection “Performance” and the latter in the “Results from Foveaglyph”.

4.3.1. Performance

The presentation of performance evaluation of the main processes, namely the image matching and the 3D foveation, aims at providing the numerics to judge if these processes are **also** suitable for real time applications such as in a dynamic VE or in an active vision application. We must bear in mind, however, that most photogrammetric tasks consist of static image sequences where the purpose would be 3D digitization and modeling of a given scene.

4.3.1.1. Why Emphasize the Image Matching?

Here we include the image matching timing separately for two reasons. First, image matching is typically done offline in stereoscopic 3D modeling tasks therefore the time it takes does not have to be considered a part of the 3D foveation process. Second, in our program, it was done by an integrated code – which means we do not claim rights to its efficiency or liability to its flaws, we merely document it, for it cannot be left out in the overall evaluation.

4.3.1.2. On the Performances of Foveation and Image Matching

Although the performance of the matching algorithm varies slightly depending on the image content, the processing times are linearly proportional to the number of pixels (resolution) in the image. This is true when the rest of the input parameters are set to be the same and the image content is highly similar.

As explained in Section 4.1.2, during the image matching process, there is a maximum disparity setting to limit the search time. This setting will affect the total time spent for the matching process. To be able to talk about the performance of the image matching process with changing input image sizes, the maximum disparity setting was kept constant for the tests.

Confirming that the program runs as expected, like it is in the matching process, foveation times are linearly related to the number of pixels in the image. However, the time spent for foveation is much (approximately 100 times) less than the time spent for image matching. Therefore, the foveation process is more suitable for real-time processing compared to image matching. Hence, as it is typically done, the image matching is better done off-line or beforehand.

All tests were run on a computer with a 1.70GHz Intel Pentium4 CPU, running Linux.

4.3.1.3. Performance Test

The CPU times for p2p alone were presented in Section 3.2.2.6. To evaluate the performance of p2p and 3D foveation comparatively, results are obtained using the calibration filed image shown in Figure (b) in Appendix 3, with a maximum disparity parameter of 100 pixels. The numerical data about the image is in Table 6 below. The number of images in the foveation pyramid (levels of detail) was 10.

The original image resolution was 3137x2084 as can be seen in Table 6, while the next two are the same image scaled by half. Time is measured in seconds and indicates the CPU time. Table 6 and Figure 63 presents the results on p2p time and foveation times.

<i>Resolution</i>	<i>P2p time</i>	<i>Foveation Time</i>
3137x2084	543.9	4.6
1569x1042	138.6	1.1
784x521	31.5	0.3

Table 6: Foveation p2p times in seconds, these are the CPU times. Note that the image pyramid was created off-line and is not a real-time process.

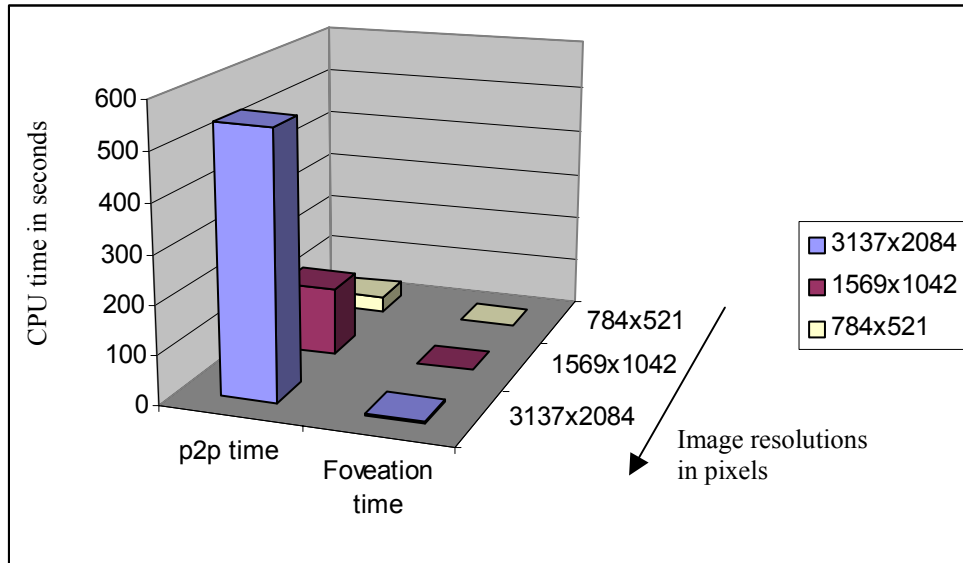


Figure 63: Foveation and p2p time comparison as a graph.

The time consuming part of our algorithm is a one-time setup cost, while the fast part of it is the run-time cost. To provide an example on the run-time cost of the foveation, we would like to note that a 320x213-pixel image can be recalculated at an interactive rate of 24Hz.

These results clearly indicate that the foveation process takes considerably less time compared to p2p times and that the time is linear to the size of the input image. Changing foveation times linearly according to image size is a positive result.

Image matching is a complex process and it is always time consuming. A presentation of CPU timings for p2p with more images can be seen in Section 3.2.2.6, where the program is introduced in a separate section.

4.3.2. Foveation Results

This section provides foveation results from *Foveaglyph* for a visual and numerical understanding. A comparison between the differences in the results with changing points of interest and scene content can also be drawn for the 2D and 3D foveated images.



Figure 64: A 2D foveated image. 8 levels of detail, POI is the center (mid-point). It can be observed that the resolution grows coarser towards the periphery in the lateral plane. Source image is by Birchfield et al., 2003, used with permission.

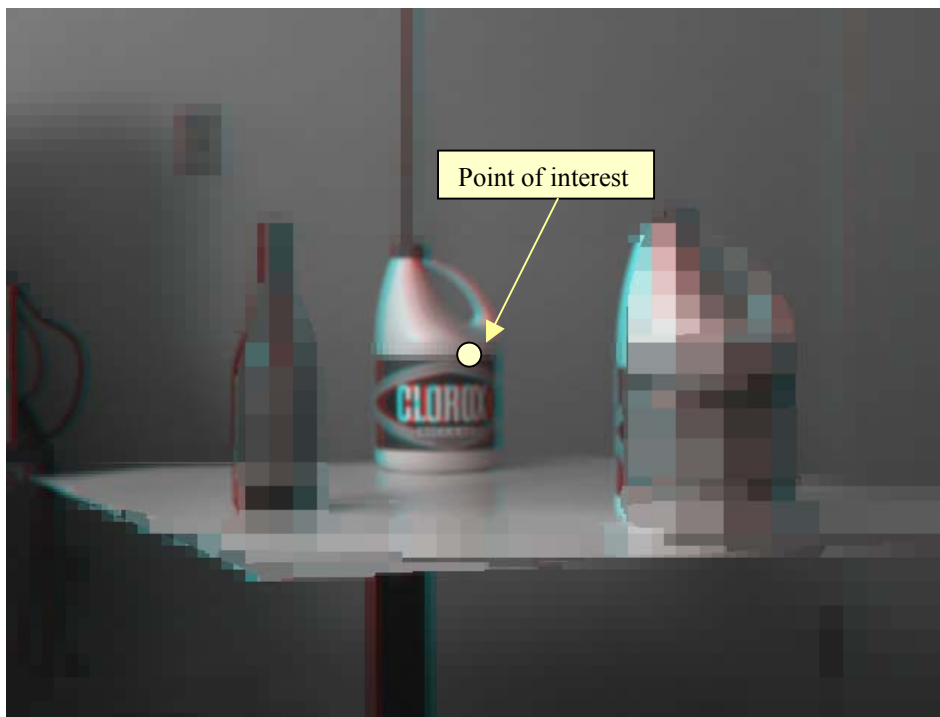


Figure 65: A 3D foveated image. POI and LOD values are the same as in 2D. The resolution grows coarser as we move away from the POI both in lateral plane and in Z direction. Source image is by Birchfield et al., 2003, used with permission.

Looking at the visual results will help the reader to judge what the program does. First a large image is presented to demonstrate the effect of 2D and 3D images separately (Figures 64 and 65). Then a set of smaller images with varying points of interest is presented in a table for comparison (Figure 66). In the table the compression rates are also noted. That is to provide a tool to compare the numerical results with their visual equivalents.

When viewing the 2D foveated image in comparison with the 3D, paying attention to closer objects will demonstrate the effect of 3D foveation. If they are closer than the point of interest their pixels will also be coarser. In a large field of view display, this effect should not bother the viewer as referred from the literature in Section 2.5.8.

The above images in figures 64 and 65 are selected from the samples (Birchfield et al., 2003) used in evaluating p2p.

It should be also noted that the results in the above images are particularly exaggerated to demonstrate the foveation effect (by changing the S parameter). Pixels towards the periphery do not have to be as coarse as demonstrated in these images. It may well be invisible in actual applications depending on several criteria such as number of levels and the scale factor. It is also possible to apply a smoothing filter for the resultant images to appear visually more pleasing.

Figure 64 and 65 demonstrate the results visually first for 2D and then for 3D. The 2D foveation results produce similar visual output to its peers as in Percept (Reddy, 1997), Foveator (UTEXAS, 2002) or Foveated Image Demo (NYU, 2003) (see Section 2.5.5). The only comparable 3D foveation is by Linde, but there is none with anaglyph, therefore an exact visual similarity cannot be drawn to a certain work. Linde's focus/foveation (Section 2.5.3.4.2, Linde, 2004) and some DOF simulations (Section 2.5.6) are similar as they give data reduction and blur along the Z-axis.

4.3.2.1. 2D versus 3D

As the concept we worked around was 3D foveation for a stereo image pair, it is in the interest of this research to demonstrate and interpret the differences between the 2D and 3D foveation. The following figure is a thumbnail table of a stereo pair foveated in 2D and 3D showing the visual and numerical results for the same POI (Figure 66).

The results in Figure 66 are obtained with an image pyramid with 4 levels. This was set as a restriction because the input images are very small. By restricting the image pyramid to a 4 level pyramid, we get low-resolution pixels big enough to be visible, but not so much that the image becomes unrecognizable.

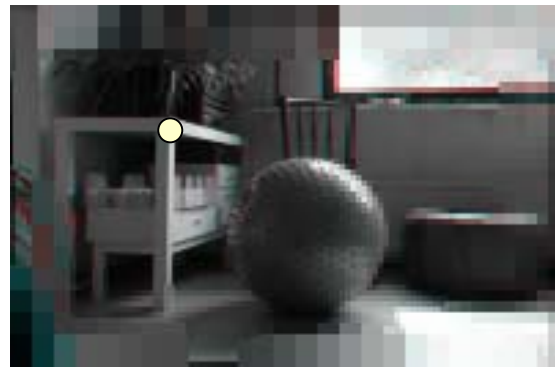
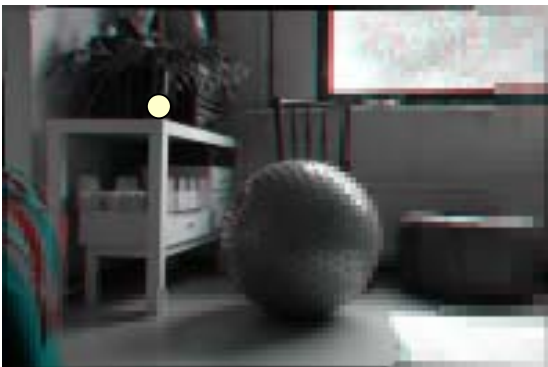
1) Original left image
200 x 132 pixels
(As seen in appendix 3d)

2) Original right image
200 x 132 pixels
(As seen in appendix 3d)



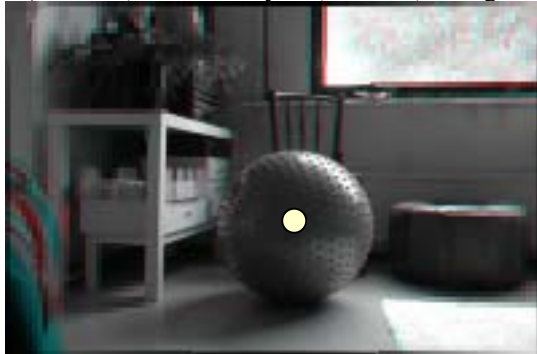
3) 2D fov., POI1 x=15 y=73, 4 levels, 72% gain

4) 3D fov., POI1 x=15 y=73, 4 levels, 98% gain



5) 2D fov., POI2 x=42 y=58, 3 levels, 62% gain

6) 3D fov., POI2 x=58, y=42, 4 levels, 70% gain



7) 2D fov., POI3 x=104, y=81, 3 levels, 59% gain

8) 3D fov., POI3 x=104, y=81, 4 levels, 63% gain



9) 2D fov., POI4 x=180, y=79, 3 levels, 71% gain

10) 3D fov., POI4 x=180, y=79 4 levels, 74% gain

Figure 66: The results as the POI changes in the gyball image. POI is marked as a circle in the images.

The above results for the gymball image (Figure 66) appears to have better compression rates and a bigger difference between 2D and 3D compression compared to the later results with clorox image (Figure 68). This is because of the scene content as explained in Section 4.1.1.

In images 3 and 4 shown in Figure 66, the POI is close to the edge of the picture and also closer to the viewer. The difference in between the two results can be clearly seen. Already in 2D foveation, because the POI is close to the edge, most of the pixels come from lower levels of detail. In the 3D foveation, the depth difference also comes into play and the gain is extreme, because not only the object is close to the edge but because there are few objects at the same depth, i.e., within Panum's fusional area.

The POI in images 5 and 6 is closer to the midpoint of the scene in comparison to the previous POI. This means less gain, because more of the pixels are taken from high LOD members of the image pyramid. The difference between 2D and 3D is still visible in front of the ball.

Images 7 and 8 give the least compression, being the closest to the mid point in the working space. The difference between the two is still visible in front of the ball. In images 9 and 10, the POI is closer to the right edge and the both 2D and 3D gain becomes higher, conforming the expectations.

These results show that 3D foveation adds a moderate, but consistent amount of compression compared to 2D and confirms the explanations given in Section 4.1. that the locations of the POI have an effect on the results.

4.3.2.2. 2D versus 3D: Compression Rates

In this section before a general evaluation of the results, we will have a particular focus on the compression rates. The results provided here do not indicate a general fixed rate of compression for the reasons explained in Section 4.1, and can be obtained only when the input conditions are exactly the same.

We will first include a graph obtained using the effective pixel counts from the results in Figure 66 to demonstrate the 2D and 3D compressions in a compact way. Table 7 and Figure 66 shows these results.

<i>POI Image Coord, Y,X</i>	<i>2D Effective Pixels</i>	<i>3D Effective Pixels</i>
POI-1: 73,15	7452/26400 (28%)	1237/26400 (4%)
POI-2: 42,58	1024/26400 (38%)	8056/26400 (30%)
POI-3: 81,104	10866/26400 (41%)	9879/26400 (37%)
POI-4: 79, 180	7664/26400 (29%)	7104/26400 (26%)

Table 7: Gymball image test results for 2D and 3D foveation. The lower percentages are better.

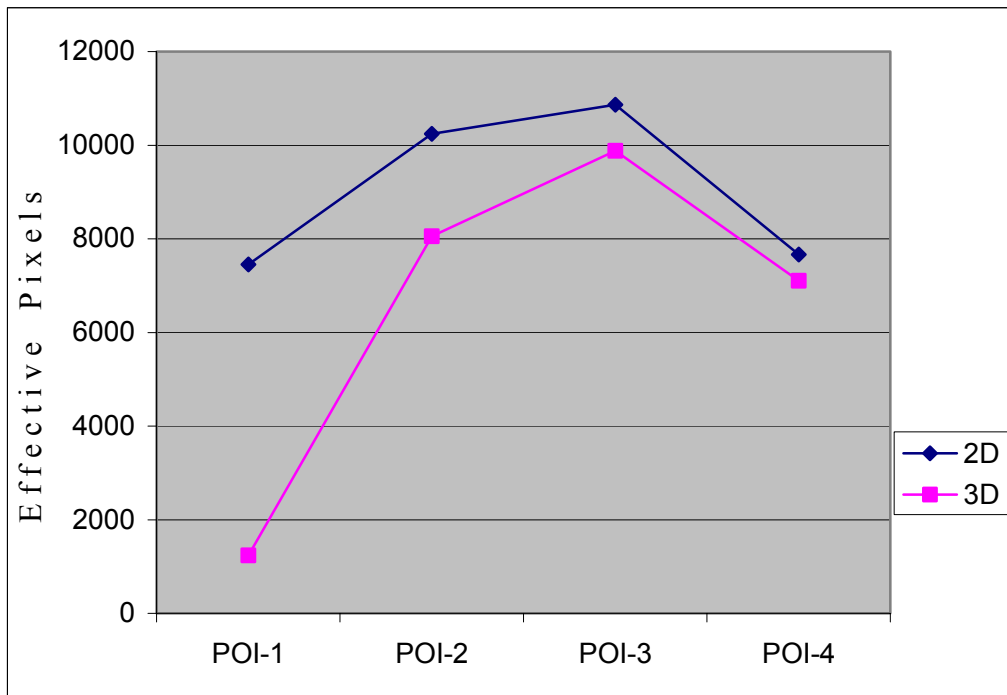


Figure 67: The graph obtained from various POIs in the gyball image shown in in Figure 66. 3D foveated image consistently uses more low resolution data, as expected. POI-1 and POI-4 lie in the eccentricities of the scene, therefore offer more compression.

Similarly, but with a different image content and size, 2D and 3D foveation were also applied to the **bircfh-clorox** image as presented in Figure 68 and Figure 69. Here we include 2D and 3D foveations for 4 different POIs for the bircf-clorox image to elaborate the results further.

These POIs are distributed throughout the space so that there is sufficient distance and depth differences between them, as was done with the gyball image. The input image and the distribution of POIs can be seen in the following Figure 68. Exact compression rates will follow in Table 8 and Figure 69 for an easier visual interpretation.

The main difference between the results obtained from birchf-clorox and from the gyball images is that the first has a much larger input size of 640x480 therefore we allowed the maximum level of detail. The results of 3D foveation appear more modest than the gyball image as will be seen in Table 8, but these are more realistic. The reason that the previous example with the gyball image had a limited number of levels was simply that we wanted to exaggerate the low-resolution pixels, so that they would show up in print and demonstrate the effect of foveation.

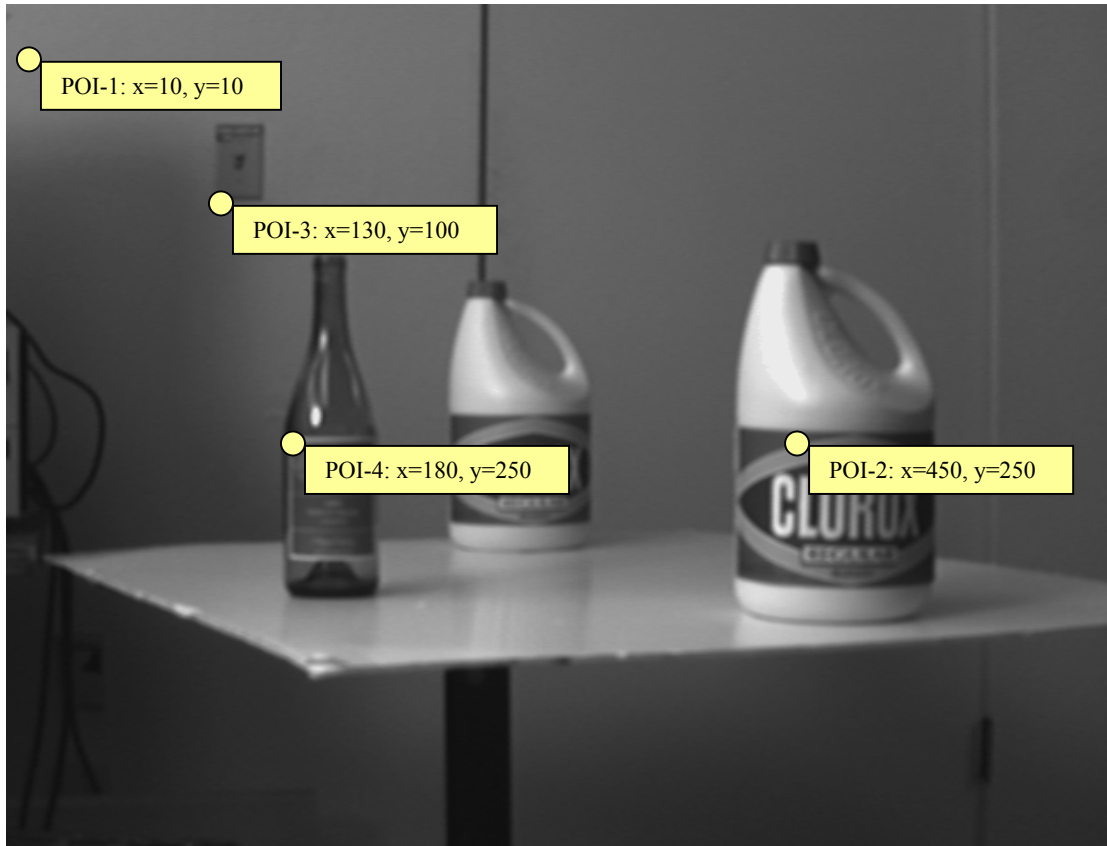


Figure 68: The original left input image and the approximate locations of the POIs as presented in the following table.

<i>POI Image Coordinates Y, X</i>	<i>2D Effective Pixels</i>	<i>3D Effective Pixels</i>
POI-1: 10, 10	35037 (11%)	34991 (11%)
POI-2: 450, 250	73189 (24%)	70888 (23%)
POI-3: 130, 100	86717 (28%)	85125 (28%)
POI-4: 180, 250	121206 (40%)	118359 (39%)

Table 8: 2D and 3D foveation results. Tests run on the same image with changing POI locations. Note that the lower percentages are better.

The input image size for was 640x480 (307200) pixels for the birchf-clorox image (seen in Figure 68). As can be interpreted from the numbers, percentages are **of the input image**, i.e. 11% means the resulting image is equal to 11% of the input image, so the compression gain is 89%.

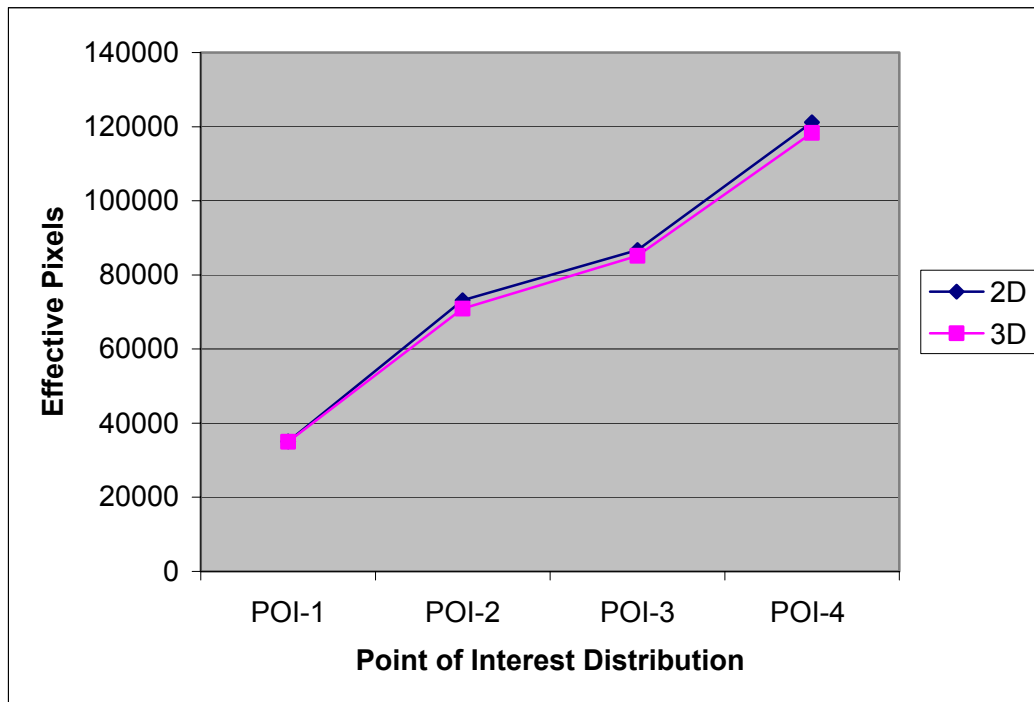


Figure 69: The results in Table 8 as a graph for an easier visual comparison. A modest but consistent extra gain in 3D foveation is observable compared to 2D foveation.

The compression gained by 2D foveation is significant, depending on the POI location and image content; there is a modest to moderate addition to the compression when 3D was added on to the 2D foveation. Based on the literature, there are also other valid reasons to research foveation for stereo images such as its potential cognitive benefits and contribution in curing VSS, which will be further discussed in Section 6.2.

4.4. Evaluation of the Results

The results from *Foveaglyph* have fulfilled our expectations both visually and in terms of process speed and compression gain. However modest, extra compression was added by taking a 3rd dimension into account and the compression rates depend on the location of the point of interest. This conforms to the findings of Linde (2004).

Our foveation program is not particularly optimized for a real time use, but image foveation has been shown to be of sufficiently low computational demand to enable real-time implementation in software (Sheikh et al., 2001; Geisler et al., 1998; Chang et al., 1997a via Linde, 2004).

We require the user to utilize a pointing device to state his/her point of interest and we then construct the model around this point. The performance of the foveation in terms of CPU time changes between 0.3 seconds for a 784 x 521 image to 4.7 seconds for a 3137 x 2084 image linearly changing according to image size. At this point we would like to remind the reader that these results were obtained on an average desktop computer of about 3 years of age with a 1.70GHz Intel Pentium4 CPU. A newer machine and some optimization would give faster results.

It is also worth noting that while we did not utilize a Graphics Processing Unit, (GPU), it would be an important asset if this work was to be optimized. The emerging field of General-Purpose computation on GPUs (GPGPU) would provide relevant references to this sort of work. For more information on GPGPU, please see (GPGPU 2005).

Whether this speed is sufficient should be evaluated depending on the task in hand. For most visualization tasks this should be sufficient given that the image matching is done offline and the disparity map is present for the stereo pair. It would also be possible to improve performance by calculating the image pyramid beforehand when the circumstances allow. It should be noted that performing these tasks beforehand is also the common practice when it is possible.

The implementation, *Foveaglyph*, was to serve as a proof of concept; therefore it was created with the most economical approach and was not refined with potentially faster models, which could give visually better results and maybe more compression. It can however serve as a test bed for comparing such techniques in the future.

A web page about *Foveaglyph* can be found at the following URL:
<http://www.foto.hut.fi/~arzu/thesis/foveaglyph/>

4.5. Summary

In this chapter, first the factors that have an effect on the results were documented and a method for measuring the pixels in the space variant image was introduced. We have presented graphic and numerical results regarding *Foveaglyph*'s tasks and provided an evaluation of these results.

CHAPTER 5: DISCUSSION

*“The map is not the territory”²¹
--Eric Bell*

While our model and the program could easily be used with any system that can provide 3D coordinates, the literature review, the implementation and the obtained results have led us to consider the potential of foveation particularly in a few different fields. This chapter is the where some of those considerations are presented. These are considerations of where this work might potentially evolve rather than a discussion of the results from the implementation.

5.5.1. Geovisualization

Foveation has a great potential for geovisualization. The datasets in many geographic and geo-scientific are often large and heterogeneous (Gahegan, 2000). This fact did not escape the attention of those who have dealt with developing methods for managing level of detail and area of interest management (e.g. Reddy et al., 2001; Been, 2002). Meanwhile in the geoinformation community, over the years a single term, ‘scale’, emerged to capture the sense of ‘level of detail’ (Goodchild, 2001). There are a number of research papers that relate geovisualization to level of detail and foveation, some of which are referred to in this section.

Conceptually similar to area of interest management, in cartography, there is a common method called generalization. In the absence of a system to know what is the interest of the viewer at the time of viewing, the theme and the audience of the map is taken into consideration. Generalization is about taking the irrelevant data out of the map and/or emphasizing the most relevant ones with symbols or colors. Three major components of the concept are classification, simplification and exaggeration. The applications of generalization have a long history, but it continues to be relevant in digital maps. Today the number of possible operations to handle the data is bigger and operations more flexible, but the data also grew more complex.

In Geographic Information Systems (GIS) there are many more features compared to a classic map. As a part of the system, there is an -often massive- database that contains as much as or more non-graphic information that is linked to the map. In complex richness, it is possible that the designer or the viewer is lost. Visualization and virtual reality offer some hope of providing an environment where many data layers can be viewed and understood concurrently (Gahegan, 2000).

²¹ An expression coined by Eric Bell and popularized by Alfred Korzybski, and used in General Semantics and Neuro-linguistic programming, **the map is not the territory** recognises that individuals may mistake a metaphorical representation of a concept for the concept itself. A specific metaphor may not capture all important facets of what it represents, and may thus limit an individual's understanding unless the two are distinguished (Wikipedia Map, 2005).

In the context of visualization and virtual reality, these complex large datasets challenge resources especially if they are to be transported over low bandwidth systems.

A generic approach to managing level of detail that is particularly suitable for terrain models can be seen in Hoppe's work on progressive meshes (Hoppe, 1996), there is also more specific research on LOD approaches to terrain visualization (Larsen et al., 2003). For the Internet, the 3D vector visualization standard has become the VRML. Under the roof of VRML, GeoVRML (Reddy et al., 2001) was particularly developed for visualizing geo-referenced 3D geographic datasets over WWW.

Stereo foveation is relevant when there is stereoscopic or 3D information, and when the data is 3D, there may also be an interest in applying the foveation *only* along the Z direction. This would give us the DOF simulation for the FOV is already very small. This can be done by taking the depth information in account and leaving the eccentricity information out. Such an application would be relevant to smaller displays such as mobile phones.

In a recent publication Kirschenbauer reported that “the true 3D²² map proved to be superior for identifying spatial phenomena” after an empirical study for the cases where the stereoscopic displays were used for geovisualization (Kirschenbauer, 2005).

Meanwhile, when working with 2D maps, foveation can still be useful. There is a good potential for 2D foveation for several geomatics applications, in particular for web mapping. Several researchers have published on foveation for geo-data (Chang et al., 1997b, Been, 2002). For example, Been presented a web-based responsive, zooming and panning visualization system using multiple levels of detail (Been, 2002).

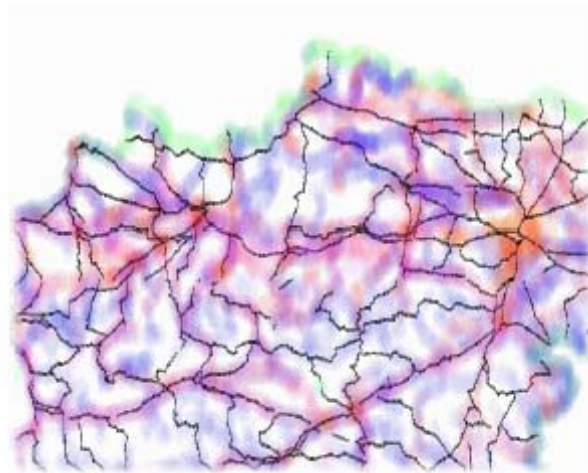


Figure 70: A GIS showing the roads layer in focus. Reprinted from Kosara et al. 2001 with permission.

Figure 70 above shows results from one interesting research paper, a “semantic depth of field”, which was presented by Kosara et al. in 2001. They used GIS as their case study where the layer of interest was in focus while the rest of the layers were presented blurred to manage the attention of the viewer and save the network resources.

²² True 3D is a popular term for stereoscopic 3D, even though some argue that it is misleading to call it “true”, it is commonly used both by professionals and media.

Relevant literature on the topic and the considerable compression gain in our implementation encourages us to suggest that foveation should be investigated further for geovisualization.

5.5.2. Visual Attention Management and Progressive Image Loading

Another use of foveation in visualization could be to manage the attention of the user.

“Focus effects are important in separating foreground objects from the background objects. Perhaps because of its role as a depth cue, simulating depth of focus²³ is an excellent way of highlighting information by blurring everything except that which is critical. Unfortunately, this technique is computationally expensive and this currently limited in utility” (Ware, 2000).

This is particularly valid if the audience is passive. In a 3D cinema, for example, an object or a set of objects of the director’s choice can easily be highlighted by employing a seamless foveation. For the static stereo pairs, this could be useful for blurring the “obstacles” that occlude the object of interest and could make it easier for the viewer to work with the area that they want to work with. In a terrestrial photogrammetry project, there may be many occluding objects to the actual area of interest depending on the scene. If the 3D environment is synthetic, the occluding object could be rendered transparent and with less polygons.

Foveation could be utilized as an alternative to progressive image loading. This idea has been exploited by several researchers (Chang 1998; Overall 1999; Sanchez et al., 2004). For example a concept called “fovea first transmission” is discussed for videos by Overall, 1999, and “prioritized region of interest transmission” by Sanchez et al., 2004.

Also, 3D cinema operates with stereoscopic media. Stereo foveation can be tested for 3D cinema both for saving bandwidth, potentially preventing VSS and for visual attention management.

3D foveation can also be a very useful tool for datasets that are consisting of point clouds, e.g. results from a laser scanning. If the occluding points were out of focus when displayed, the viewer would see his or her object of attention much better.

5.5.3. Stereo Foveation an Alternative to Stereo JPEG Compression?

In a relevant research paper, Seuntiëns et al. found that JPEG compression of a stereo pair had a negative effect on users (Seuntiëns et al., 2003). In their study they considered image quality, sharpness and eyestrain. They have not found any effect on depth perception however. The same study could be repeated with stereo foveated images to find whether quantitatively comparable sets gave different results for the perceptual quality issues.

²³ Ware uses “Depth of Focus” for what we call “Depth of Field”. See section X2.2.3 for an explanation.

One of the most common compressed image formats is JPEG. Based on this format there is also a stereo format, identical to JPEG but comes with an embedded tag, which is a stereoscopic descriptor and is called JPS. Other common formats such as GIF, BMP, PNG and TGA also have stereo formats, which, consecutively are called as GIS, BMS, PNS and H3D (see the list of abbreviations to see what these acronyms stand for).

These stereo formats would be enhanced by a progressive fovea first transmission using our stereo foveation.

5.5.4. WWW use for 3D foveation: VRML and QuicktimeVR

The potential of stereo-foveation is clearly apparent, as there is an audience for stereo imaging, both academically and commercially interesting. But 3D is not only stereo. The foveation concept easily widens to 3D graphics in general.

The ISO standard for viewing 3D vector models on the World Wide Web (WWW) is the Virtual Reality Modeling Language (VRML). The format is open and commonly used for 3D Web visualizations. It is one potential format 3D foveation can be extended and tested. It is possible to create Anaglyph VRML models, but even when working with non-stereo VRML models, as the 3D information is available, therefore there is a clear path to implement a DOF LOD in combination with an Eccentricity LOD. VRML already has a node for applying distance LOD native to the format.

Another common and popular image-based network visualization tool is AppleTM's Quicktime VR, expressed with the acronym QTVR. QTVR uses panoramic images and functions in 3D and it is to be used on the WWW with the help of a plug-in. This work can be extended to QTVR and tested for the WWW.

5.5.5. Summary

In this chapter, we have discussed the potential of foveation in several areas where stereoscopic medium was used or 3D graphics are utilized. Namely, the fields taken into consideration in this chapter were geovisualization, visual attention management, foveation as an alternative to stereoscopic compression method and foveation's potential use for 3D web graphics.

CHAPTER 6. CONCLUSIONS

*I suppose, Watson, we must look upon you as a man of letters.
--Sherlock Holmes (Arthur Conan Doyle)*

Foveation is a model of the HVS that is used for removing redundant data in computer vision and 2D image visualizations. Photogrammetry works with images and 3D modeling. So we asked ourselves, can foveation be extended to 3D and be useful? We developed an approach to apply it to a stereoscopic image pair to test it then we implemented it by programming *Foveaglyph*, tested it with stereo image pairs and found that the answer was yes. It could be extended to 3D and yield useful results.

While we have shown that it is technically feasible, we have also extracted a suggestion from the scientific literature that asking this question was valid also for other important reasons. It was not only to help the computer performance, but it was going to help the human performance in viewing virtual worlds. Simulating the DOF is reported to help the health problems attributed to VEs. In the scope of this thesis, 3D foveation includes DOF simulation, in addition to the usual FOV simulation in 2D.

By modeling the relevant biological processes of vision, we would also give a smart operation tool to the manager of a synthetic 3D world. Using this tool she or he can control the attention of the viewer.

6.1. General Remarks

The main contribution of this thesis to existing knowledge is to bring together techniques and knowledge from disparate research areas. The new knowledge is the concept of 3D foveation, which was born as a hybrid concept from these techniques and knowledge. We have demonstrated that level of detail management and foveation techniques can be useful for stereoscopic and therefore also close range photogrammetric visualization tasks.

It is a fact that there are very few works in the existing body of literature that implement a combination of foveation and depth of field simulation into one model for visualization purposes²⁴. The works on binocular foveation encountered in robot vision literature do not concentrate on visualization and displaying the information. Their focus is on designing the cameras' lens system in a foveated manner to save resources in the real time image or video capturing process. This should help the robots operate faster than if they processed the scene in a uniform fashion and still give them an equally good sense of vision.

²⁴ An example that uses depth of field simulation in 3D is by Ohshima et al. (Ohshima et al., 1996) and the focus/foveation work by Ian van der Linde (Linde 2003, 2004).

In cases of visualization and displaying the stereoscopic information, the studies on depth of field simulation are few and as demonstrated in Section 2.5.6, those that are found are not concerned with 3D foveation.

A combination of depth of field simulation and foveation for visualization on stereo displays was demonstrated by implementing an independent stereo foveation program: *Foveaglyph*. *Foveaglyph* successfully compressed images in a non-uniform nature in 3D space. While the framework introduced is open to accept different criteria, we have mainly considered large field of view displays and close range stereo applications.

The only comparable work to ours was by Linde (2003, 2004). His methods and approach are different than ours, but the results confirm one another

6.2. Perceptual Issues

In a number of research papers and recent textbooks it is stated that simulating depth of field would help with some of the problems associated with stereoscopic displays (some examples are Luebke et al., 2003; Linde, 2003 and 2004; Ware, 2000; Mulder, 2000; Blohm et al., 1997; Reddy, 1997; Martens et al., 1996). These problems are some undesirable side effects of some stereoscopic displays, such as headaches, nausea, vomiting and ataxia (Linde, 2003) and the combination is referred as virtual simulator sickness.

The statement on the assumed benefit of simulating DOF comes from the literature based on user surveys and some conclusions drawn from an understanding and observation of human visual system.

For example, Mulder et al., 2000 states that:

“A major drawback of current virtual reality display hardware is that the convergence accommodation relationship in human viewing is violated. This is a major cause for eyestrain often experienced by humans when using VR equipment. It would be an interesting research to investigate whether the application of DOF would have a positive effect in this regard.” (Mulder et al., 2000)

Linde notes that,

“Since DOF simulation has known perceptual benefits, and a pyramid is required for foveation anyway, the technique is worthwhile even though only in some cases only modest improvements in compression may be observed.” (Linde, 2004)

Based on the literature, we are encouraged to say that our research has potential to contribute to finding solutions for the problems involving diplopia and accommodation convergence conflict attributed to stereoscopic displays.

In a study for stereoscopic JPEG coding Seuntiëns et al. found that JPEG coding had a negative effect on image quality, sharpness and eyestrain levels, but had no effect on perceived depth (Seuntiëns et al., 2003). An increase in camera-base distance increased perceived depth and reported eyestrain but had no effect on perceived sharpness. Furthermore, both sharpness and eyestrain correlated highly with perceived image quality.

If this finding is correct, our approach should provide a solution to the complaints that come from uniform compression. Because it does not make a difference in the depth perception, the lower resolution levels will not disturb the user as much as it would in 2D foveation. And because we are able to use a very high resolution in the actual area of interest, this should truncate the negative effects on image quality, sharpness and eyestrain. This is however left as a future study.

6.3. Future Work

When the research question was formed and the investigation took its course to see whether foveation could be extended to 3D, several other interesting questions emerged, as usual to such projects. Naturally not all questions can be answered in one single project, so the scope was limited to evaluating the concept by testing foveation for 3D.

As this thesis was done in a photogrammetry institute, questions that required a multidisciplinary team and environment were left as future work. Some of those are listed below.

6.3.1. Usability

Although current literature provides evidence that foveation does not disturb the user and in fact depth of field simulation, hence stereo foveation, which is a combination of 2D foveation and depth of field simulation, may have positive effect on diplopia and accommodation convergence conflict, a usability test with humans should be conducted to confirm the findings in the literature on the psychophysiological effects of the application. Two clear things to look for would be:

- Whether or not the resulting foveation is distinguishable by the user,
- Whether or not there is an effect of reduced discomfort based on the accommodation convergence conflict.

These questions were asked and mostly favorably answered, but there are still relatively few studies and these results need confirmation.

6.3.2. Specific Photogrammetric Tasks

Although our implementation follows principles that would be required by photogrammetric tasks and is suitable for such systems, an additional future work would be to integrate this application into a functioning photogrammetric workstation attached to a large screen stereo display and test its real time usage for a defined photogrammetric visualization case.

6.3.3. Comparison of Alternative Foveation Techniques

This research currently uses a distance dependent model with circular and spherical geometry. Different types of LOD functions can be integrated into the system by adding the models that are documented in the literature. This would give an opportunity to compare the existing approaches to foveation and potentially bring out novel combinations and hybrid algorithms.

The potential applications of foveation in geovisualization, in 3D cinema, in 3D web graphics and visual attention management as discussed in Chapter 6 can also be listed as future investigations based on our work.

REFERENCES

Oh Lord, forgive the misprints!
Andrew Bradford, American book publisher

- Ahnelt, P. K., Kolb, H. and Pflug, R., 1987. Identification of a subtype of cone photoreceptor, likely to be blue sensitive, in the human retina. *J. Comp. Neurol.*, 255, 18-34.
- Allen, E., 2003. M.Sc. in Digital Imaging lecture material online, Lecture3: The Human Visual System , The University of Westminster web site, <http://www.wmin.ac.uk/itrg/is/msc/>, last visited 7.12.2005
- Andrews, P.R. and Campbell, F.W., Images at the Blind Spot. *Nature*. Vol. 353(6342). Pp. 308. 1991.
- Anstis, S.M., 1974. A chart demonstrating variations in acuity with retinal position. *Vision Research*, 14: 589-592
- Answers.com, 2005. <http://www.answers.com/topic/cave>, last visited 7.12.2005
- Aslin, R.N., 1993. Infant accommodation and vergence. In: Simons K ed. *Early Visual Development, Normal and Abnormal*. Oxford:Oxford University Press; 1993:30-38.
- Atkinson, K.B. 1996. Editor. *Close Range Photogrammetry and Machine Vision*, Whittles Publishing, ISBN 1-870325-46-X
- Barfield, W., Rosenberg, C., 1995. Judgements of Azimuth and Elevation as a Function of Monoscopic and Binocular Depth Cues Using a Perspective Display. *Human Factors* 37(1), (pp.173-181).
- Basu, A., Sullivan, A., Wiebe, K.J., 1993. Variable resolution teleconferencing. In *IEEE Systems, Man, and Cybernetics Conference Proceedings*, pp. 170-175.
- Basu, A., Cheng, I., Pan, Y., 2002. Foveated Online 3D Visualization. 16th International Conference on Pattern Recognition (ICPR'02), Canada, Volume 3, p.30944.
- Been, K., 2002, *Responsive Thinwire Visualization of Large Geographic Datasets*, Ph.D. Thesis, Department of Computer Science, New York University. Sept. 2002.

- Bernardino, A., Santos-Victor, J., 1999. Binocular Visual Tracking: Integration of Perception and Control, VisLab-TR 10/99 IEEE Transactions on Robotics and Automation, (15)6, December 1999
- Bernardino, A., Santos-Victor, J., 2002. A Binocular Stereo Algorithm for Log-polar foveated systems, 2nd workshop on Biologically Motivated Computer Vision, Tuebingen, Germany, Nov. 2002.
- Birchfield, S., Tomasi, C., 1996. Depth Discontinuities by Pixel-to-Pixel Stereo. Stanford University Technical Report STAN-CS-TR-96-1573, July 1996
- Birchfield, S. 2003. Depth Discontinuities by Pixel-to-Pixel Stereo. The code and the related publications are available online in this web site:
<http://vision.stanford.edu/~birch/p2p/>, last visited on 7.12.2005
- Birchfield, S., Tomasi, C. 1998. Depth Discontinuities by Pixel-to-Pixel Stereo, Proceedings of the Sixth IEEE International Conference on Computer Vision, Mumbai, India, pages 1073-1080, January 1998
- Birchfield, S., Tomasi, C., 1998. A Pixel Dissimilarity Measure that is Insensitive To Image Sampling, IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(4):401-406, April 1998
- Blake, A., Yuille, A. (Eds.), 1992. Active Vision, ISBN 0-262-02351-2
- Blanke, W. and Bajaj, C., 2002. Active Visualization in a Multidisplay Immersive Environment Eighth Eurographics Workshop on Virtual Environments (2002) S. Müller, W. Stürzlinger (Editors)
- Blohm, W., Beldie, I. P., Schenke, K., Fazel, K., Pastoor S., 1997. Stereoscopic image representation with synthetic depth of field. Journal of the Society for Information Display (SID).
- Bolduc, M., Levine, M.D., 1997. A Real-time Foveated Sensor with Overlapping Receptive Fields, Real-Time Imaging 3, pp.195-212, 1997.
- Bolduc, M., Levine Martin D., 1998. A Review of Biologically Motivated Space-Variant Data Reduction Models for Robotic Vision, Computer Vision and Image Understanding, Vol. 69, No. 2, February, pp. 170–184, 1998, Article No. IV970560
- Boyling, T. and Siebert, J. P., 2000. A Fast Foveated Stereo Matcher. Proceedings of Conference on Imaging Science Systems and Technology (CISST 2000), Las Vegas, USA, pp. 417-423.
- Boyling, T.A., Siebert J.P., 2004. Foveated Vision for Space-Variant Scene Reconstruction, Proceedings of ISPRS 2004, Istanbul

Brooker, J. P., Sharkey, P. M., 2001. An operator performance evaluation of controlled depth of field in a stereographically-displayed virtual environment, Proceedings of SPIE 2001

Bourke, P., Bannon. D. 2002. A portable rear projection stereoscopic display
A VPAC (Victorian Partnership for Advanced Computing) project
April-October 2002
<http://astronomy.swin.edu.au/~pbourke/stereographics/vpac/theory.html>, last visited on 7.12.2005.

Buser, P. & Imbert, M. 1992. Vision. Cambridge, Massachusetts: The MIT Press.

Cabral, B. 1997. OpenGL Optimizer 1.0: The Power of Silicon Graphics' Next-Generation Visualization Technology. DeveloperNews.

Campbell, F.W., and Green, D.G., 1965. Monocular versus binocular visual acuity. Nature 208: 191-192

Carr, H. A., 1935. An Introduction to Space Perception. Longmans, Green. (New York).

Chang, E.C., Yap, C., 1997a. A Wavelet Approach to Foveating Images, Proceedings of 13th ACM Symposium on Computational Geometry, pp. 397-399, 1997. For a demo see: <http://cs.nyu.edu/visual/home/demos/foveation/applet/mand.html>, last visited 7.12.2005

Chang, E.C., Yap, C., and Yen, T.-J, 1997b. Realtime visualization of large images over a thinwire. In IEEE Visualization '97 (Late Breaking Hot Topics), pages 45--48, 1997. See CD proceedings of conference. Also online at <ftp://cs.nyu.edu/pub/local/yap/visual/thinwire.ps.gz>, last visited 7.12.2005.

Chang, E.C., 1998. Foveation Techniques and Scheduling Issues in Thinwire Visualization. PhD Dissertation, Department of Computer Science, New York University.

Chang, E.C., Mallat, S., Yap, C., 2000. Wavelet Foveation, Applied and Computational Harmonic Analysis 9, 312-335 (2000). Doi:10.1006/acha.2000.0324, available online at <http://www.idealibrary.com>

Clark, J.H., 1976. Hierarchical geometric models for visible surface algorithms. Communications of the ACM 19(10), 547-554.

Clark, R.N, 1990. Visual Astronomy of the Deep Sky, Cambridge University Press and Sky Publishing, 355 pages, Cambridge.

Clark, R.N., 2005. Resolution of the Human Eye. Clarkvision web site: <http://clarkvision.com/imagedetail/eye-resolution.html>, last visited on 7.12.2005

- Constantinescu, 2001, Levels of Detail: An Overview, First NTNU CSGSC, May 2001
- Cooper, M.A.R., Robson, S., 1996. "Theory of Close Range Photogrammetry" in Close Range Photogrammetry and Machine Vision, Ed. Atkinson, K.B, Whittles Publishing, ISBN 1-870325-46-X
- Coren, S., Ward, L., Enns, J., 1999. Sensation & Perception. Harcourt Brace, New York, Ny, 1999.
- Çöltekin, A., 2005. Stereo-Foveation for Anaglyph Imaging. Proceedings of Stereoscopic Displays and Virtual Reality Systems XII Conference, IS&T/SPIE's 17th Annual Symposium, Electronic Imaging, Science and Technology, 16-20 January 2005, San Jose, California, USA. Proceedings of SPIE Vol. #5664
- Çöltekin, A., 2005. A Visualization Method Based On Foveation. Proceedings of XXII International Cartographic Conference (ICC2005), A Coruna, Spain, 11-16 July 2005 ISBN: 0-958-46093-0
- Curcio, C. A., Sloan, K. R., Packer, O., Hendrickson, A. E. and Kalina, R. E., 1987. Distribution of cones in human and monkey retina: individual variability and radial asymmetry. Science 236, pp. 579-582.
- Crowley, J. L., Bobet P. and Schmid C., 1993. Auto-Calibration of Cameras by Direct Observation of Objects, Image and Vision Computing, Vol 11, no. 2, March 1993.
- Crowley, 2005. CVOnline, online Computer Vision course, http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/ECVNET/CROWLEY1/node1.2.html, last visited on 7.12.2005.
- Cruz-Neira, C., Sandin, D.J., DeFanti T.A., 1993. Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE, SIGGRAPH 1993 Proceedings. Also online at: <http://www.evl.uic.edu/EVL/RESEARCH/PAPERS/CRUZ/sig93.paper.html> (last visited on 7.12.2005)
- CVOnline, 2005. by Price. S., Panum's Fusional Area. Online at http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/MARBLE/medium/stereo/panum.htm, last visited on 7.12.2005. A subsection of CVonline: On-Line Compendium of Computer Vision [Online]. R. Fisher (ed). Available: <http://homepages.inf.ed.ac.uk/rbf/CVonline/>, last visited on 7.12.2005.
- Debevec, P.E., 1996. Modeling and Rendering Architecture from Photographs, from Photographs, Ph.D. Thesis, University of California at Berkeley
- Diner, D.B. and Fender, D.H., 1993. Human Engineering in Stereoscopic Display Devices Plenum Press, 1993.

- Dinstein, J. R. I., Guy, G., 1988. "On stereo image coding," in Proc. IEEE 9th International Conference on Pattern Recognition, 1988, Nov. 1988, pp. 357-359.
- Dubois, E., 2001. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '01), Salt Lake City, UT, USA. Publication Date: 7-11 May 2001 Volume: 3 On page(s): 1661 - 1664
- Drascic, D. and P. Milgam. 1991. Positioning accuracy of a virtual stereographic pointer in a real stereoscopic video world. SPIE Stereoscopic Displays and Applications II, 1457:58-69.
- Drasdo, N., 1977. The neural representation of visual space. Nature 266: 554-556.
- Espiau, B., Chaumette, F. and Rives P., 1992. New Approach to Visual Servoing in Robotics", IEEE Trans. on Robotics and Automation, 8(3):313-326, June 1992
- Estes, J.E., Hemphill J., 2005. Lecture notes on Introduction to Photo Interpretation and Photogrammetry, By Remote Sensing Core Curriculum, Initiated by ASPRS, Ebook, Vol. I, can be found online at: <http://www.r-s-c-c.org/index.html>, last visited on 7.12.2005.
- Ferwerda, J. A., 2001. Elements of Early Vision for Computer Graphics, IEEE Computer Graphics and Applications, September/October 2001, pp.22-33
- Fisher, R., Perkins, S., Walker, A., Wolfart, E., 2003. Affine Transformation. Published at <http://homepages.inf.ed.ac.uk/rbf/HIPR2/affine.htm>, last visited on 7.12.2005
- Fok, S. 2002. Foveated Stereo Video Compression for Visual Telepresence, Master's thesis, University of Waterloo, Electrical and Computer Engineering
- Fryer, J., 1996. Introduction (Chapter 1) and Camera Calibration (Chapter 2) chapters in Close Range Photogrammetry and Machine Vision book, Ed. Atkinson, K.B, Whittles Publishing, ISBN 1-870325-46-X
- Gahegan, M., 2000. Visualization as a tool for GeoComputation. In GeoComputation. Edited by Stan Openshaw and Robert J. Abraham. Taylor & Francis, London. ISBN 0-7484-0900-9
- Geisler, W. S., Perry, J. S., 1998. A real-time foveation multi-resolution system for low-bandwidth video communication. Proceedings of SPIE Human Vision & Electronic Imaging, Volume 3299.
- Geisler, W.S. and J.S. Perry. 1999. Variable-resolution displays for visual communication and simulation. The Society for Information Display, 30:420-423.

- Gemshein H. and Gemshein, A., 1969. The History of photography from the Camera Obscura to the Beginning of the Modern Era. New York, NY McGraw-Hill.
- Gillam, B., 1995. The Perception of Spatial Layout from Static Optical Information. In W. Epstein & S. Rogers (Eds.), Perception of Space and Motion, pp. 23-67. New York: Academic Press.
- Godding, R., 2002. Geometric Calibration and Orientation of Digital Imaging Systems, AICON 3D Systems GmbH, Germany, 1999. Online at: http://www.falcon.de/falcon/pdf/eng/aicon/geometric_calibration.pdf, last visited on 7.12.2005.
- Goldstein, E.B., 2002. Sensation and Perception (textbook), 6th Edition, Wadsworth Group, ISBN 0-534-53964-5
- Goodchild, M.F., 2001. Models of Scale and Scales of Modelling. In Modelling Scale in Geographical Information Science. Edited by Nicholas J. Tate, Peter M. Atkinson. John Wiley and Sons, ISBN 0-471-98546-5
- GPGPU 2005, General-Purpose Computation Using Graphics Hardware, located online at <http://www.gpgpu.org>, last visited 7.12.2005
- Graham, C., 1951. Visual Perception. In S. Stevens (Ed.), Handbook of Experimental Psychology, pp. 868-920. New York: John Wiley & Sons.
- Grinberg, V.S., Podnar G., Siegel, M. W., 1994. "Geometry of Binocular Imaging," Stereoscopic Displays and Virtual Reality Systems, Vol. 2177, February, 1994, pp. 56 - 65. Also online at: <http://www.cs.ucsb.edu/~mturk/Courses/290A%20Spring%202001/papers/Geometry1.pdf>, last visited on 7.12.2005
- Haggrén, H., 2005. Stereoscopy application of spherical imaging. Videometrics VIII Conference, Proceedings of SPIE Vol. #5665, pp. 89-95.
- Hecht, 1931. The Retinal Processes Concerned with Visual Acuity and Color Vision,' Bulletin No. 4 of the Howe Laboratory of Ophthalmology, Harvard Medical School, Cambridge, Mass.
- Heipke, C., 1996. Overview of Image Matching Techniques, Proceedings of the OEEPE Workshop on Application of Digital Photogrammetric Workstations, Lausanne, 4-6 March 1996, Ed. O. Kölbl – Paper can also be found online with a discussion attachment at http://phot.epfl.ch/workshop/wks96/art_3_1.html 7.12.2005
- Helmholtz, H., 2000. Treatise on physiological optics 1867, 1924 edition reprinted Thoemmes Press, 2000

- Hendrickson, A.E. and Youdelis, C., 1984. The morphological development of the human fovea. *Ophthalmol.* 91, 603-612.
- Hitchner, L.E. and McGreevy, M.W., 1993. Methods for User-Based Reduction of Model Complexity for Virtual Planetary Exploration, *Proceedings of the SPIE – The International Society for Optical Engineering*, Vol. 1913, pp. 622– 36.
- Hoffman, D.D., 2000. *Visual Intelligence, How We Create What We See*, p.202, ISBN 0--393-31967-9 pbk. 2000, W.W. Norton & Company Ltd. New York, NY 10110
- Hoppe, H., 1996. Progressive Meshes. *ACM SIGGRAPH 1996*, pages 99-108.
- Halle, M., 1997. Autostereoscopic displays and computer graphics. In *Computer Graphics*, ACM SIGGRAPH, volume 31, pages 58-62, May.
- Holliman, N., 2005. 3D Display Systems, *Handbook of Optoelectronics*, IOP Press, ISBN 0 7503 0646 7. Also online at:
<http://www.dur.ac.uk/n.s.holliman/Presentations/3dv3-0.pdf>, last visited on 7.12.2005.
- Howarth, P.A., 1996. Empirical Studies of Accommodation, Convergence, and HMD Use Hosobunka Foundation Symposium, Tokyo, December 3 1996.
 Online at: <http://www.lboro.ac.uk/departments/hu/groups/viserg/hosobunk.htm>, last visited on 7.12.2005.
- Howarth, P.A. and Costello, P.J., 1997. The occurrence of virtual simulation sickness symptoms when an HMD was used as a personal viewing system. *Displays*, 18:107-116.
- Hubona, G.S., Wheeler, P.N., Shirah, G.W. and Brandt, M., 1999. The relative contributions of stereo lighting, and background scenes in promoting 3d depth visualisation. *ACM Transactions of Computer-Human Interaction*, 6(3):214{242, September.
- Hung, G.K., 1997, Quantitative Analysis of the Accommodative Convergence to Accommodation Ratio: Linear and Nonlinear Static Models, *IEEE Transactions On Biomedical Engineering*, Vol. 44, No. 4, April 1997
- Huk, A. “Seeing in 3D” Lecture Notes - 07/15/99. Online at:
<http://www-psych.stanford.edu/~lera/psych115s/notes/lecture8/> last visited on 7.12.2005.
- Inoue, T., and Ohzu, H., 1990. Measurement of the human factors of 3-D images on a large screen. *Large-Screen Projection Displays II*, SPIE Vol.1255, 1990.
- Intel Research, 2005, Silicon, Moore’s law,
<http://www.intel.com/technology/mooreslaw/index.htm>, last visited on 7.12.2005

ITU-R, 1990, ITU-R (International Telecommunication Union Radiocommunication) BT.709, Basic Parameter Values for the HDTV Standard for the Studio and for International Programme Exchange [formerly CCIR Rec. 709], ITU, 1211 Geneva 20, Switzerland. Also see:
http://www.poynton.com/notes/colour_and_gamma/ColorFAQ.html#RTFTtoC9. last visited on 7.12.2005

Jain, R., Kasturi, R., Schunck, B.G., 1995, Machine Vision, textbook, McGraw-Hill Inc. ISBN 0-07-113407-7

Jokinen, O., 1994. Reconstruction of quadric surfaces from disparity measurements, SPIE Vol.2298 p.593, 0-8194-1622-3/94

Joyce, T., 2003. "Saws", Trevor Joyce's collected poems, 1966-2000, published as "With the first dream of fire they hunt the cold". In 2003, Toronto press.

Kang, S.B., Webb, J.A., Zitnick, C.L., Kanade, T., 1994. An Active Multibaseline Stereo System with Real-Time Image Acquisition. Technical Report. School of Computer Science, Carnegie Mellon University Pittsburgh, PA 15213-3891 CMU-CS-94-167, also online at; <http://www-2.cs.cmu.edu/afs/cs/usr/webb/html/tech-rep94.html>, last visited on 7.12.2005

Kaufman, L., 1974. Sight and mind: An introduction to visual perception. New York: Oxford University Press.

Katz, E., 2005. The history of electrochemistry, electricity and electronics, Famous Scientists, <http://chem.ch.huji.ac.il/~eugeniik/history/wheatstone.html>, last visited 7.12.2005

Kim, W. S., Tendick, F., Stark, L., 1993. Visual Enhancements in Pick-and-Place Tasks: Human operators controlling a simulated cylindrical manipulator. In R. Ellis, M. Kaiser, J. Grunwald (Eds.), Pictorial Communication in Virtual and Real Environments. (2nd ed.). (pp. 265-282). Taylor & Francis (London).

Kirschenbauer, S., 2005. Applying "True 3D" Techniques to Geovisualization: An Empirical Study. Exploring Geovisualization, J. Dykes, A.M. MacEachren, M-J. Kraak (Editors). © 2005, Elsevier Ltd. ISBN: 0-08-044531-4 (Hardbound)

Kolasinski, E.M., 1995. Simulator sickness in virtual environments. Technical report, U.S. Army Research Institute for the Behavioral and Social Sciences, Technical Report 1027 (Public Release).

Kolb, H., Fernandez, E., Nelson, R., 2001. WEBVISION The organization of the vertebrate retina, Part XIII, Facts and Figures concerning the Human Retina, E-book, online at <http://webvision.med.utah.edu/>

Konig, 1897. Die Abhängigkeit der Schärfe von der Beleuchtungsintensität, S. B. Akad. Wiss. Berlin, 559-575.

Kortenkamp, D., Huber, E., Wasson, G., 1998. "Integrating a behavior based approach to active stereo vision with an intelligent control architecture for mobile robots", Book Chapter in Hybrid Information Processing in Adaptive Autonomous Vehicles, eds. Gerhard K. Kraetzschmar and Gunther Palm. Springer-Verlag.

Kortum, P., Geisler, W. 1996. Implementation of a foveated image coding system for image bandwidth reduction. In Human Vision and Electronic Imaging, SPIE Proceedings, Vol. 2657, pp. 350-360, 1996.

Kosara, R., Miksch, S., Hauser, H., 2001. Semantic Depth of Field Proceedings of the 2001 IEEE Symposium on Information Visualization (InfoVis 2001), pp. 97-104, IEEE Computer Society Press, 2001

Kuniyoshi, Y., Nobuyuki, K., Rougeaux, S., Suehiro T., 1995. Active Stereo Vision System with Foveated Wide Angle Lens – Proceedings of 2nd Asian Conference on Computer Vision, pp.191-200, 1995, Singapore

Krivanek J., Zara J., Kadi B., 2003. Fast Depth of Field Rendering with Surface Splatting, In Proceedings of Computer Graphics International 2003. Los Alamitos: IEEE Computer Society Press, 2003, p. 196-201. ISBN 0-7695-1946-6.

Larsen, B.D., Christensen, N.J., 2003. Real-time Terrain Rendering using Smooth Hardware Optimized Level of Detail, Journal of WSCG, Vol.11, No. 2, pp. 282-9 ISSN 1213-6972, WSCG'2003: 11th International Conference in Central Europe on Computer Graphics, Visualization and Digital Interactive Media

Linde, van der, I., 2003. Space-variant Perceptual Image Compression for Gaze-contingent Stereoscopic Displays, PhD Thesis, Faculty of Science and Technology, Department of Computing, APU. Jan 2003

Linde, van der, I., 2004. Multi-resolution image compression using image foveation and simulated depth of field for stereoscopic displays, Proceedings of SPIE Volume: 5291, Stereoscopic Displays and Virtual Reality Systems XI Editor(s): Woods, Andrew J.; Merritt, John O.; Benton, Stephen A.; Bolas, Mark T. Published: 5/2004

Lipton, L. 1982. Foundations of the Stereoscopic Cinema, A Study in Depth, by Van Nostrand Reinhold Company Inc. Library of Congress Catalog Card Number 81-1 6474, ISBN 0-442-24724-9 -- Electronic edition of "Foundations of the Stereoscopic Cinema" downloaded from the virtual library of the Stereoscopic Displays and Applications conference website <http://www.stereoscopic.org>, last visited 7.12.2005.

Lee, S., Pattichis, M. S., Bovik, A. C., (1998), Foveated video quality assessment and compression gain, Technical Report UT-CVIS-TR-98-002, Center for Vision and Image Sciences, University of Texas at Austin.

Leymarie, F.F., 2000. "Theory of Close Range Photogrammetry". Lecture notes. An extract from the book chapter of Atkinson 1996. Online at: <http://www.lems.brown.edu/vision/people/leymarie/Refs/Photogrammetry/Atkinson90/Ch2Theory.html>. Last updated May 2000. Last visited 30.7.2005.

Luebke, D., Reddy, M., Cohen, J.D., Varshney A., Watson, B., Huebner, R., 2003. Level of Detail for 3D Graphics. Textbook. Morgan Kaufmann, series in Computer Graphics and Geometric Modeling. ISBN 1-55860-838-9. URL: <http://www.lodbook.com/>, last visited 7.12.2005.

Marieb, E.N., 2000. Essentials of Human Anatomy & Physiology. Benjamin/Cummings (imprint of Addison Wesley Longman), 6 edition.

Martens, W.L, McRuer, B., Childs, C. T. and Virree, E., 1996. "Physiological approach to optimal stereographic game programming: A technical guide," in IS&T/SPIE Proceedings: Stereoscopic Displays and Virtual Reality Systems III, (San Jose, CA), pp. 261-270, 1996.

McAllister, D.F., 2002. "3D Displays," Wiley Encyclopedia on Imaging, January 2002, pp. 1327-1344. Also online at: <http://research.csc.ncsu.edu/stereographics/wiley.pdf>, last visited on 7.12.2005.

McHugh, S., 2005. Cambridge In Color, Digital Photography Tutorials. Published at: <http://www.cambridgeincolour.com/tutorials/depth-of-field.htm>, last visited on 28.8.2005.

McKay, H., 1953. Three-Dimensional Photography Principles of Stereoscopy, American Photography Book Department, New York, N. Y., Copyright 1953 The American Photographic Publishing Company. (This electronic edition of "Three-Dimensional Photography - Principles of Stereoscopy" was downloaded from the virtual library of the Stereoscopic Displays and Applications conference website <http://www.stereoscopic.org>, last visited 7.12.2005. This document was converted to electronic format by Andrew Woods, Curtin University of Technology.)

McLean, G. F., Prescott, B., Podhorodeski, R., 1994. Teleoperated System Performance Evaluation. IEEE Transaction on System, Man and Cybernetics, 24(5), (pp.796-803).

Middlebury 2005, 2005. Middlebury College Stereovision Research page. Located at: <http://www.middlebury.edu/stereo/>, last visited 7.12.2005. The work is by Scharstein and Szeliski and the document is closely tied with Scharstein et al., 2002 and Scharstein et al., 2003 which are also cited in this list.

Mikhail, E.M., Bethel J.S., McGlone, J.C., 2001. Introduction to Modern Photogrammetry. John Wiley and Sons, ISBN 0-471-30924-9

Min, P. 1994. Stereoscopy optimization for head mounted displays (MSc. thesis). Master's thesis, Leiden University, Physics and Electronics Laboratory of the Dutch Organisation for Applied Research.

Mon-Williams, M & J. P. Wann, 1998. Binocular virtual reality displays: when problems occur and when they don't *Human Factors* 40, 18-24.

Moore, G.E., 1965. Cramming more components onto integrated circuits, *Electronics*, Volume 38, Number 8, April 19, 1965. Also published online at: <ftp://download.intel.com/research/silicon/moorespaper.pdf>, last visited on 7.12.2005.

Mulder, J., van Liere, R., 2000. Fast PerceptionBased Depth of Field Rendering, *Proceedings of the ACM symposium on Virtual reality software and technology*, Seoul, Korea, Pages: 129 – 133, ISBN:1-58113-316-2

Nagata, S., 1993. How to Reinforce Perception of Depth in Single Two-Dimensional Pictures. In S. Ellis (Ed.), *Pictorial Communication in Virtual and Real Environments*, pp. 527-545. London: Taylor & Francis.

Nakayama, K., 1990. Properties of early motion processing: Implications for the sensing of egomotion. In R. Warren and A.H. Wertheim, editors, *The Perception and Control of Self Motion*, pages 69--80. Lawrence Erlbaum, Hillsdale, NJ.

Nave, C.R., 2005. HyperPhysics (©C.R. Nave, 2005)
<http://hyperphysics.phy-astr.gsu.edu/hbase/vision/rodcone.html>, last visited on 7.12.2005

Ohshima, T., Yamamoto, H. and Tamura, H., 1996. Gaze-Directed Adaptive Rendering for Interacting with Virtual Space, *Proceedings of the IEEE Virtual Reality Annual International Symposium (VRAIS)*, Santa Clara, CA, pp. 103–110.

Osterberg, G., 1935. Topography of the layer of rods and cones in the human retina. *Acta Ophthal.*, suppl. 6, 1-103.

Ostnes, R., Abbott, V., lavender, S., 2004. Visualisation Techniques: An Overview – Part 1, *The Hydrographic Journal* No. 113 July 2004

Ottoson, P. 2001. Geographic Indexing and Data Management for 3D Visualization, PhD Thesis, Royal Institute of Technology, KTH, ISBN 91-7283-174-X

Overall, B., 1999. Foveated Imaging.
<http://ise.stanford.edu/class/psych221/projects/99/wro/intro.html>, last visited on 7.12.2005.

Owens, R., 1997. Online Computer Vision Lecture Notes, University of Edinburgh, Science and Engineering Department, http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/OWENS/LECT11/node4.html, last visited on 7.12.2005.

Panerai, F., Metta G., Sandini G., 2000. Visuo-inertial stabilization in space-variant binocular systems, *Robotics and Autonomus Systems* 30, (2000) 195-214, Elsevier.

Pastoor, S. 1995. Human factors of 3d imaging: Recent research at Heinrich-Hertz-Institut, Berlin. In *Conference Proceedings of the 2nd International Display Workshop, Hamamatsu*, pages 69-72.

Pastoor, S., Wöpking, S., 1997. 3-D displays: A review of current technologies," *Displays* 17, pp. 100-110, 1997. Also online at: <http://www.dgp.toronto.edu/~gf/Research/Volumetric%20UI/3-D%20Displays%20A%20review%20of%20current%20technologies.htm>, last visited 7.12.2005.

Parkhurst, D., Niebur, E., 2004. *Proceedings of the 1st Symposium on Applied perception in graphics and visualization Los Angeles, California*, Pages: 49 - 56 ISBN: 1-58113-914-4

Patterson, R., and Martin, W.L., 1992. Human Stereopsis, *Human Factors*, 1992-34(6) 669-692.

Peters R.A., Bishay M., 1996. Centering peripheral features in an indoor environment using a binocular log-polar 4 DOF camera head, *Robotics and Autonomous Systems* 18 (1996) 271-281, Elsevier.

Perkins, M., 1992. Data Compression of Stereo pairs. *IEEE Transactions on Communications*, Vol. 40 No. 4, April 1992.

Perry, J., Geisler W. S., 2002. Gaze-contingent real-time simulation of arbitrary visual fields, *Human Vision and Electronic Imaging, Proceedings of SPIE 2002*, San Jose.

Pfautz, J.D., 2000. *Depth Perception in Computer Graphics. Doctoral Dissertation*, University of Cambridge, Cambridge, UK, May 2000.

Pfautz, J.D., 2002. *Depth Perception in Computer Graphics. Technical Report*, University of Cambridge, Cambridge, UK, May 2000.

Piranda, B., de Sorbier, F., Arquès, D., 2005. Simulation of blur in stereoscopic image synthesis for virtual reality, WSCG 2005 posters. *Proceedings of The 13th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2005*, January 31-February 4, 2005, Plzen, Czech Republic. Copyright UNION Agency – Science Press

- Pirenne, M. H., 1967. *Vision and the eye* (2nd ed.). London: Chapman and Hall.
- Piper, H.F., 1999. Peter Ludvig Panum's sensory physiological works from his years in Kiel 1853-1864. *Klin Monatsbl Augenheilkd.* 1999 Aug;215(2):73-7. Article in German. Abstract in English, online at http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10483554&dopt=Abstract. Last visited 7.12.2005.
- Pollefeys, M., 2004. *3D Photography / Multiple View Geometry in Computer Vision* Instructor: Marc Pollefeys comp290-89 Fall 2004, can be found online at: <http://www.unc.edu/courses/2004fall/comp/290/089/>, last visited 7.12.2005.
- Polyak S.L., 1941. *The Retina*. The University of Chicago Press, Chicago, Illinois.
- Pöntinen, P., 2004. *On the Creation of Panoramic Images from Image Sequences*, Licentiate Thesis, Helsinki University of Technology, 30th of January 2004
- Puolamäki, K., 2004. *Information Visualization Lecture Notes* in Helsinki University of Technology. Material based on Ware, 2000 and Goldstein 2002.
- Rao, R.P.N., G.J. Zelinsky, M.M. Hayhoe, and D.H. Ballard. 1997. *Eye movements in visual cognition : A computational study*. Technical Report 97.1, National Resource Library for the Study of Brain and Behaviour, University of Rochester, March.
- Reddy, M. 1997. *Perceptually Modulated Level of Detail for Virtual Environments*. Ph.D. Thesis. University of Edinburgh. Also online at: <http://www.martinreddy.net/percept/> last visited on 7.12.2005
- Reddy, M., 1998. *Specification and Evaluation of Level of Detail Selection Criteria*. *Virtual Reality: Research, Development and Application*, 3(2): 132-143. © Springer–Verlag, London Ltd.
- Reddy, M., Iverson, L., Leclec, L.C., Heller A., 2001. *GeoVRML: Open Web-based 3D Cartography*, In proceedings of International Cartographic Association, ICA Conference 2001, Beijing. Also see <http://www.geovrml.org>, last visited 7.12.2005.
- Reddy, M., 2001. *Lecture notes on Advanced Issues in Level of Detail, Visual Perception and Level of Detail at Siggraph 2001*, Online at: <http://lodbook.com/course/2002/talk5.reddy.perception.pps> page last visited: June 2005
- Richards. W., 1970. *Stereopsis and Stereoblindness Exp.* *Brian Res.* 10, 380-388, 1970.
- Riddell, P.M. and Bullinaria, J.A., 1999. *Incorporating Developmental Factors into Models of Accommodation and Vergence*. University of Reading Technical Report
- Robinett, W., 1999. *Exploring Virtual Worlds, Human Perception and Virtual Environments*, Lecture notes, Computer Science Department University of North Carolina at Chapel Hill

- Rougeaux, S., Berthouze, L., Chavand, F., Kuniyoshi, Y., 1996. Calibration of a Foveated Wide Angle Lens on an Active Vision Head, Proceedings Conference of Computer Vision and Pattern Recognition, San Francisco, June 1996.
- RSCC 2005. Estes, J.E. and Hemphill, J., Remote Sensing Core Curriculum, Volume 1, Presented by The International Center for Remote Sensing Education, Online at; <http://www.r-s-c-c.org/rscc/v1m7.html>, page last visited on 7.12.2005.
- RSL 2005. High-Performance Active Vision, Robotics Systems Laboratory Web Site, The Australian National University, http://wwwsyseng.rsise.anu.edu.au/rsl/rsl_active.html, last visited on 7.12.2005.
- Salmon, T.O., 2005. Vision Science III - Binocular Vision Module, Lecture notes, Online at: http://arapaho.nsuok.edu/~salmonto/VSIII_2005/Lecture6.pdf, last visited on 7.12.2005.
- Sanchez, V., Basu, A., Mandal, M. K., 2004. Prioritized region of interest coding in JPEG2000, Circuits and Systems for Video Technology, IEEE Transactions on Publication Date: Sept. 2004 Volume: 14, Issue: 9 On page(s): 1149- 1155, ISSN: 1051-8215
- Sanghoon, L., Bovik, A.C., Kim, Y.Y., 2005. High Quality Low Delay Foveated Visual Communication over mobile channels. Journal of Visual Communication and Image Representation, Vol. 16 (2005) 180-211
- Scharstein, D., Szeliski, R., 2002. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms, IJCV 47(1/2/3): 7-42, April-June 2002.
- Scharstein, D., Szeliski, R., 2003. High-accuracy stereo depth maps using structured light. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003), volume 1, pages 195-202, Madison, WI, June 2003.
- Schenk, T., 1999. Digital Photogrammetry, Vol.1, Background, Fundamentals, Automatic Orientation Procedures, TerraScience, ISBN 0-9677653-0-7 (perfect bound), ISBN 0-9677653-1-5 (case bound).
- Schermann, J., Barron, J.L., Gargantini, I.A., 2000. 3D Foveated Visualization on the Web. Proc. SPIE Vol. 4311, p. 349-360, Internet Imaging II, Giordano B. Beretta; Raimondo Schettini; Eds
- Scherman, W.R., Craig, A.B., 2003. Understanding Virtual Reality, p.10, Morgan Kaufmann Publishers, ISBN 1-55860-353-0

Schreer, O., Brandenburg, N., Askar, S., Kauff P. 2001. Hybrid Recursive Matching and Segmentation-Based Postprocessing in Real-Time Immersive Video Conferencing. Proceedings of conference on Vision, Modelling and Visualization (VMV01), November 21-23 2001, Stuttgart, Germany, p. 383-390
Paper can also be found here: <http://wwwvis.informatik.uni-stuttgart.de/vmv01/dl/papers/29.pdf>, last visited on 7.12.2005

Schwartz, E.L., 1977. Spatial mapping in primate sensory projection: analytic structure and relevance to perception. *Biological Cybernetics*, 25:181–194, 1977.

Schwartz, E.L., 1977. The development of specific visual projections in the monkey and the goldfish: outline of a geometric theory of receptotopic structure. *J.Theoret.Biol.*, 69:655–685, 1977.

Scott E., Bewley, H., 2005. Color Vision 1. Online at:
<http://www.photo.net/photo/edscott/vis00010.htm>, last visited 7.12.2005

Seuntiëns, P., Meesters, L., IJsselsteijna, W. A., 2003. SStereoscopic Displays and Applications XIIIV (2003) Proceedings of the SPIE Volume 5006

Sharp 3D 2005, <http://www.sharp3d.com/> last visited 30.7.2005

Sharp 2004, Press Release, June 10, 2004. Also see:
<http://www.i4u.com/article184.html>, last visited on 7.12.2005

Sheikh, H. R., Liu, S., Evans, B.L., Bovik, A.C., 2001. Real-Time Foveation Techniques For H.263 Video Encoding in Software. *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing*, May 7-11, 2001, vol. 3, pp. 1781-1784, Salt Lake City, UT. Paper also online at: <http://www.ece.utexas.edu/~bevans/papers/2001/foveation/>, last visited 7.12.2005

Sheikh, H. R., Evans, B.L., Bovik, A.C., 2003. Real-Time Foveation Techniques for Low Bit Rate Video Coding, *Journal on Real-Time Imaging*, vol. 9, no. 1, pp. 27-40. Feb. 2003. Paper also online at:
<http://www.ece.utexas.edu/~bevans/papers/2003/realTimeFoveation/>, last visited 7.12.2005

Skerjanc, R. and Pastoor, S., 1997. New generation of 3D desktop computer interfaces. IEE & SPIE, *Electronic Imaging*, San Jose.

Sutherland, I. 1968. A Head Mounted Three-dimensional Display. In *Proceedings of the Fall Joint Computer Conference*, Thompson Books, Washington, DC, pp. 757-764

Stereographics, 1997. *Stereographics Developers' Handbook, Background on Creating Images for CrystalEyes® and SimulEyes®, StereoGraphics Corporation*

- Stereoscopy 2004. 3D News. <http://www.stereoscopy.com/news/news-archive-6-2004.html>, last visited on 7.12.2005.
- Tan, S., Dale, J.L., Johnston, A., 2003. Performance of three recursive algorithms for fast space-variant Gaussian Filtering, *Real-Time Imaging* 9 (2003) 215–228
- Talbot, S.A., Marshall, W.H., 1941. Physiological studies on neural mechanisms of visual localization and discrimination. *American Journal of Ophtalmology*, 24: pp.1255-1263
- Tidwell, M., Johnston, R.S., Melville, D. and Furness, T.A., 1995. The virtual retinal display - a retinal scanning imaging system. In *Proceedings of Virtual Reality World*, pages 325-333.
- Turk, G., Levoy M., 1994. "Zippered Polygon Meshes from Range Images" *Siggraph* 94, pp. 311-318 – Can also be found online at <http://www.cc.gatech.edu/~turk/zipper/zipper.html>, last visited on 7.12.2005
- UTEXAS 2005. Center for Perceptual Systems, The University of Texas at Austin URL: http://svi.cps.utexas.edu/foveated_questions.htm, last visited on 7.12.2005.
- Wallace, R.S., Bederson, B.B. and Schwartz, E.L., 1992. Voice bandwidth visual communication through logmaps: The Telecortex. In *Workshop on Applications of Computer Vision*, pages 4–10, 1992.
- Wann, J.P. Rushton, S.K. & Mon-Williams M., 1995. Natural problems for stereoscopic depth perception in Virtual Environments. *Vision Research*, 35, 2731-2736.
- Wann, J.P. & Rushton S.K. & Lee D.N., 1995. Can you control where you are heading when you are looking at where you want to go? In B. Bardy, R. Bootsma & Y. Guiard. (Eds) *Proc. of the 8th International Conference on Event Perception and Action*. 171-174
- Ware C., 1995. Dynamic Stereo Displays. In *CHI'95 Conference Proceedings, Human Factors and Computing Systems*, pages 310-316.
- Ware, C., Gobrecht, C., Patton, M.A., 1998. Dynamic Adjustment of Stereo Display Parameters, *IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Humans*, Vol. 28, pp. 56-65, No. 1, January 1998
- Ware, C. 2000. *Information Visualization - Perception for Design*. Academic Press, Morgan Kaufmann Publishers. ISBN 1-55860-511-8
- Ware, C. 2004. *Information Visualization - Perception for Design*. Second edition. Elsevier Inc., Morgan Kaufmann Publishers. ISBN 1-55860-819-2

Ware, C., 2005. Interactive Data Visualization Course Notes.
URL: <http://www.ccom.unh.edu/vislab/VisCourse/VR.html>, last visited on 7.12.2005.

Ware, C. 2005. online lecture notes: OE/CS Interactive Data Visualization
HCI Course, CS760/860 Human-Computer Interaction,
<http://www.ccom.unh.edu/vislab/projects/Stereo.html>, last visited on 7.12.2005

Watson, B., Walker, N., Hodges, L.F., 1996. Effectiveness of Spatial Level of Detail
Degradation in the Periphery of Head-Mounted Displays APRIL 13-18, 1996 CH1996

Watson, B., Walker, N. and Hodges, L. F., 1995. A User Study Evaluating
Level of Detail Degradation in the Periphery of Head-Mounted Displays,
Proceedings of the FIVE '95 Conference, QMW, University of London, UK,
pp. 203–212.

Ward, M., 2005. Advanced Topics in Computer Graphics Course, Talks, RenderMan
“Introduction to RenderMan”. Online at
<http://web.cs.wpi.edu/~matt/courses/cs563/talks/renderman/ri.ov.html>, last visited
7.12.2005.

Wikipedia, 2005. The Free Encyclopedia, Online at:
http://en.wikipedia.org/wiki/Digital_divide, last visited on 7.12.2005.

Wikipedia Map, 2005. The Free Encyclopedia,
http://en.wikipedia.org/wiki/The_map_is_not_the_territory, last visited on 7.12.2005.

Webvision 2005. Perception: http://webvision.med.utah.edu/space_perception.html, last
visited on 7.12.2005.

Webvision 2005. Vieth-Muller: <http://webvision.med.utah.edu/imageswv/Space8.jpg>,
last visited on 7.12.2005.

Webvision 2005. Space Perception:
http://webvision.med.utah.edu/space_perception.html, last visited on 7.12.2005

Vincent, T., 2005. Fundamentals of Stereo Computer Vision. Online course notes.
http://egweb.mines.edu/tvincent/Welding/fundamentals_of_stereo_computer.htm, last
visited 7.12.2005

Yamamoto H., Yeshurun Y., Levine M.D., 1996. An Active Foveated Vision System:
Attentional Mechanisms and Scan Path Coverage Measures, Computer Vision and
Image Understanding, January 1996, vol. 63, no. 1, pp. 50-65(16)

Yamamoto, H., Yeshurun Y., Levine, M.D., 1996. An Active Foveated Vision System:
Attentional Mechanisms and Scan Path Coverage Measures, Computer Vision and
Image Understanding, Vol. 63, No.1, January, pp. 50-65, 1996, Article No: 0004

Yaşayan, A., 1996. Fotogrametri 1 Ders Notları (Photogrammetry-I lecture notes, in Turkish). Yıldız Teknik Üniversitesi.

Yeh, Y., 1992. Spatial Judgement with Monoscopic and Stereoscopic Presentation of Perspective Displays. Human Factors, 34(5), (pp. 583-600).

Yeh, C-P. 1993. Cyclopean stereo vision for depth perception. SPIE Intelligent Robots and Computer Vision XII, 2056.

Ziemann H. And El-Hakim S.F., 1982. System Calibration vs. Self-calibration. Internat archives of photogrammetry, 24(1):117-122

APPENDICES

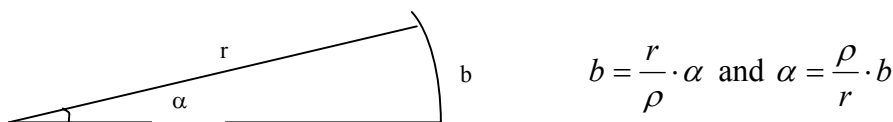
Appendix 1: An explanation of Arc Minutes

However trivial, the background of the potential readers of this thesis might be from many different fields. Therefore they might not be too familiar working with arc minutes. Here we include a brief explanation of the unit and its metric interpretation.

As referred from Riddell et al., 1999:

An arc second or a second of arc is equal to exactly 1/3600 of an angular degree or 1/1,296,000 of a circle. Sixty arc seconds comprise an arc minute; 60 arc minutes comprise an angular degree.

How to calculate the metric equivalents of arc lengths as given in Albertz and Kreiling, 1989:



R	α	1°	1'	1''
1 m		0.0175 m	0.29 mm	0.0048 mm
10 m		0.1745 m	2.91 mm	0.0485 mm
100 m		1.7453 m	29.09 mm	0.4848 mm
1000 m		17.4533 m	290.89 mm	4.8481 mm

References

Albertz J. and Kreiling, W., 1989. Photogrammetrisches Taschenbuch (Photogrammetric Guide, multilingual pocket book), ISBN 3-87907-176-4

Riddell, P.M. and Bullinaria, J.A., 1999. Incorporating Developmental Factors into Models of Accommodation and Vergence. University of Reading Technical Report

Appendix 2: Glossary

active vision a term coined for a process where the camera optics and configuration are actively controlled in order to simplify the remaining tasks in computer and robot vision.

accommodation the change in focal length or optical power of the eye produced by change in the power of the crystalline lens as a result of contraction of the ciliary muscle. This capability decreases with age (2).

area 17 (see *primary visual cortex*)

base the distance between the two viewing media such as eyes or cameras.

cone see *cone cell*

cone cell the name for photoreceptor cells in the retina which only function in relatively bright light. These cells are sensitive to color. There are about 6 million in the human eye, concentrated at the fovea and gradually becoming sparser towards the outside of the retina (also see *rods*).

convergence the coordinated turning of the eyes inward to focus on an object at close range (7).

circle of confusion closely related to *depth of field* and often abbreviated to COF, circle of confusion is a photography term that defines the limit of how fuzzy a point can be but still considered in focus. This value is often calculated as the largest circle on the film that will still be seen as a point when enlarged to 8"x10" and viewed from a normal viewing distance (2-3 feet). Anything larger is seen as a small circle, not a point and is therefore perceived as out of focus. For 35mm format the diameter of such point or circle is 0.025mm, commonly rounded to 0.03mm (6)

crosstalk describes the unwanted perspective view that is presented to each eye in a stereoscopic display system. In a perfect stereoscopic system, each eye sees only its assigned image (modified from Lipton et al., 1997).

cyclopean eye a virtual eye located exactly in between the two eyes. Modeling the sight from this viewpoint is said to have advantages in reducing the discomfort related to stereoscopic viewing.

cyclopean scale a term describing a method that scales the 3D scene from a midpoint of the two views in a stereoscopic visualization task. It is based on the *cyclopean eye* idea.

depth cue the sources of information to be found in our environment that allow us to perceive depth. "The term, cues, has been utilised to formalise the specification of stimulus conditions for space perception" (Carr 1935, via Ostnes et al. 2004).

depth of field the distance from behind an object to in front of the object within which objects appear to be in focus. In this thesis, we use depth of field to express the depth in the scene. Also see *depth of focus* and *depth range*.

depth of focus range of lens to image plane distance for which the image formed by the lens appears to be in focus. Commonly confused, this is different than *depth of field*.

depth range according to Lenny Lipton, “depth of field is an optically defined concept that applies to monocular settings. Depth range would be the range stereoscopic sight is possible”. However, the common practice in the literature shows us the term might be used for generic definition while the term depth of field, even depth of focus at times is used to express stereoscopic depth range.

diopter a unit of measure of the refractive power of a lens, equal to the power of a lens with focal distance of one meter (1).

diplopia objects in our field of view but outside the binocular fusion area (panum’s fusional area) appear double. We do not normally register this effect because the peripheral vision occurs in lower resolutions. In stereoscopic displays that may become a problem if it is not in proportion with the biological perception.

disparity the state of being different or dissimilar (as in the sensory information received) (3). This word, as a term, has been used to express what is known as parallax in photogrammetry by the computer vision community and shortly for retinal disparity by biological vision researchers. For a discussion see the entry for *parallax* in this glossary.

dynamic programming a method for reducing the runtime of algorithms exhibiting the properties of overlapping subproblems and optimal substructure (9).

epipolar geometry The geometry of stereo. Each point in the left image is restricted to lie on a given line in the right image, the epipolar line--and vice versa. This is called the epipolar constraint. Epipoles are the points at which the line through the centers of projection of each image intersects the image planes. The left epipole is the image of the center of projection of the right camera and vice versa (10).

frame cancellation this term is used for describing the phenomena that the stereo effect is cancelled by the apparent occlusion caused by the monitor frame.

ganglion cell (or **gangliocyte**) a type of neuron located in the retina that receives visual information from bipolar cells; its axons give rise to the optic nerve (11).

ghosting When an image is dragged across a computer screen, a lingering shadow of the image where it was before (12).

hyperacuity term used to describe the phenomenon that certain stimuli can be perceived which are smaller than the size of a single photoreceptor cell (Reddy, 1997).

hyperstereo a generic stereoscopic term used extensively by aerial photogrammetrists to express the idea of using larger than usual *base* to acquire stereoscopic images of remote scenes (e.g. terrain or cities). Also see virtual eye separation and hypostereo.

hypostereo opposite of hyperstereo. Where needed, the base is kept smaller than usual. Also see virtual eye separation and hypostereo.

image z -buffer see *z-buffer*.

lossy an adjective used in image compression as in “lossy compression”. This indicates loss of information during the compression and the resulting image does not have the full information to restore the original image. Also see *lossless* and *perceptually lossless*.

lossless an adjective used in image compression as in “lossless compression”. When compressing an image, while reducing the file size, if the original data was preserved precisely the resulting image is referred to as lossless. Also see *perceptually lossless* and *lossy*.

parafovea area of the retina immediately surrounding the fovea (13).

parafoveal regarding or relating to parafovea.

parallax the apparent displacement of an object caused by a change in the position from which it is viewed (Stedman 2002 via dictionary.com). This term is used interchangeably with **disparity** even though some scholars distinguish between the two by saying disparity should refer to the displacement on retina, while parallax should refer to the same thing on screen (e.g. “screen disparity”).

perceptually lossless when compressing an image, if the perceptual issues are taken into account so that the resulting image does not appear any different to the viewer, the result is referred to as perceptually lossless. The gain is like in *lossy* compressions or better because in the areas not registered by human perception, gross reductions can be done – but the result, in terms of its usability is like *lossless*, which normally is valuable for tasks when precision is needed, but does not provide as much compression gain as the *lossy* techniques.

primary visual cortex an area of the occipital lobe that performs the first stage of cortical visual processing. It receives inputs from the **retina** and sends outputs to other areas of the **visual cortex**. Also referred to as *v1*, *striate cortex*, and *area 17*.

photoreceptor(s) see *photoreceptor cells*

photoreceptor cells contained in the retina, these cells are responsible for transducing, or converting, light into signals that can be ultimately transmitted to the brain via the optic nerve. *Rods* and *cones* are photoreceptor cells (14).

psychophysics the branch of perception that is concerned with establishing quantitative relations between physical stimulation and perceptual events (15).

rods or **rod cell** rod cells are photoreceptor cells in the retina that function in less intense light. These cells are achromatic. Rods are named for their cylindrical shape. They are concentrated at the outer edges of the retina. There are about 120 million rod cells in the human retina (also see *cones*) (16).

screen disparity see *parallax*.

space variant if an image is not uniform resolution but its resolution changes throughout the spatial dimensions that image is referred to as space variant.

superacuity see *hyperacuity*.

texel the fundamental element of a texture map, a texel is a pixel on a texture. For example, an 128x128 texture has 128x128 texels. On screen this may result in more or fewer pixels depending on how far away the object is on which the texture is used and also on how the texture is scaled on the object (17).

visual angle the angle subtended at the eye by the linear extent of an object in the visual field. It determines linear retinal image size (18).

visual field the angular region of space or field of view limited by the entrance pupil of the eye, the zone of functional retina, and occlusion structures such as the nose and orbit of the eye.

virtual eye separation to control the range of disparities, adjusting the base for stereo viewing. e.g. while the human eyes in average are separated by a 65mm, “in viewing a mountain 10 km distant a virtual eye separation of 1 km might be appropriate. If viewing an object at 1 cm (as in a stereo microscope) a virtual eye separation of 1 mm will be more suitable” (Ware, 1998). Also see hyperstereo and hypostereo.

virtual simulator sickness immersion in a virtual environment can lead to adverse effects on its user including nausea, headache, dizziness and disorientation as well as detrimental effects on the eyes.

VSS see *virtual simulator sickness*





window violation when using a stereo display screen in a fixed position, the illusion of depth is destroyed where objects are occluded by the screen boundaries. This is known as **window violation**. This does not occur in HMDs. (Linde, 2003)

z-buffer the image z-buffer is matrix of values providing the depth of each of the pixels in the image, having the same dimensions as the image bitmap, commonly used in 3d graphics to avoid rendering occluded objects. The image z-buffer can also be queried to determine the depth of any screen pixel (Linde, 2004).

References for this Glossary

1. Schor, C., Optical Society Of America (Author). Handbook Of Optics, Vol. III. Blacklick, OH, Usa: Mcgraw-Hill Professional, 2000. P 309. ISBN: 0071414789 Copyright © 2000. Mcgraw-Hill Professional. All Rights Reserved.
2. Understanding And Applying Machihe Vision, Second Edition, Revised And Expanded, Nello Zuech, ISBN 0-8247-8929-6, 1. Computer Vision. 2. Applying Machine Vision. I. Zuech, Nello. Ii. Title. Iii., Series. TA1634.Z84 2000 Copyright © 2000 By Marcel Dekker, Inc. All Rights Reserved.
3. www.dictionary.com
4. Stedman 2002: The American Heritage® Stedman's Medical Dictionary Copyright © 2002, 2001, 1995 by Houghton Mifflin Company.
5. Circle of confusion: http://www.startphoto.com/learn/glossary/glossary_ch-cn.htm and http://en.wikipedia.org/wiki/Circle_of_confusion
6. Cone cell: http://www.wikipedia.org/Cone_cell
7. Convergence: <http://www.answers.com/topic/convergent-evolution>
8. Depth cue <http://psych.hanover.edu/Krantz/art/cues.html>
9. Dynamic Programming: http://en.wikipedia.org/wiki/Dynamic_programming
10. Epiplar Geometry:
http://www.cs.jhu.edu/~jcorso/classes/computer_vision/trucco_verri_outline.html
11. Ganglion cell: http://en.wikipedia.org/wiki/ganglion_cell
12. Ghosting:
<http://www.computeruser.com/resources/dictionary/definition.html?lookup=2410>
13. Parafovea <http://www.wordreference.com/definition/parafovea>
14. Photoreceptor Cells: http://www.wikipedia.org/Photoreceptor_cell
15. Pshychophysics: http://highered.mcgraw-hill.com/sites/0070579431/student_view0/chapter1/glossary.html
16. Rods: http://www.wikipedia.orgrod_cell
17. Texel: homepages.inf.ed.ac.uk/rbf/grdict/grdict.htm and www.crystalspace3d.org/docs/online/manual/cs_a.php
18. Visual angle: http://www.hfeconsulting.com/expert_witness/glossarystoz.html

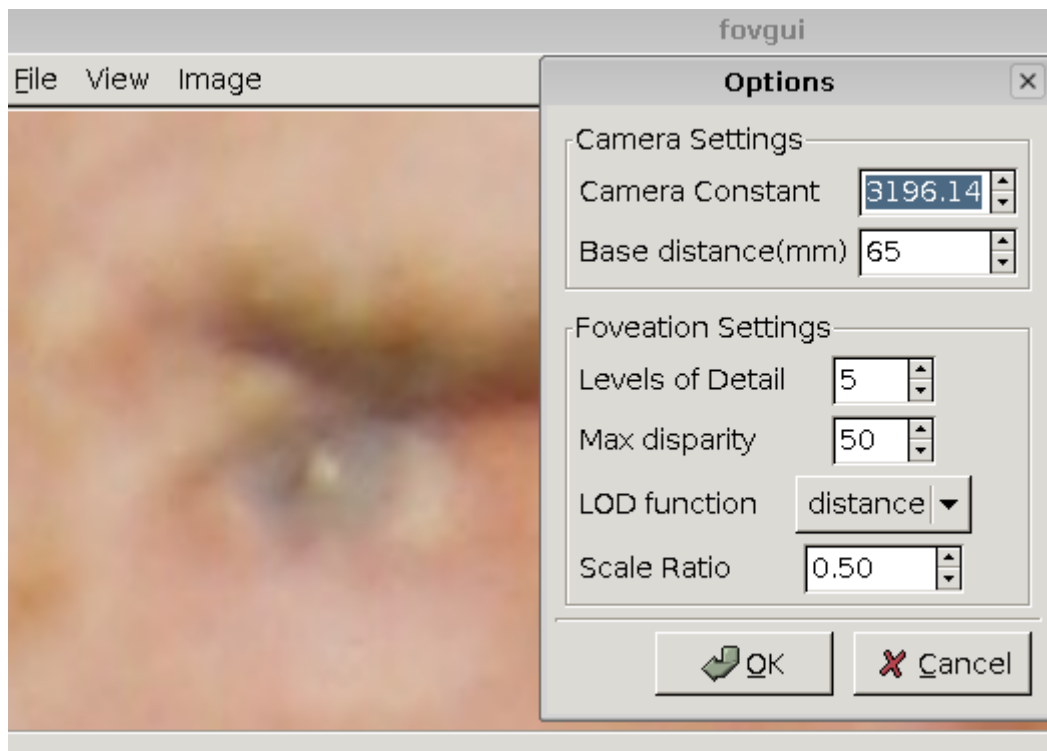
Appendix 3: Index of Test Images

<p>birchf-clorox (630x480 pixels) (a)</p> 	<p>calib-field (3137x2084 pixels) (b)</p> 
<p>furniture (3008x2000 pixels) (c)</p> 	<p>gymball (3008x2000 pixels) (d)</p> 

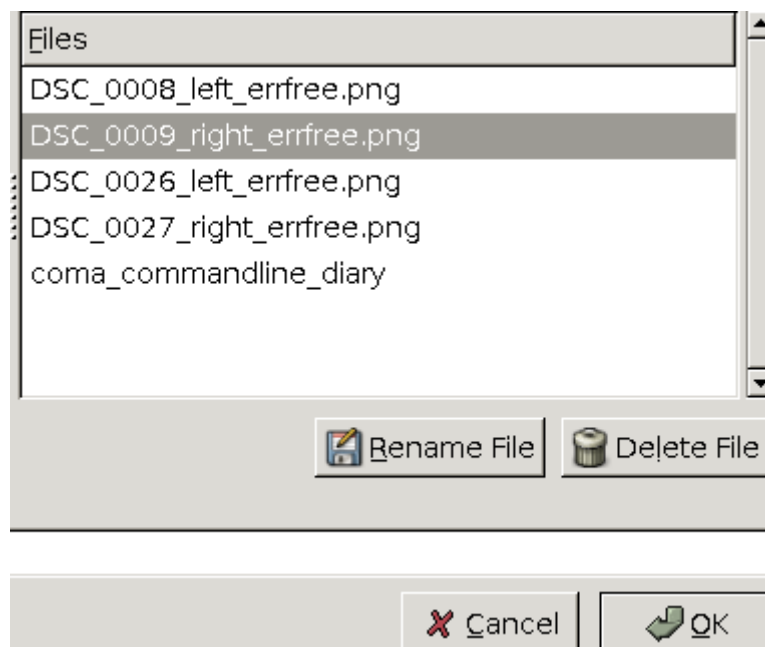
An index of the test images. Some were mentioned only by name in the text. While the original resolutions of each image is noted in the table, they were resized (down scaled) when the tests required them. If the image was used with a lower resolution, it is noted in the relevant section of the body text.

Appendix 4: The GUI Menus of *Foveaglyph*

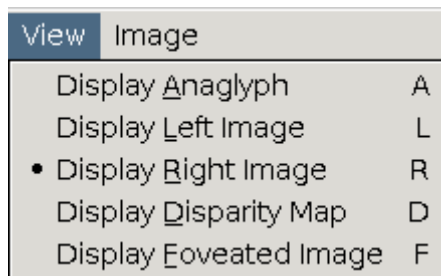
Foveaglyph's GUI version is called *FovGUI*. It looks like this:



A general look of *FovGUI*.



Loading images in *FovGUI*..



View menu pull-down list.

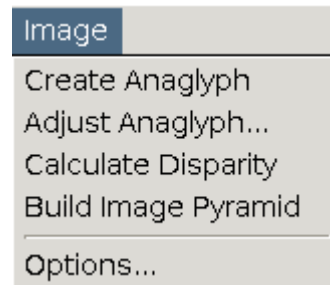


Image menu pull-down list.

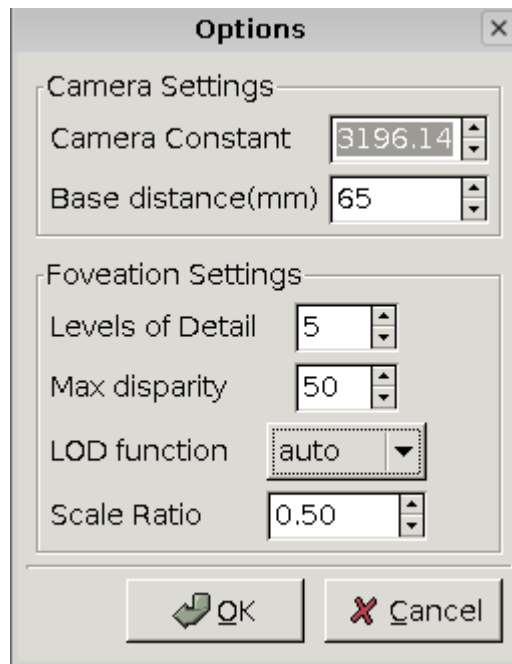
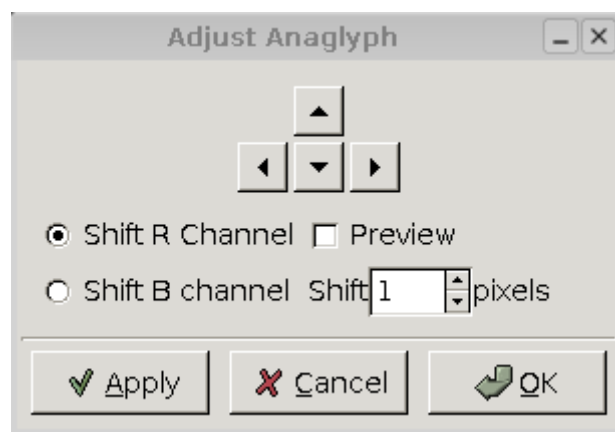


Image options dialog box.

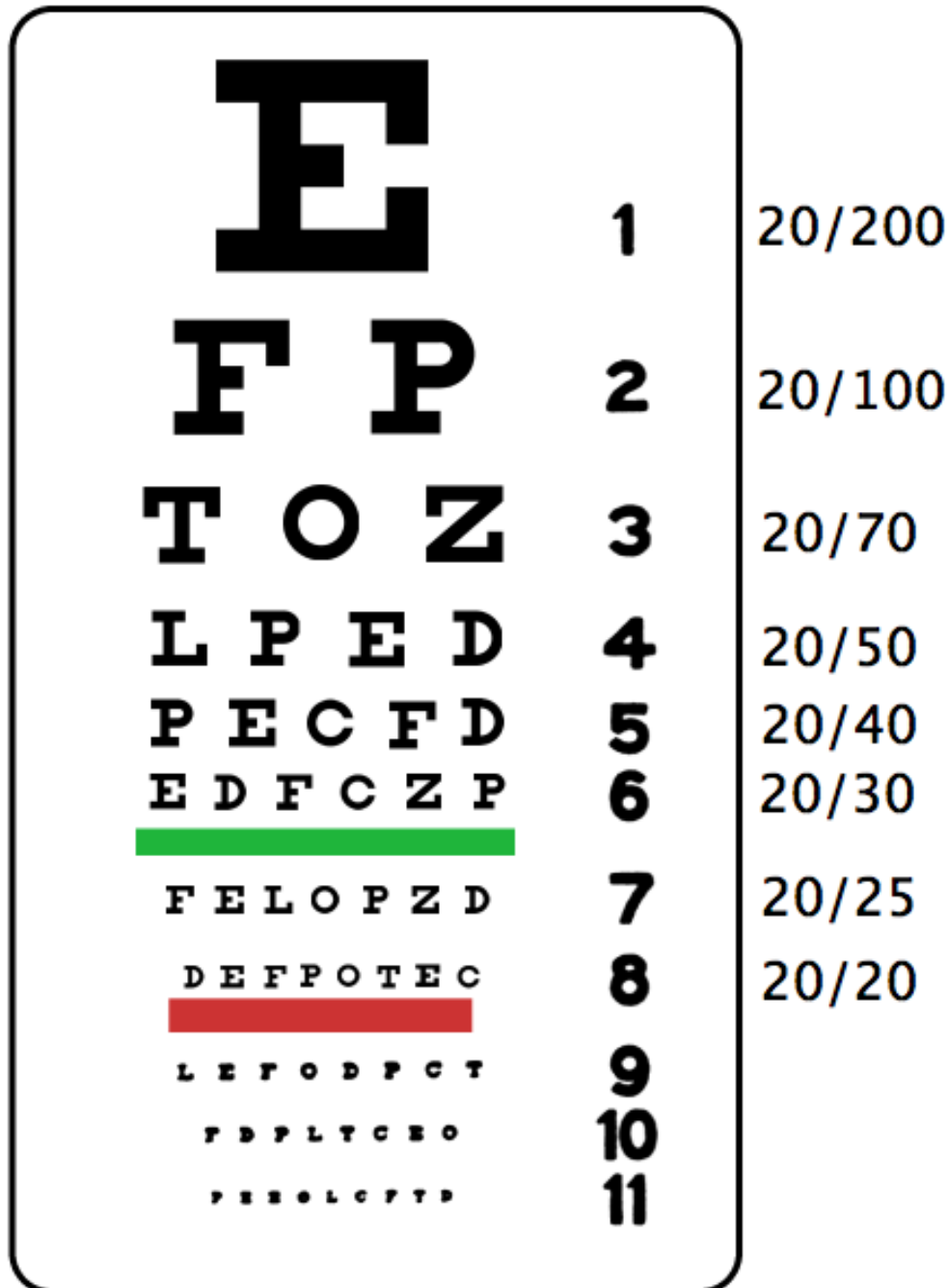


Adjusting channels to remove the extra parallax is allowed, if needed.

2032x755(#3E5858) scale: 0.30 disparity: 67 real coords: 1.97,0.73,3.10
 At the bottom bar of the window, some information about the cursor position is printed.
 The “real coords” are read from the database, they are the calculated XYZ, not the screen pixels.

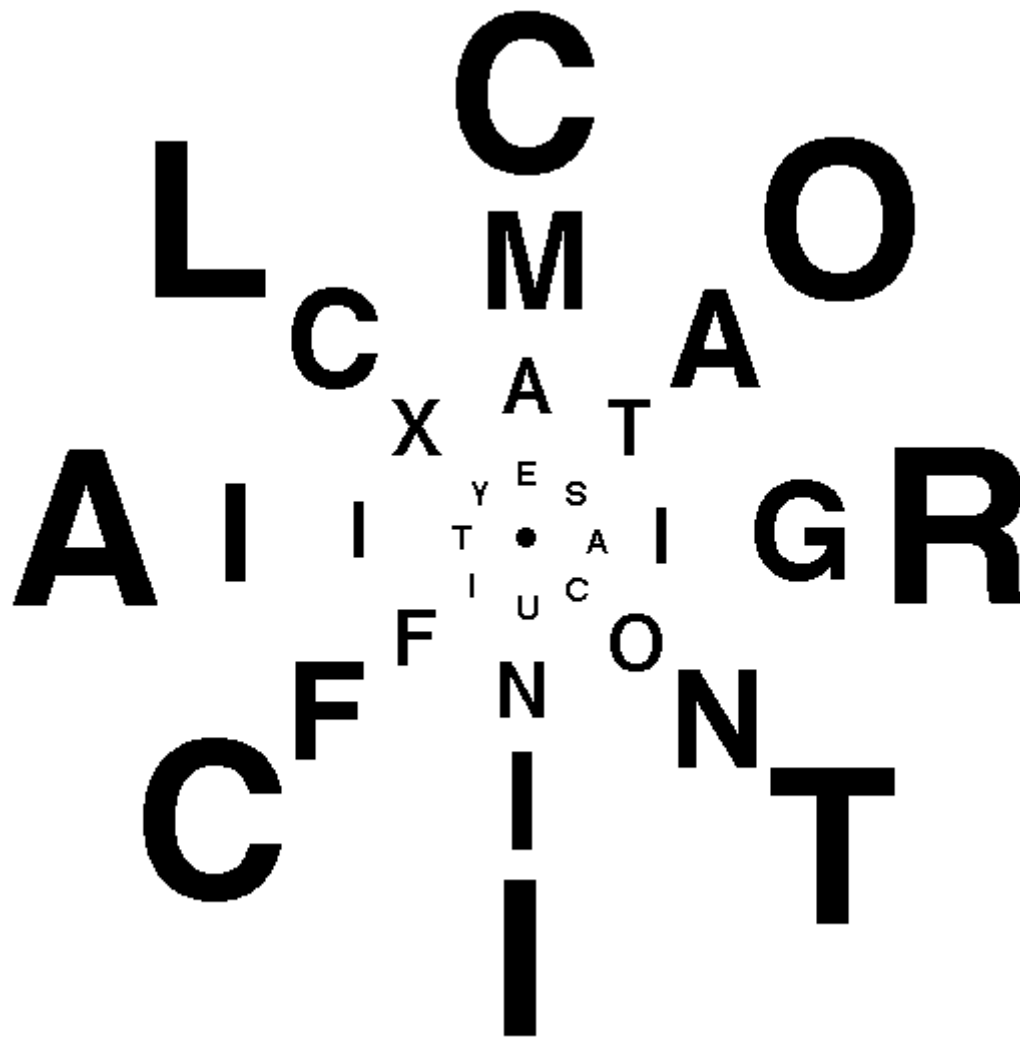
Appendix 5: Snellen Eye Chart

Developed by a Dutch opthalmologist, Hermann Snellen, Snellen Chart today is used for measuring visual acuity world-wide.



Snellen Chart. If a person can read the 8th row without optical aid at a 6-meter distance, she or he is considered to have perfect sight. This is often expressed as 20/20 (sight/feet) or 6/6 (sight/meters). This image is public domain and downloaded from http://en.wikipedia.org/wiki/Snellen_chart.

Appendix 6: Radial Eye Chart



Radial eye chart developed by Anstis, 1974. Ware (2004) uses the same figure with this caption: “Each character is about five times smallest perceivable size when the center is fixated. This is the case for any viewing distance.”

(Ware, 2004)

EPILOGUE

Vladimir: Was I sleeping, while the others suffered? Am I sleeping now? Tomorrow, when I wake, or think I do, what shall I say of today? That with Estragon my friend, at this place, until the fall of night, I waited for Godot? That Pozzo passed, with his carrier, and that he spoke to us? Probably. But in all that what truth will there be? (Estragon, having struggled with his boots in vain, is dozing off again. Vladimir looks at him) He'll know nothing. He'll tell me about the blows he received and I'll give him a carrot. (Pause) Astride of a grave and a difficult birth. Down in the hole, lingeringly, the grave digger puts on the forceps. We have time to grow old. The air is full of our cries. (He listens) But habit is a great deadener. (He looks again at Estragon) At me too someone is looking, of me too someone is saying, He is sleeping, he knows nothing, let him sleep on. (Pause) I can't go on! (Pause) What have I said?

--Samuel Beckett, *En attendant Godot* (Waiting for Godot), 1952

ISBN 951-22-8016-7 (printed)
ISBN 951-22-8017-5 (PDF)
ISSN 1796-0711