



Provided by the author(s) and NUI Galway in accordance with publisher policies. Please cite the published version when available.

Title	Advances in the testing of stereo image acquisition devices
Author(s)	Andorko, Istvan
Publication Date	2014-01-04
Item record	http://hdl.handle.net/10379/4169

Downloaded 2017-10-29T23:55:34Z

Some rights reserved. For more information, please see the item record link above.





National University of Ireland, Galway

Department of Electrical and Electronic Engineering

ADVANCES IN THE TESTING OF STEREO IMAGE ACQUISITION DEVICES

by

Mr. Istvan Andorko

Supervisor: Dr. Peter Corcoran

A thesis submitted to:

National University of Ireland, Galway

for the degree of

DOCTOR OF PHILOSOPHY (PhD)

January, 2014

Contents

1	Introduction and Overview	1
1.1	Background and Motivations	1
1.2	Contributions	3
1.3	Thesis Outline	4
2	Foundation techniques	7
2.1	Stereo concepts	8
2.1.1	Stereo Images	8
2.1.2	Stereo Baseline	11
2.1.3	Parallax	12
2.1.3.1	Measurement of the disparity caused by the parallax	14
2.1.4	Depth Determination and Depth Maps	16
2.2	Literature Review of Image Signal Processors	18
2.3	Implementation of the Stereo ISP	23
2.3.1	Design Specifications	24
2.3.2	Testing the system	26
2.3.3	Applicability Example - 3D Camera	28
2.4	Challenges in Stereo Imaging	29
2.4.1	Image Registration Issues	29
2.4.2	Asymmetrical colors	30
2.4.3	Image Acquisition Synchronization Problems	32
2.4.4	Out-of-Focus Regions	32
2.5	Conclusions	34
3	Literature Review of Depth Map Generation Techniques	35
3.1	Depth Determination using a Single Camera	36

3.1.1	Depth from Defocus	36
3.2	Depth Determination using a Stereo Camera Setup	38
3.2.1	Correlation based Techniques	38
3.2.1.1	Normalized Cross-Correlation	39
3.2.1.2	Absolute differences-based correlation	39
3.2.1.3	Squared Differences-based Correlation	41
3.2.1.4	Sum of Hamming Distances-based Correlation	42
3.2.2	Global Optimization based Techniques	42
3.2.3	Segmentation based techniques	44
3.2.4	Real-Time and Hardware-based Methods	46
3.3	Conclusions	49
4	Testing Methodology	51
4.1	Introduction	52
4.2	Description of Correlation Method-based Algorithms proposed for Test- ing	55
4.2.1	SAD Matching Cost	57
4.2.2	SSD matching cost	57
4.2.3	NCC matching cost	58
4.2.4	SHD matching cost	58
4.3	Measuring the Algorithm Performance	59
4.4	Ground Truth Generation	62
4.4.1	Methods used for Ground Truth Creation	62
4.4.2	Technique applied during the experiments	63
4.4.3	Discussion of the proposed Technique	64
4.5	Test Scene Setup	65
4.5.1	AWARD test scene setup	67
4.5.2	FACE Test Scene Setup	70
4.5.3	FACES Test Scene Setup	73
4.5.4	TEXTURED OBJECTS Test Scene Setup	75
4.6	Conclusions	78
5	Image Acquisition Methods and Preliminary Measurements	79
5.1	Devices used for Test Image Acquisition	80
5.1.1	Using the Fujifilm W3 3D Camera for Test Image Acquisition	80

CONTENTS

5.1.2	Using the uCam for Test Image Acquisition	82
5.1.2.1	Development board equipped with WLC cameras . .	82
5.1.2.2	Development Board equipped with Aptina Demo Cam- eras	83
5.1.3	Using a Single Camera for Test Image Acquisition	85
5.2	Choice of the Downsampling Method	86
5.2.1	Measurement of the MTF	86
5.2.2	Visual Assessment of the Algorithm Performance	90
5.3	Measurement of the Disparity caused by Parallax	91
5.4	Image Acquisition Issues that might influence the Results	96
5.4.1	The Influence of vertical Misalignment	96
5.4.2	Calibration of the Stereo Camera	97
5.4.3	The Influence of Noise	103
5.5	Conclusions	104
6	Experiment Results	105
6.1	Description of the Experimentation Process	106
6.1.1	Stage 1, Setting up of the Test Scene	106
6.1.2	Stage 2, Acquisition and Pre-Processing of the Images	108
6.1.3	Stage 3, Testing and Final Results	108
6.1.4	Practical Issues	111
6.2	Experiment Results using a fixed Stereo Base	112
6.2.1	Algorithm Behavior under different Light Conditions	112
6.2.2	Algorithm Behavior at different Camera-Object Distances . .	114
6.2.2.1	Initial Test using 4 Distances	114
6.2.2.2	Analysis of the Initial Results	118
6.2.2.3	Advanced Tests using 8 Camera-Object Distances . .	120
6.3	Experiment Results using a variable Stereo Base	124
6.3.1	Algorithm Behavior using variable Stereo Base Values	124
6.3.2	The Influence of Light Direction on the Algorithm Performance	126
6.3.2.1	Implementation of the Experiment	126
6.3.2.2	Analysis and Conclusions	127
7	General Test Scene Proposal	131
7.1	Introduction	132

7.2	Selection of the Objects	132
7.3	Creation of the Test Scene	134
7.3.1	Test Environment Setup	134
7.3.2	The scene setup	136
7.4	Proposed Test Scenarios	139
7.4.1	Test Image Acquisition at different Camera-Object Distances .	140
7.4.2	Test Image Acquisition under Different Light Intensity Condi- tions	143
7.4.3	Test image acquisition using different Stereo Base Length Values	146
7.5	Experiment Results	148
7.6	Ground Truth Generation	153
7.7	Explaining the Data Variations	154
7.7.1	Data Variations due to Image Registration	154
7.7.2	Data variations due to light	156
7.8	Conclusions	159
8	Conclusions and Future Research	161
8.1	Research Summary	161
8.2	Review of the Research Achievements	162
8.3	Summary of Published Papers	166
8.4	Future Research	167
8.5	Concluding Remarks	168
A	Test Images and resulting Depth Maps	169
A.1	Test images acquired for the AWARD test scene	169
A.2	Test images acquired for the FACE test scene	173
A.3	Test images acquired for the FACES test scene	177
A.4	Test images acquired for the TEXTURED OBJECTS test scene . . .	181
A.5	Depth maps generated using the AWARD test scene	185
A.6	Depth maps generated using the FACE test scene	187
A.7	Depth maps generated using the FACES test scene	189
A.8	Depth maps generated using the TEXTURED OBJECTS test scene .	191
B	Depth Map quality measurements	195
B.1	Depth map quality results. 8 distances, 5 stereo base values.	195
B.2	Depth map quality results. 12 distances, 5 stereo base values.	198

CONTENTS

C	Choice of Down Sampling algorithms measurements	201
C.1	MTF results for different down sampling algorithms	201
C.2	Downsampling of Zone Plate type of images	204
D	Details about cameras used for image acquisition	205
D.1	Fujifilm W3 3D	205
D.2	Tessera WLC Cameras	206
D.3	Aptina MT9P031	207
E	Proposed Test Scene	209
E.1	Different camera-object distances	209
E.2	Scene setup procedure	211
F	Data variation test results	215
F.1	Test results Nikon D50	215
F.2	Test results Canon EOS 600 D	218

List of Figures

2.1	Stereo image	8
2.2	Epipolar geometry	9
2.3	Convergence angle	9
2.4	Converging stereo camera system	10
2.5	Stereo imaging systems	10
2.6	Stereo base example	11
2.7	Parallax	12
2.8	Parallax effect	13
2.9	Parallax scene setup	13
2.10	Explanation of parallax and overlapped pixels	14
2.11	Horizontal field of view	15
2.12	Binocular vision in HVS	16
2.13	Disparity map example	17
2.14	Depth map calculation	17
2.15	Image Signal Processor	19
2.16	General system architecture	25
2.17	Internal system architecture	25
2.18	EMI Problem	27
2.19	EMI problem solved	27
2.20	The DLP 3D HDTV video format	28
2.21	The interlaced layer format	29
2.22	X, Y and Z planes explained	29
2.23	Camera module	31
2.24	Color artifacts	31
2.25	Depth of Field stereo camera	33
2.26	DoF example	33

3.1	ToF and stereo camera setup [152]	47
4.1	Middlebury database	53
4.2	Test pictures used in [118]	54
4.3	Corresponding pixel search	55
4.4	Combined stereo images	56
4.5	Census Transformation example	59
4.6	Depth Map quality measurement	60
4.7	Pixel selection for quality measurement	61
4.8	Middlebury image samples	61
4.9	Tsukuba image	62
4.10	Example of a Ground Truth image	64
4.11	Multiple depth layers within object	65
4.12	Vertical distance settings	67
4.13	AWARD scene setup	68
4.14	AWARD original scene, 4 distances	68
4.15	Light intensities AWARD	69
4.16	FACE scene setup	71
4.17	FACE original scene, 4 distances	71
4.18	Light intensities FACE	72
4.19	FACES scene setup	73
4.20	FACES original scene, 4 distances	74
4.21	Light intensities FACES	74
4.22	TEXTURED OBJECTS scene setup	76
4.23	TEXTURED OBJECTS original scene, 4 distances	76
4.24	Light intensities TEXTURED OBJECTS	77
5.1	Setup of the acquisition device	81
5.2	uCam using the WLC cameras	83
5.3	Stereo image pair acquired with WLC cameras	83
5.4	uCam using the Aptina cameras	84
5.5	Stereo image pair acquired with Aptina cameras	84
5.6	Using a single camera for stereo image acquisition	85
5.7	Slanted Edge image used for testing	87
5.8	MTF measurement example	88

LIST OF FIGURES

5.9	Zone plate type of test image	91
5.10	Horizontal field of view	92
5.11	Explanation of displacement and overlapped pixels	93
5.12	Parallax values for different stereo baselines	94
5.13	Alignment coordinates	97
5.14	Influence of vertical misalignment	98
5.15	Calibration coordinates	99
5.16	Calibration images	100
5.17	Calibration views	101
5.18	Calibrated Images vs Non-calibrated Images comparison	102
6.1	Tripod settings	107
6.2	View of the testing room	107
6.3	Artificial image	110
6.4	Artificial image - Disparity map	110
6.5	Light intensity comparison for the AWARD test scene	113
6.6	AWARD original scene, 4 distances	114
6.7	AWARD scene setup	115
6.8	AWARD test scene distance comparison	115
6.9	FACE scene setup	116
6.10	FACE Test Scene distance comparison	116
6.11	FACES scene setup	117
6.12	FACES Test Scene distance comparison	117
6.13	TEXTURED OBJECTS scene setup	120
6.14	AWARD and TEXTURED OBJECTS test scenes distance comparison	121
6.15	Comparison between 4 distance and 8 distance scenarios	123
6.16	Performance results for different stereo base length values, 8 distance scenario, NCC	125
6.17	Performance results for different stereo base length values, 8 distance scenario, SAD	125
6.18	Performance results for different stereo base length values, 12 distance scenario, NCC	127
6.19	Performance results for different stereo base length values, 12 distance scenario, SAD	128
6.20	Influence of the light on the test scene	129

LIST OF FIGURES

7.1	General scene layout	136
7.2	High precision rig used to set accurate stereo base lengths.	137
7.3	Proposed test scene	138
7.4	Setting up the scene	138
7.5	Physical limitation example. In the right image, parts of the table can be noticed.	142
7.6	Different camera-object distances	143
7.7	Different light intensities	144
7.8	Light direction	145
7.9	Light wall	145
7.10	Different stereo base settings	147
7.11	Example disparity maps	148
7.12	Algorithm performance using different stereo base values	149
7.13	Algorithm performance using different stereo base values, experiment 2	149
7.14	Different light intensity values	150
7.15	Different light intensity values, experiment 2	151
7.16	Influence of similar objects in a test scene	152
7.17	Influence of similar objects in a test scene, experiment 2	153
7.18	Stereo image pair and corresponding ground truth	154
7.19	Camera settings	156
7.20	Light intensity difference	157
7.21	Imatest results - normal exposure	157
7.22	Imatest results - underexposed	158
7.23	Uniformly illuminated scene	158

List of Tables

4.1	Award Scene Settings	69
4.2	FACE Scene Settings	72
4.3	FACES Scene Settings	75
4.4	TEXTURED OBJECTS Scene Settings	77
5.1	MTF measurement for a down scale factor of 2.4	89
5.2	MTF measurement for a down scale factor of 3.6	89
5.3	MTF measurement for a down scale factor of 6.2	89
5.4	MTF measurement for a down scale factor of 12.5	90
5.5	Distance table with disparity values for first test	94
5.6	Distance table with disparity values for steady scene test at 8 distances	95
5.7	Distance table with overlapped pixel values for initial test	95
5.8	Distance table with overlapped pixel values for steady scene test at 8 distances	95
5.9	Deviation values	103
6.1	Distance table for steady scene test with 8 distances	122
6.2	Stereo base values	124
6.3	Distance table with advanced test with 12 distances	127
7.1	Light intensity values	135
7.2	Light temperature values	135
7.3	Vertical position of objects	137
7.4	Distance table for general scene	142
7.5	Different light intensity setups	144

Nomenclature

2D	Two Dimensional
ASIC	Application Specific Integrated Circuit
CCD	Charge-Coupled Device
CF	Compact Flash
CFA	Color Filter Array
CMOS	Complementary Metal-Oxide Semiconductor
CPU	Central Processing Unit
CRT	Cathode Ray Tube
DCR	Device Control Register bus
DLP	Digital Light Processing
DMD	Digital Micro-mirror Device
DSLR	Digital Single Lens Reflex Camera
DSP	Digital Signal Processor
EMI	Electro Magnetic Interference
EV	Exposure Value
f	Ratio of the lens's focal length to the diameter of the entrance pupil
FOV	Field Of View

FPGA	Field-Programmable Gate Array
fps	Frames per second
GUI	Graphical User Interface
HD	High Definition
HDTV	High Definition Television
HVS	Human Visual System
I2C	IIC serial communication protocol
IR	Infra-Red
ISO	Measure of sensitivity of imaging systems
ISO chart	International Organization for Standardization
ISP	Image Signal Processor
JPEG	Joint Photographic Experts Group
LCD	Liquid Crystal Display
Lx	Unit of illuminance and luminous emittance
MPO	Multi Picture Object
MRF	Markov Random Field
MTF	Modulation Transfer Function
NCC	Normalized Cross-Correlation
PCB	Printed Circuit Board
PLB	Processor Local Bus
PMMS	Portable Multi Media Supercomputers
PTZ	Pan-Tilt-Zoom
QA	Quality Assurance

NOMENCLATURE

RAW	Unprocessed image data
RGB	Red Green Blue
ROI	Region Of Interest
SAD	Sum of Absolute Differences
SDK	Software Development Kit
SFR	Spatial Frequency Response
SHD	Sum of Hamming Distances
SoC	System-on-a-Chip
SSD	Sum of Squared Differences
ToF	Time-of-Flight
UI	User Interface
VLSI	Very Large Scale Integration
WCS	World Coordinate System
YCbCr	Family of color spaces. Y - luma component; Cb - Blue-difference chroma component; Cr - Red-difference chroma component.

Acknowledgments

First and foremost I would like to thank my mother for her unconditioned support during my student years, and especially during my study abroad. Her open-mindedness helped me a lot.

I was lucky to have Dr. Peter Corcoran as my supervisor for the PhD. Dr. Corcoran's expertise, enthusiasm and innovative thinking were extremely helpful for my work. Thank you Peter for pushing me when you felt that I needed to be pushed, and for the great help throughout my years as PhD student!

I would like to thank the people from the company that part-financed my PhD for their support, understanding and priceless advice, especially Dr. Petronel Bigioi, Alex, Lale, Piotr and Arpi. You helped me a lot in the past years, and I will be forever grateful.

I would also like to thank my better half for her understanding and support in the past two years. Thanks to her, I was able to leave my worries behind, and concentrate on the work that needed to be done.

A special thanks to Ms. Melisaratu for her support during my early years as a student.

The work presented in this thesis was part-financed by FotoNation Ireland and by the Irish Research Council through the IRCSET Enterprise Partnership Scheme.

Abstract

The core problem addressed in the current thesis is to provide a novel test bed for depth map generation and stereo image-based algorithms. The test bed comprises of elements such as (i) the choice and calibration of stereo image acquisition devices; (ii) choice and setup of objects in the test scenes; (iii) light intensity and temperature settings; (iv) camera-object and stereo baseline distance setups and (v) a repeatable test scene. In order to determine each element of the test bed, a number of experiments are presented. During the experiment process, four depth map algorithms are selected which, in conjunction with four different test scenes, are used for the initial measurements where (i) the influence of light intensity on algorithm performance is determined; (ii) the performances of the algorithms at different camera-object distances are determined and (iii) the influence of similar objects in the test scene is described. In order to acquire stereo images, a number of devices are used, from off-the-shelf stereo cameras to custom-built stereo devices. During the stereo image acquisition, a number of possible error sources are noticed, and the problems are mitigated. The experiments show that slight changes in the light intensity, light direction, camera-object distance and stereo base length can have a significant influence on the result of the depth map algorithms. With the help of these results, the details of the test bed are determined, and an on-line database is provided, containing details of the test scene and example stereo test images derived from the scene. This is available online for other researchers to download at <http://www.andorko.com/stereo.html>. The proposed test scene is provided with a view to having a standardized test scene that can be easily replicated by other researchers for testing of stereo acquisition systems and associated depth map acquisition algorithms.

Chapter 1

Introduction and Overview

1.1 Background and Motivations

In modern digital cameras all the image enhancement operations happen within what is known as the Image Signal Processor (ISP). The ISP is responsible for the transformation of the raw image acquired using a CMOS or CCD image sensor into an RGB image.

Recent digital cameras provide a large number of features, from the standard red eye reduction to face detection, face recognition, image stabilization and panorama. All of these features are implemented on dedicated image processing CPUs, and parts of the heavy lifting are performed using dedicated IP cores with a system ASIC.

As we reach the limits of standard image acquisition hardware and camera systems, it becomes increasingly evident that in order to provide the highest possible quality, more than one picture is required. For example, advanced techniques of red-eye removal use a reference image captured without flash to help localize genuine red-eyes (Petchnigg et al [104]); a blurred image can be corrected by combining with an underexposed sharp image of the same scene (Lasang et al [76], Albu et al [1]); back-lit scenes can be uniformly exposed by combining images with different exposure levels (Sun et al [129]) and the perceptual depth of a scene can be significantly enhanced by combining images with different depths of focus (Rajagopalan et al [108]).

In order to deal with information from two or more images, a dual pipeline ISP would be required, and the design and implementation of such a dual ISP, with the capabilities of acquiring stereo images, was the original focus of the research

presented in this thesis.

Around the years 2009 and 2010, 3D TV and 3D content have become very popular with researchers, with conferences such as the International Conference on Consumer Electronics and the Games Innovation Conference having special sessions and tutorials dedicated to 3D imaging. For this reason, the focus of the research has shifted towards 3D imaging. During this stage of the research, a 3D image and video acquisition device was implemented on a development board equipped with a Xilinx FPGA device, and I was invited to present a tutorial on 3D in gaming at the 2nd IEEE Games Innovation Conference. While working with 3D images, it was noticed that often 3D was used together with depth maps for various reasons ranging from 2D to 3D conversion, to the setting of the proper stereo baseline value for optimal 3D effects, and for auto-stereoscopic devices.

At this stage of the research it was decided to focus on depth map generation, and to investigate different methods and algorithms which are used for this purpose. The aim was to develop a depth map generation algorithm which uses stereo images, is suitable for hardware implementation, and provides good results.

High quality depth maps can be used in a wide area of applications such as (i) gesture detection for productivity and gaming purposes [61, 90], (ii) privacy in online communication [2], (iii) distance measurement [52, 27], (iv) 3D mapping purposes [120, 134], (v) 3D imaging [106], etc.

During the literature review about the depth map generation topic it was noticed that there was a lack of testing methods suitable for consumer electronic devices. In order to test consumer electronic devices at the initial development stage, a total control over the image acquisition condition is required, which allows the calibration of (i) the optics, (ii) different modules of the ISP and (iii) different image processing algorithms. The next stage in the development cycle of the product is the production stage, where the initial calibration of the device needs to be tested on a subset of devices. For this reason, for test and QA purposes, a test bed is required.

During the development stage of a stereo image acquisition device, a control over the stereo test scene is required, where the influence of (i) stereo baseline length and (ii) camera-object distance can be tested

In some cases, beside the image processing algorithms implemented in software or hardware, additional devices are used as well in order to process stereo images, such as (i) near-IR sensors, (ii) IR sensors or (iii) ToF cameras. In order to develop high quality depth map generation methods for real-time applications, some researchers

1.2 Contributions

[152, 70] consider that additional devices need to be used, and these devices must be considered as being part of the depth map generation method. In order to test these methods, the researchers need to create their own test scenes, and for this reason they can't compare the performance of their methods with other available methods.

Noticing the lack of such testing methods lead us to the core research theme of this doctoral thesis, which is the creation of a test bed suitable for the testing of consumer electronic devices capable of acquiring stereo images, and generate depth maps. The proposed test bed comprises the creation of a test methodology for stereo image-based depth map generation methods, the choice and setup of the image acquisition device, and the creation of a general depth map testing scene.

During the development of the test bed, a number of depth map generation algorithms were also tested. The information from the test results was used in order to design and create the standardized test scene and scenarios.

1.2 Contributions

The main contributions in this thesis are related to the development of a test bed suitable for the testing of stereo image acquisition devices.

These contributions are the (i) design, practical implementation and testing of a testbed suitable for the testing of consumer electronic devices capable to acquire stereo images and generate depth maps, (ii) the creation of a general test scene and (iii) the identification of the sources of error which have a significant effect on the quality of the depth maps. The “test bed” is a “process” to measure/quantify the ability of a particular stereo camera to create a detailed depth map.

In order for the research to get up to this point, it progressed through several stages, each stage having its own minor contributions. As presented in section 1.1, the first topic of my research was the development of a dual ISP and of a stereo camera. At this stage, the main goal was the development of a working stereo vision system, the identification of different stereo concepts, and their influence on the stereo imaging process and associated algorithms.

The second and main stage of the research was the development of the testbed, with the following contributions:

1. A study of different depth map generation algorithms.

2. The development of an initial testbed, which allows the testing under different illumination conditions and at different distances.
3. The development of several initial test scenes, which are meant to test the behavior of the depth map generation algorithms in the case of different types of objects.
4. The choice of the stereo image acquisition device, and the description of the image acquisition process, and camera setup.
5. The development of a general testbed and test scene for the performance measurement of depth maps, which are the main contributions of this thesis.

1.3 Thesis Outline

The thesis is structured in eight chapters starting with the introduction and concluding with the final chapter dedicated to conclusions and future direction of our research. Additionally, there are five appendices dedicated to detailed test results and methodologies.

Chapter 1 provides introductory and background information about the field of research, and details the goals of the research presented in this thesis.

Chapter 2 presents the foundation techniques and the initial work that has been undertaken for this research. The goal of this chapter is to present the implementation of a stereo image acquisition device, the challenges that were faced during this implementation, several potential problems that researchers might encounter when implementing such a device were identified.

In this chapter, concepts of stereo imaging are defined. These need to be taken into account when designing image processing algorithms, acquisition devices and testbeds which use stereo images.

The purpose of **Chapter 3** is to review previously published work in the area of disparity map and depth map generation. This review was very helpful in identifying problems behind the implementation of such depth map generation algorithms, and in the understanding of the fundamental problems which need to be considered when working with depth maps.

A comprehensive literature review is provided covering (i) depth map generation algorithms which only use a single camera for the purpose of depth map generation,

1.3 Thesis Outline

and (ii) depth map generation algorithms, which use two cameras and stereo images in order to generate depth maps. The selection criteria for the depth map generation algorithms which were used during the development of the proposed testbed is also described in this chapter.

Chapter 4 presents the development of the initial testbed. During the development of the initial testbed, four algorithms using different pixel matching costs were tested, which are used in depth map generation algorithms.

In order to measure the performance of the depth map generation algorithms, a test application was developed, which together with the initial test scenes developed for this purpose, are described in the second part of the chapter. The presentation of the initial test scenes contains a detailed explanation of the choice of the objects in the scenes, and the scenarios developed for each test scene.

Chapter 5 talks about the devices used for the purpose of stereo image acquisition. In this chapter, the devices which were considered for image acquisition are presented, together with their specifications, advantages and disadvantages.

Another aspect of the image acquisition process which is discussed in this chapter is the definition of the acquisition limitations, and the possible elements that might influence the test results. In order to define the limitations of the image acquisition, a parallax calculation was performed, which defines the minimum and maximum stereo base length values and camera-object values that can be used during the tests. After this calculation, the influence of vertical misalignment and noise on the depth map generation algorithm was determined.

In **Chapter 6**, the test results for different test scenarios are presented. The test scenarios from this chapter use the test scenes and depth map generation methods defined in chapter 4. The test results are explained in the form of tables and charts, and relevant conclusions are drawn, which will be used in chapter 7.

Chapter 7 presents the testbed and the test scene, which were the main goal of the research presented in this thesis. For their development, relevant information gathered during the development of the initial test scenes described in chapter 4, the choice of the acquisition device in chapter 5 and the test results in chapter 6, were used.

In this chapter, the testbed which was developed in order to test depth map generation algorithms, developed for consumer electronics purposes, is introduced. Detailed explanations and measurements are provided for the re-creation of the test scene, which can be used for the testing of a variety of depth map generation methods.

Also, relevant information about the setup of the test environment and the use of image acquisition devices are provided, which are the results of several months of testing.

The thesis concludes with **Chapter 8**, where the final conclusions are drawn, and details about the direction of the future research are given.

Chapter 2

Foundation techniques

This chapter introduces the reader to the area of stereo imaging, where the implementation of a stereo image processing pipeline and concepts related to stereo imaging are described.

The chapter starts with section 2.1, where different stereo concepts are introduced. These stereo concepts create the foundation of stereo imaging, and they will be mentioned several times throughout the thesis.

Section 2.1.1 introduces the reader to stereo images. This section defines the notion of stereo images, and describes the epipolar geometry, which is of significant importance when working with stereo images. The stereo baseline is part of the epipolar geometry, and it is described in section 2.1.2.

Section 2.1.3 describes parallax. Parallax is one of the best-known concepts in stereo imaging, and it lies at the base of depth map computation from stereo images. The depth map is introduced in section 2.1.4.

The chapter continues with the literature review of single pipeline Image Signal Processors (ISPs) in section 2.2. The information obtained during the literature review is used in section 2.3, which describes the implementation of a stereo image signal processor, performed during the early stages of the research. In the same section, an application of the stereo ISP is described in subsection 2.3.3, where the implementation of a 3D camera is presented, which was based on the stereo ISP platform.

The chapter concludes with the presentation of the challenges researchers are facing when working with stereo images, in section 2.4.

2.1 Stereo concepts

2.1.1 Stereo Images

A stereo image is meant to provide sufficient visual information, in order for the brain to be able to perceive images in 3D. A stereo image pair is a pair of images containing similar scene information, the only difference between the two images being an offset of several image columns relative to the left or right image, figure 2.1.

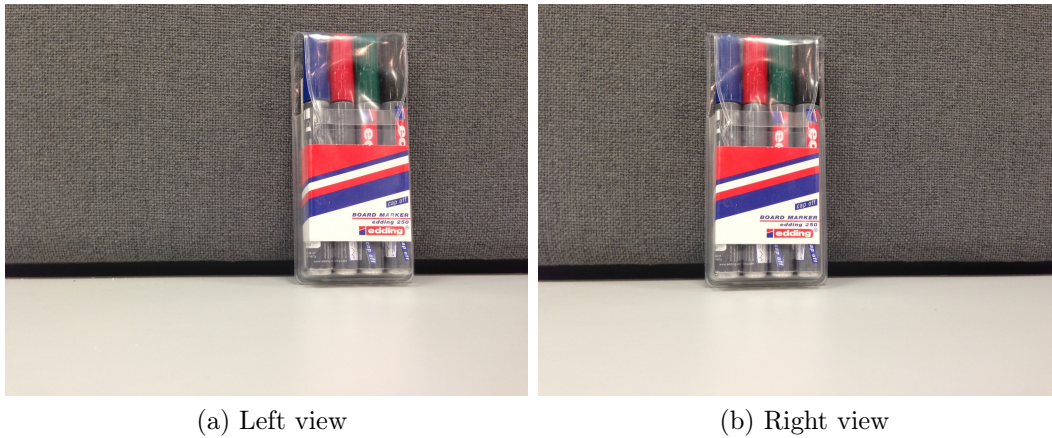


Figure 2.1: Stereo image

Depth information can be reproduced from stereo images by computing systems, as it will be presented in section 2.1.4, by finding corresponding pixels in the left and right images. The search for corresponding pixels in the left and right image is based on the epipolar geometry. Let's consider figure 2.2, where the geometry of a two pinhole camera stereo system can be seen. In this figure, O_L and O_R are the center points of the left and right cameras. The line connecting O_L and O_R is the base line, and it will be described in section 2.1.2. \mathbf{P} is a randomly selected point in the world coordinate system. The epipolar lines are the corresponding lines in the left and right image, where corresponding pixels need to be searched for. The current example presents the epipolar lines, where point \mathbf{P} is found.

2.1 Stereo concepts

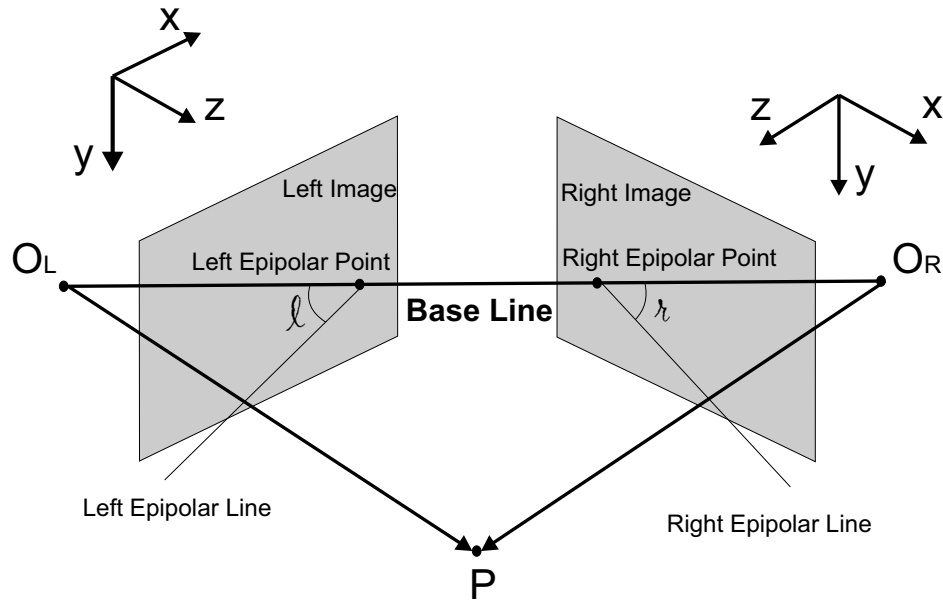


Figure 2.2: Epipolar geometry

Angles l and r determine the angle of the epipolar line relative to the base line, and they are determined based on the convergence angle of the cameras in the stereo image acquisition system, see figure 2.3.

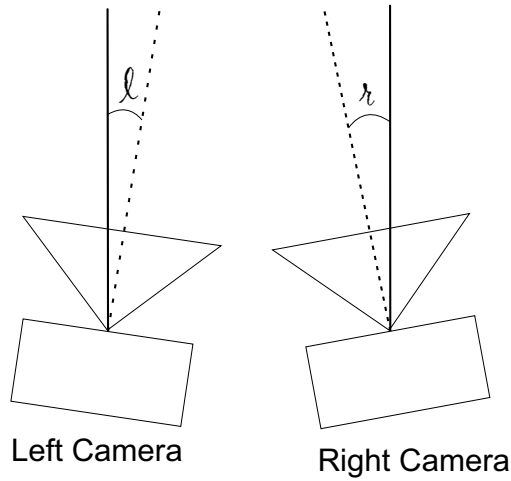


Figure 2.3: Convergence angle

A real-world example can be seen in figure 2.4, where the angle of convergence used is 30° .

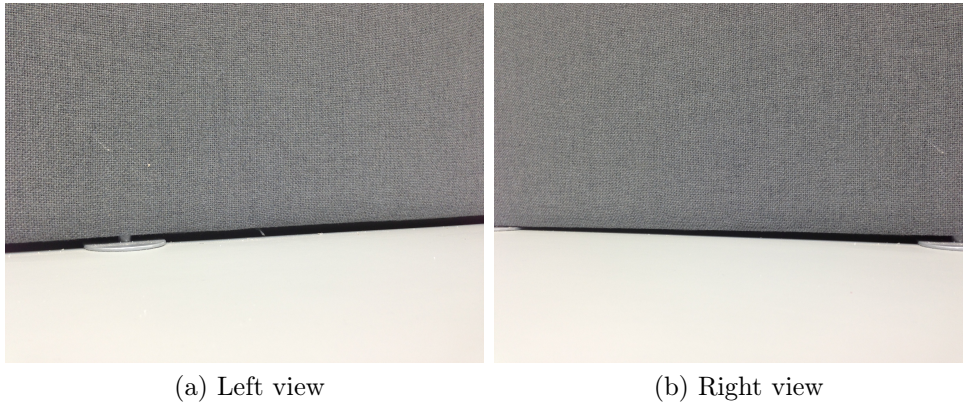


Figure 2.4: Converging stereo camera system

The convergence angle can be determined by calibrating the stereo camera system. In the case of consumer products, where real-time image processing is required, on-the-fly calibration is not an option. For this reason, the cameras in the stereo image acquisition systems should be set up in a parallel configuration, where the angles l and r have the value of 0° .

For the acquisition of stereo images, a number of devices and device setups can be used. In figure 2.5a a purpose-built stereo camera can be seen. As a second option, two similar cameras can be mounted onto a rig for the purpose of stereo image acquisition, as it can be seen in figure 2.5b. The third option is the use of the same camera mounted onto a rig, and moved slightly to the left or right for the second image acquisition, see figure 2.5c. The disadvantage of the last setup is, that the stereo image pair is not acquired synchronously. As it will be presented in section 2.4, due to un-synchronized image acquisition, data might become altered in the stereo image pair.

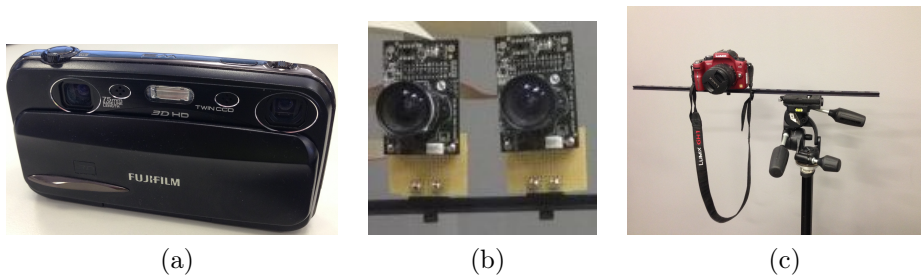


Figure 2.5: Stereo imaging systems

2.1 Stereo concepts

2.1.2 Stereo Baseline

The Stereo Baseline is the horizontal distance between the two cameras which are used for the acquisition of a stereo image, as it can be seen in figure 2.6.

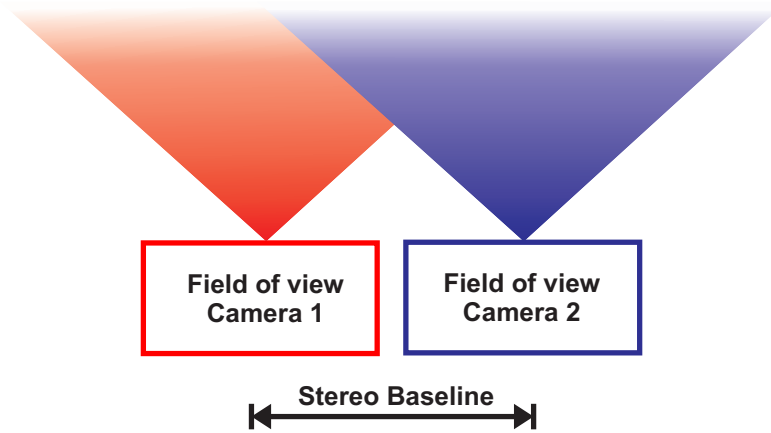


Figure 2.6: Stereo base example

One of the most important roles of the stereo baseline is the determination of the overlapping field of view for the stereo camera pair. If the baseline is larger, the overlapping field of view decreases, which means that a reduced amount of common information will be available for the cameras, but there will be a larger parallax, see section 2.1.3. In the other case, if the baseline decreases, a larger amount of common information is available, but the small parallax, especially for objects further away from the camera, might not be enough to perform the required measurements for image enhancement purposes. A detailed analysis of the relationship between the stereo baseline and the parallax value can be found at section 2.1.3.1.

In order to overcome the issue of having a small overlapping field of view, two types of cameras can be used. The first type of camera is a sensor-shift type of camera [87], which is usually used for image stabilization. By using this type of sensor in a stereo camera and shifting the sensor towards the centre of the camera, the overlapping field of view can be increased. Another type of cameras are the ones with wide sensors. In this case the field of view overlap is larger than in the case of standard sensors. For this reason, in the case of large stereo base values, the field of view overlap change is insignificant.

2.1.3 Parallax

The parallax is the difference in the position of an object against the background, when viewed from different perspectives, see figure 2.7. The closer the objects are to the observation point, the larger the effect of the parallax. The theory behind this statement is described in section 2.1.3.1.



(a) Top perspective



(b) Left perspective



(c) Right perspective

Figure 2.7: Parallax

In order to describe the effect of the parallax, let us consider a stereo image acquisition system similar to the ones described in section 2.1.1. When acquiring a stereo image of a scene similar to the one in figure 2.8, it can be noticed that different objects in the scene have different positions in the left and right images.

2.1 Stereo concepts

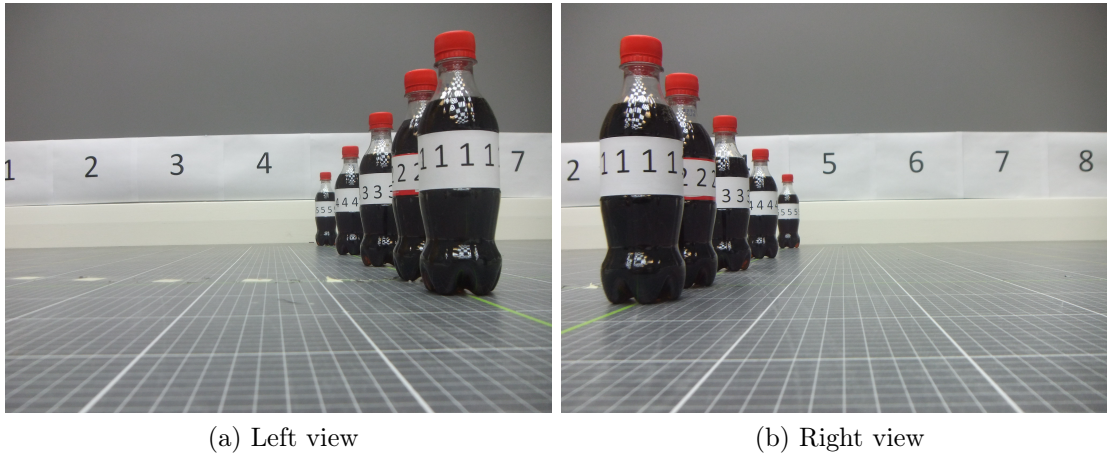


Figure 2.8: Parallax effect

The position of object 1 changes the most, while the position of object 5 only changes a little. The scene setup can be seen in figure 2.9.

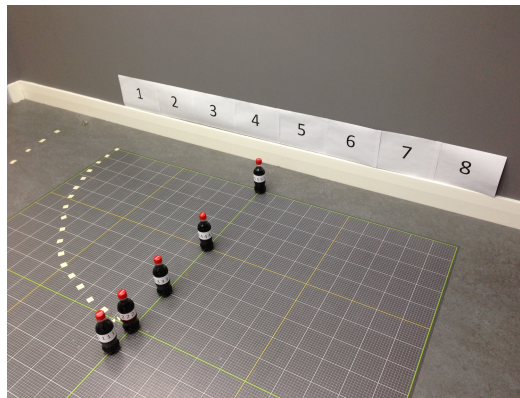


Figure 2.9: Parallax scene setup

The different locations of the objects in the left and right views create a disparity. By calculating this disparity, the depth of the scene can be determined, see section 2.1.4.

Parallax can also be used in the case of aerial picture analysis, where, by using a stereo viewer, heights of buildings can be detected, if the flight altitude and baseline distances are known. Based on the same idea, the parallax can also be used for military purposes, as a range finder, where the distance to the target is detected.

2.1.3.1 Measurement of the disparity caused by the parallax

The measurement of disparity caused by the parallax is required in order to determine the limitations of the camera, and of the test scene. These limitations define the minimum and maximum camera-object distances where objects should be placed in the test scenes, based on the specifications of the camera.

$d1$ and $d2$ in figure 2.10 present the areas which are considered as being disparity of objects caused by the parallax (different disparities for different object distances), and the area which is considered as being overlapped in the field of view. The disparity caused by parallax is important due to the fact that it is needed for depth computation. The overlapped area is important because only the overlapped areas of two views share the same information, which is required for the search of matching pixels.

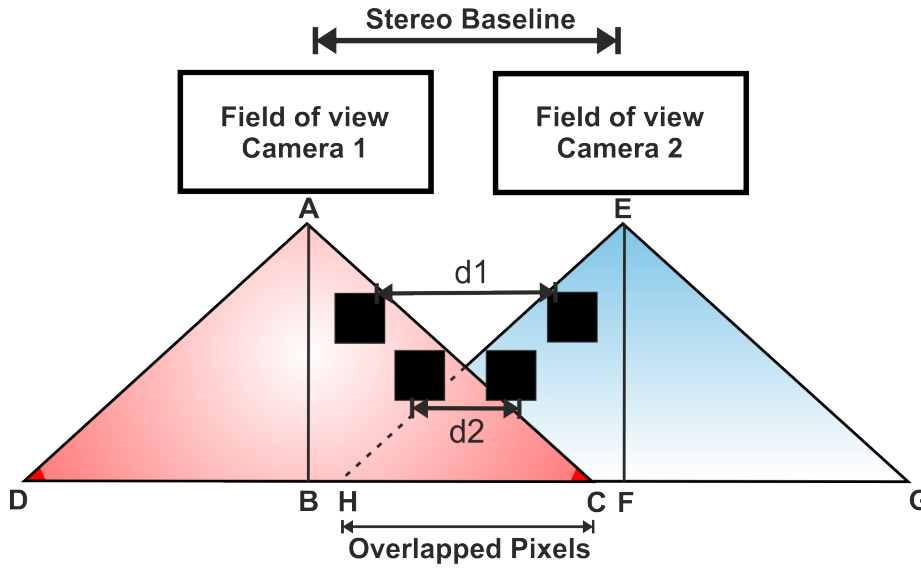


Figure 2.10: Explanation of parallax and overlapped pixels

The first step in the measurement of the displacement is the calculation of possible fields of view that the camera uses. For the calculation the following formula is used:

$$\alpha = 2 \cdot \arctan \frac{h}{2f} \quad (2.1)$$

where h = sensor size and f = focal length.

Figure 2.11 presents the horizontal field of view from a geometrical point of view, where $\alpha = \widehat{CAD}$.

2.1 Stereo concepts

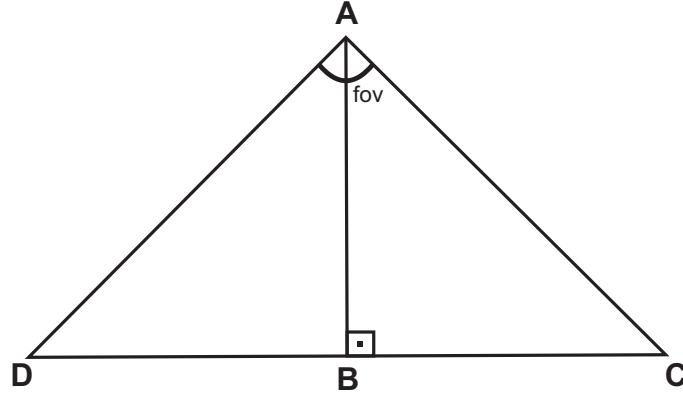


Figure 2.11: Horizontal field of view

In figure 2.11, \widehat{CAD} has the value calculated using equation 2.1. $AB \perp CD \implies \widehat{BAC} = \widehat{CAD}/2, \widehat{ACB} = 90^\circ - \widehat{BAC}$. CD has a different value for each object distance. Its value is used to calculate the disparity caused by the parallax, and the pixel overlap in two different views. In order to calculate CD , the only information needed is the value of AB , which in our case is the distance of the object from the camera. The formula used for the calculation of CD is:

$$CD = 2 \cdot (\cot(\widehat{ACB}) \cdot AB) \quad (2.2)$$

The displacement value is expressed in number of pixels. The number of pixels can differ for different resolutions settings. In order to calculate the displacement in pixel numbers, first we need to calculate the size of the pixel in cm. In order to do this, the formula $pixel = CD / HorizontalResolution$ is used.

The last step in the process is the calculation of the displacement of the objects between the two views. The only additional information that is needed at this step is the stereo baseline value. According to figure 2.10, the formula used for the displacement caused by parallax computation is:

$$displacement = HorizontalResolution - ((CD - 2 \cdot baseline) / pixel) \quad (2.3)$$

In order to calculate the number of overlapped pixels, the value of the *displacement* needs to be subtracted from *HorizontalResolution*.

A similar measurement was performed in [59] as well, where the authors develop a geometric relationship between the planar parallax displacements of pairs of points in a set of stereo images.

2.1.4 Depth Determination and Depth Maps

In order to determine depth in a scene, the Human Visual System (HVS) relies on a number of depth cues, such as occlusion [20], object size, texture and shading, parallax and stereoscopic information. The stereoscopic information, which is represented by disparity, is one of the most important depth cues, based on which the HVS can determine the distance between the object and the eyes. Figure 2.12 represents the binocular vision in the HVS [41].

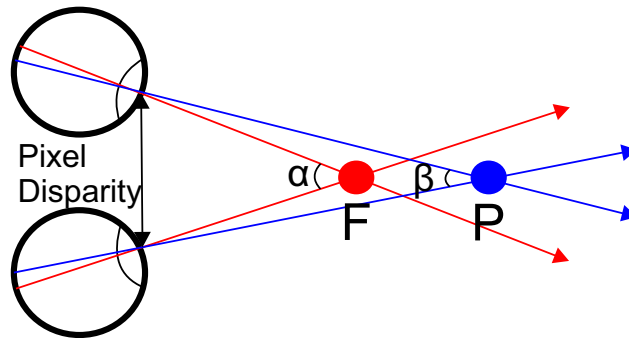


Figure 2.12: Binocular vision in HVS

Let us consider point F as the fixation point for the left and right views, and point P the point at which we want to measure the disparity. The disparity measurement in the case of the HVS is done by measuring the difference of angles α and β . This is not to be confused with the disparity in computer vision, where we are measuring horizontal distance in pixel numbers (pixel disparity).

In computer vision, depth is also determined by measuring the disparity, but this time between corresponding pixels in a pair of stereo images. The disparity map is a grayscale image, where each pixel represents the disparity (distance) between corresponding pixels in stereo image pairs. For this reason, the greater the disparity between corresponding pixels, the closer the object is to the camera. An example of an image with its corresponding disparity map can be seen in figure 2.13.

2.1 Stereo concepts

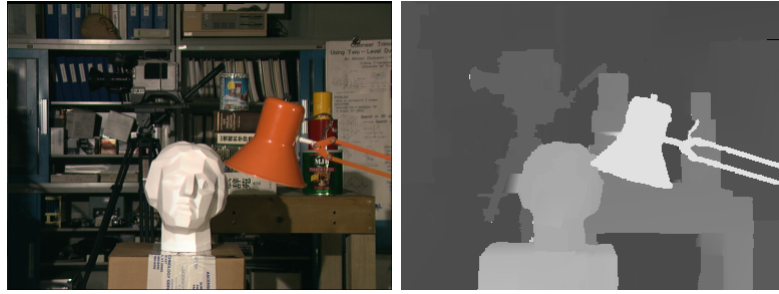


Figure 2.13: Disparity map example

A depth map is a grayscale image, where each pixel contains depth information about the corresponding pixel in the original image. The depth is inversely proportional to disparity, because objects with greater disparity are closer to cameras, and objects with smaller disparity are farther from cameras. The representation of the depth map is similar to the disparity map, the difference being, that the depth map contains exact distance values. In order to calculate the depth map from the disparity map, information such as the stereo baseline, section 2.1.2, and the focal distance, see figure 2.23, of the camera are required. Once these parameters are defined, the values of the depth map can be calculated using the formula in equation 2.4, see figure 2.14.

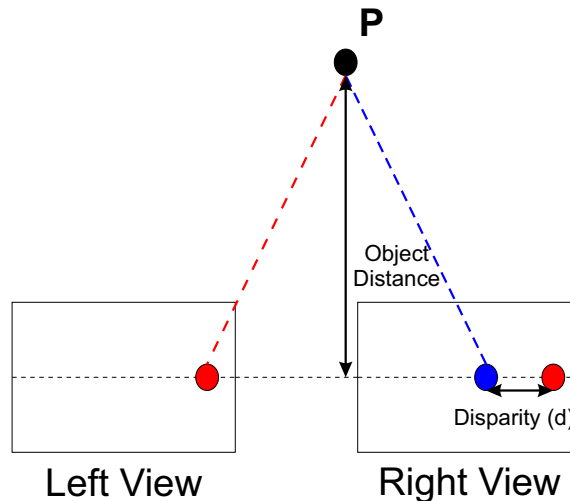


Figure 2.14: Depth map calculation

$$ObjectDistance = \frac{(b \cdot f)}{d} \quad (2.4)$$

, where b is the stereo baseline, f is the focal distance and d is the disparity, see [123].

The most difficult task in the process of depth map computation is the generation of the disparity map. For this reason, during the development process of a depth map generation algorithm, a significant amount of time is spent on the research about disparity maps.

Disparity maps can be generated using a large number of methods. Some of these methods are described in chapter 3.

2.2 Literature Review of Image Signal Processors

As mentioned in the introduction chapter of the thesis, at the early stage of the research one of the goals was the implementation of the stereo Image Signal Processor. The purpose of the current section is the review of published work in connection with the architecture of the ISP, and the algorithms used at each stage of the ISP.

The Image Signal Processor can be found in every camera, and its main purpose is the transformation of the raw data which arrives from the sensor, into an image suitable to be viewed by the user. The purpose of this section is the identification of the different stages of the ISP, and the general architecture of the ISP.

According to (Ramanath et al [111]), the typical color image processing pipeline consists of several stages before the acquired image is stored or displayed. The first stage of the pipeline, see figure 2.15 is made out of a sensors and lens system, which acquires the image. By filtering the image, setting the right exposure and making sure that the objects in the image are focused, it prepares the acquired data for the next stage of the pipeline, which is the *Preprocessing* stage.

In order to control the exposure, the characterization of the image intensity is necessary. Setting the right exposure value ensures that the image is not under or overexposed. Several algorithms were developed for the purpose of exposure control [122, 31, 67].

Exposure control is important during the image acquisition process. There is a connection between the exposure time and the size of the aperture. In order to obtain normally exposed images, there needs to be a balance between the length of the exposure and the size of the aperture. With a small aperture there is less light entering the camera system, and due to this reason the exposure needs to be longer.

2.2 Literature Review of Image Signal Processors

The size of the aperture also influences the focusing capabilities of the camera. If we want to have all the objects in the image sharp, then the aperture needs to be small.

The focus control can be performed based on two different approaches [88, 146, 83]. The active approaches usually use a pulse beam of infrared light in order to estimate the distance to the objects from the camera and this way to determine the best focus setting. The passive approaches either analyze the spatial frequency content of the image or use a phase detection technique in order to estimate the distance to the objects.

The most common filter used at this stage of the processing is the anti-aliasing filter, which reduces the artifacts that are caused by the Color Filter Array (CFA) [81, 74, 101] and the Infra Red (IR) - blocking filter, which eliminates the IR components from the acquired image.

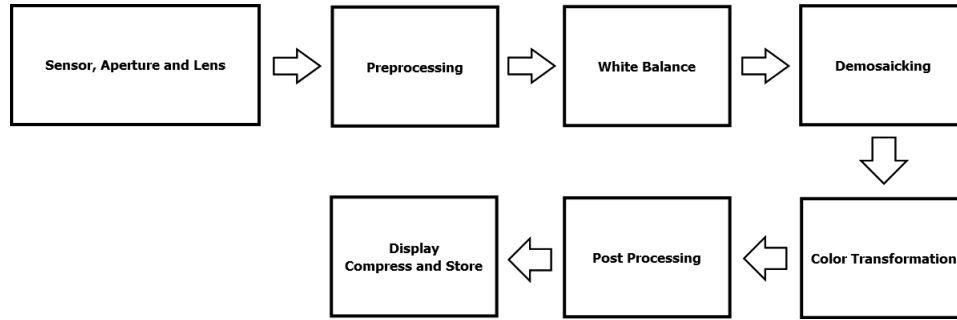


Figure 2.15: Image Signal Processor

By reviewing the work presented in this section, it was learned that the correct exposure setting is of significant importance in the case of image acquisition, and that the focusing of the image can be influenced by a number of factors, including the size of the aperture.

Preprocessing The main purpose of the Preprocessing stage is to remove noise and other artifacts from the image. One of the most common operations in this stage is the *linearization*. The acquired data in case of some cameras needs to be linearized because the captured data resides in a nonlinear space. Another operation performed at this stage of the pipeline is the *dark current compensation*. The reason is that a dark current is recorded even when the lens cap of the camera is on, and this can influence the quality of the acquired image. Several methods to overcome this problem are presented in [14, 95]. The last step that can be considered at this

stage is the *flare compensation*. It needs to be performed for images, where a bright source of light is in the field of view. The reason is that, the light that enters the optics of a camera is reflected and scattered, and this causes a nonuniform shift in the measured energy. Each camera manufacturer deals with this compensation in its own way.

White Balance The aim of the auto white balance is to guess the illumination under which the image is taken and to compensate the color shift affected by the illuminate (Kao et al [66]). The white balance problem is usually solved by adjusting the gains of the three primary colors R, G or B of the sensors to make a white object to appear white under different illuminants. There are a number of methods that were proposed for the white balance operation and some of the most important ones will be presented.

The first group of methods would be the “Gray World Methods”. These algorithms simply assume that the scene average is identical to the camera response to the chosen “gray” under the scene illuminant, (Barnard et al [11]). The second method is the “Illuminant Estimation by the Maximum of Each Channel”. This algorithm estimates the illuminant (R, G, and B) by the maximum response in each channel. The third group of methods is the “Gamut Mapping methods”. These are based on Forsyth’s gamut-mapping approach, (Forsyth [44]). The fourth method is the “Color by Correlation”. The basic idea of this approach is to pre-compute a correlation matrix which describes the extent to which proposed illuminants are compatible with the occurrence of image chromaticity, (Barnard et al [11]). The fifth type of method is the “Neural Net” method. The neural net is a multilayer Perception with two hidden layers. The general structure is pyramidal. In one of the examples the input layer consists of 2500 nodes, the first hidden layer has 400 nodes, the second hidden layer 30 nodes and the output layer has 2 nodes. The chromaticity space is divided into discrete bins. The input to each neuron is a binary value representing the presence or absence of a scene chromaticity falling in the corresponding bin. Thus, a histogram of the image is formed, (Barnard et al [11]).

We can conclude that white balance is used in order to compensate the color shift due to different color temperatures. There are 2 types of white balance algorithms, (i) white balance algorithms, where the user sets the gain values for the R, G and B colors in the sensor, and the (ii) auto white balance algorithms, where the gain values are set automatically.

2.2 Literature Review of Image Signal Processors

Demosaicking and Color Transformation The fourth, and in the same time, one of the most important stages in the pipeline is the *Demosaicking* stage. At this stage the values of the pixel colors that were not measured due to the fact that a single sensor using a CFA was used, are estimated using the pixel neighborhood information. There are a large number of methods available, with most of them being patented. This stage, without any doubt, is the most important one in any ISP, and for this reason, the camera manufacturers keep their algorithms secret. A survey of different demosaicking methods is detailed in [110, 82].

The next stage in the image processing pipeline is the *Color Transformation* to and from unrendered color spaces. The purpose of this transformation is that these color spaces are designed for the purpose of a convenient storage and calculation medium [34, 29, 39].

Post processing The last stage of the pipeline is the *Post processing* stage. The elements that can be found at this stage depend on the methods used in the previous stages of the pipeline. For example, different demosaicking methods can introduce different types of artifacts according to (Ramanath et al [111]). Some common techniques used at this stage are color-artifacts removal, edge enhancement and coring. The color artifacts removal tries to remove the artifacts possibly caused by the Demosaicking stage of the pipeline. The edge enhancement's purpose is to make the image more appealing to the user, because the human eye is known to be highly sensitive to sharp edges. By coring, the information that has no significant contribution to the image, and can be considered to be noise, is removed. Each manufacturer uses its own type of method, based on the design of their system. After the post processing stage, the acquired image can be stored or displayed.

ISP Architectures The purpose of the literature review presented in this paragraph was to learn about the approach used by other researchers when implementing a Image Signal Processor.

Several Image Signal Processors and Image Processing Pipelines have been proposed in the past two decades [149, 47, 66]. In the paper written by (Zen et al [149]), a digital signal processor that is dedicated to the Progressive Scan type CCD is described. According to the authors, the main advantage of this processor is that it adopts a new pixel interpolation algorithm which takes an average of two mean pixel values from neighboring 4 pixels, and an individual color matrix before

gamma control, which provides higher image quality. The presented processor also has the characteristic that it can switch between progressive and interlaced scanning methods without requiring additional video memory. The most important functions described in (Ramanath et al [111]) and which can also be found in the proposed ISP, are the Color Pixel Interpolation, otherwise known as Demosaicking, White Balance, Aperture Control (which also controls the exposure settings), Transformation to Unrendered Color Spaces, Auto Focus and Gamma Control for display purposes.

Another paper that presents a color image signal processor is introduced by (Kao et al [68]), where the authors propose a ISP which processes the incoming raw image that comes from a CCD or CMOS sensor, and converts it to the final color with corrected exposure. According to the authors, the presented system bridges the gap between theoretical image processing, and practical ISP implementation issues that are commonly found in digital imaging systems. According to the authors, the design of their ISP is based on the considerations that the Auto White Balance should be performed at an early stage due to the fact that, if the white balance adjustment is applied after nonlinear processing, the colors end to become incorrect; it's better to carry out color interpolation before performing any other processing; noise should be filtered out in the sensor RGB space, rather than on the Y component in the YCbCr space; noise filtering should consider edge information in order to prevent edges from being blurred; apply tone reproduction at a later stage, because it improves the rendition of the images; gamma correction should be at a latter stage in the pipeline. With the help of the test images, the authors proved that the proposed Image Signal Processor performs well under any condition, and the design allows space for further development, and eventual integration into a digital camera system.

(Gentile et al [47]) propose an implementation of an Image Signal Processor, which, similarly with the previously discussed paper is meant for the integration into digital cameras. The proposed ISP is designed for the SIMPil architecture, which is based on the Portable MultiMedia Supercomputers (PMMS) type of architecture according to the authors. The proposed ISP consists of the following processing modules: Black Clamping, Lens Distortion Compensation, Faulty Pixel Interpolation, White Balancing, Gamma Correction, CFA Color Interpolation, Color Conversion, Edge Detection, False Color Correction and JPEG compression.

After reviewing a number of papers related to the design and implementation of ISPs, a conclusion can be drawn that a large number of ISP architectures are available, each of them being designed for specific requirements. A main data flow

2.3 Implementation of the Stereo ISP

at the higher level of the ISP was identified. After the image acquisition, the first steps should be the artifact removal, which comprises of linearization, and black level removal, and the white balance. The white balance should be performed before the demosaicking stage. Noise removal can be applied before and after the demosaicking process, depending on what type of demosaicking algorithm is used. One of the last stages of the ISP should be the color transformation, which makes the data suitable for storage and processing, and the gamma correction.

2.3 Implementation of the Stereo ISP

My first challenge as a researcher was the implementation of a dual ISP on a FPGA. The goal was to use this dual ISP together with two cameras in order to improve the processing speed of the algorithms used in some of the applications mentioned in the previous chapter. The implemented dual ISP was at the base of a number of papers published during my first years as a PhD student.

In the paper by (Andorko et al [3]), the implementation of a real-time stereo imaging system was presented, which incorporates dual image acquisition chains on a single FPGA device, and is able to provide real-time synchronized video output from twin CMOS imaging sensors. The platform was designed to provide hardware support for future implementations of advanced face analysis algorithms, but it could also provide real-time capture of game players' facial actions and expressions, potential applications including 3D real-time game avatars, and employing face data for next-generation game User Interfaces (UI).

This platform was used for the application described by (Ionita et al [58]), and it was also mentioned in the patent application by (Corcoran et al [35]) where the generation of 3D face models from the use of stereo images was presented. The generated 3D face models could offer a practical real-time face model, which is suitable for a range of applications in computer gaming.

The stereo ISP was also used for the system detailed in (Andorko et al [4]), where an implementation of a real-time 3D video acquisition system was presented. The approach was based on the use of stereo image sensors, and the images were converted into 3D image format, so that they can later be displayed using 3D active shutter glasses. Two image formats were implemented. One of the formats was the DLP 3D HDTV Video format, proposed by Texas Instruments (TI), where the acquired

left and right images were sampled using the native offset diagonal format of the Digital Micro-mirror Device (DMD). The two views were overlaid and appeared as a left and right checkerboard pattern in an conventional orthogonal sampled image. The second implemented format was the interlaced layer format, where the storage of the frames was done by alternating the left and the right views. This way, the sampling frequency dropped to half, but it allowed the display of the 3D images on a standard CRT monitor with the signal coming from the development board which was used for the implementation.

Another application of the stereo image processing pipeline was proposed in (Andorko et al [2]) at the 2nd IEEE Games Innovation Conference, where a depth map generation and depth layer separation system was described with the help of which customized image information can be shared about a player on the network. The system implied the generation of a depth map from a pair of stereo images, the separation of the foreground depth map layer, the filling of the holes in this layer by using mathematical morphology, and by using this layer as a mask, separating the face of the player from the background.

2.3.1 Design Specifications

A paper was published about the tools and techniques that are required for the implementation of a FPGA-based stereoscopic camera in (Andorko et al [6]). The system was implemented on the Xilinx Virtex 4 FPGA device, which can be found on the Xilinx ML 405 development board. As development tools, (i) Xilinx ISE 10.1 was used for the design of the hardware subsystem using VerilogHDL programming language, (ii) Xilinx SDK was used for the software subsystem using C programming language and (iii) Xilinx EDK was used in order to create the HW - SW interface by using the Graphical User Interface (GUI).

The general architecture of the system is illustrated in figure 2.16. Note that the use of independent external imaging sensors enables greater flexibility in deciding their relative positioning and also enables the use of twin sensors with different capabilities, e.g. one wide angle sensor, combined with a normal field of view sensor.

2.3 Implementation of the Stereo ISP

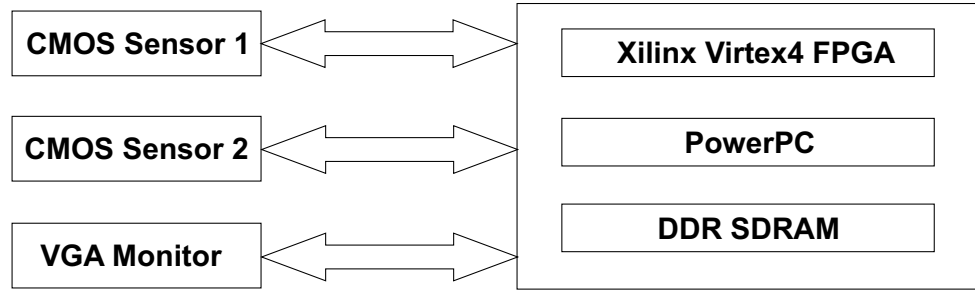


Figure 2.16: General system architecture

The internal architecture of the design is detailed in figure 2.17. The main image data is carried over the PLB while control signals are passed between the subsystems over the DCR bus.

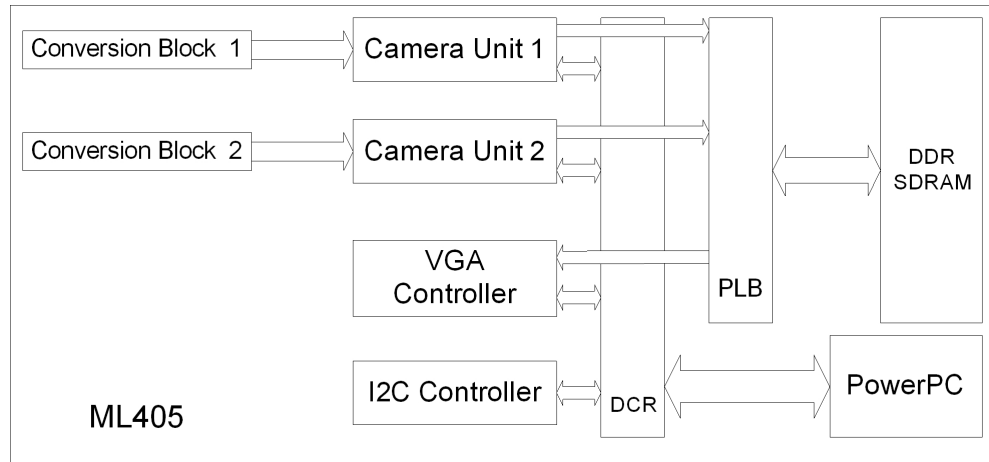


Figure 2.17: Internal system architecture

The development board is a Xilinx ML405 development board, with a Virtex 4 FPGA, and 64 MB DDR SDRAM memory. This FPGA architecture also incorporates a fully embedded PowerPC RISC processor. The chosen clock frequency of the design is 100 MHz.

The pre-processing module of the ISP, described in section 2.2, performs the color correction operations such as Lens Shading Correction and Crosstalk Correction on a dedicated SoC in the camera module. The information arriving from the camera module needs to be taken through the White Balance, Demosaicking and Gamma Correction processes only.

The camera unit requires a 25 MHz clock signal which is derived from the 100 MHz clock signal of the main FPGA system. It generates the request signals for

the PLB arbiter and generates the addresses for the DDR SDRAM. The PLB is clocked on 100 MHz and it is used for data transfers between the Camera Unit, VGA Controller and the DDR SDRAM memory. The DCR bus is clocked at 100 MHz and is used for communication between the modules in the hardware design and the PowerPC RISC microprocessor. The I2C controller is driven by the PowerPC and allows the configuration of the CMOS image sensors.

The first operation which is executed by the system is the configuration of the working parameters of the sensors using the I2C controller and the I2C bus. The resolution of the sensors is set to 640x480 as this reduces the amount of raw data generated by the imaging sensor and facilitates a real-time implementation. After this step, data is sent to the Camera Unit, which bundles the data into 64 bit registers and requests access to the PLB which directly connects the camera unit to the DDR SDRAM memory. Certain parameters of the camera unit are controlled from the PowerPC through the DCR bus. These parameters include the enabling or disabling of the camera modules, the image down scaling ratio, and the memory write start address.

2.3.2 Testing the system

For the testing and simulation of the design, a Modelsim PE 6.3 hardware simulation software has been used. Instead of the CMOS sensor a VerilogHDL model of the MT9M011 sensor was used, which simulates the functionality of the real sensor. The VerilogHDL model of the sensor was able to feed RAW images in Bayer format to the design.

For the testing of the synchronized operation of the sensors a test design was originally developed, where both sensors used the same incoming synchronization signals. This was observed to cause a physical shift between the left and right images. After some analysis and further testing it was determined that this is an ElectroMagnetic Interference (EMI) problem. Example images are shown in figure 2.18 where it can be clearly seen from the right-hand picture that the frame is shifted significantly to the right.

2.3 Implementation of the Stereo ISP

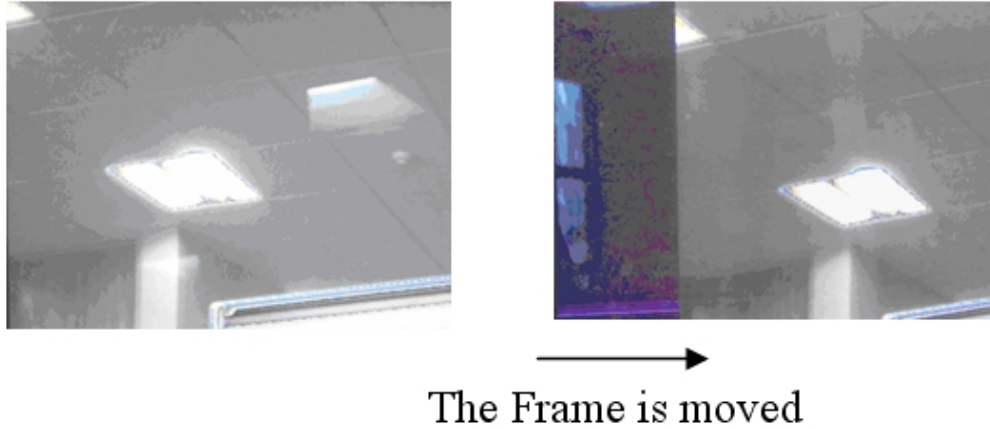


Figure 2.18: EMI Problem

This EMI problem was solved by building a custom PCB board for the connections between the sensors and the board, in particular making the connections shorter, and connecting the analogue ground pin in the close vicinity of the Vcc pin. Example images can be seen in figure 2.19 where it can be clearly seen that the frames are at the same position and this problem has been solved.

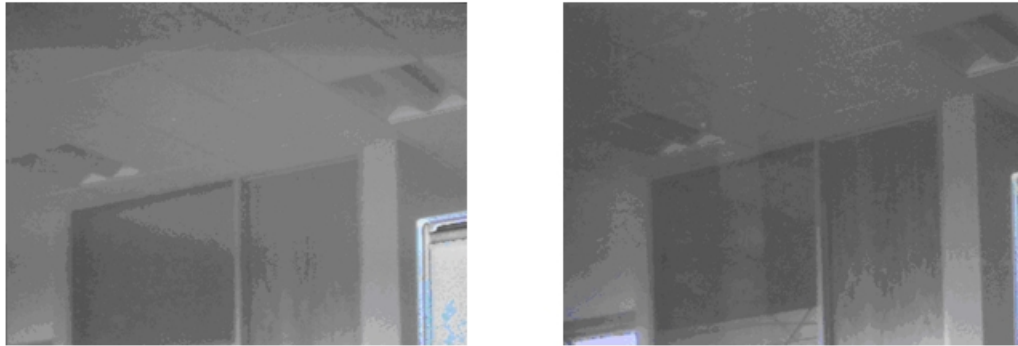


Figure 2.19: EMI problem solved

It is worth remarking that the image-offsetting effect of this EMI phenomenon is quite confusing and our initial investigations were directed to determine if some spatial, optical or timing errors were the root cause. In the end it turned out that it was the signal interconnections which were at fault due to the relatively high frequencies involved. This discussion was included so that other researchers can avoid our confusion as to the cause of this phenomenon.

2.3.3 Applicability Example - 3D Camera

There are two main parts in the 3D video and image acquisition process. The first one is the capturing of the image and the second one is the processing of the acquired images [124]. Usually, the acquisition is made by using a 2 camera system. These cameras need to be perfectly synchronized, otherwise the 3D effect could be lost or artifacts could appear in the image. Sometimes depth sensors are used together with single image sensors which can be used for 2D to 3D conversion [105], but this practice is used mostly in 3D video acquisition and it's not always the best approach to create high quality videos. When we are talking about 3D image acquisition, we need to consider the transmission and storage of this data as well.

A paper was published in (Andorko et al [4]), where the implementation of a 3D video acquisition system is described. The 3D video acquisition system was implemented on a single FPGA device, which was found on a Xilinx ML405 development board. Two different formats, the DLP 3D HDTV format 2.20 and interlaced layer format 2.21 were implemented in order to prepare the acquired images to be displayed. The 3D acquisition system is based on the FPGA-based stereoscopic camera which was presented in (Andorko et al [6]), and described in section 2.3.1.

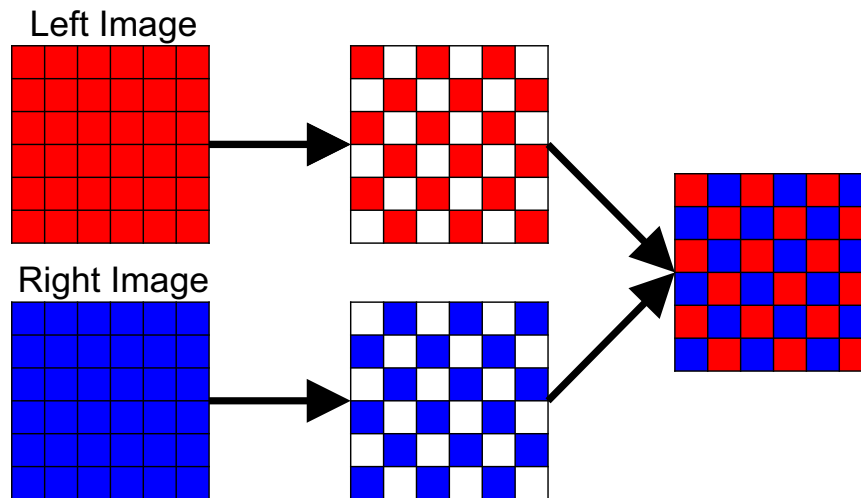


Figure 2.20: The DLP 3D HDTV video format

2.4 Challenges in Stereo Imaging

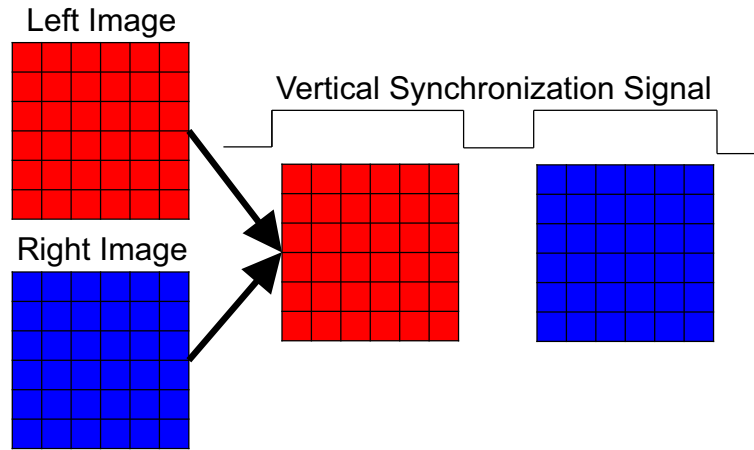


Figure 2.21: The interlaced layer format

2.4 Challenges in Stereo Imaging

When using stereo images, there are a number of challenges that can be faced by the researcher. Depending on the type of solution that the researcher needs to come up with, there are a number of ways to overcome these challenges.

2.4.1 Image Registration Issues

One common problem when using stereo cameras is the camera misalignment. The camera misalignment is defined as the offset of the left or right cameras in the x, y or z plane, figure 2.22.

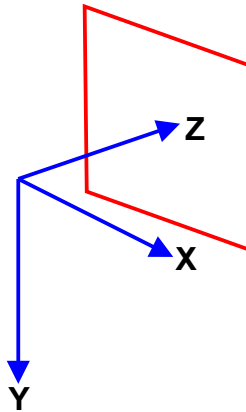


Figure 2.22: X, Y and Z planes explained

The offsets in the x,y and z planes are part of the extrinsic parameters of the

respective camera. For the purpose of distance measurement, in the case where accurate results are needed, the calibration of the cameras is required in order to find out the values of the extrinsic camera parameters. This calibration can be performed offline, in a controlled environment, where test targets are used for the calibration, or it can be done online, where an automatic calibration algorithm needs to be implemented. Beside the offset in the x, y and z planes, the yaw, pitch and roll angles of the camera are also part of the extrinsic parameters. In order to determine these parameters, the projection matrices of the cameras need to be defined. The projection matrix is a mapping between the co-ordinates of a feature in the World Coordinate System (WCS), and its corresponding position in the image. Once these matrices have been defined, and a number of corresponding points in the stereo images are determined, the calibration can be performed at sub-pixel accuracy according to (Dang et al [38]).

In the case of the stereo camera application, where the knowledge of exact distances is not required, an exact calibration is not needed. The vertical registration of the images will suffice for the tasks of disparity map generation, 3D image generation or certain multi-camera algorithms. An extensive survey of possible registration methods is detailed in (Zitova et al [153]).

The camera calibration performed during the research presented in this thesis, together with a comparison of calibrated vs uncalibrated images is described in section 5.4.2.

2.4.2 Asymmetrical colors

From a manufacturing point of view, two camera modules cannot be exactly similar. The main components of a camera module are the lens and the sensor, figure 2.23. Both of these components, due to the manufacturing process, introduce artifacts into the image in the type of (i) lens shading and (ii) cross-talk between neighboring pixels in the sensor.

2.4 Challenges in Stereo Imaging

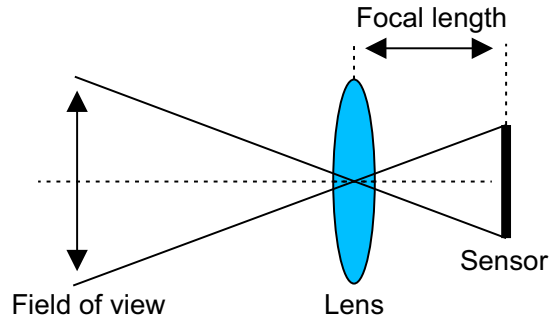
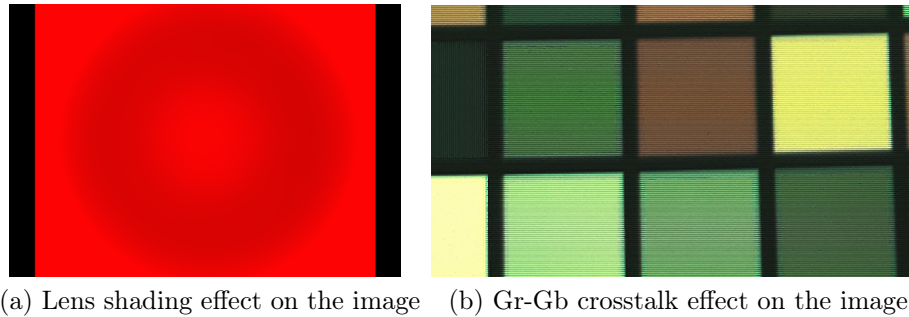


Figure 2.23: Camera module

In order to correct these artifacts, in the pre-processing section of the ISP, dedicated modules are used which need to be accurately calibrated for each batch of camera modules coming off the production line.

The calibration is usually performed on a small number of camera module samples. Based on these initial measurements, the correction modules are configured accordingly. Due to the reason, that the manufacturing process cannot be exactly the same for each camera module, the correction will only remove the artifacts up to a certain point. This leaves space for color inaccuracies when comparing images of the same scene acquired with two different camera modules. An example of lens shading artifact can be seen in figure 2.24a, whereas an example of Gr-Gb cross-talk artifact can be seen in figure 2.24b.



(a) Lens shading effect on the image (b) Gr-Gb crosstalk effect on the image

Figure 2.24: Color artifacts

In the case of the implementation of the stereo ISP presented in section 2.3, a custom-built stereo camera development board was initially used, which was equipped with two camera modules containing CMOS sensors. The pre-processing operation, which does the lens shading and cross-talk correction was performed on the camera

development board, with only the raw data being sent to the FPGA. For this reason, even if color asymmetry would have been encountered, there was no available solution to solve them.

2.4.3 Image Acquisition Synchronization Problems

One of the important features of a stereo camera is the fact, that it acquires both images exactly in the same time, making sure that the scene layout does not change from frame-to-frame. It is the system designer's job to make sure the synchronized acquisition happens.

There are several cases, where the acquisition cannot happen synchronously. The first one would be the case, when there isn't enough available bandwidth in order to save the information coming from both sensors in the same time. In cases like this, some frames might be dumped. At frame rates higher than 30 fps, dumping a frame is barely noticeable from the user's point of view, but it can influence the image processing algorithm which uses the stereo image pair.

In another case, when the exposure is controlled for both camera modules independently, if the amount of light entering one of the lenses is lower than the other one, the exposure time might be increased. This increase in exposure influences the frame rate, and data might not be saved synchronously. This is why, in the case of stereo camera systems, all the controls must be applied to both camera modules in the same way. Also it's important to leave plenty of memory bandwidth in order to be able to save the frames in the same time. In the implementation presented in section 2.3 all these aspects were considered, and for this reason, there were no image synchronization issues beside the EMI problem detailed in subsection 2.3.2.

2.4.4 Out-of-Focus Regions

Care needs to be taken when acquiring stereo images to set the appropriate depth of field, so that objects in both the left and right view are in focus. In some situations, when the focus is set independently in the camera modules, it might happen that for larger objects, placed at an angle respective to the stereo camera, the individual lenses will focus in a different way, see figure 2.25.

2.4 Challenges in Stereo Imaging

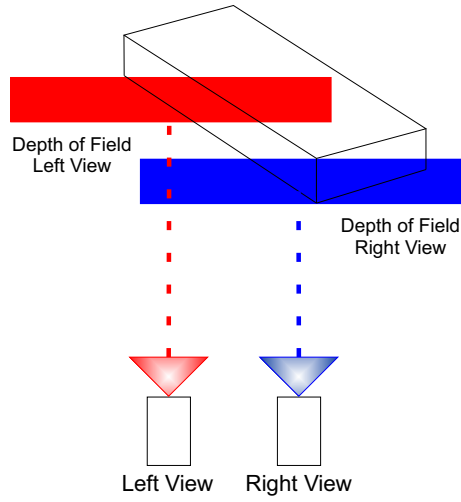


Figure 2.25: Depth of Field stereo camera

The result will be, that the in-focus and out-of-focus areas of the object will differ between the left and right views, figure 2.26.

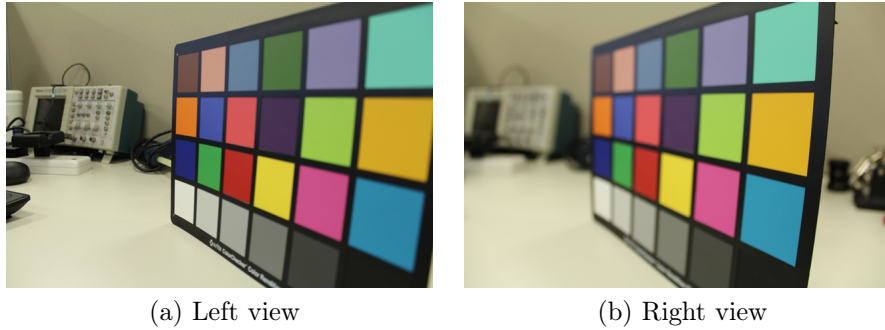


Figure 2.26: DoF example

The solution in this case is to (i) control the focus settings in a similar way for both the left and right camera module, or (ii) make sure that objects in the scene are placed past the hyperfocal distance of the lens.

When the lens is focused to infinity, the hyperfocal distance is the closest distance at which an object closer than infinity will still be in focus. The hyperfocal distance is calculated using formula 2.5, where H is the hyperfocal distance, f is the focal length, N is the f-number and c is the circle of confusion limit. The circle of confusion limit defines the amount of sharpness that is considered to be acceptable.

$$H = \frac{f^2}{N \cdot c} + f \quad (2.5)$$

In the case of all the acquired test images in this thesis, the objects in the scenes were placed past the hyperfocal distance where it was possible, or the focus was manually adjusted.

2.5 Conclusions

In section 2.1 different stereo concepts were introduced. Section 2.2 presented a literature review of image signal processors, which was followed by section 2.3, where the implementation of a stereo ISP was described. Challenges which can be encountered in stereo imaging were described in section 2.4.

The current chapter describes the basic concepts of stereo imaging such as (i) stereo images, (ii) stereo baseline, (iii) parallax and (iv) depth maps. These concepts will help the reader understand different test setups and experiments described later in the thesis. In section 2.1.3.1, a disparity measurement method is introduced, which allows the calculation of the disparity caused by parallax for different camera-object distances.

The implementation of a stereo Image Signal Processor is also presented in the current chapter. The research and challenges faced during this implementation helped determine a number of general challenges in stereo imaging, which need to be overcome by all researchers in this area. These challenges include (i) image registration issues, (ii) asymmetrical colors, (iii) image acquisition synchronization issues and (iv) out of focus regions.

Chapter 3

Literature Review of Depth Map Generation Techniques

This thesis has two different literature reviews. The first literature review was presented in section 2.2, and it reviewed work related to the design and implementation of Image Signal Processors. This review was helpful in the design and implementation process of the stereo ISP, which was described in section 2.3.

As described in the introduction chapter, the main goal of this research was the creation of a testbed which is suitable for the testing of depth map generation methods used in consumer electronic devices. For this reason, it was considered appropriate to dedicate a separate chapter for the review of depth map generation techniques.

The purpose of this review is the selection of a number of depth map generation algorithms which will be used throughout the research presented in the thesis.

In the current chapter, previous work related to the generation of depth maps is described, analyzing both methods which generate depth maps by using images acquired from a single camera in section 3.1, and methods that generate depth maps by using images acquired from a stereo camera in section 3.2. Since the stereo camera and stereo ISP was in the center point of the research from the start, previous work related to depth map generation methods from stereo images is described in more detail.

The approaches and methods used by other researchers in order to generate depth maps are important for the research presented in the current thesis, because they were taken into consideration when the testbed, test scenes and test scenarios were

designed. The conclusions drawn from the reviews are detailed in section 3.3.

3.1 Depth Determination using a Single Camera

Some of the earlier methods used for depth map computation in the consumer electronics industry were based on images acquired using single cameras [108, 56, 42, 75]. For this reason, the current section is dedicated to the review of work related to the generation of depth maps from images acquired with a single camera.

Some of the reasons for the use of single cameras for depth map generation were the costs involved in the manufacturing of stereo cameras, the lack of memory bandwidth, and the low processing speeds that were available at that time. It is only recently that we started seeing stereo cameras on the consumer market, for example the **FujiFilm W3 3D**. Due to this fact, certain methods were developed in order to create depth maps from single image sources.

The purpose of this section is to find out how these methods work, and what type of scene setups were used for the testing of the proposed methods. This information will be useful for the creation of the test bed and test scenes, which are part of the main goal of the research.

3.1.1 Depth from Defocus

One of the earlier articles that presents a number of methods that can be used to determine depth from defocus is the one described by (Ens et al [42]). The authors explain the concept of depth from defocus in detail, providing different solutions and approaches to the problem. The concept involves the calculation of distances to certain points within an image, a calculation that is based on the modeling of the effect the camera's focal parameters have on images that are acquired with a small depth of field setting. The most challenging part of the whole process is the deconvolution and modeling of the defocus operator. According to the survey, in order to recover the defocus operator, two images of the same scene need to be acquired using different focus distance settings for each image. The most common method of isolating the defocus operator is inverse filtering. In one of the first experimental works regarding depth from defocus, presented by *Pentland* [103], the author acquired one image by using a camera with a pinhole aperture, and the other image with a larger aperture. It was assumed that the defocus operator is a Gaussian,

3.1 Depth Determination using a Single Camera

and it was recovered by a division in the spatial frequency domain. Several other researchers used a similar approach, but improved Pentland's technique by using a higher order regression solution [23] or by generalizing the Gaussian spread parameter [128].

In the paper published by (Hwang et al [56]) the difference between shape from focus and shape from defocus is discussed where they also propose a depth recovery algorithm, where the defocus process is modeled as a two-dimensional Gaussian point spread function. After smoothing the input images, the depth is recovered by using a series of mathematical equations.

(Lai et al [75]) propose a generalized version of the depth estimation algorithm that was first proposed by Pentland in [103]. In this version of the depth recovery algorithm the depth from defocus is estimated using the vicinity of the edge from the input raw image. The proposed algorithm is also less sensitive to the noise resulting from the measurements due to the fact, that no differentiation operation is required on the image before the optimization process.

A different approach to depth from defocus was introduced in (Rajagopalan et al [109]), where the depth from defocus method is combined with stereo images in order to recover the depth information of a scene. In the presented method, both the depth map and the original scene are modeled as Markov random fields, and their estimates are obtained by minimizing an energy function. Although the proposed method is computationally less efficient, it provides better depth recovery results, and a space-variant restoration of the original focused image of the used scene. The method uses two pairs of corresponding stereo images, each pair blurred with a different blur parameter, and by performing an analysis of the image, once the blur is estimated, using the camera settings, the disparity can also be calculated, without solving the correspondence problem that is usually required in the case of depth from stereo.

The work reviewed in the current section describes depth from defocus methods used for the generation of depth maps. The majority of the methods use two images of the same scene acquired consecutively by a single camera, with different focus settings. The information acquired in this section was useful for the creation of the general depth map testing scene, which was part of the main goal of the research. To be more exact, care needed to be taken, that in order for the researchers to be able to test their depth from defocus methods using the proposed test scene, the objects in the scene had to be placed between the camera and the hyperfocal distance. The hyperfocal distance is described in section 2.4.4. If the objects are placed beyond

the hyperfocal distance, the proposed image processing algorithms don't work.

3.2 Depth Determination using a Stereo Camera Setup

Stereo vision and especially depth map generation are very active research areas in image processing and computer vision. A lot of research has been carried out in relation to this topic, and with the advancement of the technology, many research algorithms have become part of consumer electronics products.

As mentioned in the introduction chapter of this thesis, the main goal of the research was the creation of a testbed and test scene for depth map generation methods based on stereo images. In order to test and verify the proposed testbed and test scenes, a number of known depth map generation algorithms were selected to be used for this purpose. The selection was made based on a review of previous work published in this area. The conclusions drawn from this review, and the method used for the selection of the algorithms used for testing is presented in section 3.3 of the current chapter.

The generation of depth maps from stereo images is based on the identification of corresponding pixels in the stereo image pair. The initial corresponding pixel identification is achieved using basic pixel matching cost functions. The pixel matching cost is a formula which calculates the error when matching two pixels that are thought to be corresponding. If the error is high, it means that the cost is high, and the likelihood that the two pixels are similar is low. These matching costs lie at the base of the majority of the depth map generation algorithms, as it will be presented in the current section. Some of the most common pixel matching cost functions include cross-correlation [145, 107], absolute differences [65, 138, 16] and squared differences [100, 121, 20, 135]. Other approaches have been tried as well in the early days of depth map computation, such as gradient-based approaches [115] or cooperative algorithms that are inspired by models of the human vision [84, 132]. The research in most of these areas hasn't continued due to the main reason, that the results produced did not match the performance of the newly proposed algorithms.

3.2.1 Correlation based Techniques

As mentioned in section 3.2, a number of pixel matching cost functions lie at the base of a large number of depth map generation algorithms. For this reason, the

3.2 Depth Determination using a Stereo Camera Setup

purpose of the current section is to review algorithms which use these matching cost functions, and to determine their advantages and disadvantages. This review is required for one of the main goals of this review, which is the selection of a number of algorithms suitable to be used throughout the research presented in this thesis.

A more detailed description of these algorithms will be presented in section 4.2.

3.2.1.1 Normalized Cross-Correlation

One of the first documents that uses normalized cross-correlation as a matching method for depth map computation was a PhD thesis by *Hannah* [51]. The author chose this to be the measure of match between two areas. He also proposed the calculation area to have a rectangular shape, with odd dimensions. The proposed formula for normalized cross-correlation (NCC) is:

$$NCC(x, y, d) = \frac{\sum_{(i,j) \in W} L(x + j, y + i) \cdot R(x + d + j, y + i)}{\sqrt{\sum_{(i,j) \in W} L^2(x + j, y + i) \cdot \sum_{(i,j) \in W} R^2(x + d + j, y + i)}} \quad (3.1)$$

(Yang et al[145]) present a framework that matches a pair of stereo images, and which has a self-contained local matching module that is cascaded with a global matching module. Their novel approach is based on the decoupling of local matching from global matching, opposed to what other researchers have proposed in [9, 142] where they were using local matching embedded in the global matching algorithm. The method described in this paper is based on a divide-et-impera principle, where in the stereo matching process the global method is preceded by the local method. The proposed local method creates an image which was named by the authors spatio-disparity space image (SDSI), and which is computed using the variance-normalized correlation of different image patches using correlation windows. In the next step global matching is used which is based on a fine-to-coarse sweep and a coarse-to-fine-sweep. The objective, according to the authors, is to accumulate neighborhood support for each image point and for every disparity value. This was implemented using an adaptive non-linear filter.

3.2.1.2 Absolute differences-based correlation

Another matching cost used in correlation-based methods for stereo image matching that is worth mentioning is the Sum of Absolute Differences (SAD), which was used

in a number of papers more recently. When it comes to presenting this method, one of the most cited papers is the one by (Kanade et al [65]), where the authors present a new iterative stereo method which selects the appropriate correlation window based on the local variation of the intensity. Their method is iterative, and after each iteration and stereo matching, the window size is expanded into only one direction. If the uncertainty increases, the window is prohibited to expand into that direction, and expands into the other available directions. This way, they can find a window that has an ideal size for stereo matching. Their reason for this approach is that according to them, a larger window is good for flat surfaces, but the side-effect is that it blurs the disparity edges, whereas a smaller window gives sharper disparity edges, but the resulting image will be noisy. The formula for the SAD is:

$$SAD(x, y, d) = \sum_{(i,j) \in W} |L(x + j, y + i) - R(x + d + j, y + i)| \quad (3.2)$$

An adaptive window method similar to the one presented in the paper by (Kanade et al [65]) was also used by (Georgoulas et al [48]), where they talked about the implementation of a real-time disparity map computation module. The authors present a three-stage hardware implemented module. At the first stage, the local variation of the image is computed in order to determine the corresponding window size. At the second stage, the disparity map is computed using the SAD-based correlation method and different window sizes. At the last stage, the disparity map is filtered in order to remove the unwanted noise. The window sizes were chosen based on the local intensity variation of the pixels, with the idea that for a lower local variation value, a larger window was chosen. For the disparity map generation, the authors used SAD over other methods such as Sum of Squared Differences (SSD), or cross-correlation due to speed considerations. The final filtering was done using a Cellular Automata (CA) due to the fact that it performed better than other filters, and from a computational point of view, it's best suited for a VLSI implementation.

Another algorithm which is suitable for real-time implementation, and which uses SAD-based correlation is defined in (Di Stefano et al [40]). Beside using the SAD-based correlation method for better performance the authors also use two other techniques. The first one is called bi-directional matching, where they are looking for matching pixels both from left-to-right and right-to-left. In the case where for the same pixel a better result has been found, the previous match is rejected and the new one is considered. The second technique that is meant for performance

3.2 Depth Determination using a Stereo Camera Setup

improvement is the uniqueness constraint. The uniqueness constraint is based on the bi-directional matching and it assumes that at least one of the left-to-right and right-to-left matches is incorrect.

A similar idea to the uniqueness constraint is presented in (Muhlmann et al [91]), where the authors use a technique called *uniqueness of the minimum*. In their paper, the authors present a correlation-based disparity map computation system. As a matching cost they have chosen SAD over SSD and normalized correlation due to the fact that it performs better in the presence of outliers, and has a smaller computational complexity. After the matching was done with SAD, they also do a left-right consistency check, which in a similar way as in the previous paper, substitutes the pixels with a higher matching error. At the end, they implement the uniqueness of the minimum, which is used in regions with low texture. They are keeping track of the three smallest minima. If the third smallest minima doesn't fall under a pre-defined threshold, this means that the pixel is marked as not having a unique minimum.

3.2.1.3 Squared Differences-based Correlation

The *Sum of Squared Differences* (SSD) is a pixel matching cost which has been used and is still used in a large number of depth map computation algorithms. In a paper presented by (Kanade et al,[64]), the authors are describing the implementation of a video-rate stereo machine, which is capable of generating a dense range map. The depth computation method described in this paper uses multiple images obtained by a number of cameras, to produce different baselines in lengths and direction. By using this method, the resulting algorithm is suitable for hardware implementation. The correlation function the authors use is the SSD. For each of the available baselines, they compute the SSD for each pixel using a sliding correlation window. For each depth value, the resulting SSDs are added up, resulting a more clearer, unambiguous minimum, which will be used for the depth computation. The stereo method has three stages: the first step is the Laplacian of Gaussian filtering of the images, the second step is the SSD described earlier, and the last step is the depth and uncertainty calculation. Another paper, which shares one of the authors and uses SSD as a pixel matching cost is the one presented by (Okutomi et al [100]). In this paper, an algorithm that uses locally adaptive windows for pixel matching is introduced. The matching of two pixels is done by calculating the SSD in a correlation window who's

size can be changed based on the information from the image.

(Bobick et al, in their 1999 paper [20]) present a method for stereo matching which performs well in areas with a large occlusion. The method is based on the definition of a disparity space image (DSI), which was used also by [46, 145, 33, 98], for which the squared differences correlation method is used. In order to find matching points in the corresponding stereo images, an algorithm based on dynamic programming is proposed.

The formula used for SSD is:

$$SSD(x, y, d) = \sum_{(i,j) \in W} (L(x + j, y + i) - R(x + d + j, y + i))^2 \quad (3.3)$$

3.2.1.4 Sum of Hamming Distances-based Correlation

The *Sum of Hamming Distances* (SHD) is a matching cost which has become widespread more recently. It is used in conjunction with the *Census Transform*, which is an image texture classifying filter, as it can be seen in [86, 102].

(Miron et al [86]) present a color matching algorithm which can be used in the case of intelligent vehicles. In their paper, they introduce an alternate version of the census transform, which is named Cross-Comparison Census, and which, according to the authors, is robust to noise and illumination variations. After the generation of the cross-comparison census for both the left and right image, the sum of hamming distances approach is used for the identification of corresponding pixels in the stereo image pair, using the following formula:

$$SHD(x, y, d) = \sum_{(i,j) \in W} L(x + j, y + i) \text{ bitwiseXOR } R(x + d + j, y + i) \quad (3.4)$$

The same sum of hamming distances is used in the work presented by (Pei et al [102]) as well, where the authors create a depth image using information from a sparse census transform and segmentation, an image which is later used in a Image-to-Sound Mapping system developed for visually impaired users.

3.2.2 Global Optimization based Techniques

During the early years of computer vision for depth map generation purposes the algorithms described in section 3.2.1 were used. There was a significant develop-

3.2 Depth Determination using a Stereo Camera Setup

ment in this field when the global optimization techniques were introduced. These techniques were introduced at a time, when more powerful computing systems and dedicated DSPs were developed. For a number of years, according to the database introduced in (Scharstein et al [118]), global optimization-based algorithms were the best performing ones.

The reason why a review of these algorithms is considered necessary is that with their introduction the quality of the depth maps improved significantly, but it was not possible to use them in real-time systems due to the fact that they were computationally expensive. This drove researchers to develop new algorithms, which generate depth maps of the same quality, but in the same time are computationally less expensive.

Markov random fields were also used in the algorithm presented by (Boykov et al [25]) which is one of the earlier papers that presents a global algorithm for depth map computation. The authors show that the maximum a-posteriori estimate of an MRF can be obtained by solving a multiway minimum cut on a graph, and by formulating the visual correspondence problem as an MRF, the algorithm yields promising results on real data. In this paper, 2 algorithms are described, which are based on graph cuts, and can efficiently find a local minimum. The first algorithm that is based on expansion moves finds a labeling within a known factor of the global minimum, the second one, based on swap moves, handles more general energy functions. The goal is that for every pixel $p \in P$, a label must be assigned in a finite set L , and to find a labeling f that assigns each pixel a label $f_p \in L$, where f is both piecewise smooth, and consistent with the observed data. These problems from vision can be formulated in terms of energy minimization:

$$E(f) = E_{smooth}(f) + E_{data}(f) \quad (3.5)$$

In this case, E_{smooth} measures the extent to which f is not piecewise smooth, while E_{data} measures the disagreement between f and the observed data. The form of E_{data} is typically:

$$E_{data}(f) = \sum_{p \in P} D_p(f_p) \quad (3.6)$$

where p is the location of the pixels within the image P . D_p measures how well label f_p fits pixel p given the observed data. The choice of E_{smooth} is an important issue, and a number of different solutions have been proposed [77, 50].

The energy function considered by the authors in this paper has the form of:

$$E(f) = \sum_{\{p,q\} \in N} V_{p,q}(f_p, f_q) + \sum_{p \in P} D_p(f_p) \quad (3.7)$$

where N is the set of interacting pair of pixels, and $V_{p,q}$ is the penalty of the pixel pair $\{p, q\}$.

For the past few years, the best performing depth map computation algorithms are based on global optimization according to (Scharstein et al [117]). (Xu et al [143]) tackle the problem of partial occlusions in stereo images. They measure how likely it is for a pixel to be occluded by introducing the Outlier Confidence (OC). For the disparity estimation of both the occluded and non-occluded pixels, an algorithm based on global optimization is used. (Yang et al [144]) present a stereo matching algorithm that handles disparity, discontinuity and occlusion. The main contribution of the paper is that even though the energy minimization algorithm works in the same way with other similar approaches [143, 25], the data term in the function is updated based on the current understanding of which pixels in the image are occluded, or unstable due to low texture. The overall algorithm also integrates several other approaches like belief propagation [72, 79, 130], left-right consistency check [40, 92] and color segmentation [19, 73, 78].

(Wang et al [140]) present a global stereo model, which uses Ground Control Points (GCP) in order to generate accurate depth maps. They model the GCPs in a Markov Random Field. The GCP can be obtained from multiple sources, like accurately matched pixels, laser scan range data or user intervention. The authors do not propose a fixed way. This also allows them to integrate the information from multiple sensors. In their approach, the global energy function used differs from the previous ones, the difference being the addition of the term E_{gcp} , which refers to the GCP energy, and it encodes constraints from sparse GCPs.

$$E(f) = E_{smooth}(f) + E_{data}(f) + E_{gcp}(f) \quad (3.8)$$

3.2.3 Segmentation based techniques

According to the database proposed in (Scharstein et al [118]), segmentation-based techniques are the best performing algorithms nowadays. The quality of the generated depth maps is better than in the case of the global optimization techniques, but in the same time, some of the algorithms can be used in real-time applications.

3.2 Depth Determination using a Stereo Camera Setup

As mentioned in the introduction of this chapter, the main purpose of the literature review was to find an algorithm suitable to be used throughout the research presented in this thesis. Since the segmentation-based algorithms generate the highest quality depth map, it was considered necessary to review some of the best performing ones.

In the article written by (Tao et al [135]), a global matching framework for stereo computation is described. The basic depth map computation algorithm is based on color segmentation, and then, this approach is extended to a generalized global matching algorithm. The local matching for depth representation is done using squared differences-based correlation. The main idea of the paper is that instead of enforcing global visibility based on local matching scores, as presented in [84, 99, 24, 113], the authors employ a global match measure. The key components of the current framework are a compact region-based depth representation, efficient incremental warping algorithm, a hypothesize-and-test method for depth searching, and finally scene constraints for solution regularization. Two ideas from this paper, which are the depth map computation based on color segmentation and depth map generation based on global optimization, were widely researched in more recent papers.

The paper by (Bleyer et al [18]) proposes a stereo model which encodes the assumption that a scene is composed of a few, smooth surfaces. They claim that an aspect of a good model for a stereo image that they are using in their approach has not been expressed in any existing approach in a principled way. They are talking about the aspect of self-similar pixels that belong to the same 3D surface. In their approach, the authors propose a pixel-wise Markov Random Field (MRF) formulation, that assigns each of the pixels to a 3D surface. They say, that by doing this, they can overcome two major problems of segmentation-based stereo. These are recovered from segmentation errors by showing that it is possible to express the segmentation as a soft constraint and they can go beyond the planar world assumption. This way they make use of more B-spline surfaces.

Mr Bleyer is also the main author of two more research papers [17, 19] which are related to depth map generation with the help of image segmentation. In [17], the authors present an algorithm that simultaneously extracts disparities and alpha matting information from a stereo image pair. The approach is based on an energy minimization approach for combined stereo and matting. The algorithm is based on the partitioning the left and right images into segments of homogeneous colors. The transparency is modeled by assigning each segment pixel an alpha value and a color,

and the disparity is modeled by assigning each segment to a disparity plane. The final energy function is optimized by combining a greedy search procedure and belief propagation, which was also used in [72, 130].

A method for joint stereo matching and object segmentation is introduced in [19]. The presented model is formulated as an energy function that is optimized via fusion moves. The objects in the images are characterized by a color model, a 3D plane that approximates the object's disparity distribution, and a 3D connectivity property. Also, object-level color models are employed as a soft constraint, that are inspired by Markov Random Field type of image segmentation. This method is currently one of the best performing ones in the Middlebury ranking [119].

Another depth map generation method that is based on color image segmentation is the one described by (Hosni et al [54]), where the main author is also the co-author of the previously presented papers. In contrast with the previous papers, the current algorithm is a local depth map generation algorithm based on image segmentation. The paper presents the contribution of the authors to the segmentation of the images, based on which the stereo matching is done. The stereo matching used by the authors is a winner-takes-all approach, but other approaches such as belief propagation [72, 130] or graph-cuts [25, 114] can also be used. The authors propose a method which computes support weights that will be used in an adaptive weight approach introduced in [147], by using geodesic distances within a square window.

An over-segmentation based stereo algorithm that estimates both segmentation and depth is presented by (Taguchi et al [133]). The approach used is an over-segmentation approach, where it is assumed that all pixels in each segment have the same depth. The algorithm updates the shape and depth of these segments alternately by using a generative model of an image to update the segment shapes based on maximum a-posteriori estimation, and modeling the stereo constraints for depth estimation using Markov random fields, where the segment depths are updated using belief propagation.

3.2.4 Real-Time and Hardware-based Methods

More recently, due to the advances in consumer electronics, a number of papers have been published. The methods published in these papers, beside the information from the pair of stereo images rely on other information as well, such as depth information from a Time-of-Flight (ToF) camera [152, 70], or depth information from

3.2 Depth Determination using a Stereo Camera Setup

structured light-based systems [116, 137, 28]. The main reason for this is that global optimization techniques, which currently have the best performance when it comes to generating depth from stereo, can't be implemented in real-time applications at the moment [140]. Some other papers present algorithms that were implemented in hardware only [48, 69, 63], and for this reason, they cannot use the Middlebury testbed [118] in order to measure their performance.

(Zhu et al [152]) present a method which combines the results from a ToF camera and from passive stereo in order to achieve better overall depth map performance. Their idea is based on the fact that ToF cameras provide depth estimates in areas with low textures, where the traditional stereo techniques [17, 133, 143, 144] perform poorly. In contrast, the traditional stereo techniques provide accurate depth information on textured areas, where the ToF cameras have poor performance. In their approach, the probability distribution functions of the depth estimates from each of the sensors are fused by using a MRF. The ToF sensor provides local data, and belief propagation is used to perform global optimization on the combined sensor that will improve the accuracy of each of the depth sensors.

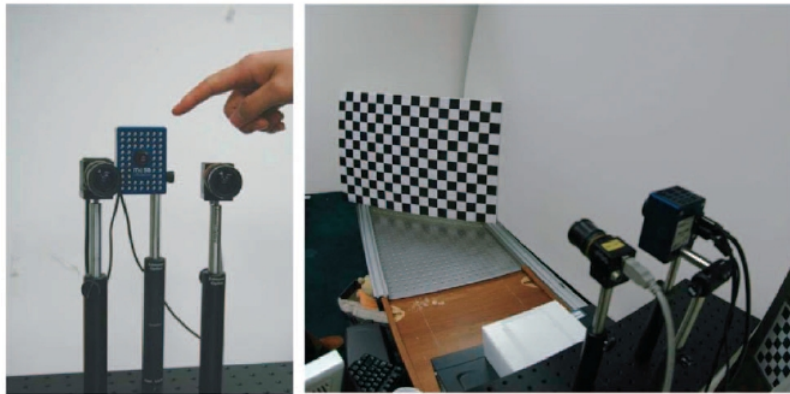


Figure 3.1: ToF and stereo camera setup [152]

A depth camera is used in (Kim et al [70]) as well, where the authors propose a scheme that generates depth maps based on regions of interest (ROI) that combines a low-resolution depth camera with high-resolution stereo cameras. Their method is based on the projection of the depth information from the depth camera onto the ROI of the left image, which was obtained from estimating initial depth information from a stereo matching algorithm, using 3D image warping.

Another method that uses a special camera is described in (Wan et al [139]),

where the authors present a depth map generation method that uses a dual-PTZ (pan/tilt/zoom) camera system. They propose to use the dual-PTZ-camera system because it can be used for large area surveillance and cooperative object tracking, features that dual-static-camera systems cannot provide. The depth information is obtained by using the alterability of the dual-PTZ-camera system parameters, and because of this, it can be regarded as an active vision system, just like [71, 136]. A coarse-to-fine iterative framework is proposed, where at each iteration, a higher resolution depth map is obtained, and a panoramic depth map is generated which allows dynamic updating using a mosaicking approach. For the depth map panorama, several depth maps are used, as the camera has a limited field of view (FOV).

(Kim et al [69]) present a hardware-based real-time stereo depth extraction module. The algorithm is based on pre-processing and parallel prediction searching using a median filter. A similar hardware implementation is presented in (Kalomiros et al [63]), where the authors present the design of a hardware co-processor for depth detection from a pair of stereo images, that is based on parallel implementation of SAD.

In the paper written by (Caspi et al [28]), the authors detail a structured light method, where the number and form of the projection patterns is adapted to the characteristics of the scene. The basic structured light system setup is similar with the stereo camera setup, the difference being the fact that one of the cameras is replaced by a projector. This projector has the form of a laser light source with cylindrical lens, and it projects light that creates a narrow stripe on the scene. With the help of these stripes, the 3D location of each point in the scene is retrieved because the intersection of a known illumination plane and a line of sight uniquely determines a point. In the proposed application, colored light is used in order to color-code each point in the acquired image. While projecting N different patterns, N images are acquired, and by analyzing these images, the depth of the scene is retrieved.

Another approach which uses structured light, but in this case has two cameras and a projector is the one presented by (Scharstein et al [116]). According to the authors, the purpose of developing such a system was to generate high quality depth maps of scenes, that can be later used to measure the performance of different depth map generation algorithms. In their setup they use one camera that is used to acquire both left and right view images, and two projectors, which illuminate the scene from different directions. The light pattern codes for each pixel are decoded in each view,

3.3 Conclusions

and these unique codes are used to compute correspondences between the left and right views, this way generating a depth map of the scene.

3.3 Conclusions

The current chapter presents a review of work published in the area of depth map generation using both single camera and stereo camera systems. The purpose of this review was the selection of algorithms to be used for testing during the development of the testbed and depth map testing scene, which are the core of the research presented in this thesis, and the selection of algorithms which will be used for future research.

In section 3.1, depth map generation methods which use a single camera for image acquisition purposes were reviewed. Since the core of the research is based on depth map generation from stereo images, this initial review was not a comprehensive one, but it was meant to provide information regarding the scenes used for the testing of these methods. While reviewing the depth from defocus methods, it was learned that in order for these methods to work, the objects in the test scene need to be placed between the camera and the camera's hyperfocal distance. If they are placed beyond the hyperfocal distance, the results will be inaccurate.

Section 3.2 presented a comprehensive review of depth map generation methods which use a stereo camera system for image acquisition. The section started with the review of several pixel matching costs, and the way they were used in different methods. These matching costs included Normalized Cross-Correlation, Sum of Absolute Differences, Sum of Squared Differences and Sum of Hamming Differences. In the following subsections, more advanced methods were reviewed such as global optimization-based methods in section 3.2.2 and image segmentation-based methods in section 3.2.3. In the case of each method, the initial disparity map generation is performed with the help of one of the pixel matching costs presented in section 3.2.1. After this initial stage, additional methods were used in order to improve the quality of the depth maps.

A third category of depth map generation methods was reviewed in section 3.2.4, where the researchers used specialized devices such as Time-of-Flight cameras and depth cameras in order to improve the quality of the depth maps. In some cases, they also combined different software-based and hardware-based methods. During

the review of this section, it was noticed that in order to test the methods which require additional devices, the researchers need to build their own test scenes due to the fact that they cannot use images from online databases as the one described in (Scharstein et al [118]). This was the point during the research, where it was decided, that beside a testbed, a general depth map testing scene should also be proposed, which can be used by any researcher in order to compare their work.

During the review presented in this chapter, a number of algorithms were selected to be used during the development of the testbed and general test scene. The purpose of the literature review from this chapter was not the comparison of the algorithms from performance point of view. Such a comparison is already available in the database proposed by (Scharstein et al [118]), where algorithms proposed in the past decade are found. The purpose was the selection of a number of algorithms which are suitable to be used in presented research in order to record their behavior under different image acquisition conditions and using different test scenes.

Due to the reason that in the case of well performing algorithms, small changes in the scene cannot be seen in the final depth map, the algorithms selected for research and development purposes should not be one of the best performing ones. In the same time, they should be used in a large number of depth map generation methods. For this reason, it was considered that the most suitable algorithms for the purpose of our research would be the correlation-based ones, presented in section 3.2.1. More exactly, algorithms based on Normalized Cross-Correlation, Sum of Absolute Differences, Sum of Squared Differences and Sum of Hamming Distances matching cost functions. This type of algorithms are used in both the correlation-based methods described in section 3.2.1 and some of the segmentation-based methods presented in section 3.2.3.

Chapter 4

Testing Methodology

This chapter describes the initial testing methodology proposed for the testing of depth map generation algorithms. There are several goals in this chapter which will be addressed. The first one (i) is the description of the disparity map generation methods which will be used during the initial testing, followed by (ii) the description of the disparity map performance measurement method, and concluding with (iii) the creation and presentation of the test scenes used for the initial measurements.

The introduction section of the current chapter, section 4.1 provides the detailed reason why a test methodology was developed. This section also contains a short literature review, which is considered necessary in order to present previous work in the field of depth map quality testing.

The chapter continues with the description of the four different pixel matching costs, Sum of Absolute Differences (SAD), Sum of Squared Differences (SSD), Normalized Cross-Correlation (NCC) and Sum of Hamming Distances (SHD) in section 4.2, which lie at the base of the depth map generation algorithms proposed for testing. Each of these pixel matching costs are used in different applications, for example the SAD is used in real-time applications, where speed is an issue due to its low complexity, whereas SSD is used in the case of applications, where quality is important. The NCC is used mostly in the case of industrial applications due to the fact, that it is more robust under different illumination changes compared to the other proposed algorithms. SHD-based algorithms use bit-level operations, and this makes them popular amongst embedded system developers. The selection criteria of the proposed depth map generation algorithms was presented in section 3.3.

The method used for the measurement of the algorithm performance is detailed in

section 4.3, where a comparison between the proposed approach and other available approaches is presented. The main difference between the approach proposed in the current work and other approaches is that in our case depth map quality is measured for each object independently, whereas in other cases, depth map quality is measured for the whole scene. As a continuity of the previous section, the method used for ground truth generation is introduced in section 4.4, ground truth which is used in the process of performance measurement.

The current chapter concludes with section 4.5, which describes the four initial test scenes that were proposed in order to measure the quality of the disparity maps generated using the selected algorithms.

All the results obtained during the research presented in the current chapter contribute towards the final goal of this research, as it will be described in chapter 7.

4.1 Introduction

As described in chapter 1, the present research evolved through several stages over the past years. When it was decided to focus on depth map generation, the first goal was the development of a depth map generation algorithm suitable for a consumer electronic device.

In order to find or develop a new depth map generation algorithm, several testbeds were looked at that were proposed for the comparison of different depth map generation methods.

(Mayoral et al [85]) present two approaches to the evaluation of the performance of the methods for the computation of a correspondence error. The first method they present is an evaluation of the compatibility of the computed error surface with ground truth data. This approach is one of the simplest approaches and is one of the first approaches used for the evaluation of depth maps. The main reason for its simplicity is that fact, that the method is based on the comparison of the resulting depth map with a real depth (ground truth). This approach lies at the base of the method proposed in the current thesis.

One of the earlier papers which talk about ground truth/true motion field is the one presented by (Barron et al [12]), where they present a survey of the performance of optical flow techniques. The ground truth method examines how well the minima

4.1 Introduction

of the correspondence errors are able to describe the disparity profile.

The second type of evaluation described in the same paper by (Mayoral et al [85]), is based on the shape of the computed error surface, and the ability to isolate the error minima from the other values is measured. Compared to other comparison methods and depth map quality measurement approaches such as the one presented in (Scharstein et al [117]), this evaluation does not provide the best results.

One of the better known baselines is the one proposed by (Scharstein et al [117]), which is referenced by almost all the important papers that are on topic of depth map generation, and were published in the past decade [72, 48, 40, 55]. The database provided by the (Scharstein et al [117]) is a very comprehensive one, and contains more than 100 different approaches to depth map generation, and for this reason, it was initially decided to use their test images and the methods from their website in order to find a suitable algorithm for our application. A screen-shot of their database can be seen in figure 4.1.

Error Threshold = 1 Error Threshold...		Sort by nonocc			Sort by all			Sort by disc			Average percent of bad pixels (explanation)			
Algorithm	Avg.	Tsukuba ground truth			Venus ground truth			Teddy ground truth				Cones ground truth		
	Rank	nonocc	all	disc	nonocc	all	disc	nonocc	all	disc		nonocc	all	disc
AdaptGCP [137]	7.9	1.03 13	1.29 6	5.60 15	0.10 4	0.14 1	1.30 6	4.63 19	6.47 9	12.5 19	1.81 1	5.70 1	5.33 1	<div><div></div></div> 3.83
ADCensus [94]	11.8	1.07 18	1.48 14	5.73 21	0.09 2	0.25 10	1.15 2	4.10 14	6.22 8	10.9 11	2.42 15	7.25 12	6.95 16	<div><div></div></div> 3.97
AdaptingBP [17]	15.3	1.11 21	1.37 9	5.79 23	0.10 6	0.21 7	1.44 7	4.22 16	7.06 13	11.8 16	2.48 19	7.92 24	7.32 24	<div><div></div></div> 4.23
CoopRegion [41]	15.7	0.87 4	1.16 1	4.61 4	0.11 6	0.21 6	1.54 11	5.16 28	8.31 17	13.0 23	2.79 38	7.18 11	8.01 41	<div><div></div></div> 4.41
DoubleBP [35]	20.6	0.88 5	1.29 5	4.76 7	0.13 10	0.45 42	1.87 19	3.53 10	8.30 16	9.63 6	2.90 42	8.78 51	7.79 33	<div><div></div></div> 4.19
RDP [102]	20.6	0.97 10	1.39 11	5.00 10	0.21 36	0.38 28	1.89 20	4.84 20	9.94 30	12.6 20	2.53 22	7.69 17	7.38 25	<div><div></div></div> 4.57
MultiRBF [153]	20.9	1.33 43	1.56 19	6.02 30	0.13 9	0.17 3	1.84 17	5.09 26	6.36 7	13.4 27	2.90 43	6.76 6	7.10 21	<div><div></div></div> 4.39
MDPM [165]	21.4	1.15 22	1.59 22	6.14 33	0.14 15	0.36 24	1.52 10	3.79 11	5.78 4	11.1 13	2.74 30	8.38 37	7.91 38	<div><div></div></div> 4.22
OutlierConf [42]	21.8	0.88 5	1.43 13	4.74 6	0.18 25	0.26 12	2.40 36	5.01 22	9.12 26	12.8 22	2.78 35	8.57 42	6.99 17	<div><div></div></div> 4.60
AdaptiveGF [151]	25.8	1.04 14	1.53 16	5.62 16	0.17 24	0.41 34	1.98 23	5.71 37	11.3 43	14.3 34	2.44 17	8.22 31	7.05 20	<div><div></div></div> 4.98
Laplace [160]	26.1	0.93 8	1.50 15	4.96 9	0.14 17	0.32 16	2.02 24	3.87 12	8.96 25	11.7 15	2.95 48	9.20 64	8.60 60	<div><div></div></div> 4.59
SOS [159]	26.2	1.45 54	1.63 26	7.83 72	0.21 34	0.32 17	2.29 35	3.13 6	8.45 19	9.74 7	2.43 16	7.10 10	7.02 18	<div><div></div></div> 4.30
SubPixSearch [127]	27.1	2.04 87	2.48 78	6.40 40	0.14 14	0.40 33	1.74 14	4.00 13	6.39 8	11.0 12	2.24 9	6.87 8	6.50 9	<div><div></div></div> 4.18
SubPixDoubleBP [30]	27.3	1.24 30	1.76 39	5.98 29	0.12 8	0.46 44	1.74 14	3.45 9	8.38 18	10.0 9	2.93 45	8.73 48	7.91 35	<div><div></div></div> 4.39
SurfaceStereo [79]	28.0	1.28 39	1.65 29	6.78 49	0.19 27	0.28 13	2.61 49	3.12 5	5.10 2	8.65 3	2.89 41	7.95 27	8.26 52	<div><div></div></div> 4.06
LLR [135]	29.7	1.05 15	1.65 28	5.64 17	0.29 59	0.81 73	3.07 60	4.56 17	9.81 29	12.2 17	2.17 6	8.02 28	6.42 7	<div><div></div></div> 4.64
LWM [164]	31.8	1.52 60	1.82 42	8.20 60	0.16 19	0.39 32	2.03 26	5.09 25	10.5 34	13.8 28	2.27 10	7.49 14	6.71 11	<div><div></div></div> 5.00
WarpMat [55]	31.9	1.16 23	1.35 8	6.04 31	0.18 26	0.24 9	2.44 40	5.02 23	9.30 27	13.0 25	3.49 63	8.47 41	9.01 67	<div><div></div></div> 4.98
ObjectStereo [98]	32.6	1.22 29	1.62 23	6.36 37	0.59 96	0.69 65	4.61 87	4.13 15	7.59 14	11.2 14	2.20 7	6.99 9	6.36 5	<div><div></div></div> 4.46
TF-ASW [154]	34.4	1.65 66	1.96 54	5.90 28	0.14 13	0.31 15	1.51 9	6.25 52	11.8 59	15.1 45	2.49 21	8.32 34	7.02 19	<div><div></div></div> 5.21
LM3C [158]	34.8	2.10 90	2.44 75	8.01 77	0.12 7	0.39 31	1.23 3	5.46 32	10.9 38	14.9 43	2.12 3	7.59 15	6.14 3	<div><div></div></div> 5.12

Figure 4.1: Middlebury database

The test images in (Scharstein et al [117]) are acquired in a controlled environ-

ment. The presented scenes are not real-world scenes, and the light and distance setup conditions do not vary. The images that were used for the tests presented in their research paper can be seen in figure 4.2. These images are ideal for academic research purposes.

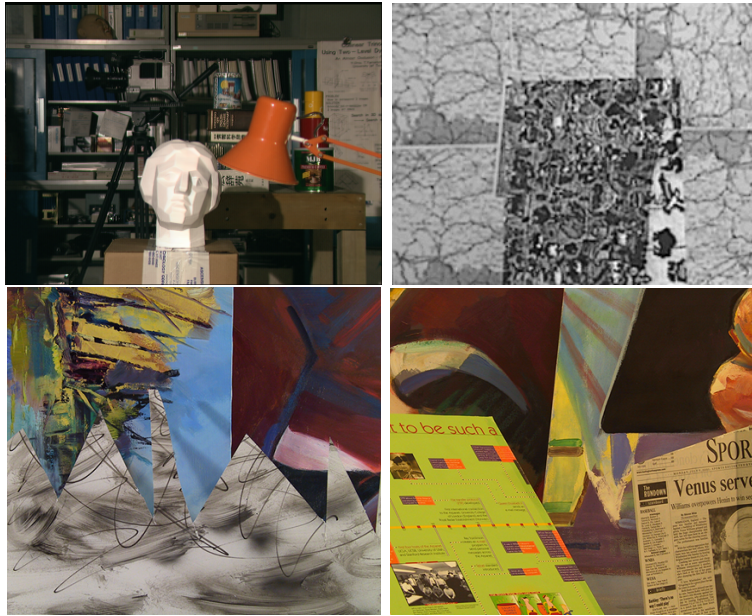


Figure 4.2: Test pictures used in [118]

In order to be used for the testing of consumer electronic devices, a test bed needs to provide the means to (i) test different types of consumer electronic devices, (ii) test the same consumer electronic device using different settings and (iii) test consumer electronic devices under different image acquisition conditions. The research papers presented in the previous paragraph did not provide the means to test consumer electronic devices due to the main reason that they only provide an online database, and the test scene and acquisition conditions can not be changed. This it led us to the main task of the current research, which was to create a new test bed, where the issues regarding the testing of consumer electronic devices can be addressed. The development of such a test bed is presented in the current chapter, where the selected pixel matching costs used for the initial disparity map generation, and the measurement methods are presented.

4.2 Description of Correlation Method-based Algorithms proposed for Testing

The majority of the stereo image - based depth map generation algorithms described in [117] use a **Pixel Matching Cost** in order to find the corresponding pixels in the stereo image pair, and to create the disparity map. After this initial stage, additional methods such as uniqueness constraint, bidirectional matching [40] and segmentation [18] are used in order to improve the performance of the algorithm.

Four different matching costs have been chosen for the current tests based on the selection criteria defined in section 3.3. At this stage, the purpose was to analyze the quality of the disparity maps generated using matching-cost based correlation. These matching costs are Sum of Absolute Differences (SAD), Sum of Squared Differences (SSD), Normalized Cross Correlation (NCC) and Sum of Hamming Distances (SHD).

The pixel matching costs proposed for testing were incorporated into a simple correlation-based disparity map generation algorithm. As described in section 2.1.4, the depth map is inversely proportional to the disparity map. The algorithm looks for corresponding pixels in the stereo image pair, by scanning the corresponding line in the right image, as it can be seen in figure 4.3. The locations of the pixels in the left image are considered as being the *origin*. Once the corresponding pixel in the right image is found, the disparity value d between the locations is recorded as a pixel value in the disparity map. During the experiments, the maximum disparity value up to which the scanning is performed was set to 60, 90 and 120. In all three cases similar performance results were achieved. These experimental results will be presented in chapter 6.

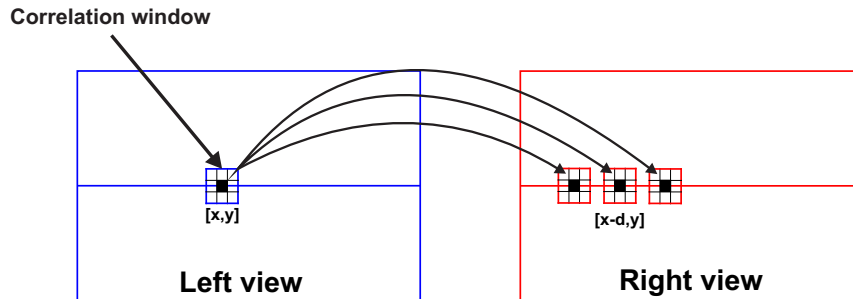


Figure 4.3: Corresponding pixel search

The search for corresponding pixels is not performed at a pixel level. A correlation windows of size 9x9 pixels is used, where the currently processed pixel lies in the

center of the window. As explained in (Kanade et al [65]), the correlation windows can have multiple sizes. Based on the experiments shown in (DiStefano et al [40]) the 9x9 sized correlation window was shown to be the most effective, and for this reason a window of the same size was selected for use in our experiments. The proposed **Pixel Matching Costs** are used within the **Correlation Window** of the algorithm.

An example of corresponding pixel search in real-world images can be seen in figure 4.4, where the left and right views of the stereo image are combined into a single image.

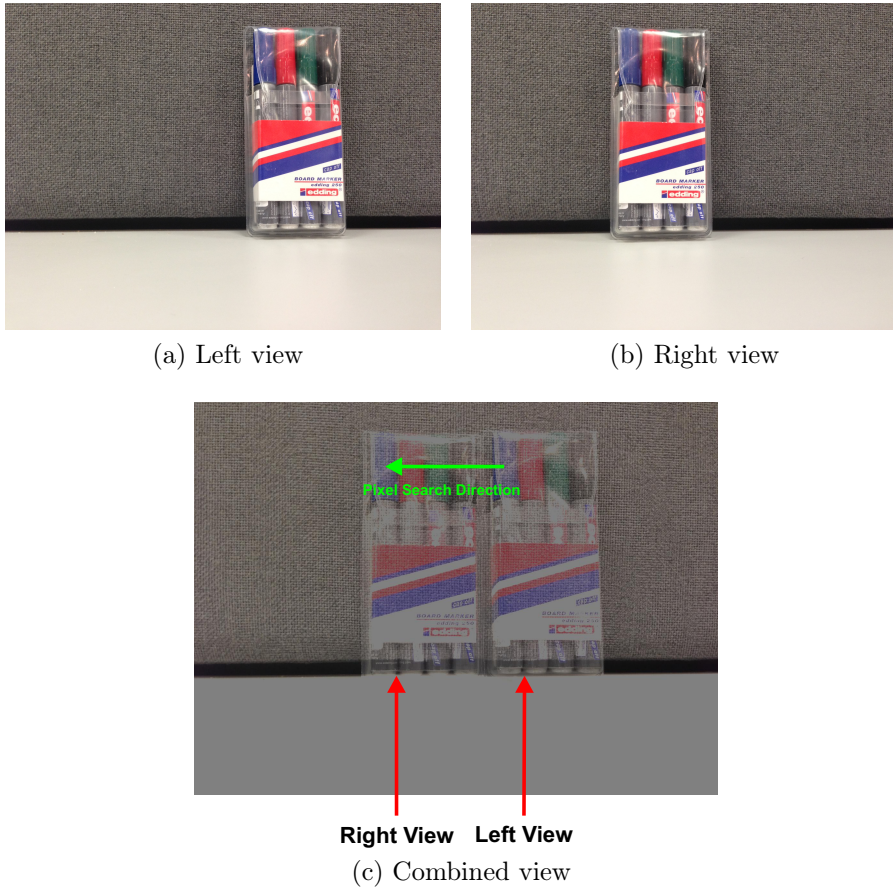


Figure 4.4: Combined stereo images

A detailed comparison of the performance of these algorithms will be presented in chapter 6. Here the algorithms themselves are described and their benefits and drawbacks are commented.

4.2.1 SAD Matching Cost

With the help of the SAD-based correlation method, the values of the pixels within the correlation window are analyzed, and the corresponding pixel is chosen based on the smallest error value, which is computed as the absolute difference between the two correlation windows. This is a simple method, but one that was proved useful by several authors [40, 48]. Due to its simplicity, it can be easily implemented in hardware, and this is another reason why it proved to be popular.

The formula for the SAD correlation method is:

$$SAD(x, y, d) = \sum_{(i,j) \in W} |L(x + j, y + i) - R(x + d + j, y + i)| \quad (4.1)$$

where x and y are the coordinates of the image, d is the search distance on the line and i and j are the coordinates within the correlation window W . The method scans the corresponding horizontal lines and finds the best match using the absolute value of two pixel luminance difference.

The disadvantage of this method is, that it cannot handle depth discontinuities well. Depth discontinuities can be noticed at the edges of the objects. For this reason, additional pre- or post-processing methods are usually used in the algorithms. Pre- or post-processing methods include image segmentation, right-left and left-right search for corresponding pixels, and uniqueness constraints, see section 3.2.1.

4.2.2 SSD matching cost

The SSD-based correlation method works in a similar way with the SAD-based method. The only difference is that instead of computing absolute differences of the pixel luminance, square differences are computed, and this improves the performance by a small amount, as presented in [117]. The major disadvantage is, that when implemented in hardware, the gate count is double compared to the implementation of the SAD due to the fact, that squared values are used.

In the case of the SSD method, the matching cost is the squared difference of intensity values at a given disparity. In this case, the formula is:

$$SSD(x, y, d) = \sum_{(i,j) \in W} (L(x + j, y + i) - R(x + d + j, y + i))^2 \quad (4.2)$$

This method was used in (Zhao et al [151]) and (Kanade et al [65]), where beside

the SSD, the authors also used a variable sized correlation window.

4.2.3 NCC matching cost

In the case of the NCC-based method, the formula applied in the correlation window is:

$$NCC(x, y, d) = \frac{\sum_{(i,j) \in W} L(x + j, y + i) \cdot R(x + d + j, y + i)}{\sqrt{\sum_{(i,j) \in W} L^2(x + j, y + i) \cdot \sum_{(i,j) \in W} R^2(x + d + j, y + i)}} \quad (4.3)$$

The NCC is more robust than SAD and SSD under uniform illumination changes, and for this reason it has been widely used in object recognition and industrial inspection [141].

The major disadvantage is, that even though it performs well in a controlled environment, its performance is worse when it comes to working with natural images. (Wei et al [141]) use the NCC for a pattern-matching algorithm, whereas (Sun et al [131]) are using NCC for real-time tracking of imaging targets in video sequences.

4.2.4 SHD matching cost

The formula of the SHD uses bit-level operations on the input data, and for this reason, the SHD method is usually used for matching **Census-Transformed** images, but it can also be used in images that have not been transformed. The **Census Transformation** used in this case is presented in the following paragraph.

The SHD formula applied in the correlation window is:

$$SHD(x, y, d) = \sum_{(i,j) \in W} L(x + j, y + i) \text{ bitwiseXOR } R(x + d + j, y + i) \quad (4.4)$$

SHD was used as a pixel matching cost in (Zabih et al [148]) with promising results. Due to the fact that the SHD uses bit level operations, and that the Census Transformation itself is based on bit level operations, it is efficient to implement in hardware.

The **Census Transformation** is a filter for classifying image texture, and it's main uses are in optical flow calculation, image segmentation and face detection.

When performing the Census Transformation, local windows of odd sizes are used. The value of each pixel in the image is replaced by the value of a register

4.3 Measuring the Algorithm Performance

which is calculated by comparing each pixel in the window with the center pixel. During this comparison, if the intensity of the center pixel is larger than the one of the neighboring pixel, the bit value is set to 0. Otherwise it is set to 1, algorithm 4.1.

Algorithm 4.1 Census Transformation

```
for(i=0; i<=windowSize; i++) {  
  for(j=0; j<=windowSize; j++) {  
    (pixelVal[i,j] < centerPixel) ? censusVal[i+j] = 0 : censusVal[i+j] = 1;  
  }  
}  
censusImage[i,j] = censusVal;
```

The original test image, and the resulting Census Transformation using a window of 5x5 can be seen in figure 4.5.

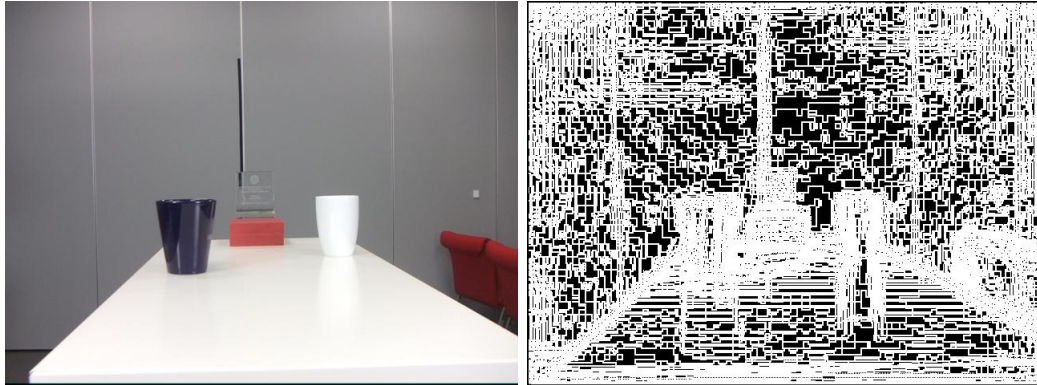


Figure 4.5: Census Transformation example

4.3 Measuring the Algorithm Performance

The development of the depth map generation algorithm was one of the later stages of the research presented in this thesis. The algorithm was developed for use in a foreground-background separation and gesture detection project. In the case of gesture detection and foreground-background separation it is important to have well defined boundaries and objects within the depth map. Due to this reason, a depth map quality measurement method which measures the depth quality of the objects, rather than the one of the entire image was developed. Since the visual representation of the depth map is similar to the one of the disparity map, as it was explained in

section 2.1.4, the same quality measurement technique applies to disparity maps as well.

In order to measure the quality of the depth maps, a ground truth is required. An example of the ground truth can be seen in figure 4.6. The generation of the ground truth is described in section 4.4 of the current chapter.

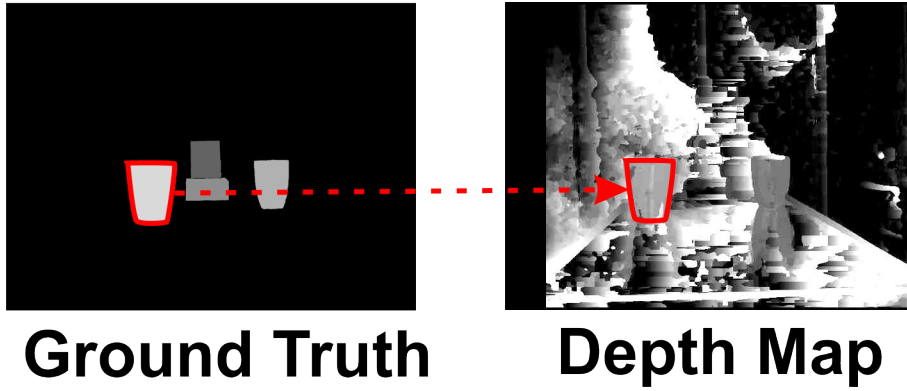


Figure 4.6: Depth Map quality measurement

In order to measure the quality of a depth map, the following steps are required:

1. Based on the information from the ground truth image, identify the number of depth layers that are found in the image. In our case, each object represents one depth layer.
2. Select regions of interest (ROI) from the depth map, based on the shape of the depth layer identified at the previous step, as it can be seen in figure 4.6.
3. Within the ROI selected from the depth map, form groups of pixels which have the same depth value. The group with most members defines the shape and depth of the current object, figure 4.7. The members of this group are considered to be true positives. All the members outside this group, but still belonging to the defined ROI are considered to be false negatives. Within a defined ROI, by using the current approach, false positives cannot be determined. False positives can only be determined if the whole scene is analyzed.
4. The ratio of majority depth pixels to the total number of pixels found within the ROI gives the depth map quality for that specific object, as in can be seen in equation 4.5.

4.3 Measuring the Algorithm Performance

$$ObjAcc = (DepthMapPixObject / TotPixObject) \cdot 100(\%) \quad (4.5)$$

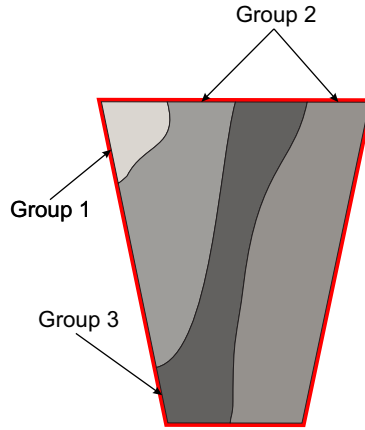


Figure 4.7: Pixel selection for quality measurement

Based on the literature reviews, no other authors used the measurement approach proposed in this section, mostly because of the fact that they wanted to use standardized approaches, such as that presented in (Scharstein et al [117]).



(a) Real-world image



(b) Artificial image

Figure 4.8: Middlebury image samples

According to the authors of (Scharstein et al [117]), the two approaches for stereo algorithm performance evaluation are (i) the computation of the error statistics based on some ground truth data, and (ii) the evaluation of synthetic images which are obtained by warping the reference images by the computed disparity map. The stereo image database proposed by (Scharstein et al [117]) contains two types of images (i)

images of real-world scenarios, and (ii) artificial images, figure 4.8. This is the reason why they propose two different types of evaluation.

Evaluation (i) is similar to the one used in the current case. The most important difference between (Scharstein et al [117]) and the implementation proposed in the thesis is that in the case of *Scharstein*, the error measurements are performed on the entire image, whereas in our case, they are performed on individual objects. The images used for our tests are real-world scenes, and for this reason the approach proposed in (ii) does not apply in our case.

4.4 Ground Truth Generation

The ground truth represents the real depth of the scene, where each pixel in the image is assigned with a value which represents the distance of that pixel from the camera in the real world. The ground truth is used for comparison with the generated depth map, and based on this comparison, the performance of the depth map generation algorithm can be measured.

4.4.1 Methods used for Ground Truth Creation

One of the most often used pairs of stereo images is the **Tsukuba** [93] image pair. The image, together with its ground truth can be seen in figure 4.9 and it was first introduced in the paper written by (Nakamura et al [93]). The reason why it is often used is that the ground truth provided for the image pair was created manually, with each pixel having a depth value individually assigned, making it very accurate. There are other stereo image pairs that are also used quite often, but in those cases the images were created artificially and because of that, the creation of ground truth was a straight-forward operation.

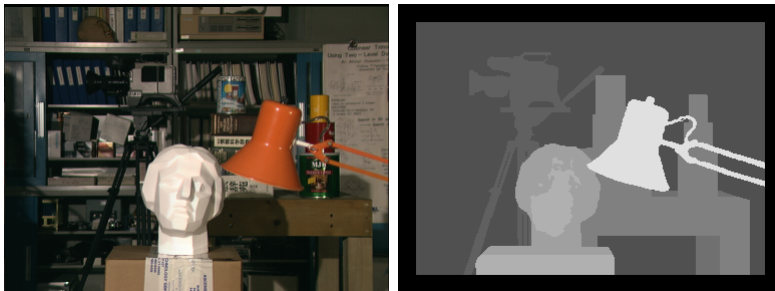


Figure 4.9: Tsukuba image

4.4 Ground Truth Generation

One of the most common approaches for the generation of ground truth is presented in (Scharstein et al [116]), where a method for acquiring high-complexity stereo image pairs with pixel accurate correspondence using structured light is detailed. In their approach, the authors use a stereo camera pair, and one or more light projectors, which can project structured light patterns onto the test scene. Based on a sequence of structured light patterns, each camera determines a unique label for each pixel. Based on these unique labels, the exact correspondence of each pixel can be found in the corresponding images, and a ground truth can be generated. The same method is used in (Zhu et al [152]) as well.

Beside the common method of structured light, other methods have been used for the generation of ground truth as well, such as the use of a station to measure the coordinates of the objects [62], using a robotic arm in order to determine the true depth [45], or using a 3D scanning device in order to scan the scene [70]. In other cases, the researchers didn't generate a ground truth, they only used subjective evaluation instead [150, 79].

4.4.2 Technique applied during the experiments

For the generation of ground truth images used during the experiments presented in the current thesis, an image processing application was used. The steps required for the generation of the ground truth are the following:

1. Import the original image in the form of a layer.
2. Add an extra layer of the same size, and make it invisible.
3. Paint the new layer black.
4. Select the contour of each object independently.
5. Fill the newly selected object with a color corresponding to the real depth (brighter if close to the camera, darker if further away from the camera). The real depth is decided by a visual inspection of the test scene.
6. Delete the layer containing the original image.
7. Export the resulting ground truth image.

An example on the ground truth image can be seen in figure 4.10.

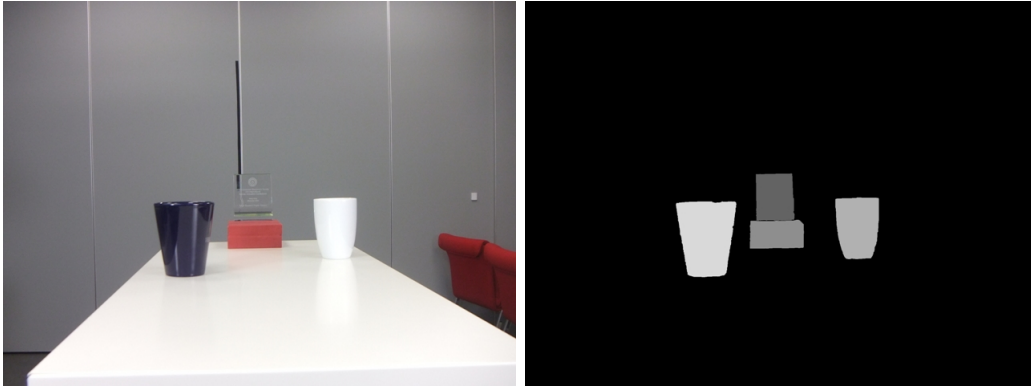


Figure 4.10: Example of a Ground Truth image

4.4.3 Discussion of the proposed Technique

The main difference between the ground truth generated for the initial testing and the one presented in (Scharstein et al [116]) is that in our case a similar depth pixel value is assigned to the entire object. This is OK for objects with flat surfaces, but in the case of objects with round surfaces, it might introduce some errors. The ground truth generation technique proposed in [116] generates more accurate ground truth images, but they are not as accurate as the ones generated for artificial images or the one described in [93].

The main reason why the method presented in the current section was chosen instead of other methods was that in order to generate high quality and reliable ground truths, state of the art and expensive equipment is required. Since the disparity map generation algorithms tested during the current experiments didn't generate high quality depth maps, it was considered that the ground truth images generated with the proposed approach will be satisfactory for the performance measurement.

Figure 4.11 represents an example of the number of depth layers that could be found in a conic shaped object. In figure 4.11a, the top view of the conic shaped object can be seen, where due to the round shape of the object, multiple depth layers can be noticed. By looking at the same object from one side, in figure 4.11b, beside the depth layers identified in the top view, another 2 depth layers can be seen due to the conic shape of the object. This example shows us a simple example of the number of depth layers that can be found in an object which is not flat.

4.5 Test Scene Setup

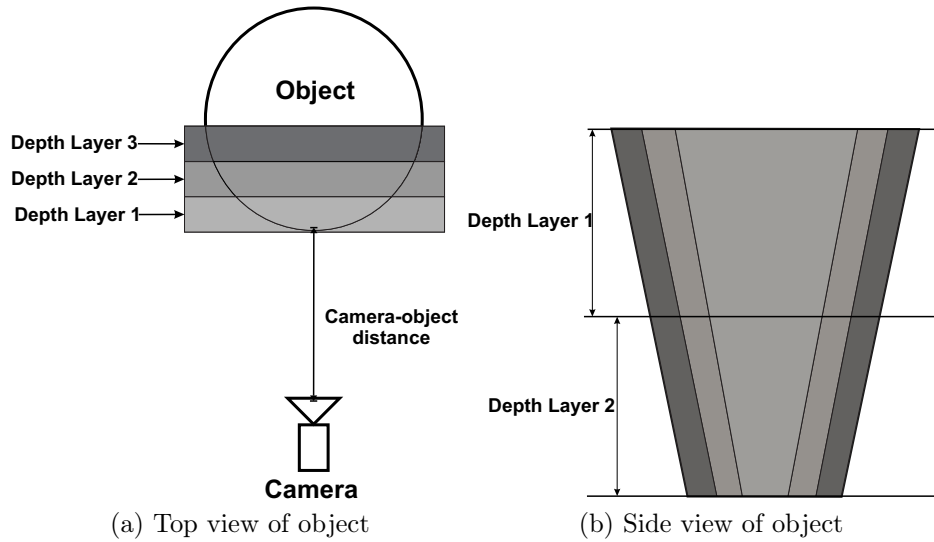


Figure 4.11: Multiple depth layers within object

The number of depth layers that are found in an object is influenced by the parallax, introduced in section 2.1.3.1, which is influenced by the length of the stereo base, and the camera-object distance. For larger camera-object distances, the value of the parallax decreases, section 5.3, this causing the decrease of the number of depth layers. The decrease of the stereo base length also causes the decrease of the depth layer number within an object.

Another factor which influences the number of depth layers is the camera resolution. For a large camera resolution, the number of depth layers increases, whereas in the case of a low resolution image, the number of depth layers decreases. By increasing the number of depth layers, the algorithms will have more layers to “choose” from, and this way, the performance of the algorithms will drop. It has been shown in [7], that in the case of low resolution images the depth map generation algorithms perform much better than in the case of higher resolution images.

A number of alternative performance measurement methods to the ones using ground truth images are described in section 7.6.

4.5 Test Scene Setup

The purpose of the test scenes presented in this section was to analyze the behavior of the disparity map generation methods over a number of camera-object distances and by using different light settings, the goal being to determine the limitations and the

ideal working conditions of these methods. During the literature review, similar tests were not found, which were carried out using disparity map generation algorithms. The test images available in online databases are acquired at fixed distances, so they could not be used for the current tests.

For our initial tests it was decided to use four principle camera-object distances. For later tests some intermediate distances were selected to confirm the form of the relevant performance graphs.

We selected 4 equally spaced setting levels for our tests, and the light intensity was measured individually on each of the objects, in the case of each test scene. This offers a sufficient variation in lighting level to demonstrate variations in algorithm performance due to different lighting levels.

For the measurement of the light intensity the **Sekonic L-758D** light meter was used. Based on the setting of the ISO to 100, it returns a value in **EV** that can be converted into **Lx** using formula 4.6.

$$Lx = 2.5 * 2^{EV} \quad (4.6)$$

Initially three different test scenes were created. The test scenes were named AWARD, FACE and FACES. The reason why the first test scene was named AWARD is, that the test scene contains a transparent item, which is an award made out of crystal. This scene is presented in section 4.5.1. The second test scene, presented in section 4.5.2, was named FACE because it contains a single human-face shaped object. The third test scene was name FACES, because it contains a human-face shaped object and a picture of a human face, and it is presented in section 4.5.3.

In total,16 test cases were created for each test scene: 4 different light conditions were created for each of the 4 test distances.

The camera settings which include the values of the **ISO** and the **f** number, as well as the **light intensity** measured on each object, have been recorded into a spreadsheet in the case of each test scene. The spreadsheets are presented in section 4.5. The test results for the 3 initial test scenes are presented in chapter 6.

A fourth test scene was created after the initial experimental results presented in section 6.2.2. This test scene was named TEXTURED OBJECTS, see section 4.5.4. During the experiments described in section 6.2.2, it was noticed that the highest depth map quality was achieved for objects at a certain distance from the camera. The purpose of this test scene was to re-create the initial findings by using a test

4.5 Test Scene Setup

scene with objects of different colors. More information is provided in section 6.2.2.2.

The vertical distance settings used in the case of all test scenes are illustrated in figure 4.12. The horizontal distance settings are presented in the case of each test scene individually, scenes defined in sections 4.5.1, 4.5.2, 4.5.3 and 4.5.4.

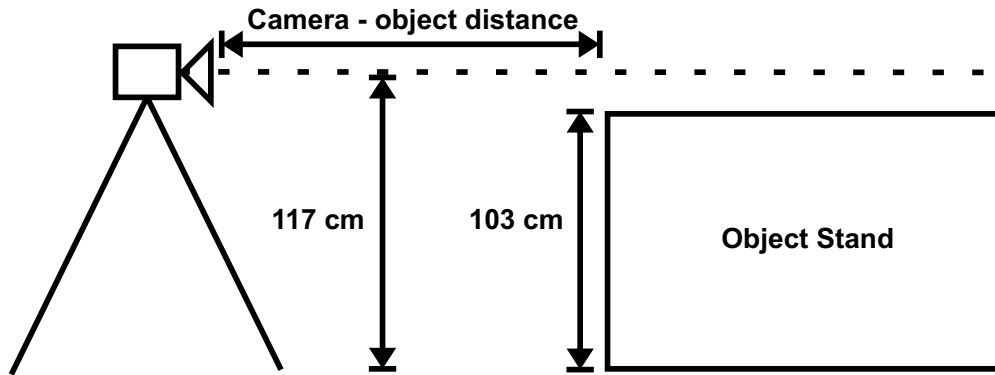


Figure 4.12: Vertical distance settings

4.5.1 AWARD test scene setup

The scene presented in figure 4.13 contains four different objects. The objects used in this scene were chosen randomly, with the purpose to create a difficult scenario for the correlation-based depth map generation algorithms that were going to be tested. This scene included textureless objects combined with nearly transparent objects. The scene was named AWARD due to the fact that the object at the back of the scene is an award made out of crystal. The original image of the scene can be seen in figure 4.14.

The original distance between the camera and the stand where the objects were placed was 500 mm. After this, the distance was increased by 500 mm for each set of tests. The respective distances were 1000 mm, 1500 mm and 2000 mm. Since this was the first stage of the testing process, it was decided to use four different camera-object distances for the start. The distances were chosen based on the size of the testing room. For camera-object distances larger than 2 m, additional objects from the room can be seen in the test images.

Figure 4.15 shows the four different light intensities for the AWARD scene. The AWARD scene settings used during our experiments can be seen in table 4.1.

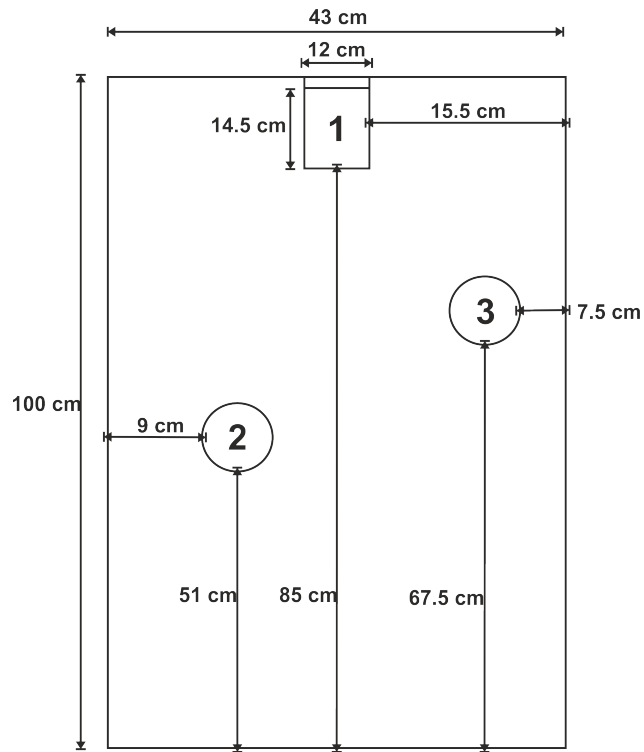


Figure 4.13: AWARD scene setup



Figure 4.14: AWARD original scene, 4 distances

4.5 Test Scene Setup

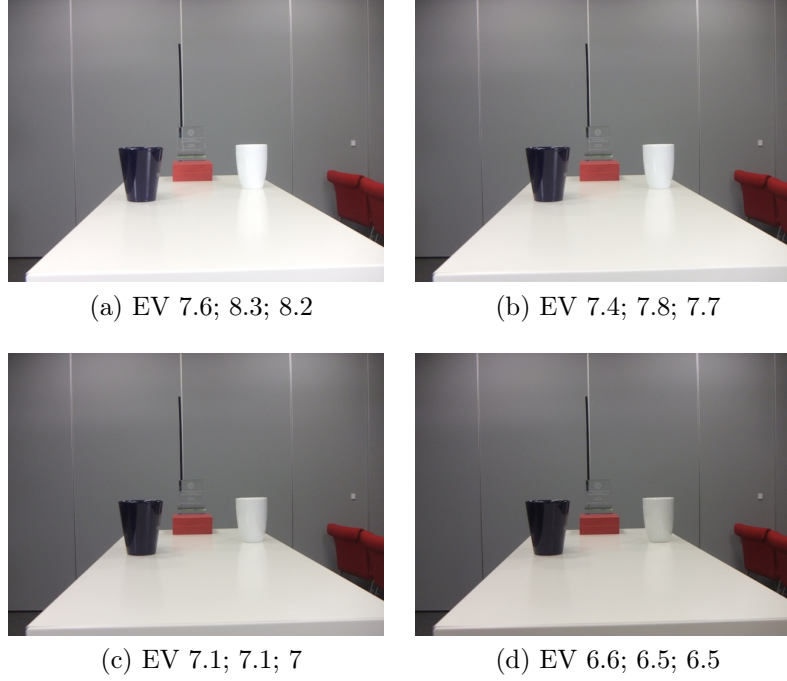


Figure 4.15: Light intensities AWARD

Table 4.1: Award Scene Settings

Image	Resolution	Dist. (cm)	Light (EV)			Shutter	f	ISO
			Obj. 1	Obj. 2	Obj. 3			
1	2048 x 1536	50	7.6	8.3	8.2	30	3.7	200
2	2048 x 1536	50	7.4	7.8	7.7	30	3.7	200
3	2048 x 1536	50	7.1	7.1	7	30	3.7	200
4	2048 x 1536	50	6.6	6.5	6.5	30	3.7	200
5	2048 x 1536	100	7.6	8.3	8.2	30	3.7	200
6	2048 x 1536	100	7.4	7.8	7.7	30	3.7	200
7	2048 x 1536	100	7.1	7.1	7	30	3.7	200
8	2048 x 1536	100	6.6	6.5	6.5	30	3.7	200
9	2048 x 1536	150	7.6	8.3	8.2	30	3.7	200
10	2048 x 1536	150	7.4	7.8	7.7	30	3.7	200
11	2048 x 1536	150	7.1	7.1	7	30	3.7	200
12	2048 x 1536	150	6.6	6.5	6.5	30	3.7	200
13	2048 x 1536	200	7.6	8.3	8.2	30	3.7	200
14	2048 x 1536	200	7.4	7.8	7.7	30	3.7	200
15	2048 x 1536	200	7.1	7.1	7	30	3.7	200
16	2048 x 1536	200	6.6	6.5	6.5	30	3.7	200

4.5.2 FACE Test Scene Setup

The purpose of the second scene was to test the behavior of the depth map generation algorithms in the case of human faces, because (i) faces are complex textured objects and (ii) the consumer camera market at the moment evolves around features developed for face enhancement. Faces are in the center point of a large number of application in the field of consumer cameras. Some applications where faces are used are (i) face detection and tracking [15, 126, 127, 125], which can be used as part of the auto-focus algorithm, or for the enhancement of the images; (ii) face beautification [32], which is used for the enhancement of the images; (iii) smile and blink detection [96], which can be used in order to determine when to acquire certain images; (iv) eye tracking [94] can be used for gesture detection, or in the case of auto-stereoscopic devices; (v) face recognition [36, 57] is useful for security applications, or for different consumer camera features.

In the case of this scene, a custom engineered mannequin-head was used, which has the shape and color of a human head. The test setup can be seen in figure 4.16, and the original image can be seen in figure 4.17.

The original distance between the camera and the stand where the objects were placed was of 700 mm. The reason why the distance of 700 mm was chosen instead of 500 mm as in the case of the AWARD test scene was that this distance allowed us to have the entire face visible in the acquired image. After this, the distance was increased by 500 mm for each case. The respective distances were 1200 mm, 1700 mm and 2200 mm.

In figure 4.18, the four different light intensities for the FACE scene are illustrated. The FACE scene settings used during our experiments can be seen in table 4.2.

4.5 Test Scene Setup

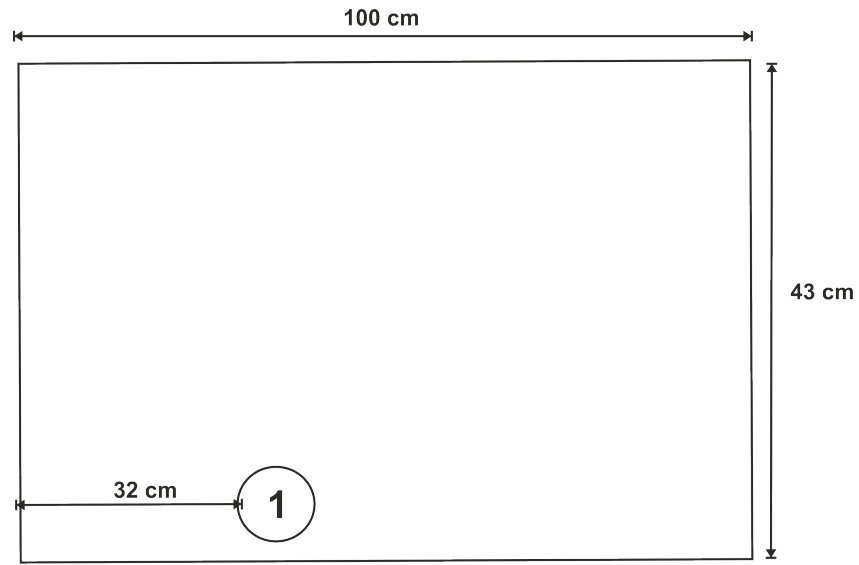


Figure 4.16: FACE scene setup



Figure 4.17: FACE original scene, 4 distances

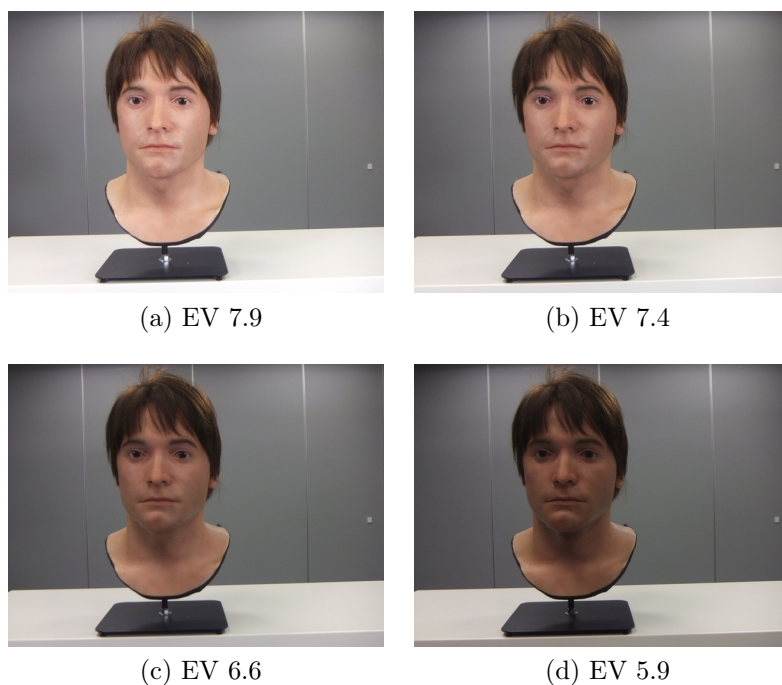


Figure 4.18: Light intensities FACE

Table 4.2: FACE Scene Settings

Image	Resolution	Dist. (cm)	Light (EV)	Shutter	f	ISO
			Obj. 1			
1	2048 x 1536	70	7.9	30	3.7	200
2	2048 x 1536	70	7.4	30	3.7	200
3	2048 x 1536	70	6.6	30	3.7	200
4	2048 x 1536	70	5.9	30	3.7	200
5	2048 x 1536	120	7.9	30	3.7	200
6	2048 x 1536	120	7.4	30	3.7	200
7	2048 x 1536	120	6.6	30	3.7	200
8	2048 x 1536	120	5.9	30	3.7	200
9	2048 x 1536	170	7.9	30	3.7	200
10	2048 x 1536	170	7.4	30	3.7	200
11	2048 x 1536	170	6.6	30	3.7	200
12	2048 x 1536	170	5.9	30	3.7	200
13	2048 x 1536	220	7.9	30	3.7	200
14	2048 x 1536	220	7.4	30	3.7	200
15	2048 x 1536	220	6.6	30	3.7	200
16	2048 x 1536	220	5.9	30	3.7	200

4.5 Test Scene Setup

4.5.3 FACES Test Scene Setup

The last scene had two purposes. The first one was to test the behavior of the algorithms on 2D and 3D objects combined in the same scene. This is why a human head replica was used together with a picture of a human head. The second purpose was to test how, having the same skin color in different places of the frame, would influence the behavior of the algorithm. For this purpose, an object which has the shape and color of a human face was used, together with a real-size picture of a human face. Even though the colours are not identical, they are similar. The scene setup can be seen in figure 4.19, and the original image can be seen in figure 4.20.

The original distance between the camera and the stand where the objects were placed was of 700 mm, which is the same distance as in the case of the original FACE test scene. After this, the distance was increased by 500 mm for each case. The respective distances were 1200 mm, 1700 mm and 2200 mm.

In figure 4.21, the light intensities for the FACES scene are illustrated. The FACES scene settings used during our experiments can be seen in table 4.3

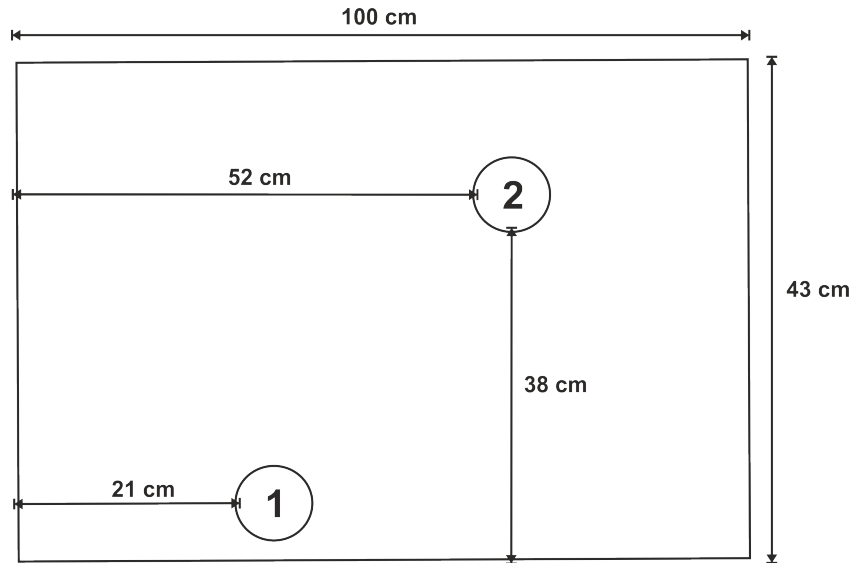


Figure 4.19: FACES scene setup



Figure 4.20: FACES original scene, 4 distances



(a) EV 7.9; 7.9



(b) EV 7.4; 7.3



(c) EV 6.7; 6.7



(d) EV 6.1; 6.4

Figure 4.21: Light intensities FACES

4.5 Test Scene Setup

Table 4.3: FACES Scene Settings

Image	Resolution	Dist. (cm)	Light (EV)		Shutter	f	ISO
			Obj. 1	Obj. 2			
1	2048 x 1536	70	7.9	7.9	30	3.7	200
2	2048 x 1536	70	7.4	7.3	30	3.7	200
3	2048 x 1536	70	6.7	6.7	30	3.7	200
4	2048 x 1536	70	6.1	6.4	30	3.7	200
5	2048 x 1536	120	7.9	7.9	30	3.7	200
6	2048 x 1536	120	7.4	7.3	30	3.7	200
7	2048 x 1536	120	6.7	6.7	30	3.7	200
8	2048 x 1536	120	6.1	6.4	30	3.7	200
9	2048 x 1536	170	7.9	7.9	30	3.7	200
10	2048 x 1536	170	7.4	7.3	30	3.7	200
11	2048 x 1536	170	6.7	6.7	30	3.7	200
12	2048 x 1536	170	6.1	6.4	30	3.7	200
13	2048 x 1536	220	7.9	7.9	30	3.7	200
14	2048 x 1536	220	7.4	7.3	30	3.7	200
15	2048 x 1536	220	6.7	6.7	30	3.7	200
16	2048 x 1536	220	6.1	6.4	30	3.7	200

4.5.4 TEXTURED OBJECTS Test Scene Setup

The TEXTURED OBJECTS test scene was created at a later stage of the research process, after the initial experiments presented in 6.2.2 were finished. The purpose of this additional test scene was to replace some of the textureless objects in the AWARD test scene with textured objects, of a similar shape, in order to check if the tested algorithms perform in a similar way. For this reason, Object 2 in the AWARD test scene was substituted with a textured object of a similar shape and size. A second textured object, of a different shape and size than any of the objects in the AWARD scene was also added to the new test scene. The scene setup can be seen in figure 4.22, and the original image can be seen in figure 4.23.

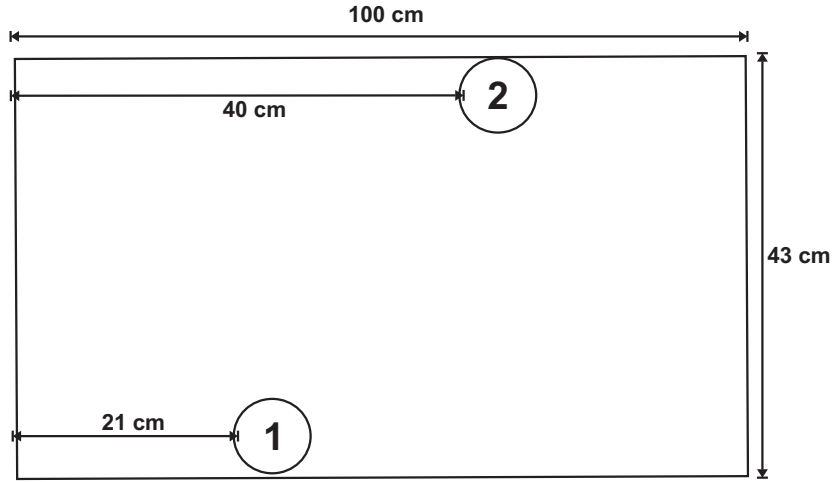


Figure 4.22: TEXTURED OBJECTS scene setup



Figure 4.23: TEXTURED OBJECTS original scene, 4 distances

The original distance between the camera and the stand where the objects were placed was 500 mm. After this, the distance was increased by 500 mm for each case. The respective distances were 1000 mm, 1500 mm and 2000 mm.

The light intensities for the TEXTURE OBJECTS test scene are illustrated in figure 4.24. The scene settings used during the experiments can be seen in table 4.4.

4.5 Test Scene Setup

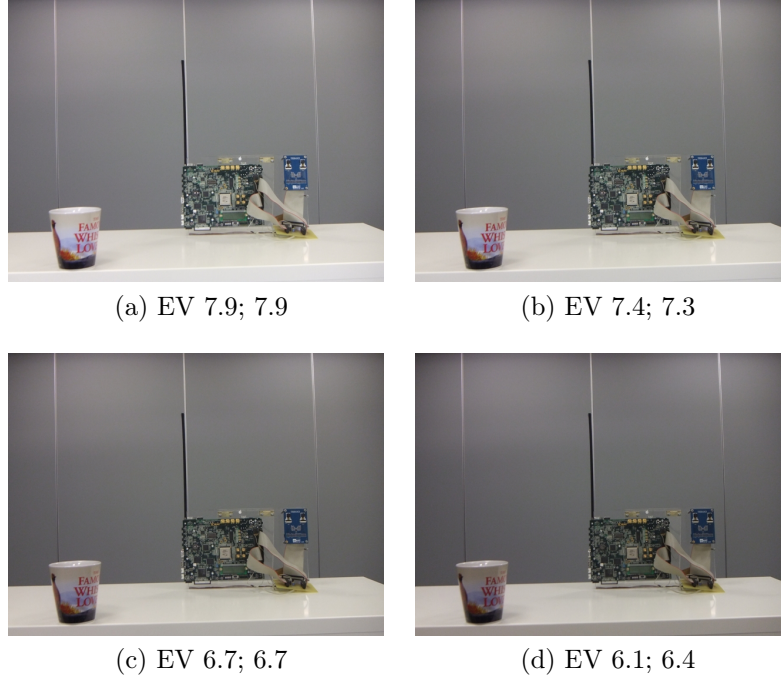


Figure 4.24: Light intensities TEXTURED OBJECTS

Table 4.4: TEXTURED OBJECTS Scene Settings

Image	Resolution	Dist. (cm)	Light (EV)		Shutter	f	ISO
			Obj. 1	Obj. 2			
1	2048 x 1536	50	7.9	7.9	30	3.7	200
2	2048 x 1536	50	7.4	7.3	30	3.7	200
3	2048 x 1536	50	6.7	6.7	30	3.7	200
4	2048 x 1536	50	6.1	6.4	30	3.7	200
5	2048 x 1536	100	7.9	7.9	30	3.7	200
6	2048 x 1536	100	7.4	7.3	30	3.7	200
7	2048 x 1536	100	6.7	6.7	30	3.7	200
8	2048 x 1536	100	6.1	6.4	30	3.7	200
9	2048 x 1536	150	7.9	7.9	30	3.7	200
10	2048 x 1536	150	7.4	7.3	30	3.7	200
11	2048 x 1536	150	6.7	6.7	30	3.7	200
12	2048 x 1536	150	6.1	6.4	30	3.7	200
13	2048 x 1536	200	7.9	7.9	30	3.7	200
14	2048 x 1536	200	7.4	7.3	30	3.7	200
15	2048 x 1536	200	6.7	6.7	30	3.7	200
16	2048 x 1536	200	6.1	6.4	30	3.7	200

4.6 Conclusions

An introduction to the proposed testing methodology was presented in section 4.1, which was followed by a description of correlation-based depth map generation methods in section 4.2 and the algorithms measurement technique in section 4.3. The chapter concludes with the presentation of a proposed ground truth generation technique in section 4.4 and the setup of the initial test scenes in section 4.5.

This chapter introduces the reader to the core algorithms, test scenarios, test scenes and measurement methods used throughout the experiments. Four different correlation-based depth map generation algorithms are presented, which are used in during experiments. An alternative performance measurement method is also described and the differences from other methods are discussed. Four different test scenes are introduced, with the help of which certain experiments are performed later in the thesis. These scenes and experiments provide priceless information for the main contribution of the work.

Chapter 5

Image Acquisition Methods and Preliminary Measurements

After the introduction of the proposed test methodology in chapter 4, where a number of correlation-based depth map generation algorithms and test scenes were presented, the current chapter talks about the technical elements of the test environment.

In section 5.1 the devices used for the acquisition of stereo images for testing purposes are described with details about the reasons why they were chosen and their specifications. From the stereo baseline point of view, explained in section 2.1.2, the test images acquired during the research can be split into two categories: (i) stereo cameras with a fixed stereo baseline and (ii) stereo cameras with a variable stereo baseline. In the case of the stereo cameras with fixed baseline, the Fujifilm W3 3D camera was used due to the fact that the same camera body contains two image sensors, as shown in section 5.1.1. For the second category of cameras, with variable stereo baseline, an alternative option had to be found, and for this reason, a custom image acquisition device was developed and tested, as described in section 5.1.2. In section 5.1.3 a third approach is described, namely the setup of a single camera for stereo image acquisition purposes.

Sections 5.2 and 5.3 present the downsampling method used in the current tests, as well as the parallax calculations that were performed in order to define the limitations of the test environment setup. Downsampling was necessary in order to reduce the size of the images used for the experiments. With the help of the parallax calculations, the distance limitations of the test scenes were determined.

The chapter concludes with section 5.4, where the influence of vertical misalign-

ment, the influence of noise and the lack of camera calibration, which might influence the quality measurement are analyzed, and relevant conclusions are drawn.

Vertical misalignment could appear in the case, when the cameras acquiring stereo images are not perfectly aligned. In this case, depending on the type of depth map generation algorithm, the quality of the depth map might drop. For this reason, the influence of vertical misalignment is discussed in section 5.4.1.

In the cameras used for stereo image acquisition, different noise levels can appear. The main reason is, that two devices cannot be mechanically built in exactly the same way. Another reason for the appearance of the noise is the quality and sensitivity of the image sensors. The influence of the noise of the quality of the depth maps is explained in section 5.4.3.

When acquiring stereo images, the cameras should be calibrated both intrinsically and extrinsically. If they are not calibrated, the measurements based on disparity maps might be erroneous. For this reason, a comparison of depth and disparity maps generated with un-calibrated and calibrated cameras is presented in section 5.4.2, where the scenarios where un-calibrated cameras can be used are discussed.

5.1 Devices used for Test Image Acquisition

In this section, attempts of using different devices for the purpose of stereo image acquisition are presented. In the case of each device, the specifications of the device are presented, together with the advantages and disadvantages of each device.

5.1.1 Using the Fujifilm W3 3D Camera for Test Image Acquisition

As a stereo image acquisition device initially a **Fujifilm W3 3D** camera was used. This camera was chosen because it incorporates two standard cameras, and it offers the capability to acquire stereo images. The distance between the two cameras is fixed at 75 mm. More details about the camera can be found in Appendix D.1.

The camera acquires stereo images in Multi Picture Object (“mpo”) format, which is a standard of the Camera & Imaging Products Association. This format provides a method to store multiple JPEG images, and in the case of this camera, it is used for the purpose of storing stereo images. To separate the views, an application called “StereoPhoto Maker” was used. The camera was placed on a standard tripod, making

5.1 Devices used for Test Image Acquisition

sure that it was perfectly aligned both horizontally and vertically. The setup of the image acquisition device can be seen in figure 5.1.



Figure 5.1: Setup of the acquisition device

During the test image acquisition process, the camera settings were set manually. Since the camera used was an off-the-shelf camera, the lowest f number that we could choose was 3.7. The f number represents the size of the aperture of the lens. A smaller f number represents a larger aperture.

In order to keep the noise that can be induced by the sensitivity of the sensor at a minimum, it was decided to use an **ISO** value of 200. The ISO represents the sensitivity of the sensor of the film in the case of non-digital cameras. The higher the ISO value, the smaller the exposure time needs to be.

With the f number set to 3.7, and ISO set to 200, a light intensity meter was used in order to decide the value of **exposure time** to be used. The light meter returned the value of 30, which means an exposure time of 1/30 seconds.

Right from the start of the research, it was decided to work with images at VGA resolution. The main reasons were that (i) it takes less time to process information from a smaller image than from a larger image, (ii) major image features such as object boundaries and object structure are adequately preserved at VGA resolution; (iii) many algorithms are better down-sampled from HD or higher resolutions before applying image algorithms, but the down-sampling process can itself introduce artifacts; thus acquiring images at the ‘right’ rather than the highest available resolution is preferable.

For this reason, it was decided to acquire images at the lowest resolution allowed by the camera, 2048x1536, and use a down sampling method, detailed in section 5.2, to reduce its size to VGA resolution. The VGA resolution is 640x480 for an aspect ration of 4:3.

In order to overcome the focusing problems described in section 2.4.4, the camera focus was set to infinity.

5.1.2 Using the uCam for Test Image Acquisition

After the initial experiments described in 6.2, it was considered necessary to acquire a new set of images, where the length of the stereo base can be set manually. During these experiments, it was noticed that at a certain camera-object distance, the depth map generation algorithm performs quite well. We wanted to check if the same results can be achieved for images acquired with a different stereo base value.

The camera which was used for the acquisition of the original test scene, and was presented in 5.1.1, had a fixed stereo base length of 75 mm. Due to this reason, it was considered necessary to develop a new stereo image acquisition rig.

In the custom-built stereo image acquisition system, the functionality of the camera was substituted by a custom-built FPGA development board. The stereo ISP used on this board was the one presented in section 2.3, and published in (Andorko et al [6]). The reason why this custom-built ISP was initially chosen was, that it allowed manual control of all the stages of the ISP, providing a theoretically perfect synchronization between the pipelines. Practically, a perfect synchronization is not possible due to the manufacturing process of the integrated circuits, which do not guarantee perfect similarity between components.

5.1.2.1 Development board equipped with WLC cameras

As a choice of sensors, two different approaches were tried. In the first approach, a custom-built camera board was used, which allowed the adjustment of the stereo base length and was equipped with Wafer-Level Cameras (WLC). This camera board can be seen in figure 5.2. A description of the WLC cameras can be found in Appendix D.2.

5.1 Devices used for Test Image Acquisition

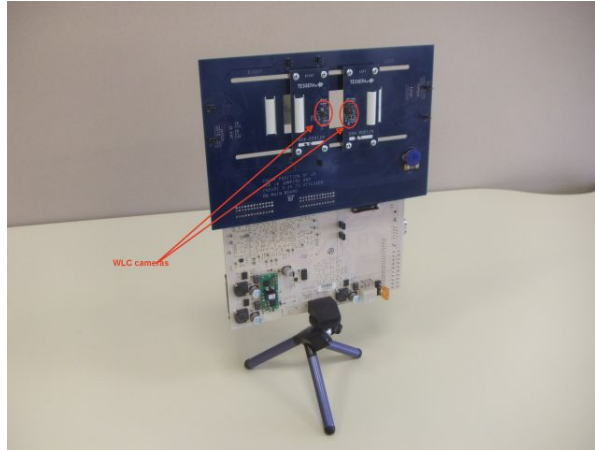


Figure 5.2: uCam using the WLC cameras

The problem with this camera board was that only one of the cameras had a built-in IR filter. It was attempted to fit an IR filter over the lens of the other camera, but this changed the optical path of the light, and the specific image became blurred. Another problem was that the filter which was applied over the lens didn't have the same properties as the one built into the WLC. For these reasons, the images acquired with this camera board were not used for tests. An example of the stereo image pair acquired with the WLC cameras can be seen in figure 5.3.

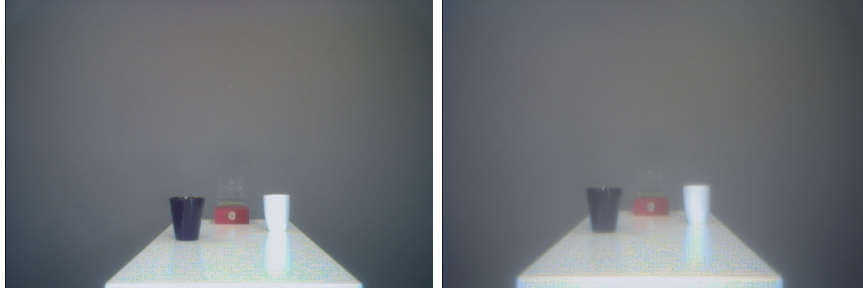


Figure 5.3: Stereo image pair acquired with WLC cameras

5.1.2.2 Development Board equipped with Aptina Demo Cameras

In the second approach, two camera demo heads, developed by Aptina, were used. The sensor was MT9P031, and Navitar 1.4 f lenses were used. A description of the cameras can be found in Appendix D.3.

These cameras were fitted onto a horizontal rig, which allowed their movement, and consequently, the adjustment of the stereo base length. The image acquisition

system setup can be seen in figure 5.4.

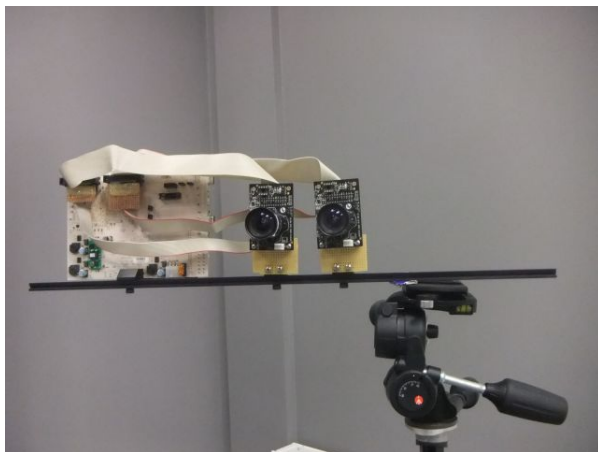


Figure 5.4: uCam using the Aptina cameras

Problems were encountered with this setup as well due to dirty lenses and image quality. Attempts were made to clean the lenses, but it was discovered that the dirt was on one of the inside elements of the lens, and for this reason, a working solution was not found. During the attempt to clean the lens it was noticed that the dirt was only visible when the aperture of the shutter was reduced. The aperture had to be reduced because (i) by not having auto-exposure the images would have been over-exposed and (ii) in order to focus to infinity, the aperture of the shutter needs to be reduced. The acquired images can be seen in figure 5.5.

A problem regarding image quality was also noticed. In certain areas of the image, an overflow was noticed. It was discovered, that this problem was due to contamination in the sensors of the camera. Even though the lenses might have been changed, changing the camera bodies was not an option at this stage of the research.

For to these reasons, the use of this custom built stereo image acquisition setup was abandoned.

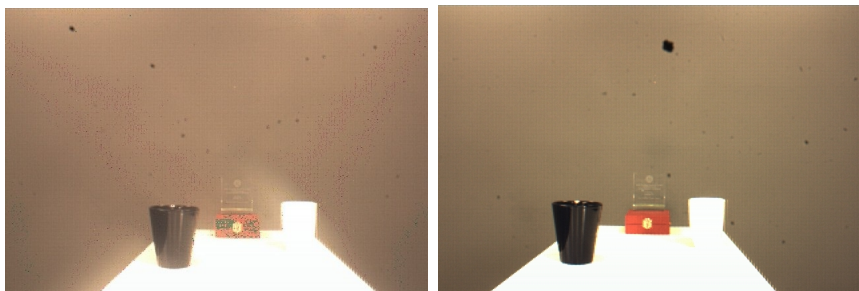


Figure 5.5: Stereo image pair acquired with Aptina cameras

5.1 Devices used for Test Image Acquisition

5.1.3 Using a Single Camera for Test Image Acquisition

For the purpose of stereo image acquisition with a variable baseline, section 5.1.2 described the development of a stereo image acquisition system. Due to various reasons, this system could not be used.

As an alternative, a simple solution was proposed, which allowed us to acquire stereo images with variable stereo baseline. The solution was to use a single camera mounted onto a rig, and acquire the images consecutively while moving the camera horizontally for each new acquisition, as shown in figure 5.6.



Figure 5.6: Using a single camera for stereo image acquisition

The choice of camera was the Fujifilm W3 3D described in section 5.1.1. In the case of the Fujifilm camera, only one of the image sensors was used. The camera settings were the same as the ones described in section 5.1.1.

The disadvantages of using a single camera for stereo image acquisition were that (i) the images were not acquired at the same time, (ii) the required horizontal motion of camera along the rig increased the chances of acquiring misaligned images.

In order to overcome the first disadvantage of using a single camera, the scene layout and light settings were kept the same for both image acquisitions (left and right). The images were acquired in a controlled environment, where only artificial light was used. Prior to each image acquisition, the intensity of the light was measured using a light meter.

The problem of the camera misalignment during stereo image acquisition will be discussed in section 5.4.2.

5.2 Choice of the Downsampling Method

As mentioned in section 5.1 of this chapter, in order to capture the images the Fujifilm W3 3D camera was used. The lowest resolution setting available on the camera was 2048 x 1536. In order to decrease the computation time while testing the depth map generation approaches, it was decided to downsample the images to a VGA resolution of 640 x 480, for a 4:3 aspect ratio.

The choice of downsampling method can have a significant influence on the image quality according to (Mrak et al [89]), and for that reason, it was decided to run several tests in order to find the appropriate downsampling method. At the same time, the company that part-financed this research was interested in the outcome of these tests.

It was decided to test 5 different downsampling methods: Lanczos [60], Bicubic [97], Bilinear [49], Nearest Neighbor [80] and Averaging [83].

In order to test the proposed DS algorithms, two different test were carried out:

- Measurement of the Modulation Transfer Function (MTF), section 5.2.1.
- Visual inspection using a zone plate type of image, section 5.2.2.

5.2.1 Measurement of the MTF

Modulation Transfer Function (MTF) measurement was used on the resulting image, in order to find the method which performs the best. The MTF is the magnitude response of the optical system to sinusoids of different spatial frequencies [21]. The measurement algorithm used is an edge-gradient algorithm, which follows the intent of the standard ISO 12233 [43].

In order to test these algorithms, a 12 Mega Pixel image of a slanted edge was artificially created, with a 16:9 aspect ratio, and a resolution of 4616 x 2600, figure 5.7.

5.2 Choice of the Downsampling Method



Figure 5.7: Slanted Edge image used for testing

Four downsampling scales were chosen based on the current industry requirements:

- Scale of 2.4, for a 1920 x 1080 (Full HD) resolution.
- Scale of 3.6, for a 1280 x 720 (HD) resolution.
- Scale of 6.2, for a 744 x 416 (VGA) resolution.
- Scale of 12.5, for a 368 x 208 (QVGA) resolution.

The MTF measurement application that was used for the tests [26] returns the values of the Spatial Frequency Response(SFR) for frequencies ranging between 0 and the Nyquist Rate, as it can be seen in figure 5.8. The MTF measurement is performed based on the quality of the slanted edge in the test image after the downsampling operation has been performed. The Nyquist Rate is the minimum sampling rate required in a system to avoid aliasing, and its value is twice the value of the highest frequency within the signal.

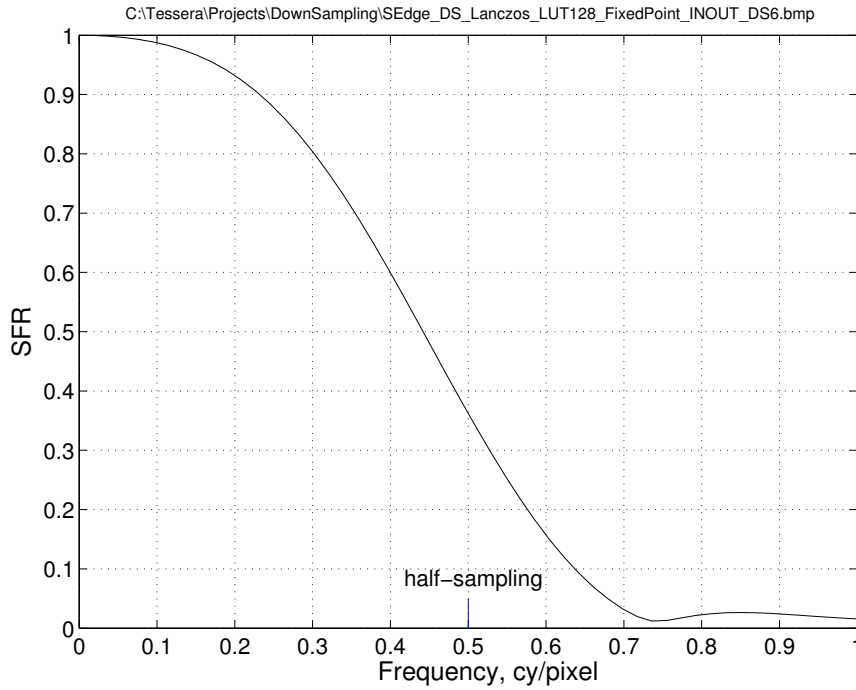


Figure 5.8: MTF measurement example

The areas of interest in the resulting plot are (i) the SFR values for frequencies between 0 and the Nyquist Frequency, which is $Nyquist\ Rate / 2$, where the SFR should have values as high as possible and (ii) the SFR values for frequencies between the Nyquist Frequency and the Nyquist Rate, where SFR should have values as low as possible. In the case of (i), high SFR values represent sharp, un-blurred images, whereas in the case of (ii), low SFR values represent the lack of aliasing.

For the reasons described in the previous paragraph, the value of the frequency was recorded at two of the SFR values of interest. These values of interest were 0.9, in order to check the loss of details in the final image and 0.1 in order to check the amount of aliasing in the final image. In the case of the values at 0.9, the larger frequency meant better performance and in the case of the values at 0.1, the smaller frequency meant better performance.

The value of the SFR was recorded at one frequency of interest. The value of interest in the case of the frequency was the Nyquist Frequency, which is half the sampling rate, in order to check the performance of the anti-aliasing filter.

The values were recorded in tables corresponding to each down sampling scale, and they can be seen in tables 5.1, 5.2, 5.3 and 5.4. The MTF graphs can be found

5.2 Choice of the Downsampling Method

in appendix C.1.

Table 5.1: MTF measurement for a down scale factor of 2.4

Down Scale factor 2.4			
Method	Nyquist Rate/2	SFR < 0.9	SFR < 0.1
	[SFR]	[Nyquist Rate]	[Nyquist Rate]
Lanczos	0.21	0.21	0.57
Bicubic	0.35	0.23	0.61
Bilinear	0.21	0.15	0.6
Nearest Neighbor	0.81	0.36	1
Averaging	0.75	0.35	1

Table 5.2: MTF measurement for a down scale factor of 3.6

Down Scale factor 3.6			
Method	Nyquist Rate/2	SFR < 0.9	SFR < 0.1
	[SFR]	[Nyquist Rate]	[Nyquist Rate]
Lanczos	0.31	0.23	0.6
Bicubic	0.31	0.23	0.61
Bilinear	0.25	0.15	0.59
Nearest Neighbor	0.91	0.51	1
Averaging	0.85	0.45	1

Table 5.3: MTF measurement for a down scale factor of 6.2

Down Scale factor 6.2			
Method	Nyquist Rate/2	SFR < 0.9	SFR < 0.1
	[SFR]	[Nyquist Rate]	[Nyquist Rate]
Lanczos	0.35	0.25	0.6
Bicubic	0.31	0.26	0.61
Bilinear	0.29	0.15	0.61
Nearest Neighbor	0.89	0.39	1
Averaging	0.83	0.45	1

Table 5.4: MTF measurement for a down scale factor of 12.5

Down Scale factor 12.5			
Method	Nyquist Rate/2	SFR < 0.9	SFR < 0.1
	[SFR]	[Nyquist Rate]	[Nyquist Rate]
Lanczos	0.42	0.25	0.62
Bicubic	0.31	0.27	0.61
Bilinear	0.22	0.18	0.58
Nearest Neighbor	0.9	0.41	1
Averaging	0.9	0.41	1

Based on the values presented in the previous tables, the Bicubic downsampling method provides marginally better results than the Lanczos method. When compared to the other methods, the final image retains more details, and the amount of aliasing is lower. For this reason, the decision on the optimal downsampling was taken after the visual assessment of the algorithm performance, described in section 5.2.2.

5.2.2 Visual Assessment of the Algorithm Performance

After measuring the MTF in the previous section, a visual inspection was also considered necessary in order to decide which downsampling algorithm to use for the experiments. For this reason, a zone plate type of image was used, such as the one that can be seen in figure 5.9. A zone plate type of image is an image, which contains alternating white and black circles. The size of the circles in the middle of the image is clearly visible, but towards the edges, their size becomes smaller, and their frequency increases. Due to this high frequency, this type of images are ideal for the visual inspection of aliasing artifacts in images. The downsampled zone plate images can be found in appendix C.2, and a sample image can be seen in figure 5.9.

These images are automatically downsampled by the text editor and their quality is affected by the printer used to print the thesis. For this reason, their quality cannot be properly assessed based on the printed images. The assessment of the resulting images was performed on LCD computer monitors at full resolution. Based on the visual assessment of the resulting images, and the results presented in section 5.2.1, it was concluded that the Lanczos method provides the best results.

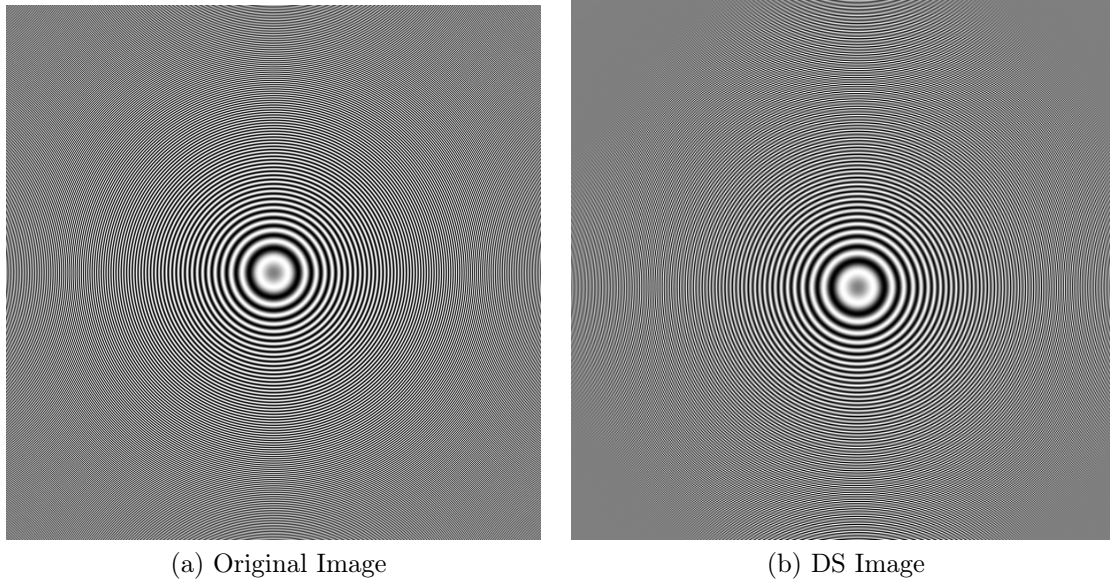


Figure 5.9: Zone plate type of test image

5.3 Measurement of the Disparity caused by Parallax

Before tests with stereo images were carried out, it was considered necessary to perform a measurement of the disparity caused by the parallax, as introduced in section 2.1.3.1, using the parameters of the camera used for image acquisition. The purpose of these calculations was to find the minimum and maximum distances where the objects should be placed according to the camera when using different stereo base values. The minimum distance where the object should be placed is the one, where common information is still available in the stereo image pair. The maximum distance is where we have a disparity of at least 2 pixels between objects in the stereo image pair.

For the disparity measurement, the following parameters are required:

- Stereo base value of the camera used for image acquisition.
- Camera-object distance used in the test scene.
- Size of the sensor used in the camera.
- Focal length used during the image acquisition.

The stereo base values and camera-object distances that were considered for the measurements can be found in tables 5.5 and 5.6. The reason why two different sets of measurements were carried out is, that the initial tests were carried out only for the distances which can be seen in table 5.5. During the experiments presented in section 6.2 it was decided that further tests needed to be done for the distances in table 5.6.

The calculations were made using the information from the Fujifilm W3 camera specifications, described in section 5.1.1.

According to the specifications of the Fujifilm W3 camera, the size of the sensor is $1/2.3$ inch, which in metric notation is 11 mm. The focal distance between lens and sensor is between 6.3 mm and 18.9 mm. All these information were used in order to compute the disparity values. This information was used in order to calculate the possible fields of view that the camera uses, with the help of the following formula:

$$\alpha = 2 \cdot \arctan \frac{h}{2f} \quad (5.1)$$

where h = sensor size and f = focal length. The resulting field of view was between 80° for a focal length of 6.3 mm, and 32° for a focal length of 18.9 mm. Since during the test image acquisition process detailed in section 4.5, the camera was set to a focal length of 6.3 mm, it has been decided to use the 80° horizontal field of view for our future calculations.

Figure 5.10 presents the horizontal field of view from a geometrical point of view.

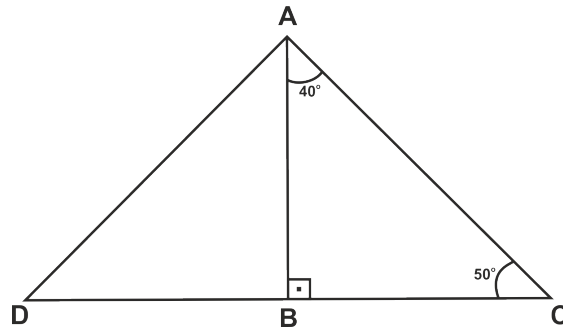


Figure 5.10: Horizontal field of view

In figure 5.10, \widehat{CAD} has a value of 80° . $AB \perp CD \implies \widehat{BAC} = 40^\circ, \widehat{ACB} = 50^\circ$. CD has a different value for each object distance. Its value is used to calculate the disparity caused by the parallax, and the pixel overlap in two different views. In order to calculate CD , the only information needed is the value of AB , which in our

5.3 Measurement of the Disparity caused by Parallax

case is the distance of the object from the camera. The formula used for calculation of CD is:

$$CD = 2 \cdot (\cot(50^\circ) \cdot AB) \quad (5.2)$$

It was decided to express the disparity value in *number of pixels*. The reason behind this decision was that in image processing we are measuring distances in pixel numbers rather than cm. The number of pixels can differ for different resolutions settings. Since all the test were performed at a VGA (640x480) resolution, it was decided to use the horizontal resolution value of 640 in the calculations. In order to calculate the displacement in pixel numbers, first we need to calculate the size of the pixel in cm. In order to do this, the formula $pixel = CD/640$ was used.

The last step in the process is the calculation of the displacement between the two views. The only additional information that is needed at this step is the stereo baseline value. The formula used for the displacement caused by the parallax computation is:

$$displacement = 640 - ((CD - 2 \cdot baseline)/pixel) \quad (5.3)$$

The formula was deducted from figure 5.11. The baseline is measured in centimeters

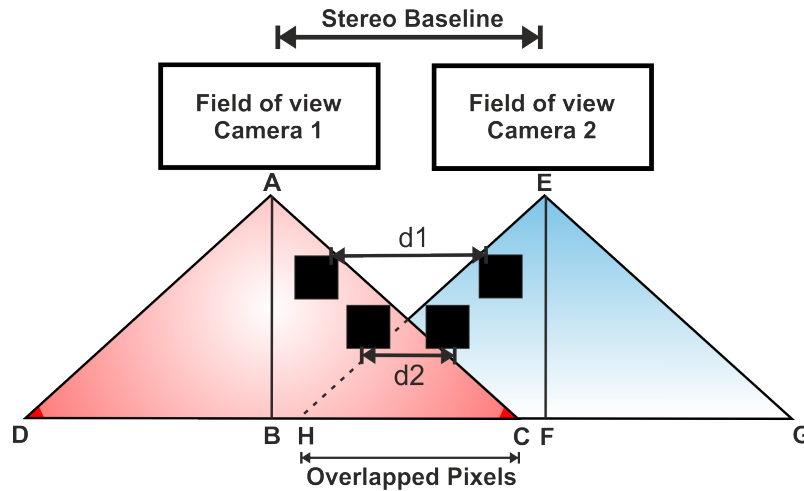


Figure 5.11: Explanation of displacement and overlapped pixels

In order to calculate the number of overlapped pixels, the value of the *displacement* needs to be subtracted from 640 in our case.

In figure 5.12, the plots of the disparity evolution are presented for each of the stereo base values that were considered in the tests. The start distance is 10 cm from the camera, there are 30 samples in total, each sample incrementing the object distance from the camera by 10 cm. The purpose of the plots is to have a general idea of the disparity values. Exact values will be provided in detailed tables later in this section.

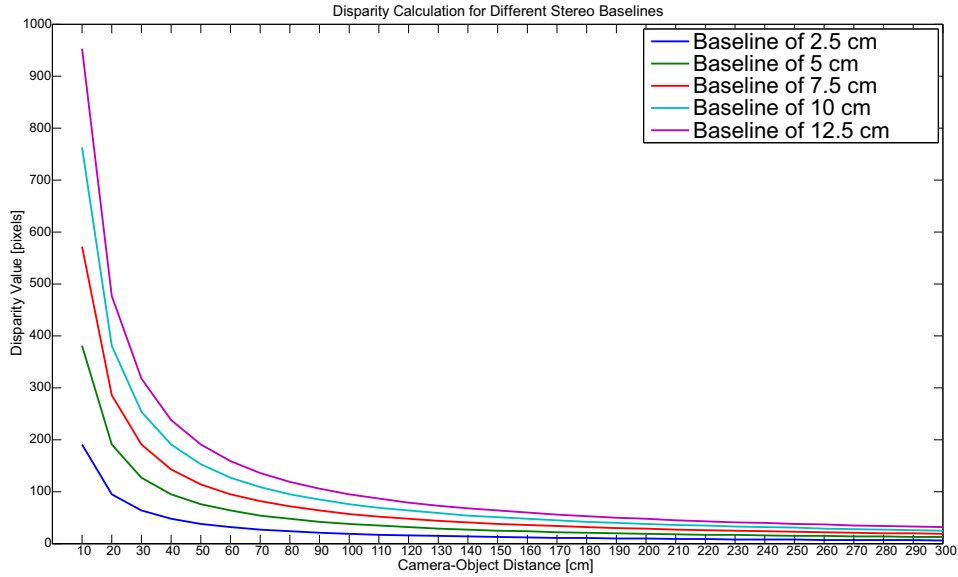


Figure 5.12: Parallax values for different stereo baselines

The exact displacement values have also been computed for the distances used in the experiments. The values can be seen in tables 5.5 and 5.6.

Table 5.5: Distance table with disparity values for first test

Distance [cm]	Stereo base values				
	2.5 cm	5 cm	7.5 cm	10 cm	12.5 cm
46	41	83	124	166	207
96	20	40	60	79	99
146	13	26	39	52	65
196	10	19	29	39	49

5.3 Measurement of the Disparity caused by Parallax

Table 5.6: Distance table with disparity values for steady scene test at 8 distances

Distance [cm]	Stereo base values				
	2.5 cm	5 cm	7.5 cm	10 cm	12.5 cm
46	41	83	124	166	207
66	29	58	87	116	144
86	22	44	67	89	111
116	16	33	49	66	82
136	14	28	42	56	70
156	12	24	37	49	61
176	11	21	33	43	54
196	10	19	29	39	49

For the same stereo base values, and distances of the objects from the camera, the number of overlapping pixels in each view have also been computed. These values can be seen in tables 5.7 and 5.8.

Table 5.7: Distance table with overlapped pixel values for initial test

Distance [cm]	Stereo base values				
	2.5 cm	5 cm	7.5 cm	10 cm	12.5 cm
46	599	557	516	474	433
96	620	600	580	561	541
146	627	614	601	588	575
196	630	631	611	601	590

Table 5.8: Distance table with overlapped pixel values for steady scene test at 8 distances

Distance [cm]	Stereo base values				
	2.5 cm	5 cm	7.5 cm	10 cm	12.5 cm
46	599	557	516	474	433
66	611	582	553	524	496
86	618	596	573	551	529
116	624	607	591	574	558
136	626	612	598	584	570
156	628	616	603	591	579
176	629	619	607	597	586
196	630	621	611	601	591

Based on the calculation technique above, in equation 5.3 it has been observed, that for the stereo base values of 10 and 12.5, the overlapping of the pixels start at

the distances of 12 cm, respectively 15 cm from the camera. At distances closer than this, there is no common information in the two separate views. This means that for large stereo base values, bringing the camera closer to the objects would result in inaccurate depth map results.

5.4 Image Acquisition Issues that might influence the Results

When using stereo cameras a number of problems can arise, such as vertical misalignment, color inconsistency in the acquired image pair, and other problems which were presented in section 2.4.

For the acquisition of stereo images, there are a number of available methods that can be used, which were described in section 2.1.1. Due to the fact that the cameras used for image acquisition are not perfectly aligned, there can be a vertical shift, or a difference in the rotation between the two images. For the alignment of the images in the Y plane, a simple algorithm is used, section 5.4.1. In order to measure the rotation angle of the images around the Z axis, the images need to be calibrated, as explained in section 5.4.2.

During the acquisition of consecutive sets of stereo images, even if the test scene and environment is not changed, it is not guaranteed that the acquired images will be exactly the same. This might be due to several reasons, such as (i) the quality of the light sources in the room, (ii) the temperature of the image sensor, (iii) the dust in the air, etc. The influence of all these elements, which were generically called “noise”, are analyzed in section 5.4.3.

5.4.1 The Influence of vertical Misalignment

During the experiments, the acquired test images are aligned in the Y plane, described in figure 5.13, before they are used for depth map computation. This alignment is performed by an algorithm described in [10], but it does not guarantee precise alignment. The proposed algorithm calculates horizontal profiles of the frames by summing the values of each pixel in corresponding rows. In order to align the frames, the algorithm calculates the error between the lines using Sum of Absolute Differences. A survey of alternative image registration algorithms can be found in a work

5.4 Image Acquisition Issues that might influence the Results

published by (Zitova et al [153]).

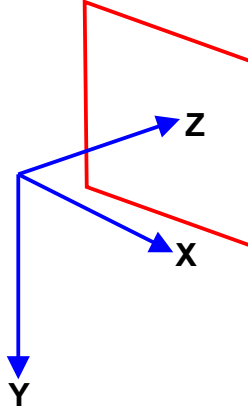


Figure 5.13: Alignment coordinates

In order to determine the influence of the Y displacement on the test results, a number of tests were carried out with the same set of pictures, but the right image was vertically shifted on the Y axis by -3, -1, +1 and +3 pixels. After the pixel shift process, in order not to have rows of black pixels in the image, the images were cropped.

The resulting images were used as input images to the disparity map generation algorithms selected for testing in section 4.2. In the case of each disparity map generation algorithm, test images acquired at 4 different camera-object distances were used. The vertical pixel shift described in the previous paragraph was applied to each of these images.

Figure 5.14 represents the test results, where in the case of each algorithm the performance of the algorithm at different shift values is presented. Based on the information from the test results, it can be concluded, that (i) even the smallest Y displacement in the image plane can have a significant impact on the test results and (ii) even though the impact on the results is significant, the shape of the plot is similar in the case of all displacement values.

5.4.2 Calibration of the Stereo Camera

As presented in 2.4, when a stereo camera is used for measurement purposes, calibration is required. The purpose of this section is to analyze whether the use of an uncalibrated camera influences the quality of disparity and depth maps.

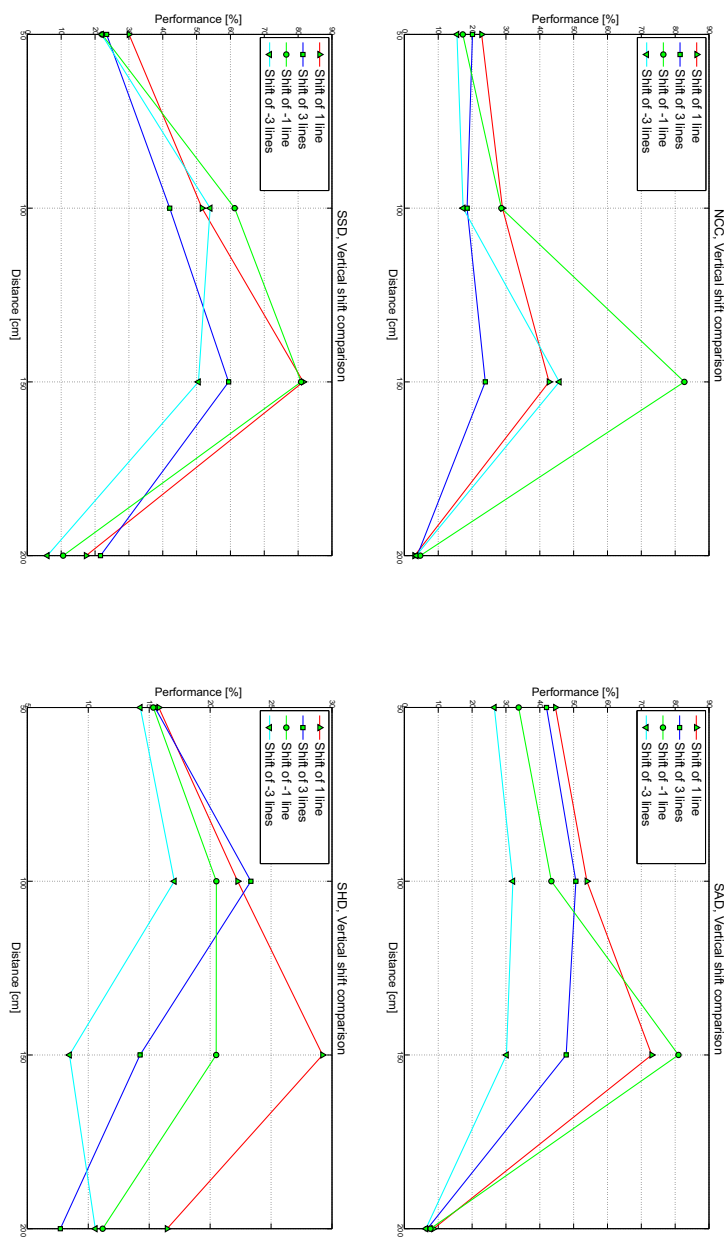


Figure 5.14: Influence of vertical misalignment

5.4 Image Acquisition Issues that might influence the Results

In order to analyze the influence of camera calibration in the case of the disparity and depth map generation, a comparison of the algorithm performance was performed in the case of disparity maps generated before calibration, and disparity maps generated after calibration. The test scene used for this purpose was the AWARD test scene, which was introduced in chapter 4.5.1. The results correspond to the performance measured for Object 2 in the test scene, figure 4.13. As mentioned in chapter 4.3, the performance of an algorithm is determined individually for each object in the scene.

In order to calibrate the stereo camera, which in our case was the Fujifilm W3 3D, presented in section 5.1.1, the *Camera Calibration Toolbox for Matlab* was used, which was developed by Jean-Yves Bouguet [22]. This toolbox calibrates both the intrinsic and extrinsic parameters of the cameras. During the first stage of the calibration, the intrinsic parameters of the cameras are determined, which include (i) focal length, (ii) principal point, (iii) skew and (iv) distortion. The intrinsic parameter calibration is performed during the manufacturing process and inside the ISP. After the individual intrinsic parameter determination, the extrinsic parameters of the stereo camera are determined, which define the (i) rotation vector and the (ii) translation vector. The rotation vector contains the rotation angles around the x, y and z coordinates of the image, see figure 5.15. The translation vector contains the displacement values in the x, y and z directions of the right image compared to the left image.

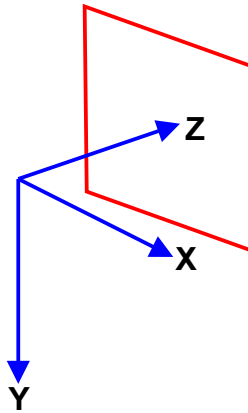


Figure 5.15: Calibration coordinates

The following steps are required for the calibration:

- A set of at least 15 images of a planar checkerboard have to be acquired, figure

5.16, using each of the cameras individually, while not altering the camera position.

- Each of the cameras from the stereo camera pair needs to be calibrated intrinsically, and the calibration results are stored in a special parameter file.
- Using the parameter file generated after the intrinsic calibration, and the stereo images, the toolbox creates a diagram with the views that are used when acquiring the calibration images, see figure 5.17, and at the end of the calibration process, it generates a set of extrinsic parameters, which represent the rotation and translation vectors between the stereo cameras.

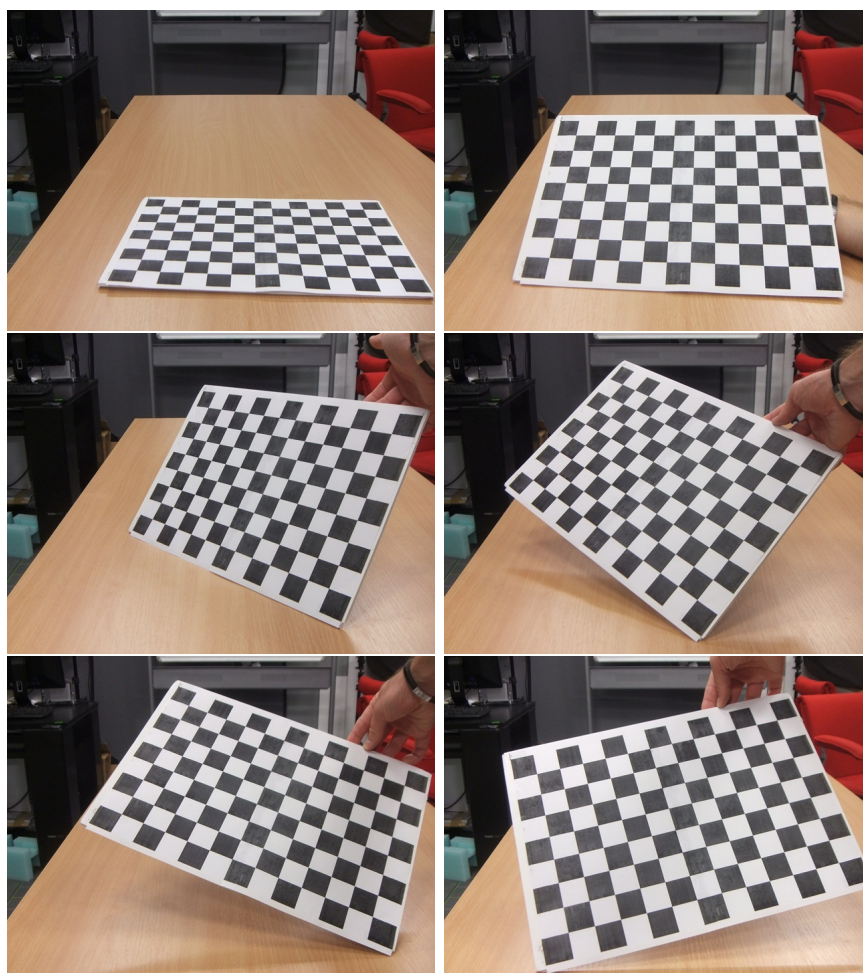


Figure 5.16: Calibration images

5.4 Image Acquisition Issues that might influence the Results

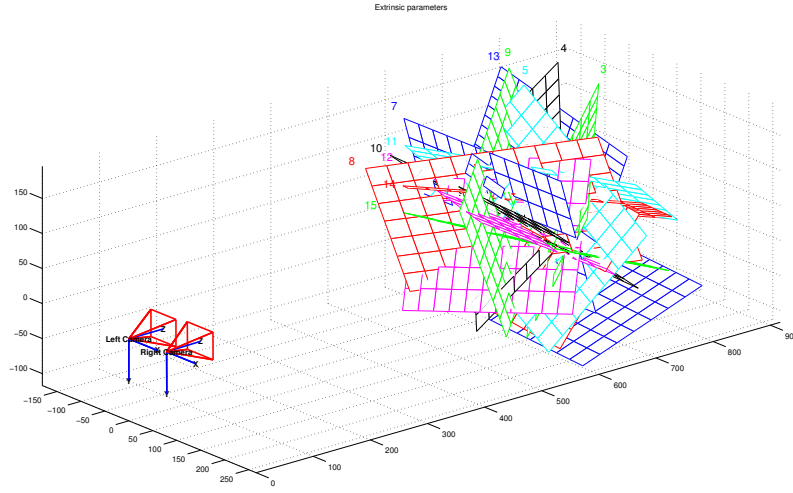


Figure 5.17: Calibration views

The extrinsic parameters used for calibration are:

Extrinsic parameters

(position of right camera wrt left camera):

Rotation vector: $\text{om} = \begin{bmatrix} -0.01189 & 0.00997 & -0.00584 \end{bmatrix}$

Translation vector: $T = \begin{bmatrix} -73.32710 & 1.24772 & -3.86140 \end{bmatrix}$

The world coordinates for which the calibration values are provided can be seen in figure 5.15. Based on this figure, the values in the Rotation vector correspond to the rotation angles for X, Y and Z coordinates, whereas the values in the Translation vector correspond to the translation values of the right image in the X, Y and Z directions.

In order to calibrate the images, only the information from the rotation matrix was used. The information from the translation vector only contains information about (i) the size of the stereo baseline, which does not need to be corrected, and (ii) information about the Y displacement, which was corrected using a different algorithm integrated into the script used to run the tests.

As it was mentioned in the introduction part of the current section, disparity maps were generated using calibrated and non-calibrated images. For the generation of the disparity map, the algorithms described in section 4.2 were used. In order to determine the difference in performance, disparity map generated with calibrated and non-calibrated images were compared, as it can be seen in figure 5.18.

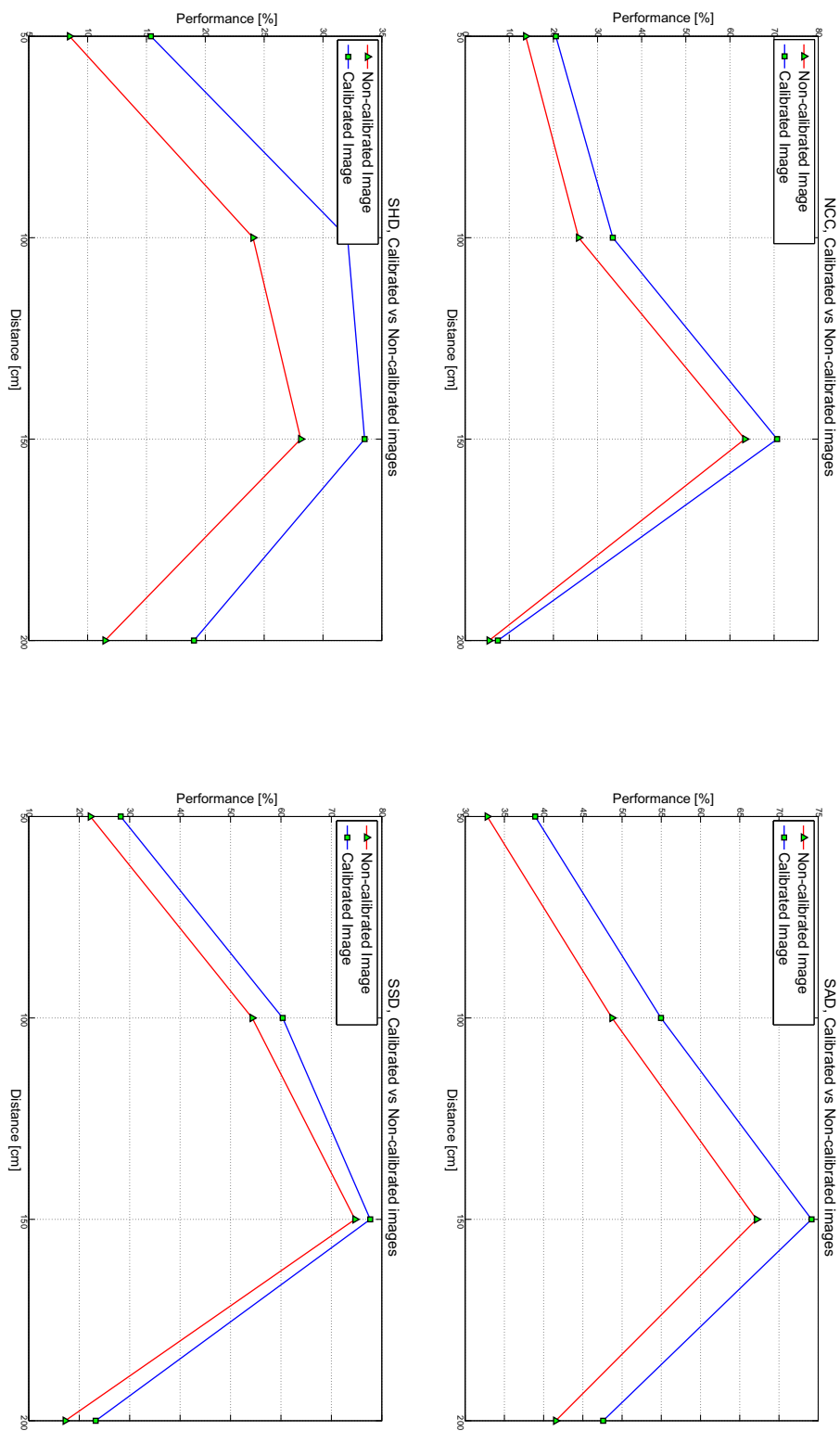


Figure 5.18: Calibrated Images vs Non-calibrated Images comparison

5.4 Image Acquisition Issues that might influence the Results

By looking at figure 5.18, it can be notice that the performance measurements for the calibrated images provide higher values than the ones for the non-calibrated images. It can also be noticed that even though the values are different, the shape of the plot is the same. This means that in if the purpose of the research is only to determine which correlation-based algorithm performs better, then calibration is not necessary. If the purpose of the research is to create an algorithm which is able to generated high quality depth maps, then calibration is a must.

5.4.3 The Influence of Noise

Due to the quality of the camera elements used for testing, especially the CMOS/CCD sensor and lens system, certain differences can be noticed in consecutive images acquired of the same scene. These differences were described in the introduction part of section 5.4. It was decided to name these differences generically as “noise”.

In order to test the possible influence of the noise in the test results, it was decided to acquire 10 pairs of stereo images, without changing any of the setup parameters such as light intensity, light temperature or scene layout.

The acquired stereo image pairs were used to generate disparity maps, by using the algorithms described in section 4.2. As a results, 10 different disparity maps were generated in the case of each algorithm. In the case of each disparity map generation algorithm, our purpose was to determine the difference in quality between the resulting disparity maps. For this reason, the standard and mean deviation was calculated for each algorithm as it can be seen in table 5.9.

Table 5.9: Deviation values

Standard Deviation		Mean Deviation	
Algorithm	Value [%]	Algorithm	Value [%]
NCC	2.5047	NCC	2.0285
SAD	1.3272	SAD	1.1226
SHD	1.1154	SHD	0.9166
SSD	2.2928	SSD	1.8283

As a conclusion, the noise does influence the results by a small amount, but this amount is negligible.

5.5 Conclusions

The Fujifilm W3 3D camera was chosen for test image acquisition purposes over the custom-made stereo image camera due to image quality reasons which manifested themselves as dust inside the camera lens, blurred and non-symmetrical images from photometric point of view, as presented in section 5.1.3.

Due to the reason that we wanted to keep computation times low during the tests, it was decided to use images at VGA (640x480) resolution. The minimum resolution that could be set on the camera was of 2048x1536, and for this reason, we needed to use a down sampling algorithm in order to reduce the resolution. A number of tests were carried out, and based on these tests, the Lanczos downsampling algorithm was chosen to be the most suitable one from image quality point of view.

In order to determine the limitations of the image acquisition device, and to decide which uncontrolled elements might influence the final results, a number of calculations had to be performed, such as (i) measurement of the parallax, (ii) the influence of vertical displacement on the depth map quality and (iii) the influence of non-calibrated images on depth map quality and (iv) the influence of noise.

Based on the camera specifications and basic trigonometrical formulas, the minimum camera-object distance and the maximum stereo base length that can be used for our test cases were determined. It was determined that for the stereo base values of 10 and 12.5, the overlapping of the pixels start at the distances of 12 cm, respectively 15 cm from the camera. At distances closer than this, there is no common information in the two separate views, and for this reason, there isn't enough available information for depth map generation.

In this chapter, we also concluded, that (i) a slight Y displacement might have significant influence on the quality of the depth maps, as presented in section 5.4.1 and (ii) stereo images need to be calibrated in order to be able to measure exact distances, and generate accurate depth maps, as described in section 5.4.2.

The noise in the image acquisition process influences the results by a small amount, but compared to the final values, this amount is negligible.

Chapter 6

Experiment Results

So far in the thesis, the development of the test methodology and the description of the test scenes were described in chapter 4, and the devices used for image acquisition purposes together with their limitations were presented in chapter 5.

The goal of the current chapter is to describe the experiments that were performed in order to measure the performance of the depth map generation algorithms proposed in section 4.2. The experiments can be divided into two different categories: (i) the measurement of the algorithm performance under different light conditions, and (ii) the measurement of the algorithm performance at different camera-object distances.

The chapter starts with the description of the experimentation process, which is described step-by-step, and time guidelines are given.

The experiments for (i) are described in section 6.2.1, where the intensity of the light is varied, according to the test scene specifications from section 4.5. The experiments for (ii) can be sub-divided into two additional categories, where a set of experiments are performed for images with a fixed stereo baseline in section 6.2, and for images with a variable stereo baseline, section 6.3.

After the initial experiments on the AWARD, FACE and FACES test scene, in section 6.2.2, preliminary conclusions are drawn, and further experiments are introduced.

The chapter concludes with a discussion about the possible influence of the light direction on the quality of the disparity map, in section 6.3.2, a discussion which is based on the experiments carried out in section 6.3.1.

6.1 Description of the Experimentation Process

The purpose of the current section is to provide details about the process required to perform certain experiments throughout the research presented in the thesis. This description includes details from the test scene setup to the result being analyzed.

The complete experiment process can be divided into several stages such as (i) setting up of the test scene, (ii) acquisition and pre-processing of the images and (iii) testing and final results. Each of these stages includes several sub-stages, which will be introduced in the following paragraphs. Other researchers can select the relevant steps suitable for their own experiments.

6.1.1 Stage 1, Setting up of the Test Scene

- Identify a space large enough for the testing scene to the set up.
- Based on the test scene specifications, section 4.5, re-create the test scene, making sure that all the measurements are accurate.
- Set up the studio lights, making sure that (i) the intensity of the light is the same as in the case of the initial test scene, and (ii) make sure the lighting equipment is not visible in the acquired images.
- Set up the tripod and the camera, making sure that (i) the initial camera-object distance is accurate, and (ii) the camera is not rotated and tilted, figure 6.1.
- Figure 6.2 represents a view of the testing room after stage 1 of the experimentation process.

6.1 Description of the Experimentation Process



Figure 6.1: Tripod settings



Figure 6.2: View of the testing room

6.1.2 Stage 2, Acquisition and Pre-Processing of the Images

- When using the camera, make sure that auto-shutter is used in order to eliminate the rotation caused by triggering the image acquisition manually. At least two sets of test images had to be discarded due to this reason.
- Before each image is acquired:
 - Make sure that the camera settings, which include the **f number**, **aperture**, **ISO** and **focus** are the same as in the previous cases.
 - Measure the intensity of the light on all the objects in the test scene.
 - If the camera is moved to a different camera-object distance in the room, make sure to check the tripod settings.
 - Make sure, that you, as a camera operator, are not in the way of the light.
- Move the camera to a different camera-object distance and/or decrease the intensity of the light, according to the purpose of the experiment.
- As presented in section 5.1.1, the camera used for image acquisition saves the stereo image pair in .mpo format.
- After the image acquisition process, the MPO files need to be transformed into JPEG files which contain the left and right views of the images.
- Once the JPEG files are obtained, the images need to be downsampled to VGA resolution, as described in section 5.2.
- Before using the images for tests, they have to be re-named, so that they can be used with the MATLAB script. A file name example is: “awardP1_l.jpg” and “awardP1_r.jpg”.

6.1.3 Stage 3, Testing and Final Results

- The current stage has two main sub-stages, (i) the generation of the disparity maps using the proposed pixel matching methods in section 4.2 and (ii) the performance measurement of each algorithm.

6.1 Description of the Experimentation Process

- A MATLAB script was written, which reads each of the test images and takes them through the disparity map generation process, which includes (i) the vertical alignment of the images, and (ii) the generation of the disparity map using different pixel matching costs. The disparity map generation algorithm was described in section 4.2. The names given to the resulting disparity maps are in the form of: “depthSSD_awardP1.jpg”.
- After the generation of the disparity maps, the quality needs to be measured and for this task, ground truth images are needed, as presented in section 4.3. For each camera-object distance, a ground truth is generated using the technique described at section 4.4.
- The quality measurement script reads the disparity map and the the corresponding ground truth. For each object in each disparity map, a TXT file is created, which contains the quality measurement in percentage.
- At the end of the quality measurement process, a Microsoft Excel “.csv” file is created, which contains the quality measurement values of each object in the test scene. Each algorithm has it’s own CSV file.
- In order to generate the CSV files, Cygwin was used, which is a “collection of tools which provide a Linux look and feel environment for Windows”. The command used for each algorithm is the following:

Algorithm 6.1 Cygwin command

for name in SAD*.txt; do echo -n ‘cat \$name | sed -e ‘s/\s//g’;’ ’; done > SAD.csv

- After the generation of the CSV files, the values were read, and placed into (i) Microsoft Excel spreadsheets, and (ii) copied into MATLAB for graph generation purposes.
- A MATLAB script was written, which generated each graph separately.

In order to determine whether the disparity map generation script and the quality measurement scripts work properly, and that there are no implementation errors, 2 tests were performed. In the case of the first test, the same image was used for both left image and right image inputs of the script. The result was a black image, which is the expected result.

For the second test, an artificial image was created, where pixels on each line in the image had a different color, figure 6.3.

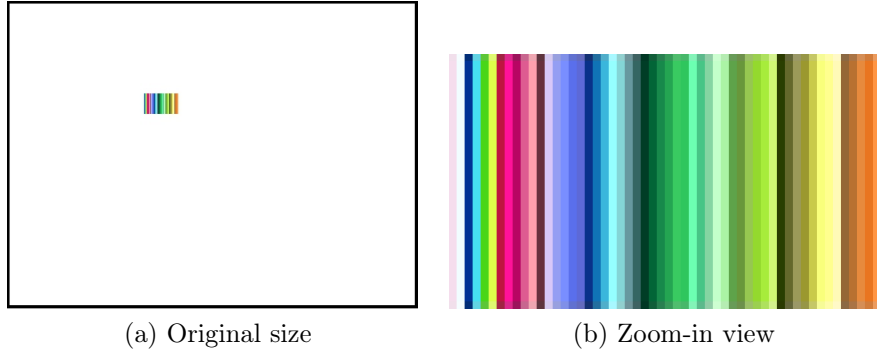


Figure 6.3: Artificial image

The results can be seen in figure 6.4, where it can be noticed that the artificial shape was almost perfectly re-created. A small shadow can be noticed on the left of the result in figure 6.4. This shadow is a side-effect of the correlation-window based algorithms. Due to the fact that correlation windows are used, and in this case the search was performed from right to left, this shadow was created on the left edge of the object.

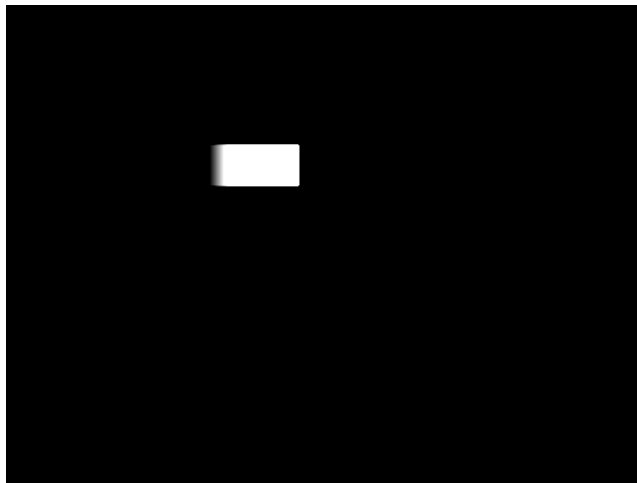


Figure 6.4: Artificial image - Disparity map

Both these test prove that the disparity map generation script works fine.

6.1 Description of the Experimentation Process

6.1.4 Practical Issues

In order to complete the first stage it takes on average between 1.5 and 2 hours, depending on the complexity of the test scene. The completion time of stage 2 depends on the number of camera-object distances, on the number of stereo baselines that are used, and on the number of light settings. The least amount of time it took me was 3 hours, and the most amount of time was 7-8 hours. The final stage is the most time-consuming, due to the MATLAB code which generates the disparity maps, the time it takes to create the ground truth images, and the time it takes to measure the quality. On average, the total time for this stage was between 8 and 10 hours. This time depends on the number of images that are being processed. The main reason why MATLAB was used is, that it provides an Image Processing Toolbox which is very helpful in the implementation of the algorithms.

It needs to be mentioned that when using the uCam for image acquisition purposes, section 5.1.2, instead of the FujiFilm, the time for stage 2 doubles or triples. The reason is, that in the case of the uCam, the images are not acquired with the push of a button. For each individual image, a different file name has to be given in the C code that was written using the Xilinx SDK. The acquired images are written onto a Compact Flash card. After acquiring 3 or 4 consecutive images, the CF card has to be re-formatted, because the SDK crashes quite often.

In the current chapter and thesis, only a limited number of experiments are described, in the case of which conclusions were drawn. A large number of image sets were acquired, but some of them had to be discarded due to different image acquisition mistakes. 3 sets of test images were acquired using the uCam, and 15 sets of test images were acquired using the FujiFilm camera. A total of 22 complete tests were run, which include stage 3 of the experimentation process, by using different test images, correlation window sizes, disparity limits, etc.

6.2 Experiment Results using a fixed Stereo Base

6.2.1 Algorithm Behavior under different Light Conditions

As presented in chapter 4.5, one of the purposes of the test scenes was to determine the way light intensity influences the algorithm performance. For this reason, for each test distance defined in chapter 4.5, 4 different light intensity values were used, figure 4.15. The light intensity values are can be seen in tables 4.1, 4.2 and 4.3.

The test results obtained in the current section are the ones corresponding to the AWARD test scene, which was presented in chapter 4.5.1. The algorithm performance values illustrated in the test result plots correspond to Object 2 in the scene 4.13. As mentioned in section 4.3, the algorithm performance is measured for each individual object in the test scene.

The performance comparison between different light intensity values can be seen in figure 6.5. The figure illustrates the performance of a certain algorithm, at different camera-object distances, for different light intensity values. For each camera-object distance (x axis) 4 different algorithm performances are presented based on the change in light intensity. The performances are displayed using different colors for each light intensity. During the tests performed using different light intensity values, it was noticed that for lower intensity values, the algorithm performance drops. By analyzing information from figure 6.5, it can also be noticed that the biggest difference in performance is found at a camera-object distance of 150 cm, where the difference in performance between intensity 1 and intensity 4 is of 35.98% on average for the presented algorithms. The performance difference of 35.98% at the camera-object distance of 150 cm can be explained with the help of the results presented in section 6.2.2.1. The results described in this section show that the highest algorithm performance is recorded at a camera-object distance of 150 cm. If the performance at a certain camera-object distance is high, it means that each change in the testing conditions will have a larger effect on the result at that specific distance, compared to other camera-object distances. In our case, the change in the testing condition was the light intensity.

The average difference in light intensity between intensities 1-2, 2-3 and 3-4, which were used during our experiments, is 170 lx. The best performing matching costs, based on the information retrieved from the test results, are the SAD and SSD.

6.2 Experiment Results using a fixed Stereo Base

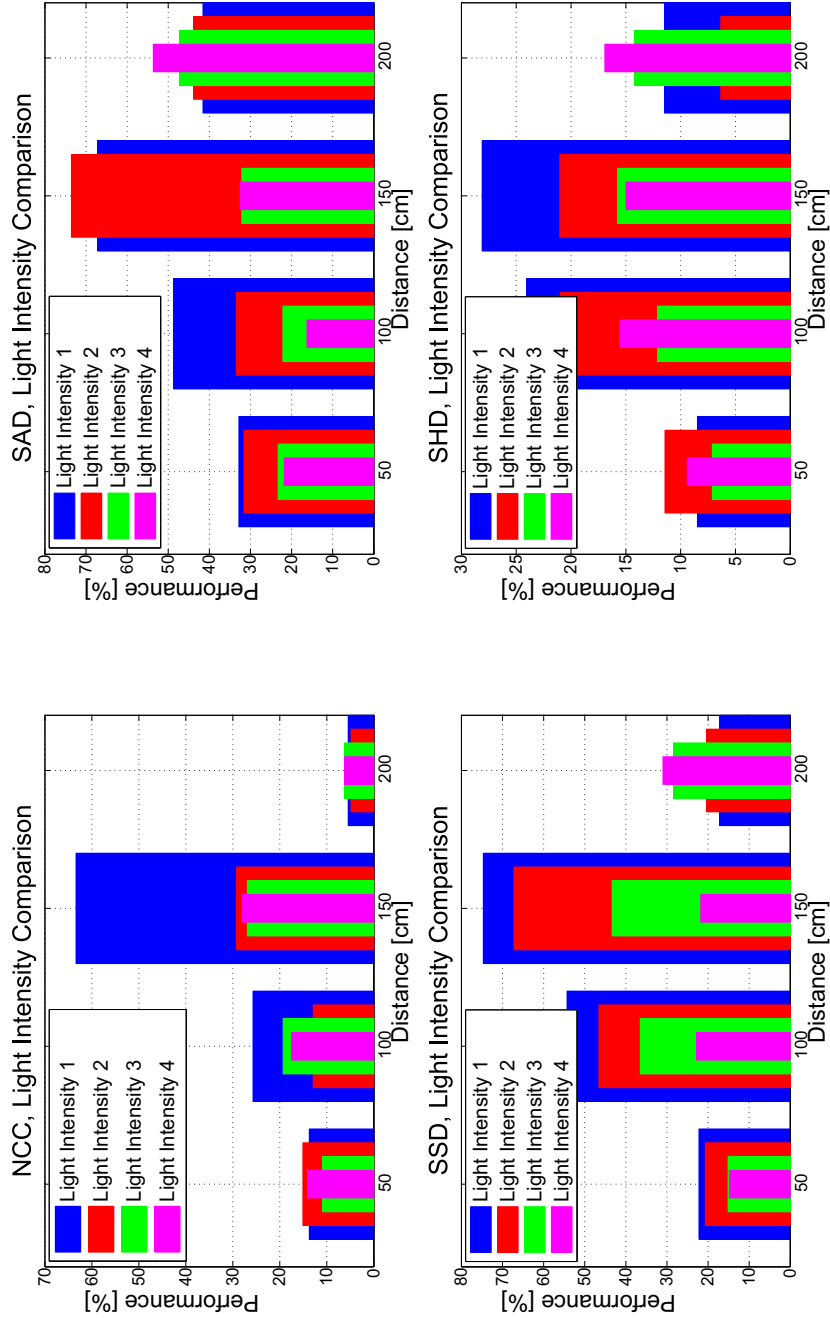


Figure 6.5: Light intensity comparison for the AWARD test scene

6.2.2 Algorithm Behavior at different Camera-Object Distances

6.2.2.1 Initial Test using 4 Distances

AWARD test scene As presented in section 4.5, another purpose of the test scenes was to show the behavior of the proposed algorithms, using the same test scene, with test images acquired at different distances, which can be seen in figure 6.6.



Figure 6.6: AWARD original scene, 4 distances

The test results described in this section are the ones corresponding to the AWARD test scene, introduced in section 4.5.1. The algorithm accuracies from the test results are the ones corresponding to Object 2 of the test scene 6.7.

6.2 Experiment Results using a fixed Stereo Base

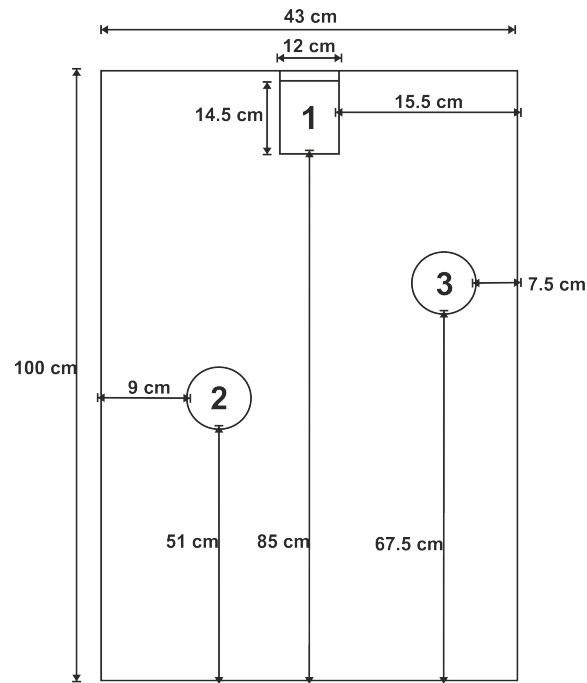


Figure 6.7: AWARD scene setup

The resulting algorithm performance values, and the corresponding distances can be seen in figure 6.8.

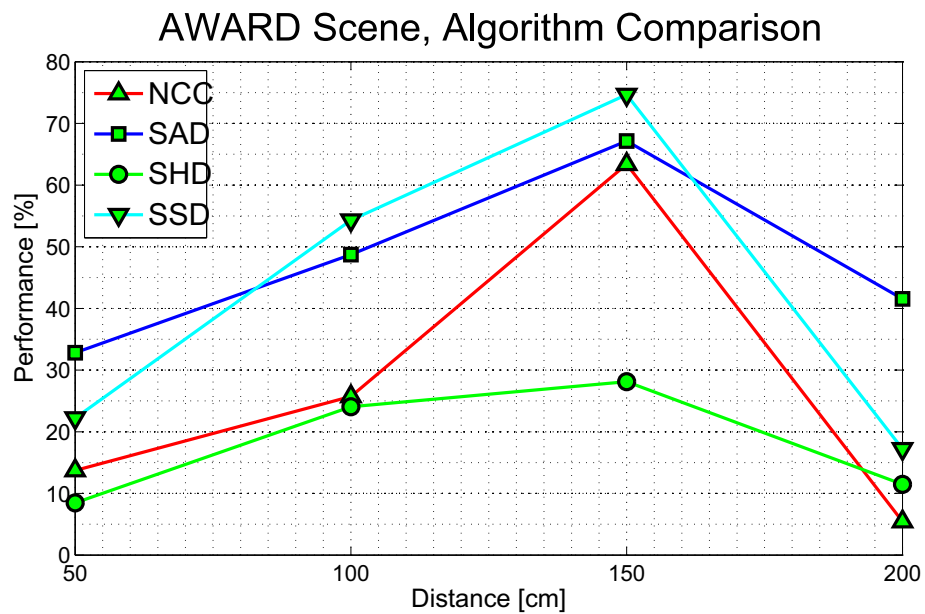


Figure 6.8: AWARD test scene distance comparison

FACE test scene The FACE test scene was one of the initial test scenes that were created for the purpose of algorithm testing, and it was presented in chapter 4.5.2. Similarly with the AWARD test scene, test images with different camera-object distances were acquired in order to test behavior of the proposed algorithms.

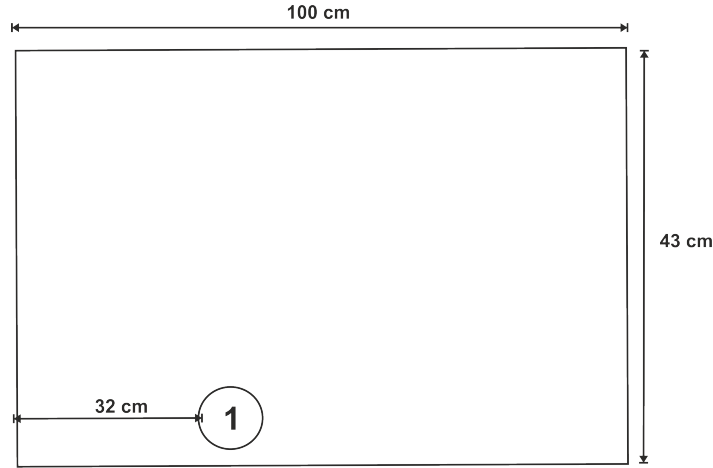


Figure 6.9: FACE scene setup

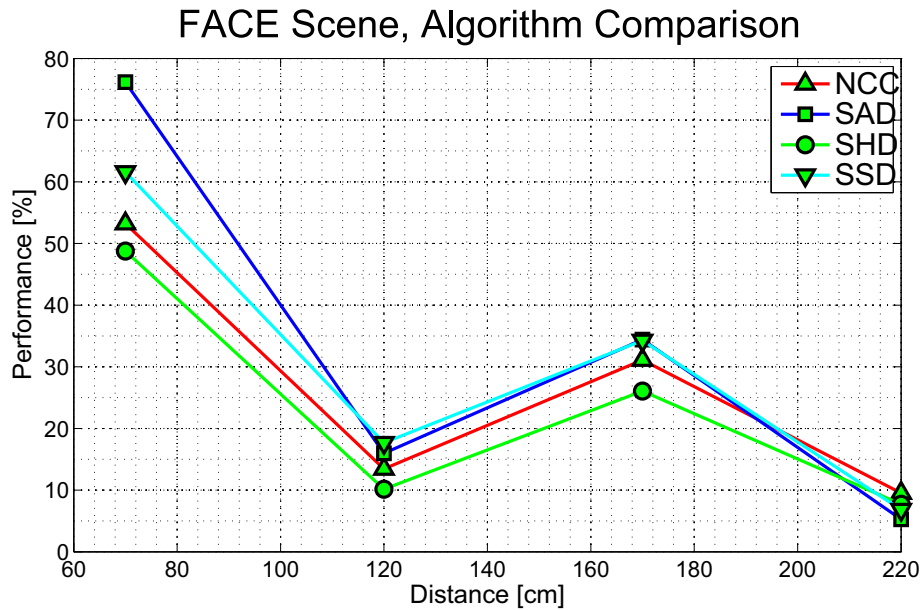


Figure 6.10: FACE Test Scene distance comparison

The performance values obtained in this section correspond to Object 1 in the test scene 6.9. The resulting performance values, together with the corresponding distances can be seen in figure 6.10.

6.2 Experiment Results using a fixed Stereo Base

FACES test scene The third and last one of the initial test scenes was the FACES test scene, introduced in chapter 4.5.3. The main purpose of the creation of the test scene was explained in 4.5.3. Test images were acquired of this scene at difference camera-object distances in order to test the behavior of the algorithms.

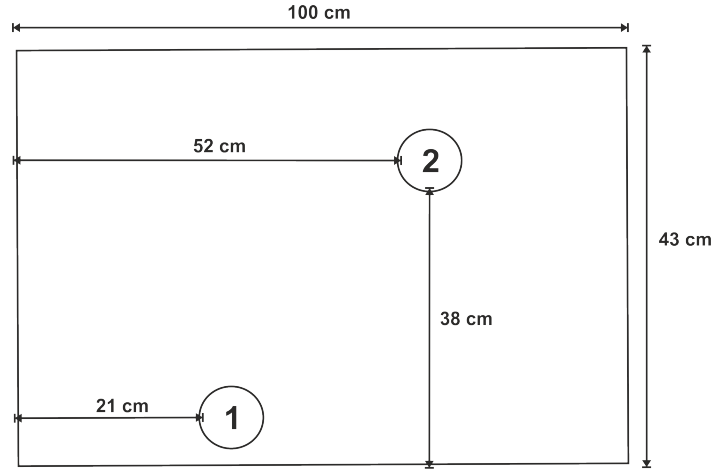


Figure 6.11: FACES scene setup

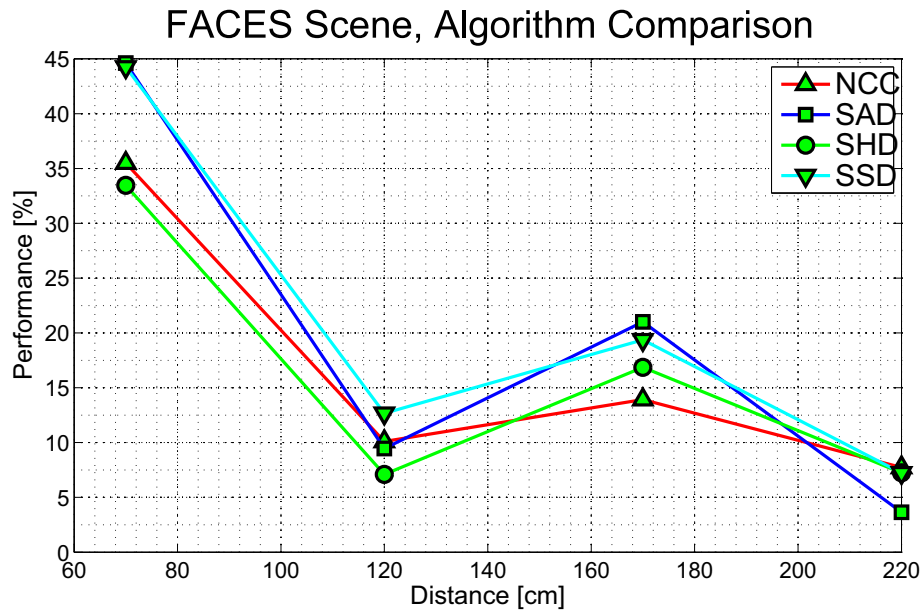


Figure 6.12: FACES Test Scene distance comparison

The test result values obtained in this section correspond to Object 1 in the test scene 6.11. The performance values resulting from the tests, together with the corresponding distances can be seen in figure 6.12.

6.2.2.2 Analysis of the Initial Results

During the tests performed on the AWARD test scene by acquiring test images at different distances it was noticed that with the increase of the camera-object distance the algorithm performance increases as well. This observation is in contrast with the observations from the FACE and FACES test scenes, where, with the increase of the distance, the performance decreases. In the case of the AWARD test scene it was noticed that at a camera-object distance of 150 cm, the recorded algorithm performance is the highest. The increase in the performance was noticed in the case of all the algorithms that were tested. The best performing correlation-based algorithms, decided by analyzing the test results, are the SAD-based and SSD-based, with maximum performance values of 67.144% and 74.68%.

Based on the results in the case of the FACE scene, it can be noticed that the highest performance is achieved at the closest camera-object test distance which was used, 70 cm. This is in contrast with the AWARD test scene, where the highest performance was recorded at a camera-object distance of 150 cm, but the performance is similar to the FACES test scene, where the maximum performance is achieved at the same distance. The best performing algorithms in the case of the FACE test scene are the SAD-based and SSD-based, with maximum performance values of 76.156% and 61.579%.

Based on the information from the test results in the case of the FACES test scene the best performance of the algorithms was recorded at a camera-object distance of 70 cm, the same distance as in the case of FACE test scene, which is in contrast with the AWARD test scene. The best performing algorithms are the SAD-based and SSD-based, with maximum performance values of 44.615% and 44.251%.

It was mentioned in section 4.5.3 that one of the purposes of the FACES test scene, which contains 2 faces, was to test how having multiple faces or multiple objects of the same color in a scene might influence the performance of the algorithms. By looking at the maximum performance values of the best performing SAD and SSD algorithms, comparing them to the results achieved in the case of the FACE test scene, it can be noticed that multiple faces of the same skin color in the same image influence the performance of the algorithms in a negative way. In the case of the SAD-based correlation, a drop in performance of 31.541% can be noticed, whereas in the case of the SSD-based, a drop in performance of 17.328% is noticed.

Based on the information presented in the case of the AWARD, FACE and FACES

6.2 Experiment Results using a fixed Stereo Base

test scenes it was noticed that there is a contrast between the performance of the algorithms in the case of the AWARD scene and their performance in the case of FACE and FACES scene. This contrast manifests itself in the fact that in the case of the AWARD test scene, the maximum performance values achieved by the best performing algorithms can be noticed at a camera-object distance of 150 cm. In the case of the FACE and FACES test scenes, the maximum accuracies achieved by the same SAD-based and SSD-based correlation can be noticed at a distance of 70 cm.

The contrast in the results obtained for the AWARD, FACE and FACES test scenes can be explained by the influence of light direction on the test results, as presented in section 6.3.2. The light reflects from object of different sizes and shapes in a different way. This means, that for different objects, different type of information is recorded by the image acquisition device. During the test image acquisition, care was taken in order to illuminate the objects in the test scene in an uniform way. The main problem was that the light equipment used during the initial experiments was not as advanced as the one used in chapter 7, section 7.4.2. For this reason, a uniform illumination was not guaranteed.

After these initial results it was decided that further experiments need to be carried out, and it should be checked, (i) if the occurrence of the high performance in the case of the AWARD test scene at a camera-object distance of 150 cm was influenced by the shape and color of the object, (ii) or whether it was due to the fact that there is an “ideal” location. In order to do this, the TEXTURED OBJECTS test scene, detailed in section 4.5.4 was created.

TEXTURED OBJECTS test scene This test scene contains an object of a similar shape with the one that was used for performance measurement in the case of the AWARD test scene, but of different color. To be more precise, the object in the initial test scene had the color navy blue, whereas the object in the second test scene has multiple colors. During this experiment images at different camera-object distances were acquired, and were used to generate disparity maps.

The test results presented in this section correspond to Object 1 in the scene 6.13.

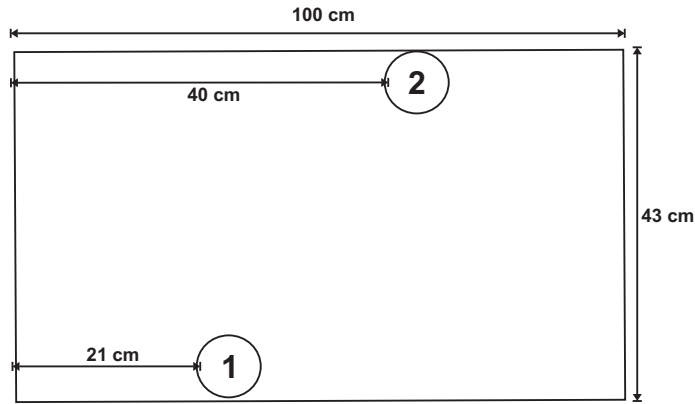


Figure 6.13: TEXTURED OBJECTS scene setup

The performance results, together with the corresponding distances can be seen in figure 6.14, where a comparison of the AWARD and TEXTURED OBJECTS test scenes is illustrated. It can be noticed from these results that the algorithms behave in a similar way for both test scenes, and the difference in performance for the two separate test scenes is of 2.939% for the SAD-based algorithm and 9.743% for the SSD-based algorithm. This provides a strong indication of the fact that the results in the case of the AWARD test scene were not accidental, and that the results can be reproduced by using different objects as well.

6.2.2.3 Advanced Tests using 8 Camera-Object Distances

After noticing the unexplained increase of the performance of the algorithms in the case of the AWARD and TEXTURE OBJECTS test scenes, which apparently was caused by the increase of the camera-object distance, it was decided that more tests need to be performed in order to find the reasons behind this increase in performance.

In order to do this, the AWARD test scene was used, but the number of test distances were doubled, from 4 to 8. In the case of this scenario, the initial camera-object distance was of 50 cm, and the distance between the camera and the test scene was increased by 20 cm at each new acquisition. The distances can be seen in table 6.1. During the acquisition of the test images at different distances, the scene was not moved in order to minimize the possible influence of light direction on the scene, an influence which is presented in detail in 6.3.1. Instead, the camera mounted on a tripod was moved to different locations.

6.2 Experiment Results using a fixed Stereo Base

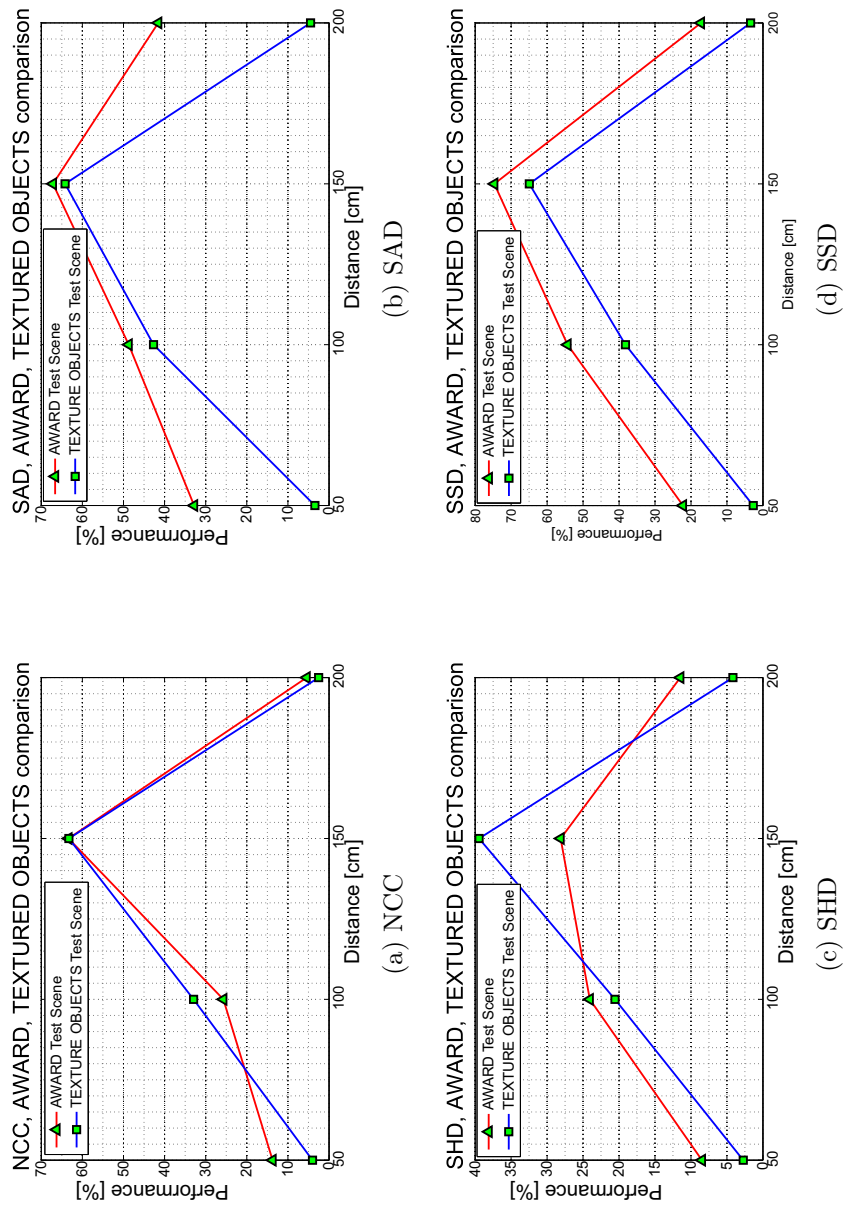


Figure 6.14: AWARD and TEXTURED OBJECTS test scenes distance comparison

Table 6.1: Distance table for steady scene test with 8 distances

Distance location	Distance [cm]
1	50
2	70
3	90
4	110
5	130
6	150
7	170
8	190

The test results in this section correspond to Object 2 in the AWARD test scene, section 4.5.1. A comparison between the results of the tests performed in the 4 distance test scenario and the ones performed in the 8 distance scenario can be seen in figure 6.15. For the purpose of simplicity, only 2 correlation-based algorithms are used in this comparison. A full comparison can be found in Appendix B. It was decided to present the NCC-based and SAD-based algorithms, because the SAD was shown to be a top-performer in our previous tests, whereas the NCC was one of the worst performing pixel matching costs.

In the comparison, there are two facts that need to be observed. The first is, that in the case of the 8 distance scenario, the increase of the camera-object distance causes an increase in the performance as well. This observation is valid for both the NCC- and SAD-based algorithms. The second observation is, that in both the 4 distance and 8 distance scenarios, maximum performance values are recorded at a camera-object distance of around 150 cm. In the case of the NCC algorithm, even though peak results are recorded at a distance of 150 cm, there is a big difference between the recorded values. We cannot make the same observation about the SAD algorithm, where, even though the results were not perfectly reproduced, the difference between the highest recorded values is smaller than in the case of the NCC.

6.2 Experiment Results using a fixed Stereo Base

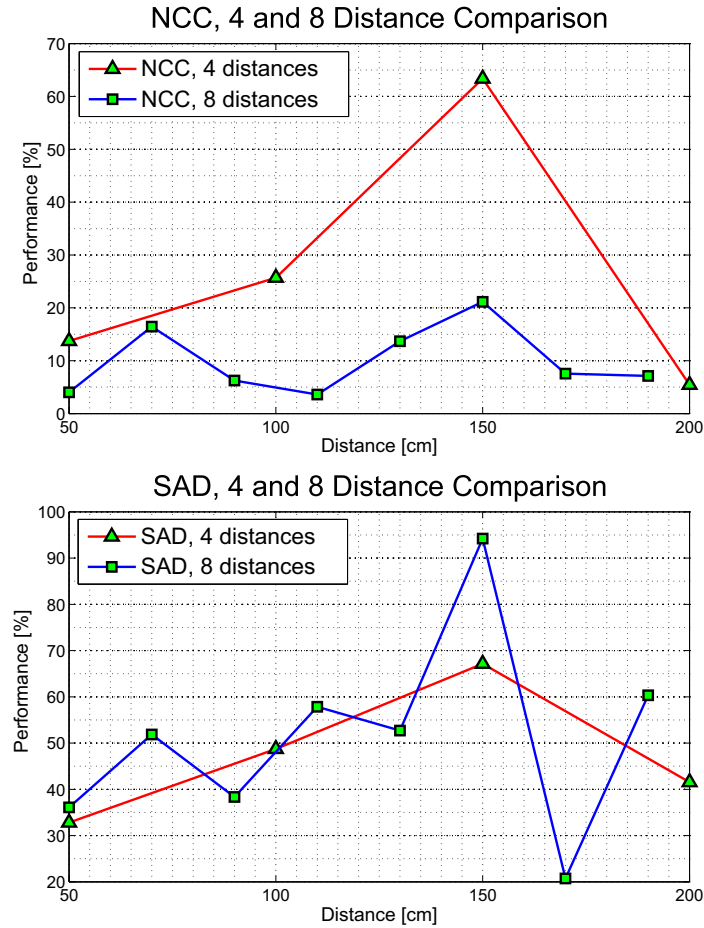


Figure 6.15: Comparison between 4 distance and 8 distance scenarios

By looking at the comparison in figure 6.15, it can also be observed, that in the case of the SAD matching cost, the difference in the maximum performance values obtained during the 4 distance experiment and 8 distance experiment is 27.063%. The same difference in the case of the NCC matching cost is 42.238%. The interesting fact is, that in the case of the SAD-based method, the 8 distance experiment provided better performance results, whereas in the case of the NCC, the 4 distance experiment provided better results. In order to explain this difference in performance, we need to look at how the test scenes were created. As it was explained at the start of this chapter, in section 6.1, the objects were placed at exactly the same locations, the camera settings were kept the same, and the light intensity was kept the same. The experiment errors that might have influenced the results are the camera misalignment, and most likely, the direction of the light, which is described in section 6.3.2 of the current chapter. The amount of difference in performance also

shows that the NCC is more likely to be influenced by experiment errors than the SAD pixel matching cost.

6.3 Experiment Results using a variable Stereo Base

6.3.1 Algorithm Behavior using variable Stereo Base Values

In section 6.2.2.3, the results in the case of 8 distance scenario were presented. During those experiments it is noticed that the influence of the camera-object distance on the performance of the proposed algorithms is not accidental, and that the results can be replicated up to a certain point, but are not fully repeatable. The reason why the maximum performance appears in the same place was still not found.

At this stage, one of the only parameters that wasn't adjusted during the tests was the length of the stereo base, see section 2.1.2. Due to this fact it was considered necessary to acquire a new set of test images, where for each test distance shown in table 6.1, five different stereo base values should be used. The stereo base length values can be seen in table 6.2.

Table 6.2: Stereo base values

Number of the stereo base setup	Length of the stereo base [cm]
1	2.5
2	5
3	7.5
4	10
5	12.5

The test scenario involving different stereo base length values uses the AWARD test scene 4.5.1. All the test results presented correspond to Object 2 in the test scene 4.13. The results corresponding to the same two algorithms which were described in section 6.2.2.3, SAD-based and NCC-based are presented in this section as well, where a comparison between different stereo base values is illustrated in figure 6.17. A full comparison can be found in Appendix B. Only 3 out of the 5 stereo base length values are described in the comparison due to the fact that these three values are sufficient to distinguish the main characteristics of variation due to different stereo bases, and the graphical data is easier to understand and interpret with this number of datasets. The stereo base length values are 2.5 cm, 7.5 cm and 12.5 cm.

6.3 Experiment Results using a variable Stereo Base

By analyzing the results from figure 6.17, it can be noticed that in the case of the NCC algorithm there are large differences between maximum performance values for different stereo base lengths. The same large differences in the case of the NCC algorithm were noticed in section 6.2.2.3 as well. By looking at figure 6.16, in the case of the NCC algorithm a change in the camera-object distance is also noticed, where the highest performance values are achieved.

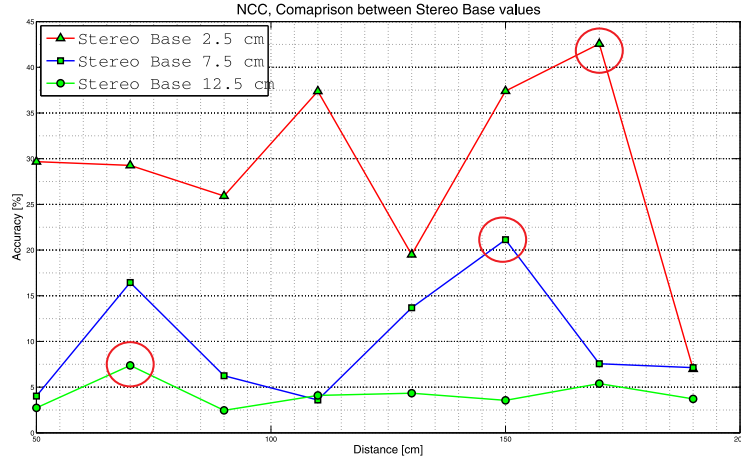


Figure 6.16: Performance results for different stereo base length values, 8 distance scenario, NCC

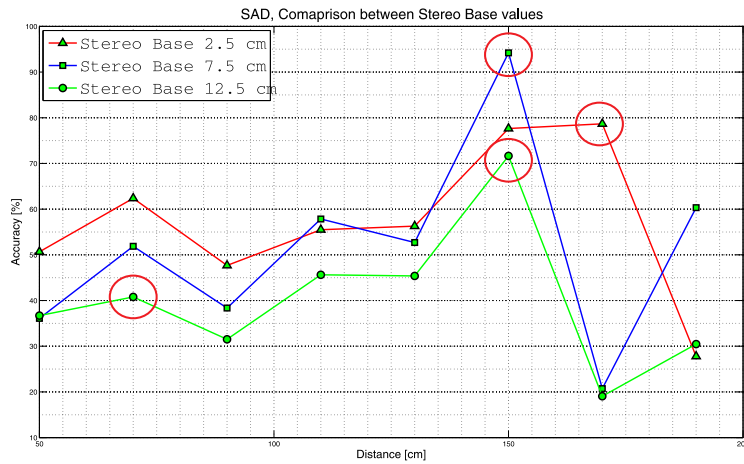


Figure 6.17: Performance results for different stereo base length values, 8 distance scenario, SAD

The same change in camera-object distance for high performance values cannot

be noticed in the case of the SAD algorithm, see figure 6.17, where the highest performance values are recorded at approximately the same camera-object distance. An important thing to notice here is, that the difference between recorded performance values is not as large as in the case of the NCC algorithm. This shows that the SAD-based algorithm is more robust than the NCC-based algorithm.

6.3.2 The Influence of Light Direction on the Algorithm Performance

6.3.2.1 Implementation of the Experiment

In section 6.2.2.3, results for the 8 distance scenario were presented. Even though by analyzing the results it was noticed that with the increase of the camera-object distance the performance of the algorithms increases too, the results oscillate between higher and lower values. For this reason, in order to overcome some possible problems that might have been caused by the movement of the camera during the acquisition of the test images, it was decided to acquire a new set of images, where the camera was kept still, and the scene was moved to different locations.

It was decided that the distance between two acquisition locations should be decreased to 10 cm. Since in the tests detailed in section 6.2.2.3, the longest distance with a maximum recorded value was in the area of 160 cm, it was decided that for the new set of test images, the test locations should be between the values of 50 cm and 160 cm, as it can be seen in table 6.3.

Three different stereo base values were selected in order to present the test results. The selected stereo base values are 2.5 cm, 7.5 cm and 12.5 cm. These values are the same as the ones presented for the 8 distance scenario in section 6.2.2.3. All the performance results correspond to Object 2 in the AWARD test scene 4.13. The results can be seen in figure 6.19, where only the results from the NCC-based and SAD-based algorithms are illustrated. The full set of test results can be found in Appendix B.

6.3 Experiment Results using a variable Stereo Base

Table 6.3: Distance table with advanced test with 12 distances

Distance location	Distance [cm]
1	50
2	60
3	70
4	80
5	90
6	100
7	110
8	120
9	130
10	140
11	150
12	160

6.3.2.2 Analysis and Conclusions

By analyzing the result, we were hoping to record similar performance values as in the case of 6.17. Unfortunately, a similarity in performance was not noticed.

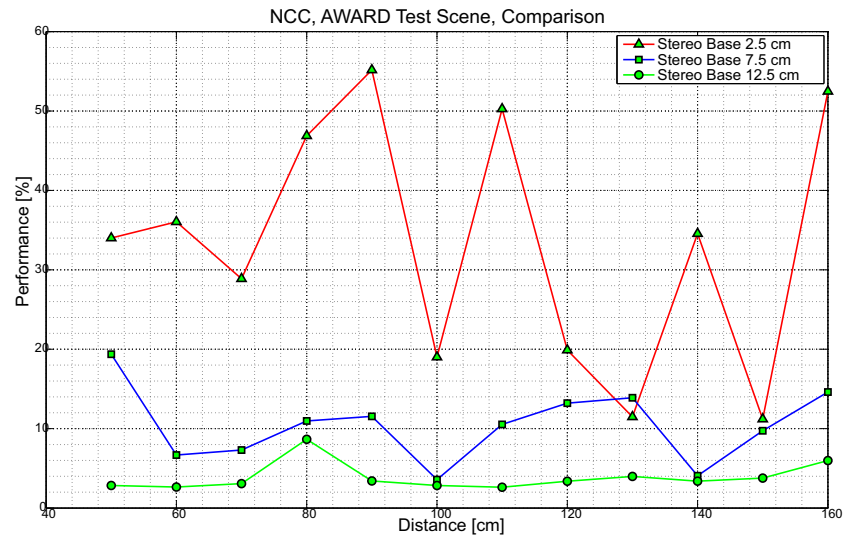


Figure 6.18: Performance results for different stereo base length values, 12 distance scenario, NCC

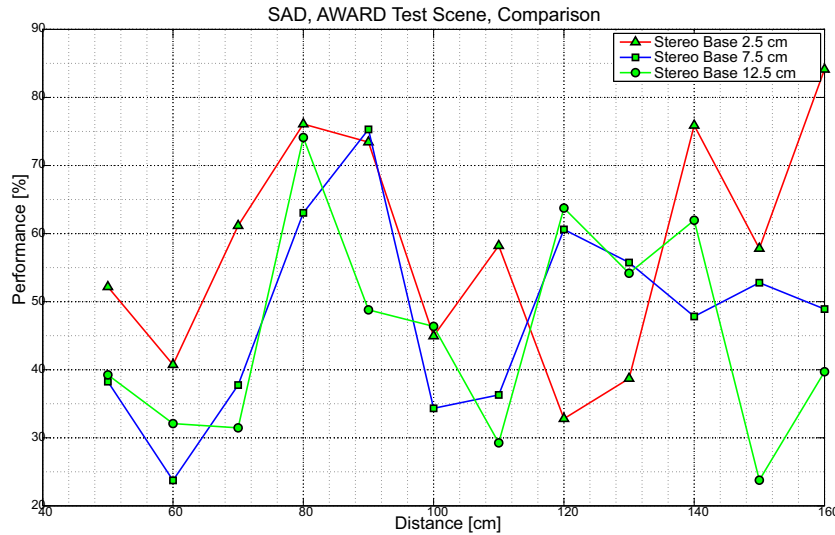


Figure 6.19: Performance results for different stereo base length values, 12 distance scenario, SAD

One of the main reason for this inconclusiveness is that during the test image acquisition, the scene was moved to different locations. In the test room, the light sources are on the ceiling, and by moving the test scene underneath them, the objects in the test scene can be illuminated in a different way for each test location, as it is presented in figure 6.20. By having different parts of the objects illuminated in a different way within the pair of stereo images, the task of the algorithm in finding corresponding pixels is much more difficult, especially in the case of the objects with similar color. This is why both the light intensity, as presented in section 6.2.1 and the light direction should be kept unchanged through the process of test image acquisition.

The conclusions described in the previous paragraph can be confirmed with the help of the research paper published by (Basri et al [13]). In the paper, the authors describe a photometric stereo method, where the shape of objects is recovered from images acquired using a fixed viewpoint, and variable lighting conditions. The proposed method uses a first order harmonic approximation operating in a 4-dimensional space. The fact, that the shape of the image can be recovered from images illuminated in a different way, shows us that care needs to be taken when the test scene is set up, so that objects are illuminated in a uniform way.

6.3 Experiment Results using a variable Stereo Base

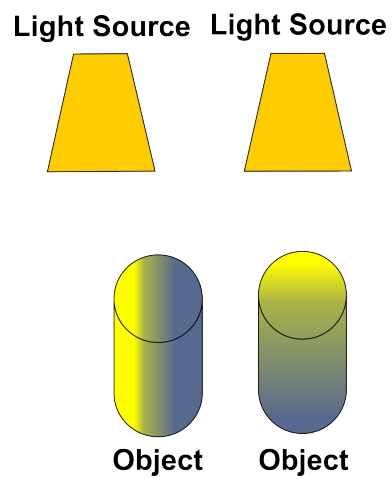


Figure 6.20: Influence of the light on the test scene

Chapter 7

General Test Scene Proposal

In the previous chapters, a proposed test methodology was presented. In chapter 4 the usefulness of different depth map testing scenes was described. A number of devices used for test image acquisition together with their limitations were described in chapter 5, and test results based on a number of test scenarios were presented in chapter 6.

The purpose of the current chapter is to introduce a general depth map testing scene and a number of depth map testing scenarios. During the development of the newly proposed test scenes and scenarios, research findings from the previous chapters are used. The purpose of the proposed test scene and scenarios is to provide a means for depth map generation algorithm comparison, algorithms which can be software or hardware based, or a combination of the two.

For this purpose, section 7.2 presents the selection of the objects that are used in the test scene, and section 7.3 presents the creation of the test scene, where the layout of the objects, and the test environment settings are described. Experiments carried out using the proposed test scene are introduced in section 7.5, where the results are similar to the ones presented in chapter 6.

The chapter concludes with section 7.4, where different test scenarios are analyzed, and section 7.6, where different methods for performance measurement are proposed.

7.1 Introduction

The majority of the available depth map generation methods only rely on image processing algorithms that are implemented on off-the-shelf CPUs, DSPs or SoCs. In order to test these methods, the researchers can either acquire their own test images, or access test images from a specific database, as the one presented in (Scharstein et al [118]). This database also allows researchers to compare their method with other methods by using the same test images.

But, there are some cases, where beside the image processing algorithms implemented in software or hardware, some additional devices are used as well, devices such as near-IR sensors, IR sensors, ToF cameras, as described in section 1.1. For this reason, it was considered necessary to propose a test scene, which can be used by any researcher, no matter what type of method they use. In case of the methods that are only implemented in software or hardware, they will be able to download the stereo image pair together with the ground truth from our on-line database (<http://www.andorko.com/stereo.html>). In case of the methods which use special devices, the researchers will be able to build a similar scene to the one in the proposed database, only by following the specifications of the scene provided by us.

The major advantage of the proposed test scene is, that the objects used in the scene are common objects which can be purchased no matter where the researchers are geographically located, and by following the instructions provided, the test scene can be re-built.

7.2 Selection of the Objects

Since the database presented by (Scharstein et al [118]) is arguably one of the most often used databases for depth map testing, it was decided to follow their comments for our object selection phase. Based on the Tsukuba stereo image pair defined in chapter 4.4 , the most interesting regions of a test image are presented as being:

1. Specular surfaces, which can cause difficulty in depth computation due to the reflected motion of the light surface.
2. Textureless regions, which are locally ambiguous and they are a challenge for stereo algorithms.

7.2 Selection of the Objects

3. Depth discontinuities, which can be seen at the border of all the object within the test image.
4. Occluded pixels, which can cause algorithms to give incorrect depth results for some objects.

Based on the specifications, it was decided to search for objects that have at least one of the characteristics described in the previous paragraph. The objects were chosen based on their global availability and their standardized shape and color characteristics. The following objects were used in the proposed test scene:

1. ISO chart - The purpose of ISO test chart that was used is to provide a certain texture to the otherwise gray background. The size and colors are standard, and the design is available online. The name of this chart is **Siemens star chart**, and it can usually be used for MTF measurement, Moire patterns and the detection of other image processing artifacts. The size of the chart used in the test scene is 32 x 28 cm.
2. Tennis balls - The tennis balls have a matte surface, so they don't reflect light, but they are textureless. The size and color of the tennis balls are standard, and they are defined by the governing body of this sport.
3. 0.5l Coca Cola bottle - The bottle's surface is shiny, while the material is transparent. It also has a standard color, and the bottle has a standard size and transparency factor defined by the manufacturer. Also, the same material is used for the manufacture of all the plastic bottles, and for this reason, the reflexion and transparency indexes are the same. According to [30], translucent objects pose a difficult problem for traditional structured light 3D scanning techniques.
4. Macbeth color checker - The purpose of the color checker is to have a matte and textured object in the test scene. The colors used for the color checker are standard. The only thing that might differ is the size of the chart, but charts of different sizes are available. The size of the color chart used in the test scene is 29 x 21 cm.
5. 18% gray background, which is used in the majority of image quality test scenes. The 18% gray color is used mostly for exposure measurement and

white balance settings. Due to the fact that it has only 18% reflectance over the visible spectrum, it does not interfere with the colors of the objects in the test scene.

7.3 Creation of the Test Scene

7.3.1 Test Environment Setup

Based on the experiment results presented in 6, and the test methodology in chapters 4 and 5, there are certain aspects that need to be considered when creating a scene for image quality testing:

1. Light conditions. The white balance in the Image and Signal processing Pipeline (ISP) is mostly influenced by the temperature of the light. Several algorithms have been developed that are aiming to provide color consistency over a wide range of light temperatures, like the one introduced in (Barnard et al [11]). In order to overcome the possibility of different ISPs generating different colors due to the difference in the light temperature, it has been decided to set the light temperature of the test scene to 5500 K. Most of the studio lights available on the market at the moment have this color temperature.
2. Camera setup. Another important aspect of the proposed test scene is the camera choice and settings. We propose the use of a DSLR camera fitted with a standard 18-55 mm lens. The reason is that generally DSLR cameras have a good sensor quality, they have high performance ISPs, and give the user the option of full manual settings. The following camera settings are recommended: f number is 16 in order for the picture to be sharp at any focus distance. Exposure to be 1/4 s because of the small aperture, and the ISO should be set to 200 in order to avoid noisy pictures in the case of lower quality sensors.
3. Scene setup. The scene should be set up in a way to challenge even the most advanced depth map generation methods. Some of these settings include the use of overlaid textureless objects, a number of similar objects placed one behind the other, half-transparent objects, a number of occluded objects, etc. The measurements of the object distance relative to the camera lens are provided for each object in the scene, and they can be found in section 7.3.2.

7.3 Creation of the Test Scene

These include the horizontal and vertical distances from the camera. These measurements provide all the information needed in order to re-create the proposed test scene. All the objects in the scene need to face the camera. In the case of the bottle, the writing needs to face the camera. In the case of the tennis balls, their orientation needs to be similar to the one in the proposed test scene. Care needs to be taken, so that the scene is set up on a perfectly flat surface.

For the image acquisition, the following camera settings were used:

- F number: 16
- Exposure time: 1/4 s
- ISO: 200
- Aspect ratio: 4:3
- Lens zoom setting: 35 mm

The intensity (illuminance) and temperature of the light were measured on each of the objects in the test scene, see figure 7.1. The corresponding values can be seen in tables 7.1 and 7.2.

Table 7.1: Light intensity values

Light Intensity Values [Ev]	
Bottle	8.5
Ball 1	8.3
Ball 2	8.3
Ball 3	8.3
Color Chart	8.6
ISO Chart	8.3

Table 7.2: Light temperature values

Light Temperature Values [K]	
Bottle	5350
Ball 1	5370
Ball 2	5210
Ball 3	5270
Color Chart	5290
ISO Chart	5280

7.3.2 The scene setup

The proposed layout of the scene can be seen in figure 7.1.

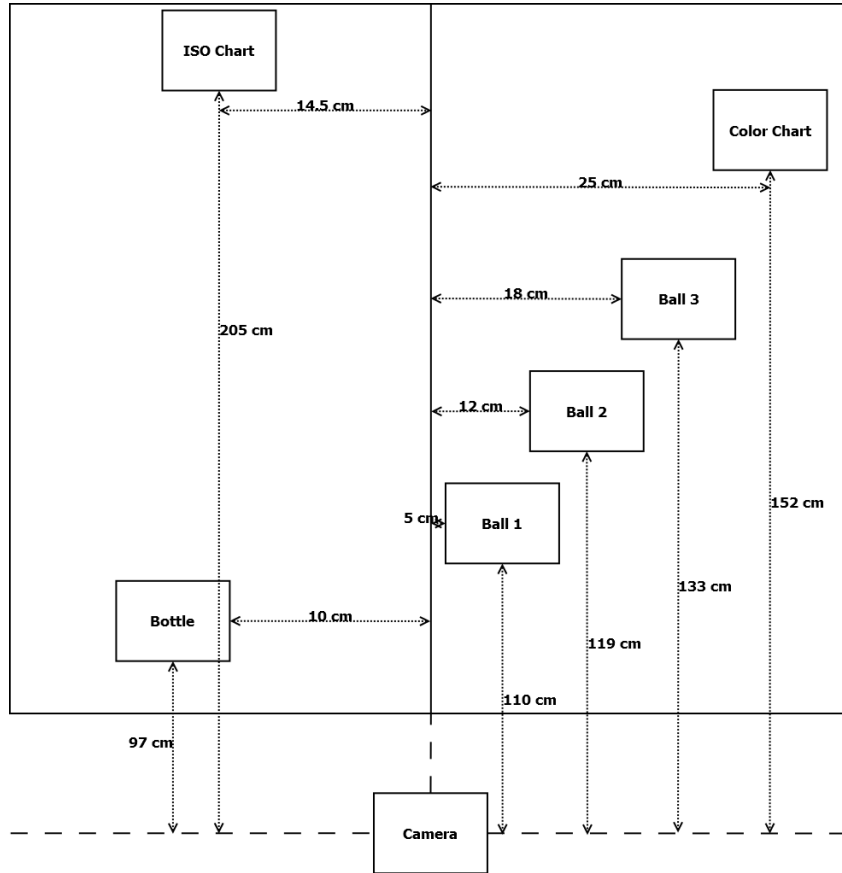


Figure 7.1: General scene layout

The values in the figure correspond to horizontal values, from a top view of the scene. All the distances were measured relative to the center of the camera lens. The camera rotation was set to 0 degrees on all x, y and z coordinates by using a high precision rig, which can be seen in figure 7.2. The rig allows the accurate setting of different stereo base values, and accurate camera-object distances. This rig was not available for the experiments described in chapter 6.

7.3 Creation of the Test Scene



Figure 7.2: High precision rig used to set accurate stereo base lengths.

The vertical positions of the objects relative to the center of the camera lens can be seen in table 7.3. The center of the camera lens is considered to be positioned at 0 cm.

Table 7.3: Vertical position of objects

Object	Positions [cm]
Bottle	-19
Ball 1	-19
Ball 2	-19
Ball 3	-19
Color Chart	-19
ISO Chart	-14.5

The proposed test scene can be seen in figure 7.3.

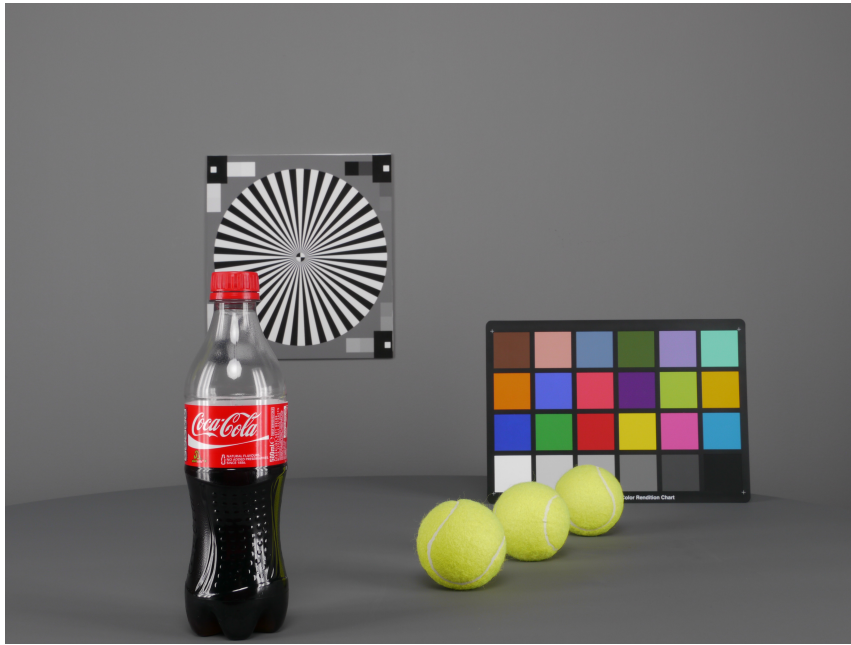


Figure 7.3: Proposed test scene

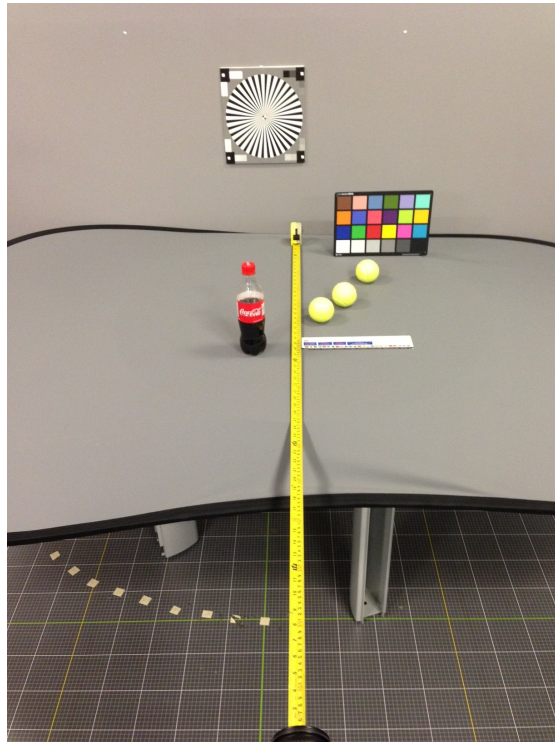


Figure 7.4: Setting up the scene

It was found that the easiest way to set up the objects in the scene is to use

7.4 Proposed Test Scenarios

two different rulers, see figure 7.4. The first one should be a tape measure which beside measuring the distance from the camera, also determines the center of the scene. The second one should be a simple geometrical ruler, which can be used to determine the distance from the center of the scene.

The reason why the half-empty bottle was chosen to be in the foreground is because half of the bottle has a distinctive color, but the other, empty half, blends in with the background. Both parts belong to the same object, but it would be difficult for the algorithm to recover the full shape.

In section 6.2.2.1, it was shown that by placing two faces into the test scene one next to the other, the performance of the depth map computation algorithm drops compared to the case where there is only one face in the test image.

For this reason, it was decided to replicate this issue in the general scene proposal, where three tennis balls were used, one behind the other, each of them having certain parts occluded. When looking from the front, due to the reason that they have the same color, they look like belonging to the same object. The algorithm should be able to differentiate between the different depth levels.

The Macbeth color checker is a multi-colored object, partially occluded by two of the tennis balls. Since it has multiple colors, it should challenge algorithms which use segmentation as a refining method [18, 79, 72].

The ISO chart is placed on the same depth level with the background. Some algorithms can eliminate the background with the help of segmentation [18, 79, 72], but by using the chart on the same depth level, the role of the algorithm would be to determine if both the chart and the background are on the same level.

7.4 Proposed Test Scenarios

The aim of this section is to present the proposed test scenarios which will be found in the on-line database. Based on the experiment results presented in chapter 6, it was considered necessary for the acquisition of the test images to be performed at different camera-object distances, under different light intensity conditions and by using different stereo base values. During the acquisition of the test images, the test camera was mounted onto a tripod, and it was moved to different locations in order to acquire test images at different camera-object distances, and with different stereo base length settings.

7.4.1 Test Image Acquisition at different Camera-Object Distances

In section 6.2 it was shown that by increasing the the camera-object distance, the performance of the depth map generation algorithms can improve or it can deteriorate. For this reason, it was considered necessary to acquire a number of test images, by using different camera-object distance settings.

The camera-object distance settings can have certain limitations regarding the minimum distances and the maximum distances which can be used. These limitation can be categorized into two types: theoretical limitations and practical limitations.

The theoretical limitations are based on mathematical and geometrical calculations performed in chapter 5.3, which are based on the type of the camera and camera settings used. The camera specifications needed for the calculations are the sensor size, and the focal length of the lens. The camera settings used are the scale factor and the resolution size.

The theoretical limitations define the minimum and maximum camera-object distances which can be used for a certain stereo base length value. The minimum camera-object distance is the smallest distance where common information can still be found in the corresponding images in the stereo image pair. In this case, if the stereo base length is too big, and the camera-object distance is too small, some objects in the test scene won't be visible in one of the views. The maximum camera-object distance which can be used is the largest distance, where disparity can still be noticed between the corresponding images in the stereo image pair. The depth map generation algorithms are based on disparity measurement [40], and by increasing the camera-object distance, the disparity between the views drops. In the current case, as presented in section 5.3, the disparity is measured in pixel numbers. The maximum distance is considered as being the one, where a disparity of 1 pixel can still be identified between the corresponding stereo image pair. The maximum distance is influenced by the length of the stereo base and the resolution of the camera used for image acquisition.

The practical limitations, beside the elements described in the case of the theoretical limitations, also include the equipment used for the test environment, which are the size of the studio lights, the intensity of the studio lights, the size of the objects in the test scene, the layout of the test scene, and the size of the room which is used for test image acquisition.

7.4 Proposed Test Scenarios

When acquiring test images, care needs to be taken to make sure that for each acquired image all the test objects are visible. This limits the minimum camera-object distance which can be used, because if the camera is placed too close to the test scene, some larger objects might not be visible. Another issue which might arise is that when acquiring test images, no additional objects, such as studio lights or tables, should be visible. For this reason, the maximum camera-object distance which can be used is limited.

In order to acquire the test images which can be found in the database, a Nikon D50 DSLR camera was used, which has a 23.7 x 15.6 mm CCD sensor. Based on the sensor size and lens focal length specifications presented in section 7.3.1, the theoretical limitations of the test scene were calculated using the calculations introduced in section 5.3. Based on these calculations, the minimum camera-object distance which can be used for a stereo base value of 100 mm is 27 cm. If the distance goes lower than this, there is no common information in the two separate views. Based on the same calculations, the maximum camera-object distance which can be used for a stereo base value of 25 mm is 1800 cm. Above this distance, the disparity is similar, and this makes the job of basic depth map generation algorithms difficult.

Considering the values imposed by the theoretical limitations, the acquisition of the images was done based on the practical limitations of the test scenario, which set the minimum camera-object distance to 97 cm, and the maximum distance to 207 cm.

Based on the limitations presented in this section, it was decided to acquire a number of test images at the distances from table 7.4. The distance between two test locations is 10 cm, and it is the same distance which was used in chapter 6.2.1.

The maximum distance had to be limited to 207 cm from the “Bottle” object in the scene due to the limitation presented in the previous paragraphs. By using a larger distance camera-object distance, parts of the equipment used for the scene setup were visible, as it can be seen in figure 7.5.

A number of test images acquired at different camera-object distances can be seen in figure 7.6. A full set of test images acquired at different camera-object distances can be found in Appendix E.1.

Table 7.4: Distance table for general scene

Distance location	Distance [cm]
1	97
2	107
3	117
4	127
5	137
6	147
7	157
8	167
9	177
10	187
11	197
12	207

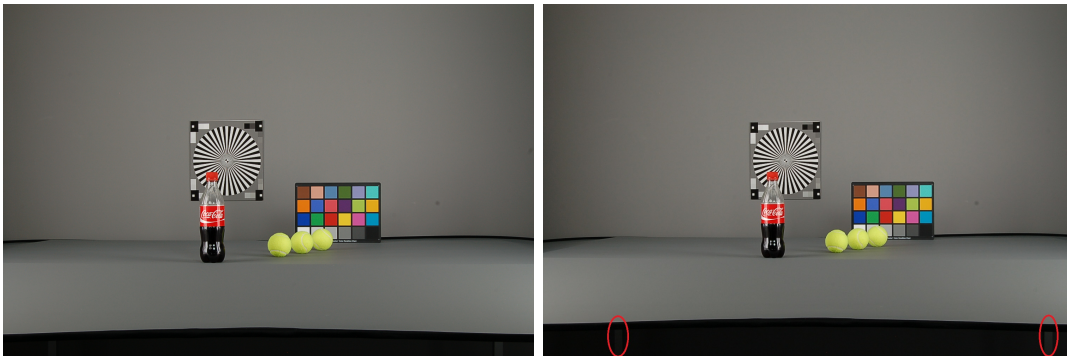


Figure 7.5: Physical limitation example. In the right image, parts of the table can be noticed.

7.4 Proposed Test Scenarios

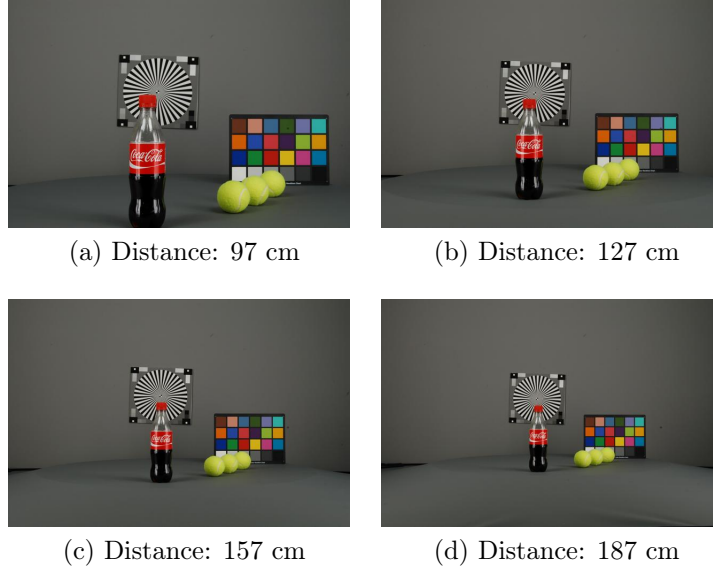


Figure 7.6: Different camera-object distances

7.4.2 Test Image Acquisition under Different Light Intensity Conditions

One of the purposes of the test scenes introduced in chapter 4 was to decide if a change in the intensity of the lights influences the performance of the depth map generation algorithms. It was shown in chapter 6, that a decrease in the intensity of the light does influence the performance of the algorithms. For this reason it was decided, that a set of images need to be acquired by using different light intensity values.

The limitations in the case of the light intensity setup are the equipment used, and the scenario which is proposed. The equipment influences the maximum intensity of the light, and the amount by which the intensity can be reduced at each new acquisition. If the change in the light intensity is used in the same test scenario with the different camera-object distances, care needs to be taken for the equipment not to be visible in the acquired image.

Based on the the information presented in this section, and the tests carried out in chapter 4, it was decided to use 3 different light intensity values for the proposed test scenario. During the acquisition of the test images for each camera-object distance and each stereo base length, the intensity of the light was changed. For each test case, the intensity of the light was measured on each object independently, and these

values can be seen in table 7.5. The light temperature was kept the same for each test setup, and the temperature values for each object can be seen in table 7.2.

Table 7.5: Different light intensity setups

Light Intensity Values [EV]			
	Light 1	Light 3	Light 3
Bottle	8.2	8.0	7.4
Ball 1	8.3	8.2	7.6
Ball 2	8.3	8.0	7.5
Ball 3	8.3	8.0	7.4
Color Chart	8.5	8.1	7.6
ISO Chart	8.2	7.9	7.1

For the measurement of the light intensity, a **Sekonic L-758D** light meter was used. For the measurement of the light temperature, a **Sekonic C-500** color meter was used. As studio lights, two different devices were used. The first one was a pair of **GTI GLE-M4/32/CBA** light walls, placed close to the test scene. The second one was a pair of **Interfit Super Coolite 6** studio lights placed above the light walls. A number of test images acquired by using different light intensities can be seen in figure 7.7.

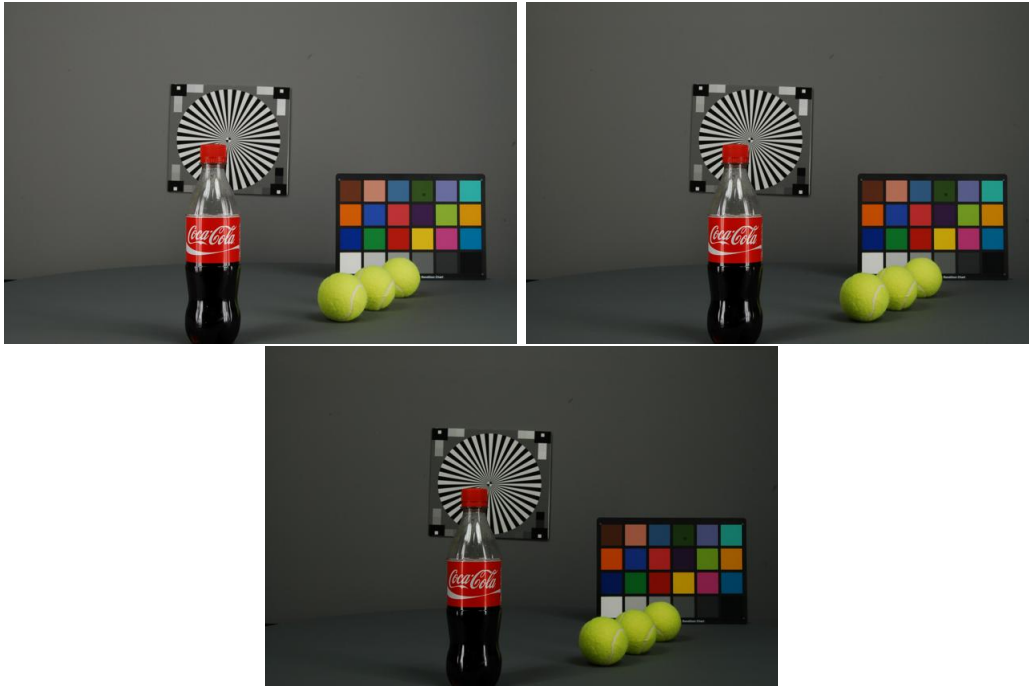


Figure 7.7: Different light intensities

7.4 Proposed Test Scenarios

Beside the intensity and temperature, the light also has a direction. From a direction point of view, the light can be split into two main groups, (i) directional and (ii) diffused light. In the case of the diffused light, the light rays travel in multiple angles from the light source. By doing this, the objects are evenly illuminated, and no shadows are created. An example of diffused light can be seen in figure 7.8a. Opposite to the diffused light, the light rays in the case of directional light travel in parallel. For this reason, the objects are not evenly illuminated, lighter and darker areas being visible on them. Beside un-evenly illuminated objects, in the case of directional light, shadows are also visible in the test scene. An example of directional light can be seen in figure 7.8b.



(a) Diffused light



(b) Directional light

Figure 7.8: Light direction



Figure 7.9: Light wall

During the experiments described in the current chapter, only diffused light sources were used in order to keep the test scene evenly illuminated. For this purpose, professional light walls were used, such as the one in figure 7.9.

7.4.3 Test image acquisition using different Stereo Base Length Values

In chapter 6.3.1, a set of results were presented, which showed a relationship between the performance of the tested depth map generation algorithms, and the length of the stereo base. For this reason, it was decided to acquire a set of stereo image pairs using stereo base lengths of different values.

Similarly with the scenario of different camera-object distances, there are certain limitation that need to be considered in this case as well. The limitations are the same as the ones described in section 7.4.1, and they are based on theoretical limitations which consider the sensor size, focal length, the scale factor and resolution size of the camera, and practical limitations which are based on the objects in the test scene, the equipment and the scene layout.

For these reasons, it was decided to acquire test images using 4 different stereo base lengths. The stereo base lengths used during the image acquisition are 25 mm, 50 mm, 75 mm and 100 mm. For each camera-object distance, 4 different stereo base length settings were used during the stereo image acquisition process. A set of images acquired using different stereo base values can be seen in figure 7.10.

7.4 Proposed Test Scenarios



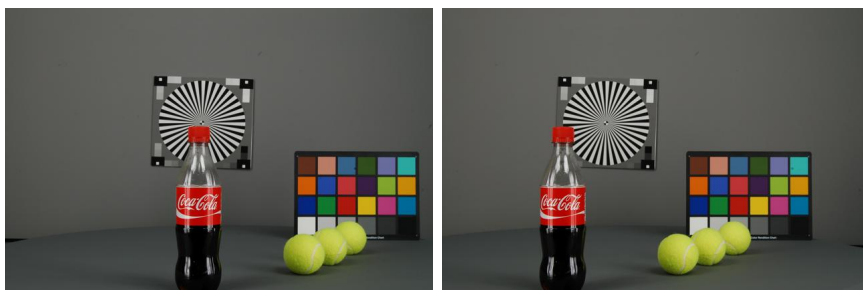
(a) Stereo Base: 25 mm



(b) Stereo Base: 50 mm



(c) Stereo Base: 75 mm



(d) Stereo Base: 100 mm

Figure 7.10: Different stereo base settings

7.5 Experiment Results

Based on the proposed test scenarios in section 7.4, several experiments were carried out using the acquired images. All the images that were used for the experiments presented in the current section, together with the instructions needed for the scene setup, can be found in the proposed database: <http://www.andorko.com/stereo.html>. Images acquired during the setup of the test scene can be found in Appendix E.2.

During the experiments, disparity maps were generated using the algorithms introduced in section 4.2. For the measurement of the algorithm performance, the methods described in section 4.3 were used. A number of resulting disparity maps can be seen in figure 7.11.

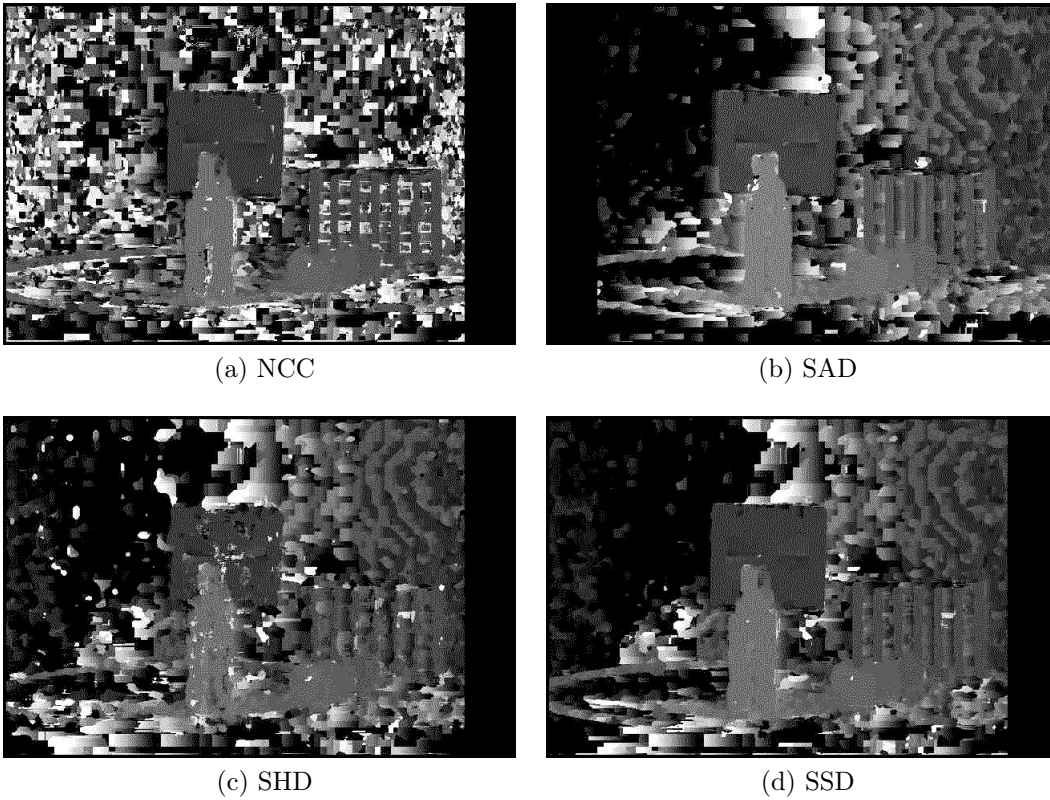


Figure 7.11: Example disparity maps

Three different experiments were carried out using the proposed general depth and disparity map testing scene. The first experiment compares the performance of the SAD-based algorithm, when used with images of different stereo base values. The same experiment was carried out in section 6.3.1 as well.

7.5 Experiment Results

Figure 7.12 presents the performance of the SAD-based algorithm at different camera-object distances, and using different stereo base values. The measurements in the figure correspond to the “Bottle” in the test scene.

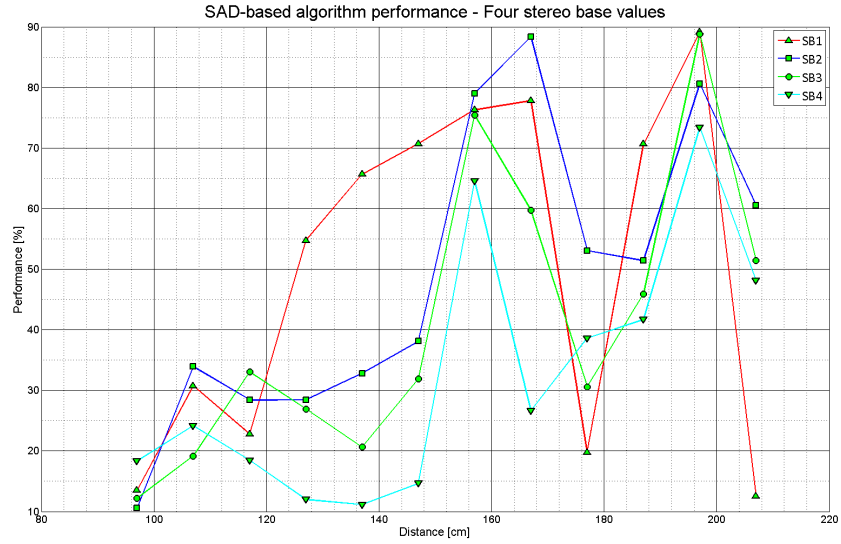


Figure 7.12: Algorithm performance using different stereo base values

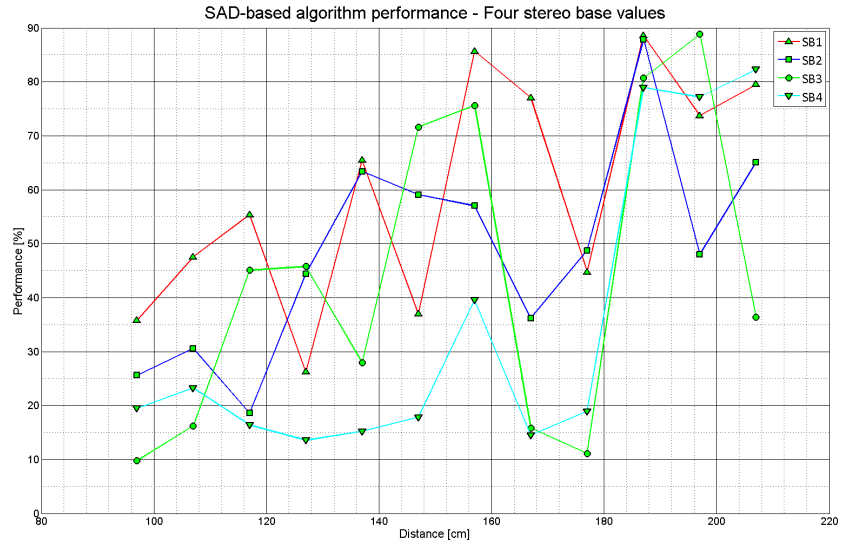


Figure 7.13: Algorithm performance using different stereo base values, experiment 2

By analyzing the figure, several things were noticed. This first one is, that the

highest overall performance is achieved in the case of the stereo images acquired using a stereo base value of 25 mm. The second observation is, that two camera-object distances are noticed where best algorithms performance is recorded. These distances are 167 and 197 cm. Also, by looking at the results, it can be noticed that the performance results tend to be higher between the camera-object distances of 140 cm and 200 cm. In order to double-check these results, it was decided to acquire a new set of images and perform the same experiments. The results of the follow-up experiment can be seen in figure 7.13.

The results of the follow-up experiment tend to confirm the fact, that (i) the highest overall performance is achieved for a stereo base value of 25 mm, and (ii) the performance results tend to be higher for camera-object distances between 140 cm and 200 cm.

The aim of the second experiment carried out with the help of the proposed test scene, was to measure the influence of the light intensity on the algorithm performance. Figure 7.14 presents the performance of the SAD-based algorithm, at different camera-object distances, using a fixed stereo base value of 25 mm, with 3 different light intensity settings. The measurements in the figure correspond to the “Bottle” in the proposed test scene.

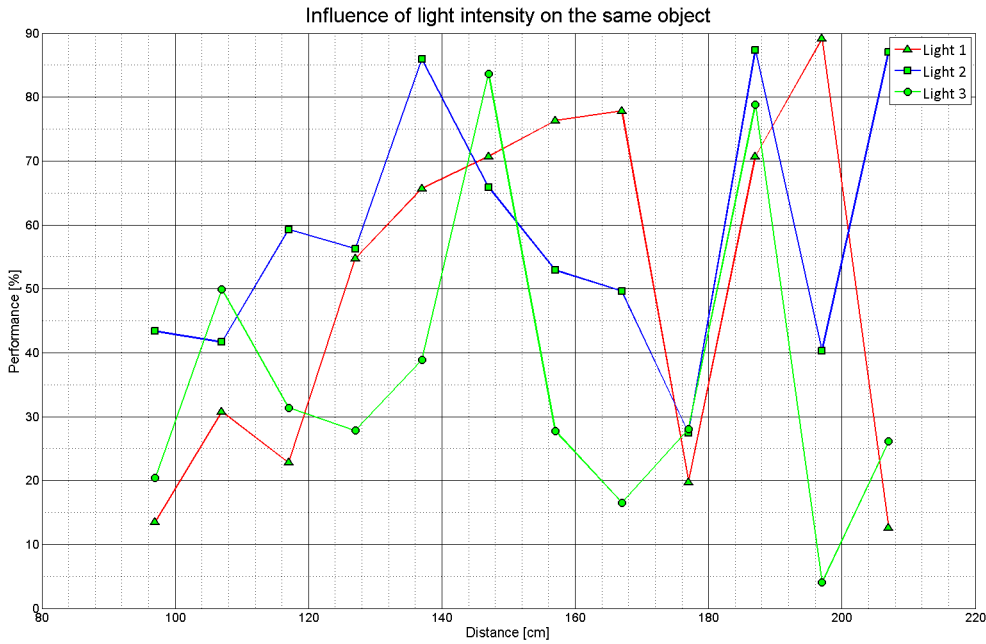


Figure 7.14: Different light intensity values

7.5 Experiment Results

Similarly with the experiments presented in section 6.2.1, it can be noticed from figure 7.14, that the intensity of the light does influence the performance of the algorithm. More specifically, by decreasing the intensity of the light, the performance of the algorithm is influenced in a negative way.

In order to confirm these results, the newly acquired set of images used for the previous experiment were used in this case as well. The results of the follow-up experiment can be seen in figure 7.15.

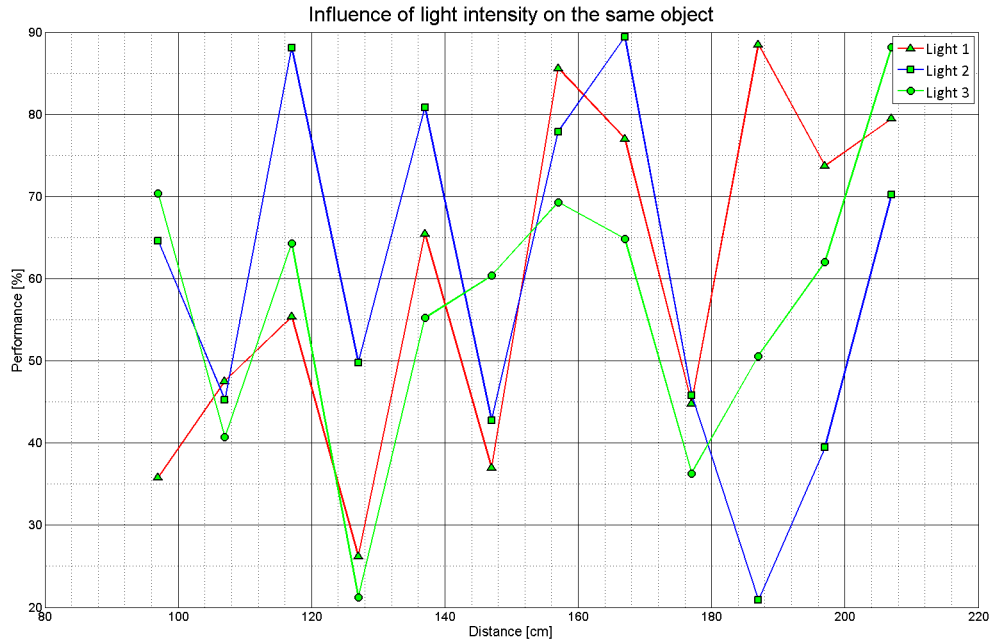


Figure 7.15: Different light intensity values, experiment 2

From the follow-up experiment it can be seen that the intensity of the light does influence the performance of the algorithm. While for the first two intensities the performance values are similar, for the third light intensity, the performance of the algorithm drops significantly.

The results of the third experiment carried out using the proposed test scene can be seen in figure 7.16. The aim of the experiment was to analyze in what way do similar objects in the same scene influence the performance results of each other. As it was mentioned earlier, the performance is measured for each object in the test scene. The aim here is to see if the depth quality of the object in the background is influenced by the object in the foreground. The same experiment was carried out

in section 6.2.2 as well, by using the FACES test scene, section 4.5.3, where it was noticed, that having two different faces in the test scene influences the performance results in a negative way.

Figure 7.16 presents the depth quality of 3 different objects (Ball 1, Ball 2 and Ball 3), at different camera-object distances, by using a fixed stereo base value of 25 mm, and a fixed light intensity.

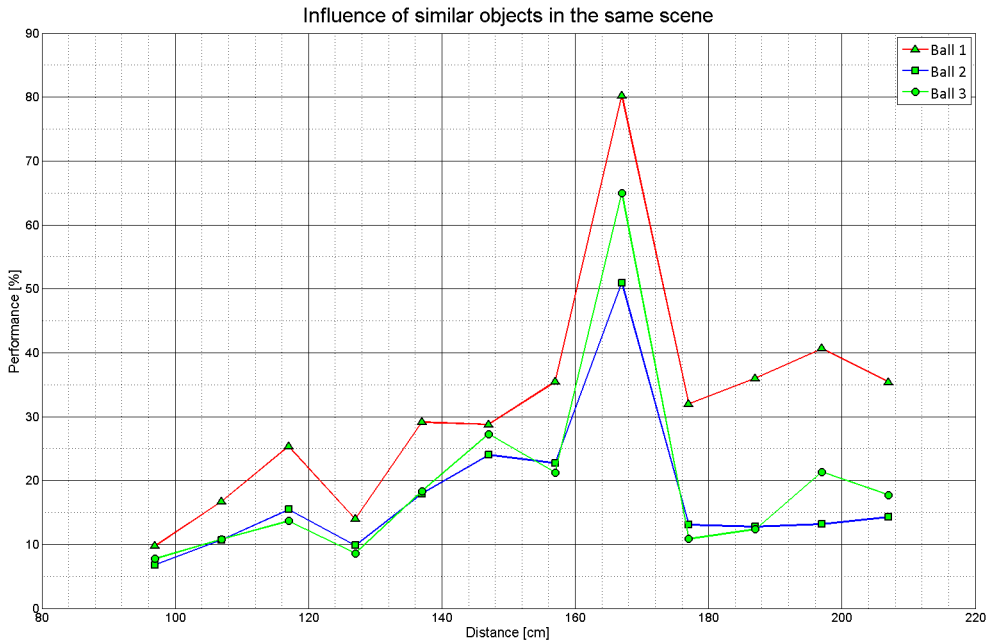


Figure 7.16: Influence of similar objects in a test scene

It can be noticed from figure 7.16 that the highest depth quality is achieved by Ball 1, which is closest to the camera. The overall depth quality of balls 2 and 3 seems to be similar. The conclusions that can be drawn are that having similar objects in a test scene influences the performance of the algorithm in a negative way.

In order to double-check these result, and to confirm the conclusions, the newly acquired set of images was used. The results of the follow-up experiment can be seen in figure 7.17.

7.6 Ground Truth Generation

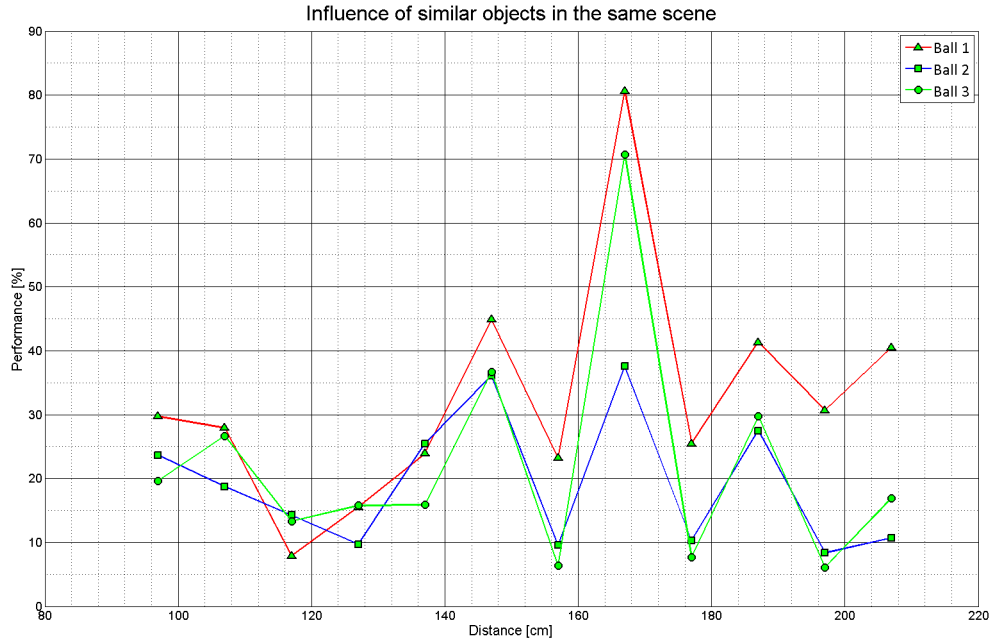


Figure 7.17: Influence of similar objects in a test scene, experiment 2

By analyzing the follow-up experiment, it can be noticed that the depth quality of Ball 1 is higher than the quality of Ball 2 and Ball 3 for 83% of the data points. It can also be noticed that by comparing the experiments performed on two different sets of images, the highest algorithm performances are achieved at the same camera-object distance of 167 cm.

7.6 Ground Truth Generation

In the previous sections of the current chapter, the design of the test scene and test scenarios was presented. The aim of the current section is to provide information about the methods which can be used to measure the performance of the depth map generation algorithms.

As defined in chapter 4.3, one way of measuring the performance of a depth map generation algorithm is by comparing the resulting depth map with a ground truth image. The method used in the current research uses ground truth images, and is presented in chapter 4.3. For this reason, it was decided to provide ground truth images for all the proposed stereo image pairs. The ground truth images are created

manually, and for this reason, their accuracy is not at a pixel-level. In order to have pixel-level ground truth images, the test images need to be created artificially. Figure 7.18 shows a ground truth example of one of the proposed stereo image pairs.

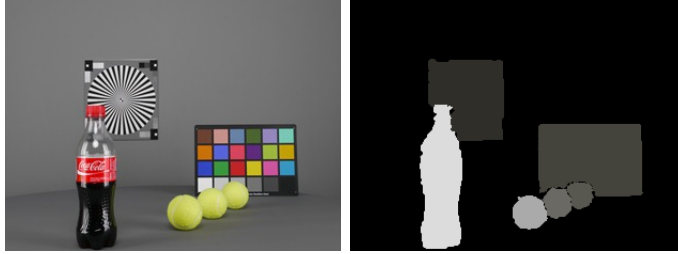


Figure 7.18: Stereo image pair and corresponding ground truth

Other testing methods have recently also become available, [53, 37]. The authors of these papers present novel techniques for the measurement of the depth map quality, when these depth maps are used to help rendering of views in 3D video. These methods do not require ground truth images in order to determine the performance of a depth map generation algorithm.

These methods have not been widely adopted yet, but they are a solution of the problem which lies in the fact, that for real-life scenes, accurate depth maps cannot be created, which can be an issue for future research projects.

7.7 Explaining the Data Variations

As noticed in the current chapter, as well as in chapter 6, there are variations in depth quality between consecutive camera-object distances or different stereo base values which cannot be explained by using an empirical method.

The performance of the algorithms can be influenced by a number of elements. These elements were discussed throughout the thesis, but it is considered necessary for these to be mentioned once more in the current section.

7.7.1 Data Variations due to Image Registration

The measured performance of the algorithm can be influenced by the generation of the ground truth. As described in section 4.4, the ground truth is created manually for each test image setup. If the ground truth does not have the exact shape as the object in the scene, the performance won't be exact either. Creating a perfect

7.7 Explaining the Data Variations

ground truth image manually is an almost impossible task, as it was discussed in section 4.4.1. The measurement error introduced by the ground truth is influenced by the size of the object in the test scene, and the amount of inaccuracy between the ground truth and the scene. In order to provide an estimate of the error introduced by an inaccurate ground truth, let us consider an object which has the size of 38400 pixels. This size is $1/8$ of the total size of a VGA (640 x 480) image. Let us assume that by marking the ground truth of this object manually, we will have more than 1000 inaccurate pixels. In this case, 1000 inaccurate pixels will introduce a measurement error of approximately 2.6%.

Another performance measurement error that might occur is the one caused by the acquisition devices. As presented in section 5.4.3, acquisition devices can introduce noise into the image, and this noise can influence the measured performance of the algorithm. The effect of the noise on the performance is negligible, but when adding it to the errors introduced by other factors, it adds up. Based on our measurements, the devices can introduce an error of up to 3%.

Section 5.4.2 introduced a discussion about the calibration of the stereo cameras. In this section, the performances of different depth map generation algorithms were compared by using un-calibrated and calibrated images. In the case of all algorithms, an improvement in performance was recorded in the case of calibrated images. The performance improvement ranges between 3% and 10%, where the quality of the depth maps generated using calibrated images is better. For the experiments described in the thesis, the acquired images were not calibrated. The calibration of stereo images was performed at a later stage during this research. The calibration was performed during the confirmation of the final test scene. Since most of the experiments had already been done using un-calibrated images, there wasn't enough time to repeat all the measurements due to time constraints and the additional burden of calibration for each case. Stereo cameras are calibrated during the manufacturing process. Additional calibration in real-time consumer electronic devices is not used at the moment due to processing speed issues.

Just as a reference, in order to acquire 180 images it takes approximately 3 hours. Let us consider the acquisition of a set of test images, where 4 different stereo base settings and 12 different camera-object distances are used. In this case, 60 images would need to be acquired. If we need to acquire 15 calibration images for each test image, it would give us a total of 900 calibration images. In order to acquire 60 test images + 900 calibration images, it would take 18 hours only for acquisition.

After this step, each calibration image needs to be marked manually. Even though calibration images were not acquired, their effect was taken into consideration when drawing conclusions.

7.7.2 Data variations due to light

As explained in section 6.3.2, the direction of the light can also influence the performance of the algorithm. This influence can mostly affect the results in the case of the scenes which are re-created in order to confirm certain results, such as in the case of the experiments presented in section 7.5. In section 6.3.2 it was shown that the direction of the light can influence the performance of the depth map generation algorithms up to the point, where the results are inconclusive. When setting up test scenes, even though exact specifications are given regarding the location of each object, the light sources used are not guaranteed to be in the same exact place every time. For this reason, the angle of the incident light can be different from the previous cases, and this might influence the final results.

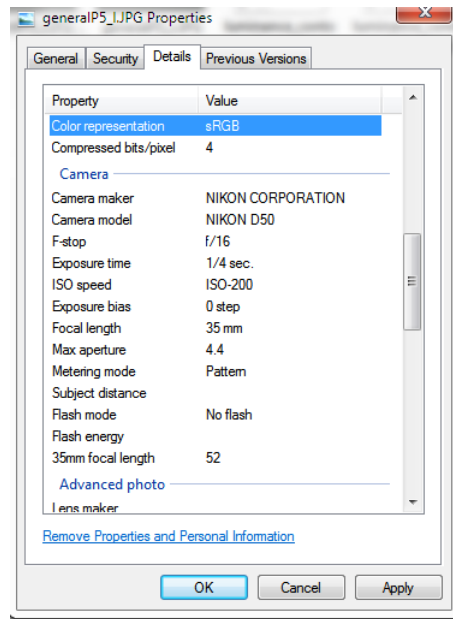


Figure 7.19: Camera settings

During the acquisition of the test images used for the experiments described in section 7.5, it was noticed that even though the camera settings were applied manually, in some cases there was a difference of light intensity between consecutively

7.7 Explaining the Data Variations

acquired images. For both images, all the camera settings and light intensity settings in the test room were kept the same. The settings of the camera can be double-checked by looking at the details of the acquired image, figure 7.19.

The difference in intensity can be noticed visually. In order to measure this change in intensity the Imatest image quality analysis application was used. With the help of this application, a uniformly illuminated area of the test image was selected and analyzed, figure 7.20.

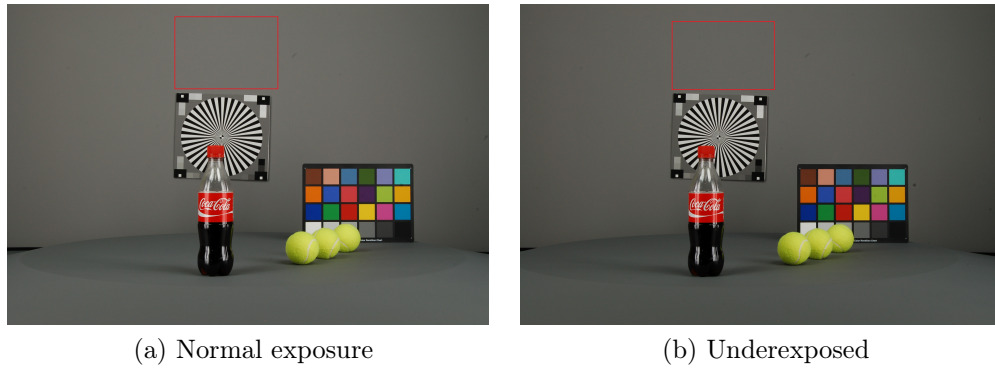


Figure 7.20: Light intensity difference

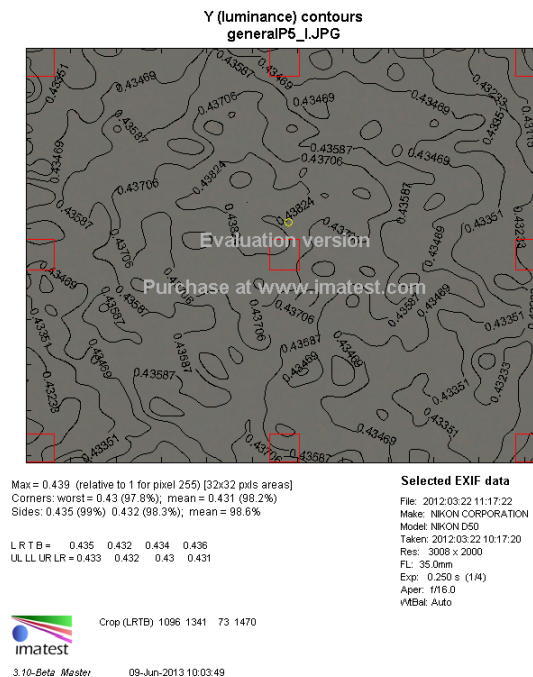


Figure 7.21: Imatest results - normal exposure

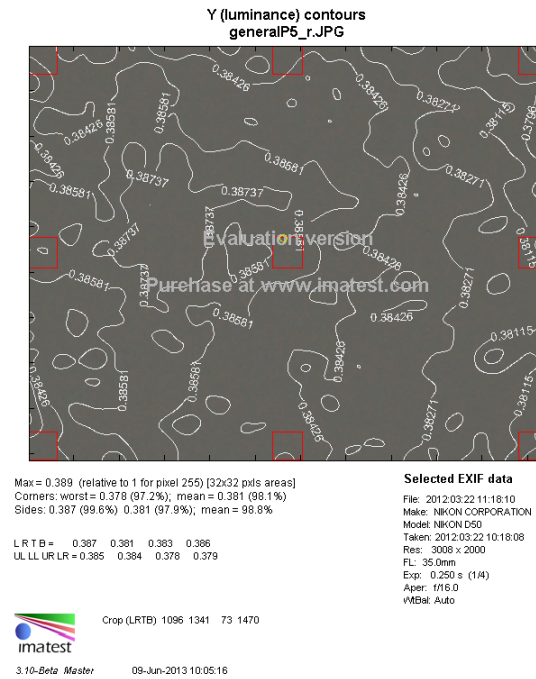


Figure 7.22: Imatest results - underexposed

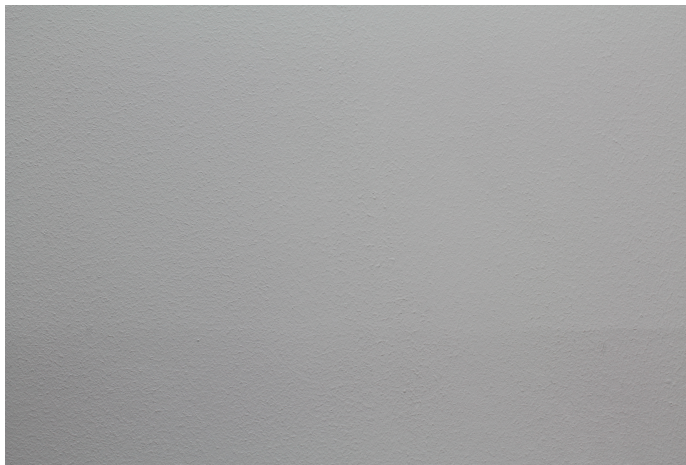


Figure 7.23: Uniformly illuminated scene

The analysis results can be seen in figures 7.21 and 7.22, where a difference in peak luminance value of 0.05 can be noticed. The values are calculated relative to 1 for a maximum pixel value of 255.

In order to determine the frequency at which differences in the light intensity occur during image acquisition, the DSLR camera used for image acquisition was put

7.8 Conclusions

to the test. 5 different images were acquired of a textureless and evenly illuminated scene, figure 7.23.

The images were analyzed using the Imatest image quality application previously mentioned, where the luminance values were measured. The results show, that even though the difference in intensity cannot be noticed visually, in the 5 different images, 3 different luminance values were recorded. Ideally there should have been a single luminance value for all 5 images. The results can be seen in Appendix F.1. In order to make sure that the device used for image acquisition was not faulty, a similar test was performed with a state-of-the-art DSLR camera manufactured by a different company. In this case, out of 5 different images, 4 different luminance values were recorded. Similarly with the previous case, ideally a single luminance value should have been recorded. The results can be seen in Appendix F.2.

The purpose of these test was to show that differences in scene illumination can appear even though all precautions are taken. These differences can interfere with the functionality of the algorithms, and cause data variations.

In the case of the test images, used for depth map generation, where a difference in illumination was noticed, the images were edited. During editing, the brightness of the under-exposed images was increased. In the case of the images where a difference in illumination was not noticed visually, the images were not edited.

7.8 Conclusions

A novel depth map testing scene was introduced in this chapter, the purpose of which is to serve as a mean of comparison between methods which (i) simply rely on image processing algorithms and (ii) beside the image processing algorithms also rely on special types of camera, like ToF and IR.

For this purpose, a number of objects were selected to be part of the test scene. The main ideas behind this selection process were (i) for these objects to be universally available and (ii) for them to have special characteristics which make them less detectable by the depth map generation algorithms.

The selected objects were placed in a test scene, for which all the geometrical measurements, such as the distance from the camera, the vertical position, and light measurements such as the intensity and temperature of the light, are provided. Based on these measurements, any researcher interested in the re-production of this scene,

will be able to do so.

After the details regarding the selection of the objects and creation of the scene, different test scenarios were also proposed. The proposal of the test scenarios relies on the experiments detailed in chapter 6, and includes scenarios for (i) different camera-object distances, (ii) different light intensity conditions and (iii) different stereo base length values. For each of the proposed test scenario, certain experiments were performed, where the conclusions are similar to the ones drawn in chapter 6.

The chapter concludes with a discussion regarding the possible reasons why data variations are noticed during the experiments, and the proposal of ground truth generation for the general test scene.

Chapter 8

Conclusions and Future Research

In this concluding chapter the outcome of this research will be reviewed by analyzing the importance of the proposed solutions, and how they influence the future research.

In section 8.1, the work is summarized, with emphasis on the more important outcomes of the research.

The main contributions, together with a summary of all the research achievements are presented in section 8.2.

The chapter concludes with the summary of the papers published in relation to this research, in section 8.3, and a discussion about our future work, in section 8.4.

8.1 Research Summary

The work published in this thesis is the result of research performed in an area of high interest in computer vision and image processing: depth and disparity map generation. The topic of depth map generation is and has been a highly researched one for the past decades, especially due to the novelties that it can introduce to consumer cameras. Even though, one might think that it's quite difficult to introduce serious contributions to this area, I consider that my contribution is significant for the purpose of depth map quality measurement and testing methodology. I state this, because the testing methodology introduced in the current thesis defines test scenarios for (i) different camera-object distances and (ii) different light conditions. The test scene proposed in the thesis, based on our literature review, is the only one of its kind, which allows comparison of all the depth map generation algorithms, not only the software-based ones.

In order to achieve the contributions described in the previous paragraph, the current research went through several stages, each stage having its own goals. The first stage of the research was the development of a stereo ISP, presented in chapter 2, which was the starting point of the work on depth and disparity map generation. The development of the stereo ISP was also one of my first achievements, together with a demo application of a 3D camera.

After the commencement of the work on disparity map generation, the first goal was the generation of the preliminary test scenarios and test scenes, presented in chapter 4. In this chapter, several disparity map generation algorithms were selected for preliminary testing purposes, where the selection was made based on the literature review from chapter 3. Four different test scenes were created, each one having its own purpose in the final goal of the research.

Chapter 5 dealt with the selection of the image acquisition device, and the preliminary measurements required for the validation of the tests. The major achievements in this chapter were (i) the development of a stereo image acquisition device implemented on an FPGA, (ii) the selection of a downsampling algorithm, and (iii) the measurement of the parallax in order to define the limitations of the image acquisition device.

After the presentation of the experiments in chapter 6, the most important achievements are presented in chapter 7, where the well defined test scenarios and the general depth map testing scene are detailed. Beside the presentation of the test scenarios and test scene, practical advices are also given about their implementation.

8.2 Review of the Research Achievements

In this section I will provide more details about the evolution of the research, and the work of the past 4 years, which led to the achievements presented in the current chapter.

The main goal of the research was the development of a general testbed and test scene suitable for the quality measurement of depth maps used in consumer electronics. The research had several stages, each stage having its own set of goals.

The first stage of the research was the design and implementation of a stereo Image Signal Processor (ISP), described in chapter 2. In order to design such a system, the knowledge of basic concepts was important, and for this reason, chapter

8.2 Review of the Research Achievements

2 starts with the presentation of the basic stereo concepts. The next step in the design of the stereo ISP was a review of work previously published in this area, which was useful for several reasons: (i) different modules of a standard ISP were identified, (ii) architectures of standard ISP were reviewed, and (iii) previously published work about stereo ISPs was reviewed in order to prove the feasibility of the system.

The major achievements of the current chapter are the design and implementation of the stereo ISP, and the use of this ISP for the development of a 3D image acquisition device, which uses two different display techniques.

During the development of the system mentioned at the previous paragraph several challenges were identified while working with stereo images. These challenges are presented at the end of the chapter, and they were considered during the development of the proposed test bed.

After the initial development of the stereo ISP and the identification of possible problems and challenges, the next stage of the research was the actual work towards the final goal. For this reason, chapter 3 presents work previously published in the area of depth map generation. Depth map generation methods can be divided into two main categories: (i) depth maps generated from single source images, and (ii) depth maps generated from stereo images. The purpose of this literature review was to gather knowledge about the algorithms that are currently used for depth map generation purposes, knowledge which was used for the development of the test scenes and scenarios.

The description of the work towards the main goal of the research spans over 3 chapters, where research regarding (i) algorithms to be used for preliminary testing, (ii) the development of the initial test scenes and methodology, (iii) the choice of image acquisition devices and (iv) the initial experiments are presented.

Chapter 4 starts with the selection of the disparity map generation algorithms which were proposed for testing. Even though the purpose of the research was not the comparison of different depth and disparity map generation algorithms, several tests were required in order to make sure that the proposed test scenes and scenarios were suitable for our purpose. The algorithms were selected based on a criteria detailed in chapter 3. The selected algorithms were based on the Sum of Absolute Differences, Sum of Square Differences, Normalized Cross-Correlation and Sum of Hamming Distances pixel matching costs. The algorithms were selected on the basis, that they should not be the best performing ones, but instead they should be used in a large number of currently available methods.

After the selection of the algorithms, the following sections in the chapter describe the quality and performance measurement measurement and ground truth generation techniques proposed. These techniques are different from the ones used by other researchers, the reason for this being, that they were adjusted for our needs. The initial test scenes are presented at the end of the chapter. Experiments carried out using these test scenes contributed towards the development of the general test scene proposed in chapter 7. This chapter describes the importance of having accurate ground truths when measuring image quality. Small inaccuracies in the ground truth can cause significant measurement errors, as described in section 7.7.

Chapter 5 presents several research achievements, such as (i) the choice of stereo image acquisition device, (ii) the selection of a downsampling method, (iii) the measurement of the parallax in order to determine the limitations of the image acquisition device, and (iv) details about the camera calibration.

In order to acquire stereo image pairs for testing purposes, several devices were proposed. The initial device used was the FujiFilm W3 3D camera, which has 2 built-in camera modules. The images with a fixed stereo baseline which were used for the initial experiments were acquired using this device. In order to acquire stereo images with a variable stereo baseline, a custom device was built, which used a uCam development board and 2 different camera modules. During the tests carried out using this device, it was noticed that the quality of the acquired images was not satisfactory, and for this reason, a single camera mounted on a horizontal rig was used instead.

When processing large resolution images, it takes a lot of time for the results to be generated. For this reason, it was decided to use VGA (640 x 480) images for all the tests. The smallest image acquisition resolution setting on the camera was 2048 x 1536, and because of this, a downsampling algorithm had to be used. After several tests, it was decided that the most suitable algorithm is the Lanczos downsampling algorithm. The choice was made based on the quality of the image after the DS operation was performed.

The limitations of the stereo image acquisition device were determined by measuring the parallax for different stereo baseline and camera-object distances. During these measurements it was determined, that in the case of the FujiFilm W3 3D camera, for stereo baseline values of 10 and 12.5 cm, the overlapping of the pixels starts at a camera-object distance of 12 cm, respectively 15 cm. For distances closer than that, there is no common information in the two acquired image frames. This

8.2 Review of the Research Achievements

information was helpful during the development of the test scenes, and test scenarios.

The same chapter also describes the fact that by calibrating the stereo cameras, higher quality depth maps can be generated. The calibrations process is not a simple one though, and it can only be performed for certain stereo image acquisition devices.

The experiments carried out with depth map generation algorithms are presented in chapter 6. Several research achievements can be mentioned in this case, such as (i) the influence of light intensity on the performance of the depth map generation algorithm, (ii) the behavior of depth map generation algorithms at different camera-object distances, (iii) the comparison of algorithms in the case of stereo images with different stereo baselines, and (iv) the influence of light direction on the performance of the algorithms.

During the experiments performed in order to determine whether the intensity of the light influences the performance of the algorithms, it was noticed that with the decrease of the light intensity, the performance of the algorithm decreases as well. There was nothing new in this case, because the outcome in this case was predictable. The results of the comparison between algorithms at different camera-object distances were not predictable though. In the case of the AWARD test scene, it was noticed that the highest performance is achieved at a camera-object distance of 150 cm. This was in contrast with the experiments in the case of the FACE and FACES test scenes, where the highest performance values were achieved at a camera-object distance of only 50 cm. This unexplained discovery led us to the development of several other experiments, such as the acquisition of stereo images with variable stereo base sizes.

During the experiments with images using different stereo base sizes, the influence of the light direction on the performance of the algorithm was noticed. Chapter 6 concludes with a discussion regarding this important topic.

The results and findings of all the tests and experiments carried out in chapters 4 - 6 are used in chapter 7, where the newly developed test bed and proposed test scene are presented, together with factors and sources of error that can have significant effect on disparity measurement. The achievements described in this chapter were the main goals of the research presented in this thesis.

In this chapter, a general depth map testing scene is introduced, which contains 6 objects, 3 of which are the same. The objects were chosen based on their purchase availability, and shape and color characteristics.

Three different test scenarios were proposed (i) stereo images acquired at different

camera-object distances, (ii) stereo images acquired using different light intensity settings, and (iii) stereo images acquired using different stereo base values. Several experiments, similar to the ones presented in chapter 6 were carried out, and similar results were noticed.

This chapter also explains the factors which can cause data variation when generating depth map, or measuring algorithm performance. These variations can happen due to image registration issues, or light condition variation.

All the stereo images and depth ground truth images used in this chapter can be downloaded from our online database: <http://www.andorko.com/stereo.html>.

8.3 Summary of Published Papers

We published several papers at the beginning of my research in connection with the implementation of the stereo ISP on an FPGA and its applications.

The first paper was presented at the IEEE Consumer Electronics Society's First Games Innovation Conference which was held in London in 2009 [3]. The paper described the implementation of the stereo ISP, and its proposed applications in computer gaming which included the generation of real-time 3D avatars, and the employment of face data for game UIs. At the same conference, I co-authored a paper by (Ionita et al [58]), where we presented the applicability of stereo images for the generation of 3D face models based on AAMs, which are useful for real-time face models in computer gaming.

At the OPTIM conference which was held in Brasov, Romania in 2010, we presented the implementation of a 3D video acquisition system [4], which was based on the same stereo ISP described in the paper at the Games Innovation Conference.

The tools and techniques that we used for the implementation of the stereo ISP were first presented at the RoEduNet conference held in Sibiu, Romania in 2010, and were later published in the UBICC Journal [6]. The purpose of this paper was to present each step of the implementation process, together with the tools that were used, so that other researchers can duplicate our experiments.

After this stage, I started doing research in the depth map generation area, and I published a paper at the IEEE Consumer Electronic Society's Second Games Innovation Conference held in Hong Kong in 2010 [2]. For this paper I received the Best Student Paper Award. The paper presents a system, where depth map

8.4 Future Research

information is used in order to separate different depth layers from images used for online gaming. This separation has two main advantages. The first one is, that by only selecting a certain part from the image, in the case of on-line games which send webcam information to other parties, the bandwidth occupied by the image can be lowered. The second advantage is the implementation of privacy in case the user does not want to share the background of his environment.

The advantages of a stereo ISP camera were presented at the International Conference on Consumer Electronics in 2011 [5]. In this papers, the possible applications of a stereo ISP camera were detailed.

A paper was written about the general test scene for depth maps described in chapter 7, which was presented at the International Conference on Consumer Electronics in 2013 [7], and an extended version was published in the IEEE Transaction on Consumer Electronics [8].

A third journal paper was submitted to the IET Image Processing Journal, where details about the generation of the test scenes in chapter 4, and the results in chapter 6 are presented. We received the comments from the reviewers, and at the moment we are working on the modification of the paper.

8.4 Future Research

In section 7.7.2 a difference in light intensity was noticed in the case of consecutive images acquired with a state-of-the-art DSLR camera. It would be interesting to determine the exact reasons which cause this difference in illumination. The light source has a visually invisible flicker which alternates the intensity of the light at a frequency of 50 Hz. Usually, cameras have flicker removal algorithms. If the flicker removal algorithm could be disabled, we could notice if the difference in intensity is due to this algorithm, or due to other reasons. In the case of the light it's not only the difference in intensity which influences the performance of the algorithms, but also its direction and reflection coefficient of the objects in the scene. A study should be conducted in order to determine how significant these influences are.

In section 7.7.1 it was mentioned that one of the possible reasons for measurement errors is the lack of calibration in the case of stereo image acquisition cameras. There are several calibration methods available, which provide good results [112], but the aim in the case of devices capable of acquiring stereo images would be to come up with

a self-calibration algorithm, which wouldn't require pre-defined objects in a scene. It should detect random objects which can be found in both views, and based on their location, perform a device calibration. For the extrinsic parameter calibration of the cameras, gyroscope and accelerometer data could also be used, since gyroscopes and accelerometers can be found in a large number of devices nowadays.

The ultimate goal in the case of my future research would be the integration of the developed test bed into the sponsoring company's testing process. This would probably be done once more devices will be capable of acquiring stereo images.

8.5 Concluding Remarks

As mentioned in previous sections, the research presented in the current dissertation progressed through several stages. At the initial stage, when working on the acquisition of stereo images, everything looked much easier than it turned out to be after several years of work.

After the acquisition of stereo images, in order to be able to process the information from the images, the stereo image pair need to be registered. This sounded easy at the beginning, but we realized the complexity of this method, when beside the vertical alignment of the images, the rotation also had to be corrected. Beside the geometrical calibration of the images, we also need to make sure that the information in the images is similar. This is not a trivial task either, due to the fact that different light sources, image acquisition devices can behave in different ways, and they don't always acquire the same information, as it was described several times in the thesis.

When designing a test scene, every aspect of the scene needs to be taken into consideration, from the position of the objects, their colour, distance in relation to the camera, light intensity in the room, light temperature in the room, etc. By designing these test scenes and related scenarios for the experiments, I realized how important it is to be meticulous when doing research.

The end-point of a research can be different from that originally envisaged. This was proven in my case as well. The original direction of my research was hardware development-based, and the aim was the development of a depth map generation engine. In the following years, after reading different research papers, and identifying different issues, the research has become image quality-oriented. This is a good thing, because research needs to fill the gaps, and not re-invent the wheel.

Appendix A

Test Images and resulting Depth Maps

A.1 Test images acquired for the AWARD test scene



Figure A.1: AWARD, light 1, distance 1



Figure A.2: AWARD, light 2, distance 1



Figure A.3: AWARD, light 3, distance 1



Figure A.4: AWARD, light 4, distance 1

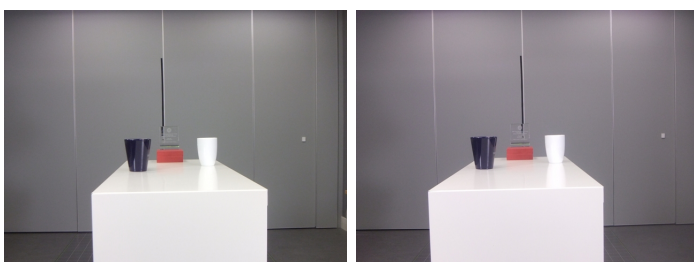


Figure A.5: AWARD, light 1, distance 2

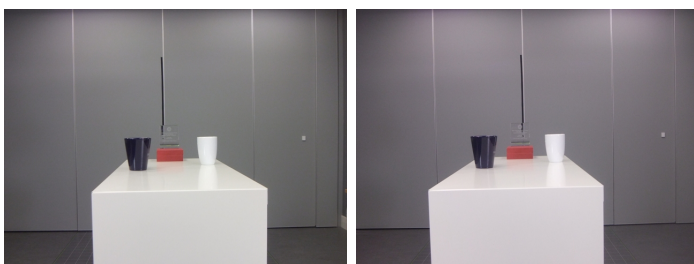


Figure A.6: AWARD, light 2, distance 2

A.1 Test images acquired for the AWARD test scene

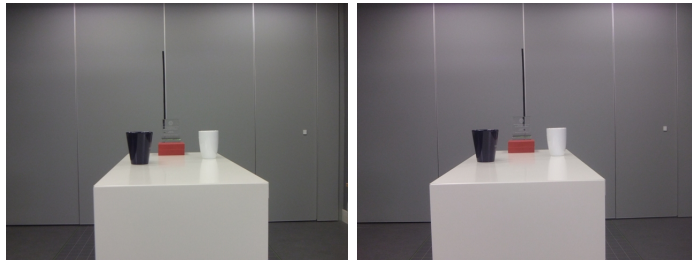


Figure A.7: AWARD, light 3, distance 2

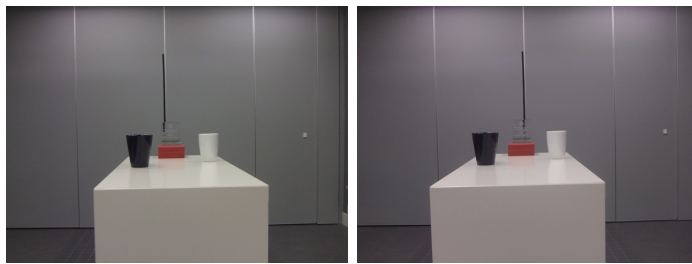


Figure A.8: AWARD, light 4, distance 2

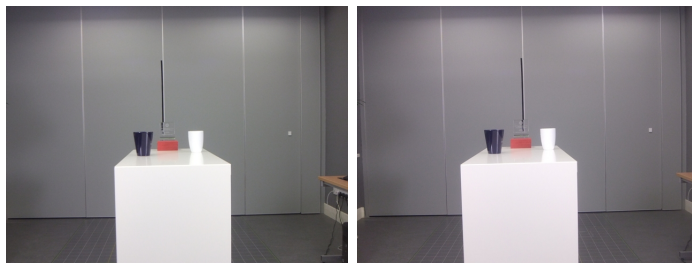


Figure A.9: AWARD, light 1, distance 3

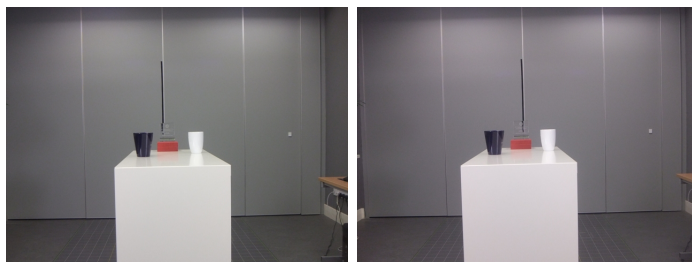


Figure A.10: AWARD, light 2, distance 3

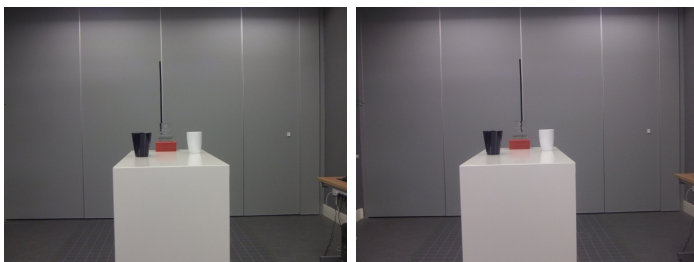


Figure A.11: AWARD, light 3, distance 3

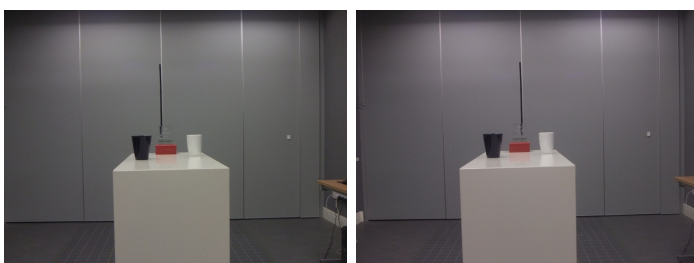


Figure A.12: AWARD, light 4, distance 3



Figure A.13: AWARD, light 1, distance 4



Figure A.14: AWARD, light 2, distance 4

A.2 Test images acquired for the FACE test scene



Figure A.15: AWARD, light 3, distance 4



Figure A.16: AWARD, light 4, distance 4

A.2 Test images acquired for the FACE test scene



Figure A.17: FACE, light 1, distance 1



Figure A.18: FACE, light 2, distance 1



Figure A.19: FACE, light 3, distance 1



Figure A.20: FACE, light 4, distance 1

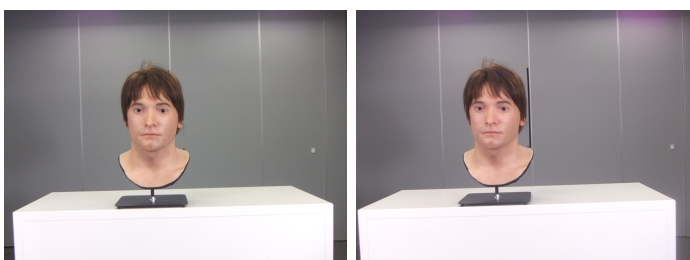


Figure A.21: FACE, light 1, distance 2



Figure A.22: FACE, light 2, distance 2

A.2 Test images acquired for the FACE test scene



Figure A.23: FACE, light 3, distance 2



Figure A.24: FACE, light 4, distance 2



Figure A.25: FACE, light 1, distance 3



Figure A.26: FACE, light 2, distance 3



Figure A.27: FACE, light 3, distance 3



Figure A.28: FACE, light 4, distance 3



Figure A.29: FACE, light 1, distance 4



Figure A.30: FACE, light 2, distance 4

A.3 Test images acquired for the FACES test scene



Figure A.31: FACE, light 3, distance 4



Figure A.32: FACE, light 4, distance 4

A.3 Test images acquired for the FACES test scene



Figure A.33: FACES, light 1, distance 1



Figure A.34: FACES, light 2, distance 1



Figure A.35: FACES, light 3, distance 1



Figure A.36: FACES, light 4, distance 1



Figure A.37: FACES, light 1, distance 2

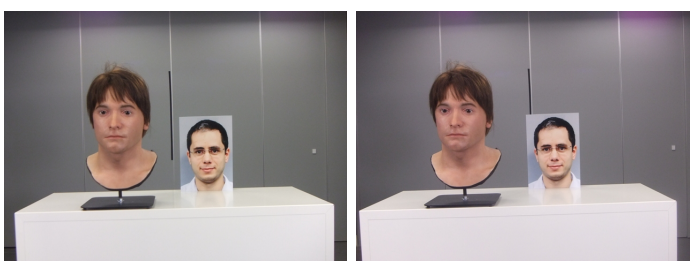


Figure A.38: FACES, light 2, distance 2

A.3 Test images acquired for the FACES test scene



Figure A.39: FACES, light 3, distance 2



Figure A.40: FACES, light 4, distance 2



Figure A.41: FACES, light 1, distance 3



Figure A.42: FACES, light 2, distance 3



Figure A.43: FACES, light 3, distance 3



Figure A.44: FACES, light 4, distance 3



Figure A.45: FACES, light 1, distance 4



Figure A.46: FACES, light 2, distance 4

A.4 Test images acquired for the TEXTURED OBJECTS test scene



Figure A.47: FACES, light 3, distance 4



Figure A.48: FACES, light 4, distance 4

A.4 Test images acquired for the TEXTURED OBJECTS test scene

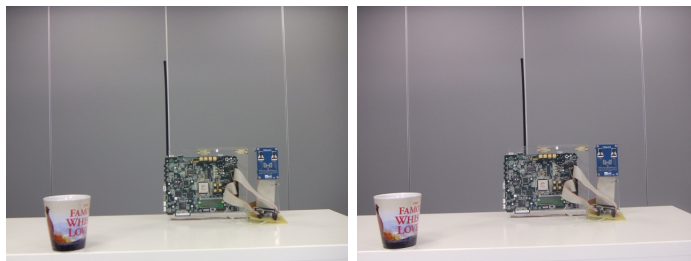


Figure A.49: TEXTURED OBJECTS, light 1, distance 1

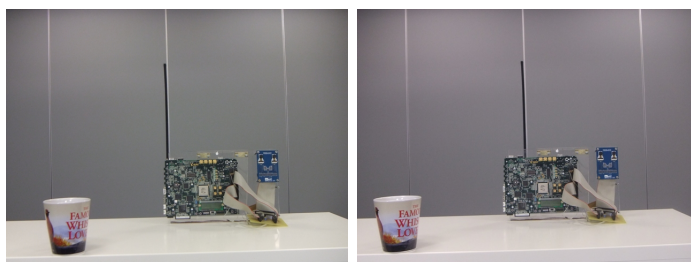


Figure A.50: TEXTURED OBJECTS, light 2, distance 1

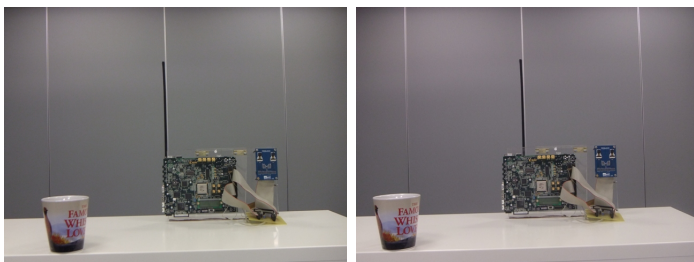


Figure A.51: TEXTURED OBJECTS, light 3, distance 1

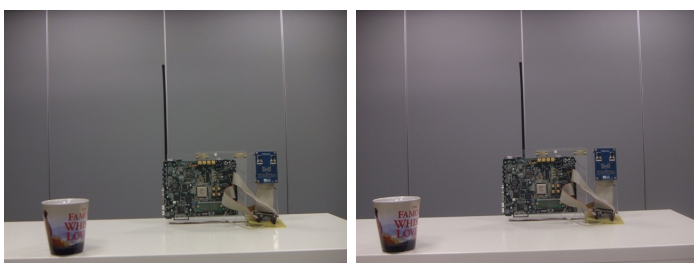


Figure A.52: TEXTURED OBJECTS, light 4, distance 1



Figure A.53: TEXTURED OBJECTS, light 1, distance 2



Figure A.54: TEXTURED OBJECTS, light 2, distance 2

A.4 Test images acquired for the TEXTURED OBJECTS test scene



Figure A.55: TEXTURED OBJECTS, light 3, distance 2



Figure A.56: TEXTURED OBJECTS, light 4, distance 2

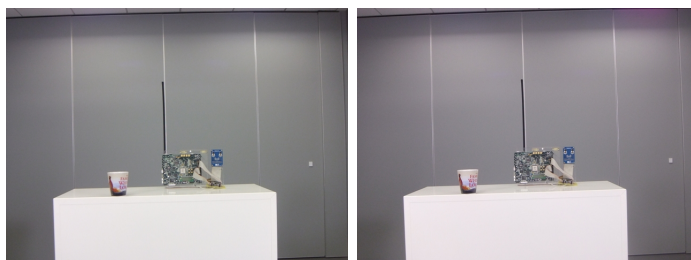


Figure A.57: TEXTURED OBJECTS, light 1, distance 3



Figure A.58: TEXTURED OBJECTS, light 2, distance 3

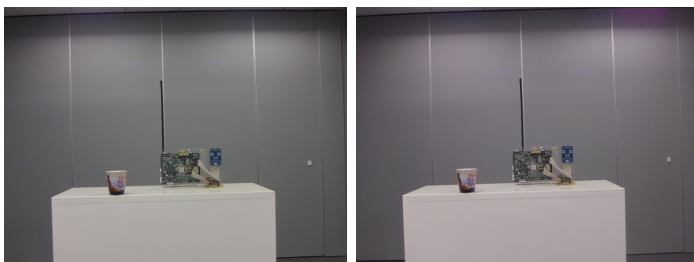


Figure A.59: TEXTURED OBJECTS, light 3, distance 3



Figure A.60: TEXTURED OBJECTS, light 4, distance 3

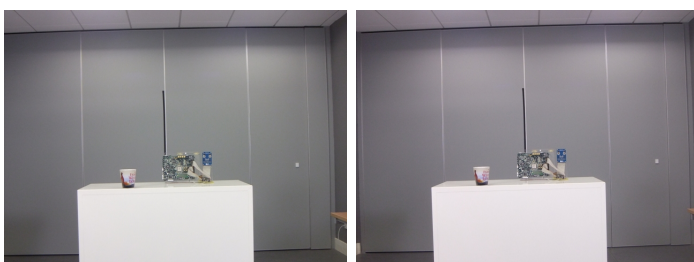


Figure A.61: TEXTURED OBJECTS, light 1, distance 4



Figure A.62: TEXTURED OBJECTS, light 2, distance 4

A.5 Depth maps generated using the AWARD test scene

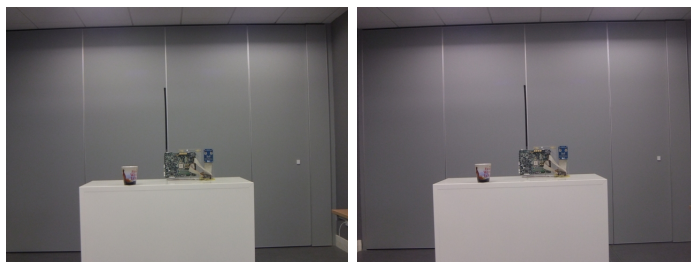
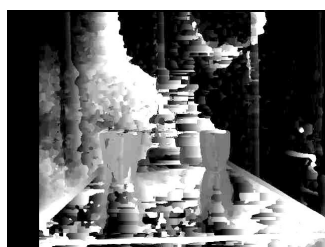


Figure A.63: TEXTURED OBJECTS, light 3, distance 4



Figure A.64: TEXTURED OBJECTS, light 4, distance 4

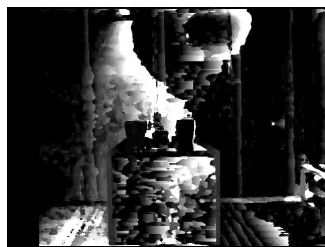
A.5 Depth maps generated using the AWARD test scene



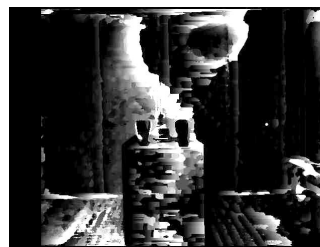
(a) SAD, Distance 1



(b) SAD, Distance 2



(c) SAD, Distance 3



(d) SAD, Distance 4

Figure A.65: SAD-based algorithm

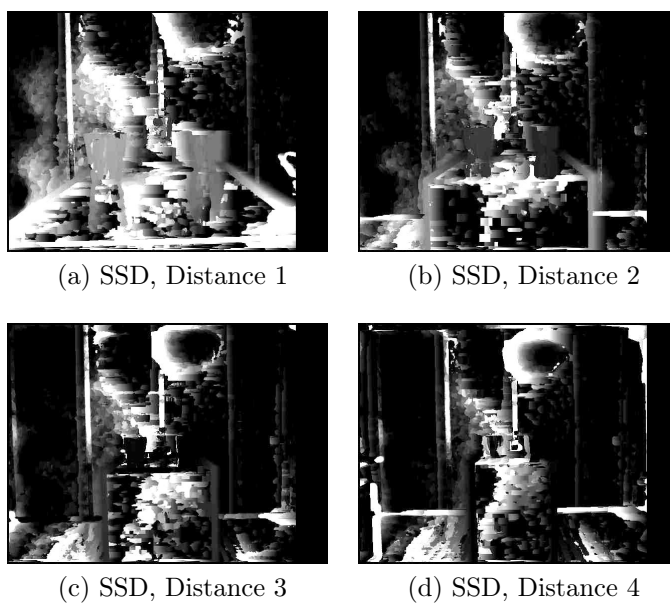


Figure A.66: SSD-based algorithm

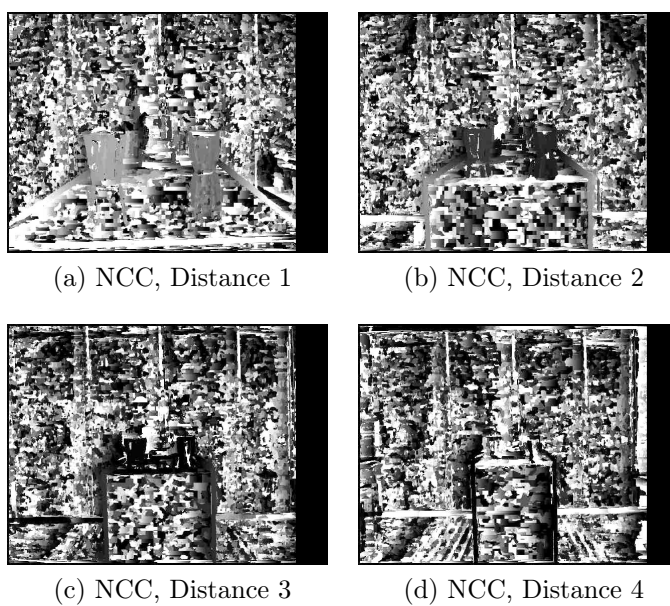


Figure A.67: NCC-based algorithm

A.6 Depth maps generated using the FACE test scene

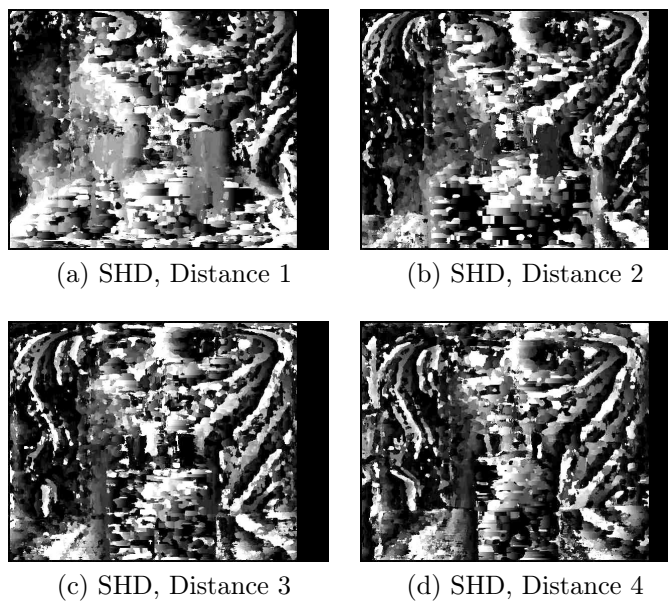


Figure A.68: SHD-based algorithm

A.6 Depth maps generated using the FACE test scene

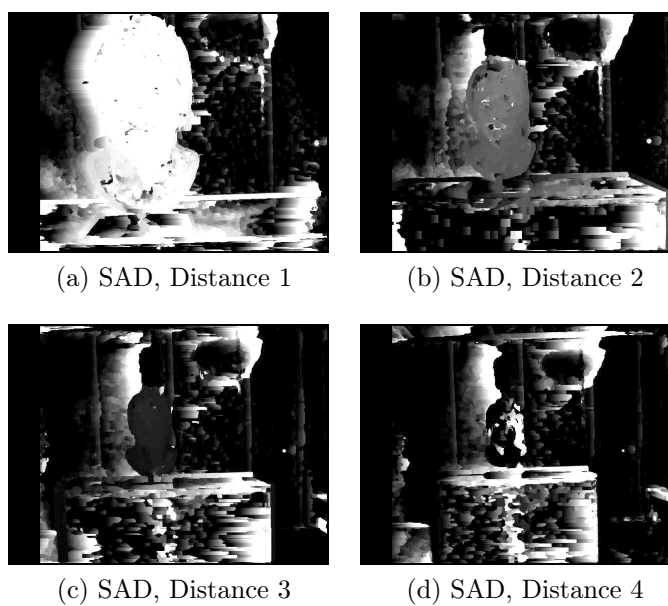


Figure A.69: SAD-based algorithm

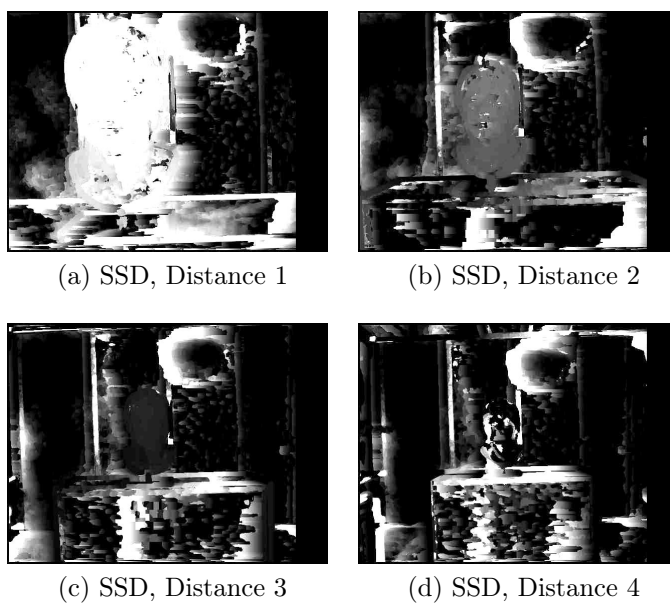


Figure A.70: SSD-based algorithm

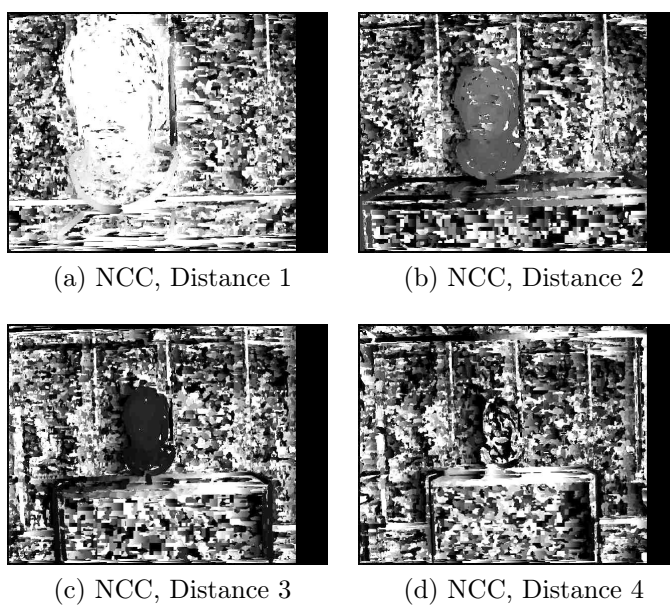


Figure A.71: NCC-based algorithm

A.7 Depth maps generated using the FACES test scene

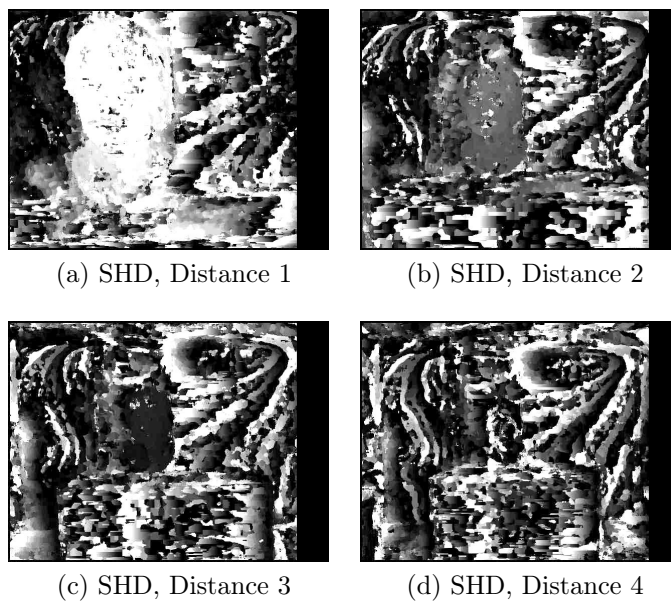


Figure A.72: SHD-based algorithm

A.7 Depth maps generated using the FACES test scene

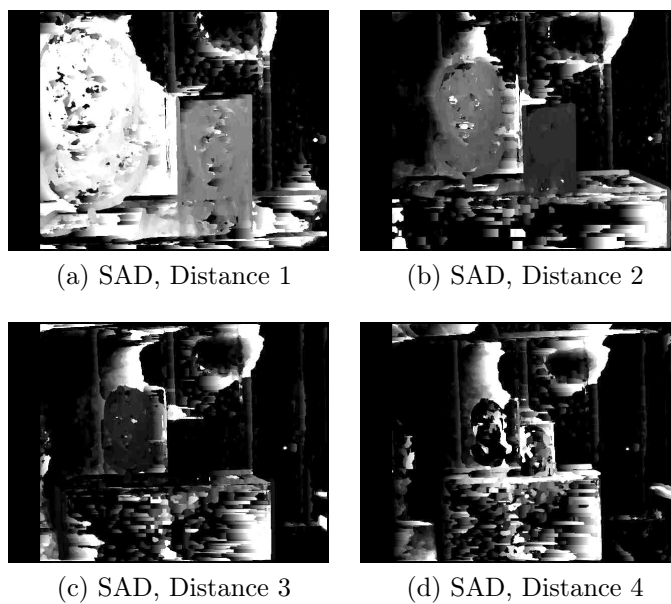


Figure A.73: SAD-based algorithm

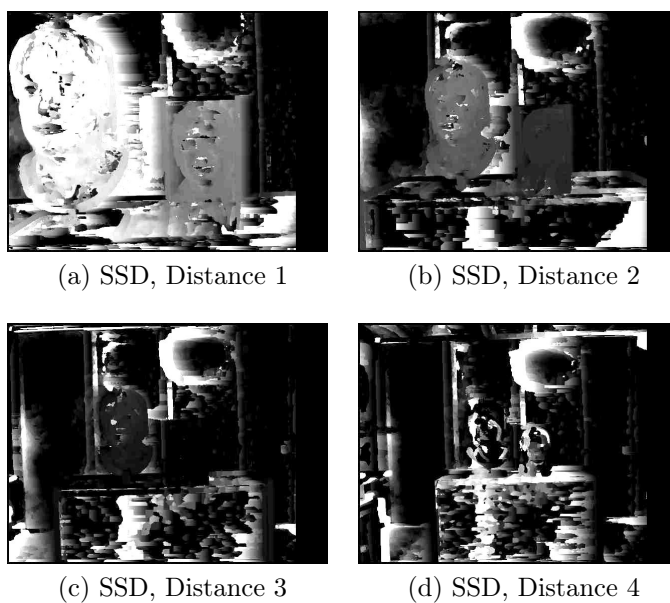


Figure A.74: SSD-based algorithm

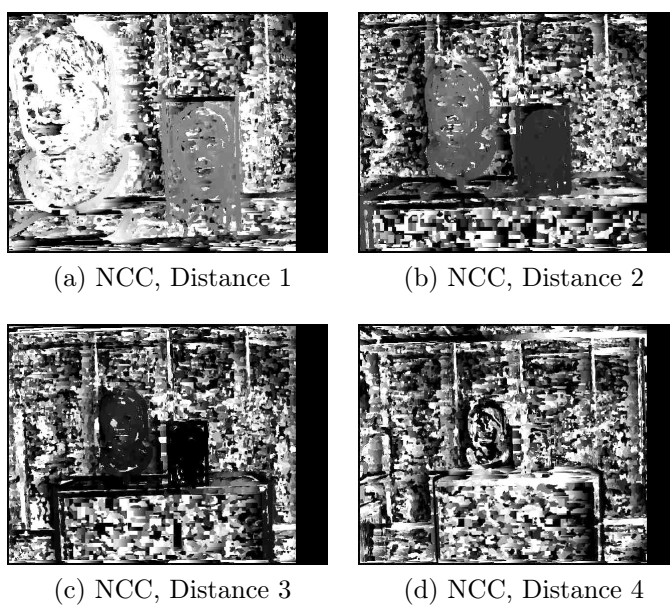


Figure A.75: NCC-based algorithm

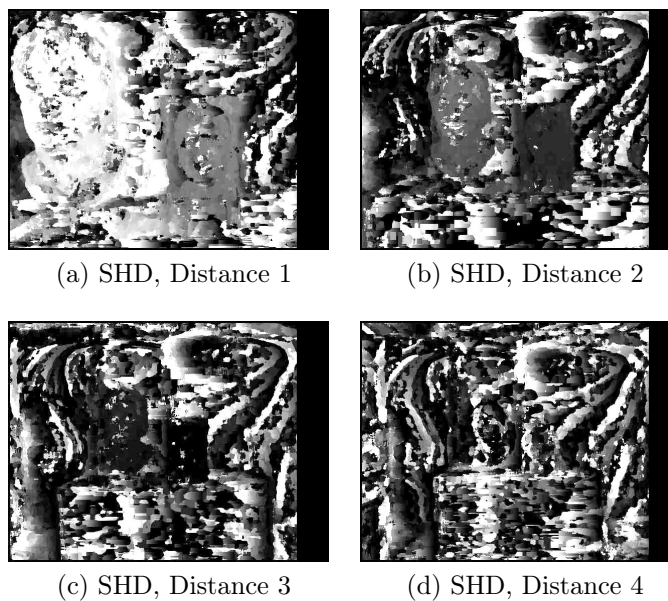


Figure A.76: SHD-based algorithm

A.8 Depth maps generated using the TEXTURED OBJECTS test scene

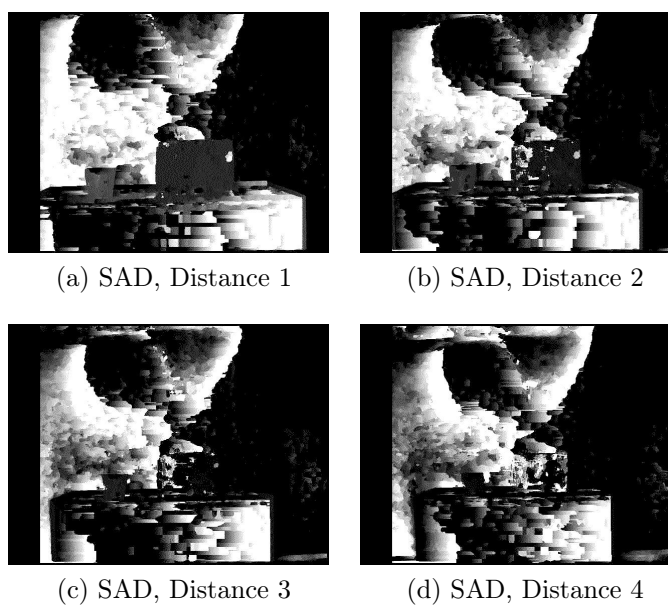


Figure A.77: SAD-based algorithm

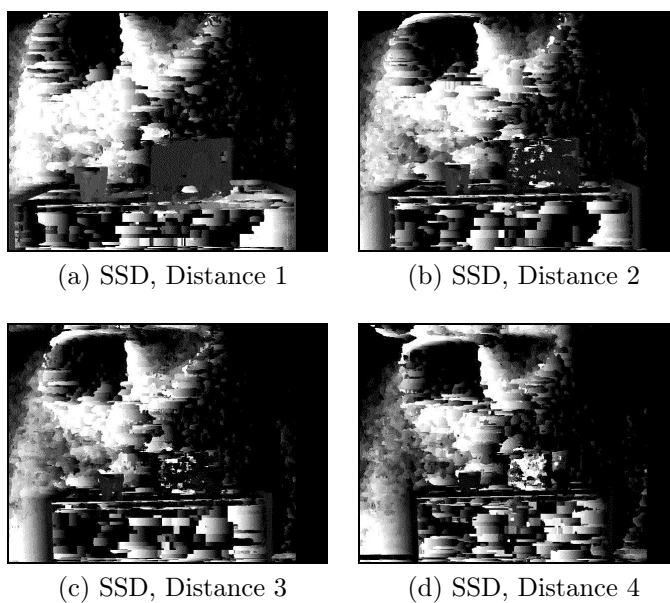


Figure A.78: SSD-based algorithm

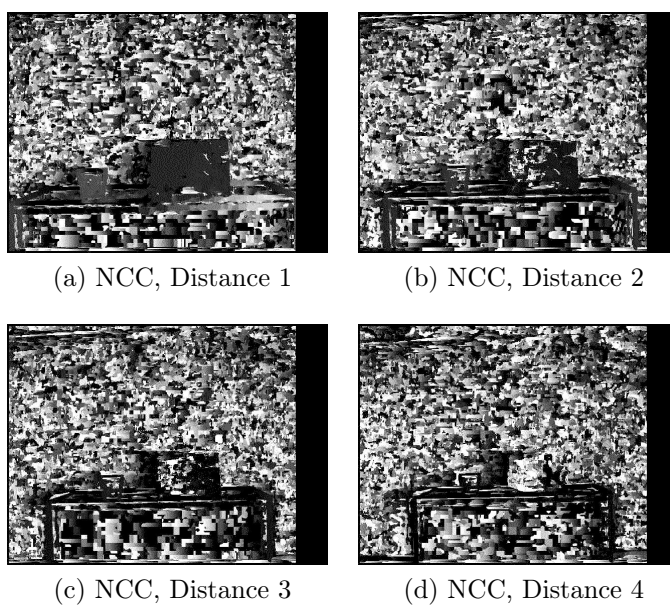
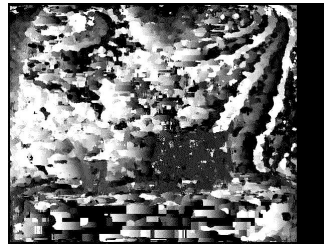


Figure A.79: NCC-based algorithm



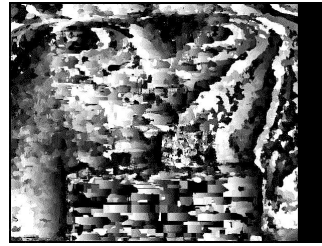
(a) SHD, Distance 1



(b) SHD, Distance 2



(c) SHD, Distance 3



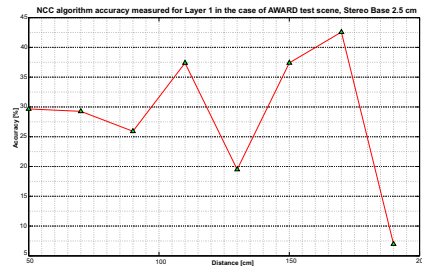
(d) SHD, Distance 4

Figure A.80: SHD-based algorithm

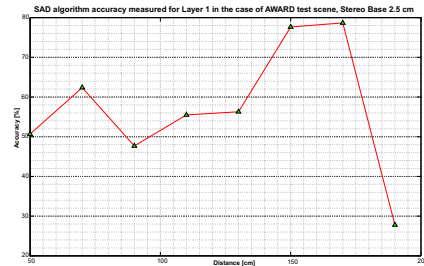
Appendix B

Depth Map quality measurements

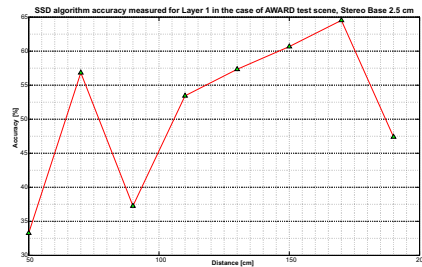
B.1 Depth map quality results. 8 distances, 5 stereo base values.



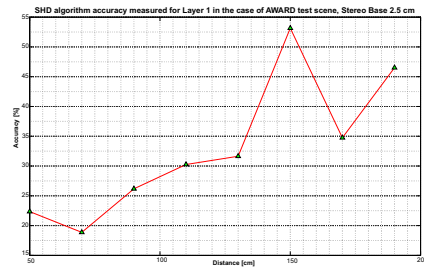
(a) NCC



(b) SAD



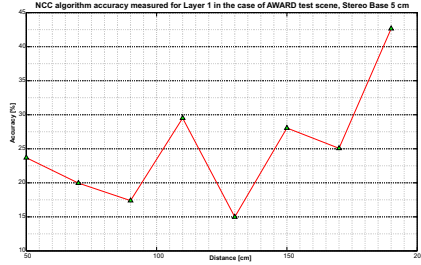
(c) SSD



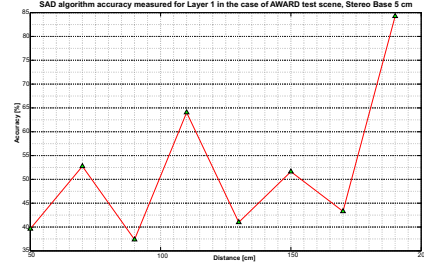
(d) SHD

Figure B.1: Stereo base length of 2.5 cm

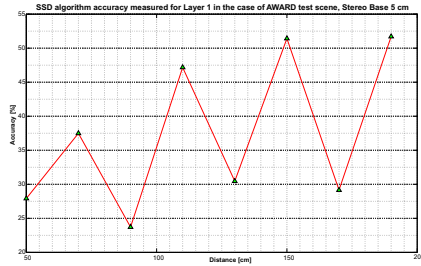
Depth Map quality measurements



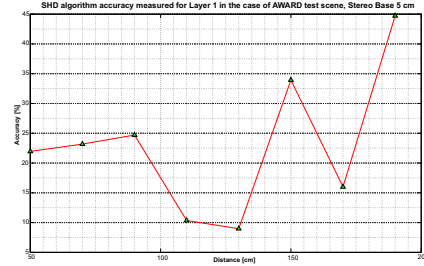
(a) NCC



(b) SAD

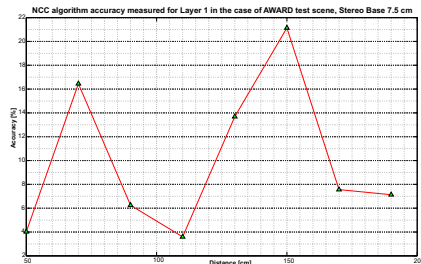


(c) SSD

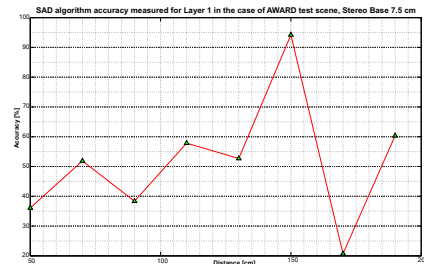


(d) SHD

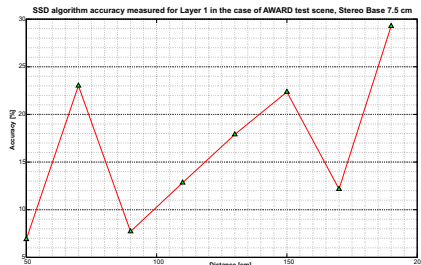
Figure B.2: Stereo base length of 5 cm



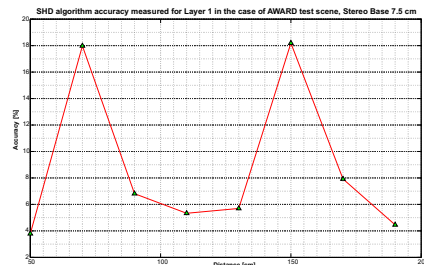
(a) NCC



(b) SAD



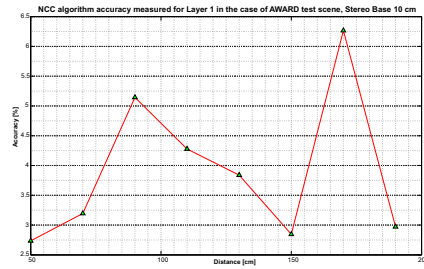
(c) SSD



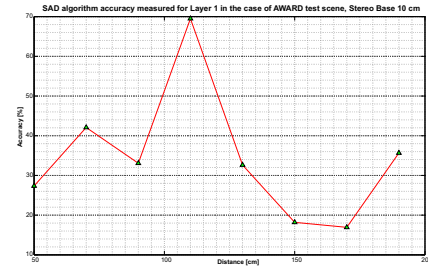
(d) SHD

Figure B.3: Stereo base length of 7.5 cm

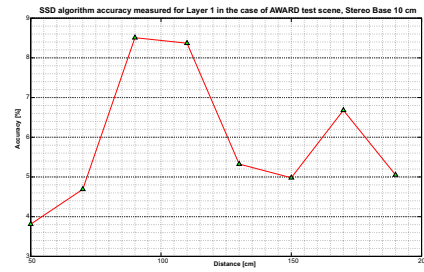
B.1 Depth map quality results. 8 distances, 5 stereo base values.



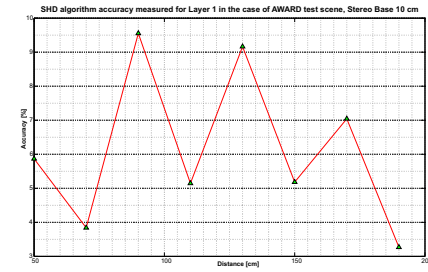
(a) NCC



(b) SAD

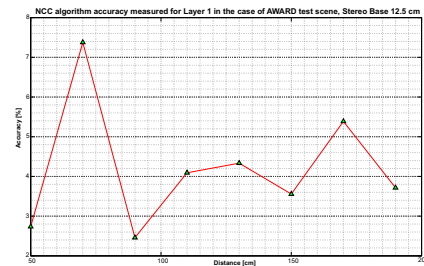


(c) SSD

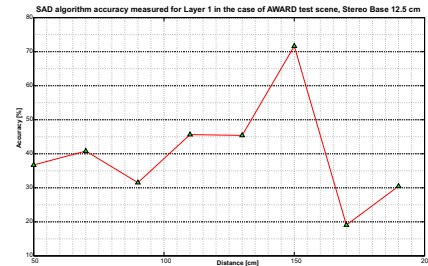


(d) SHD

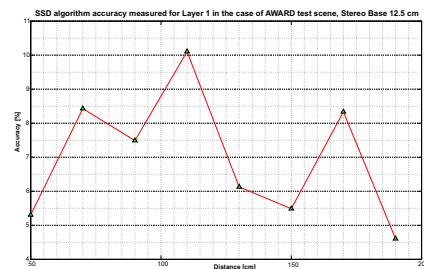
Figure B.4: Stereo base length of 10 cm



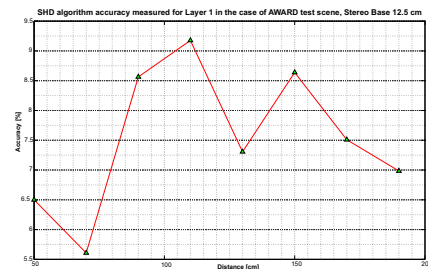
(a) NCC



(b) SAD



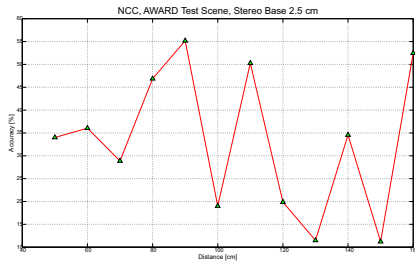
(c) SSD



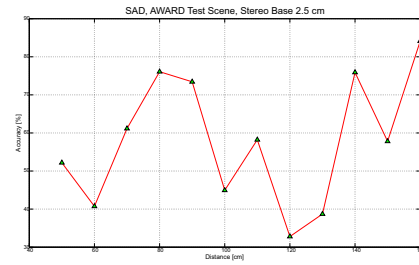
(d) SHD

Figure B.5: Stereo base length of 12.5 cm

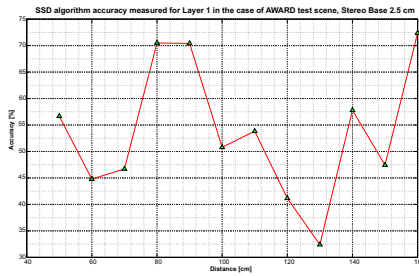
B.2 Depth map quality results. 12 distances, 5 stereo base values.



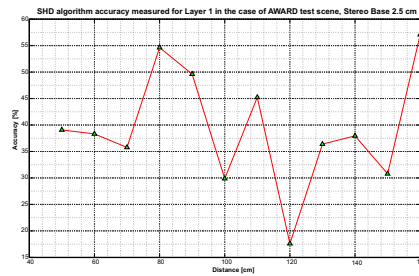
(a) NCC



(b) SAD



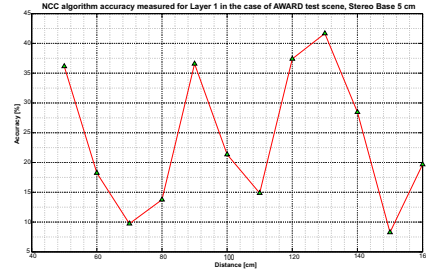
(c) SSD



(d) SHD

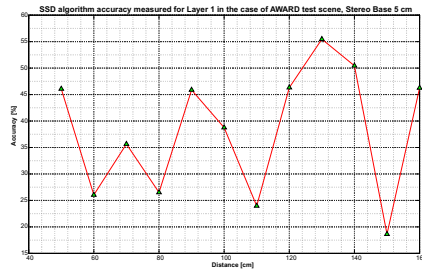
Figure B.6: Stereo base length of 2.5 cm

B.2 Depth map quality results. 12 distances, 5 stereo base values.

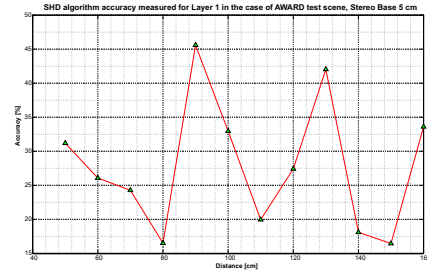


(a) NCC

(b) SAD

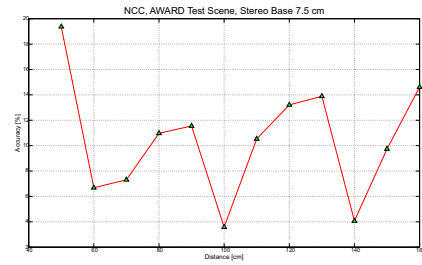


(c) SSD

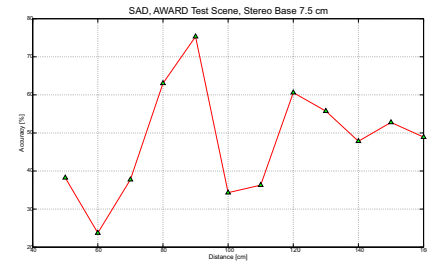


(d) SHD

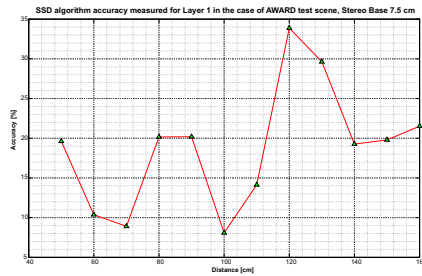
Figure B.7: Stereo base length of 5 cm



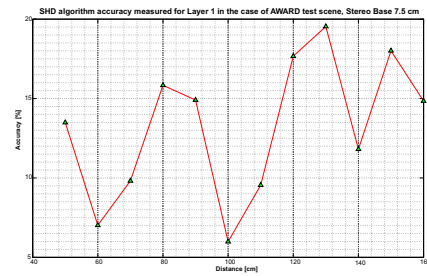
(a) NCC



(b) SAD



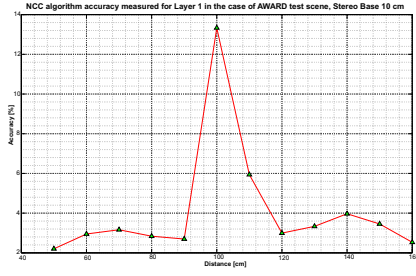
(c) SSD



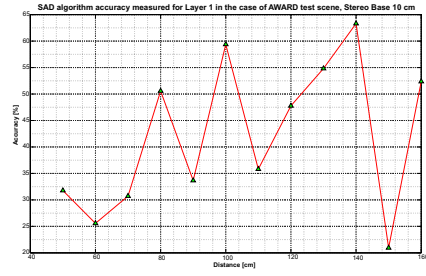
(d) SHD

Figure B.8: Stereo base length of 7.5 cm

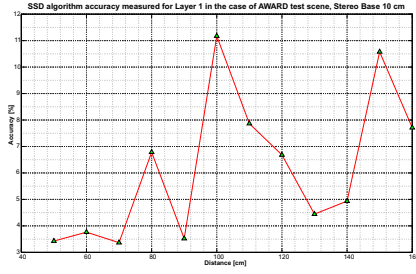
Depth Map quality measurements



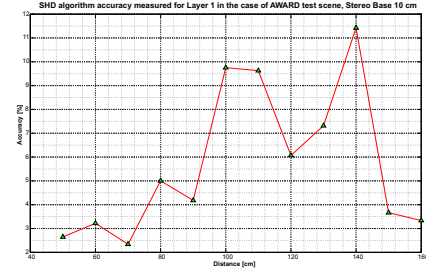
(a) NCC



(b) SAD

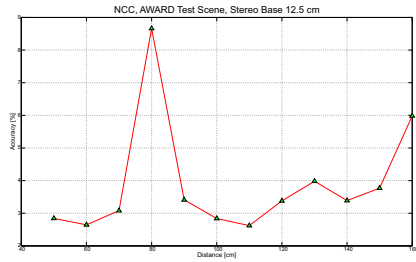


(c) SSD

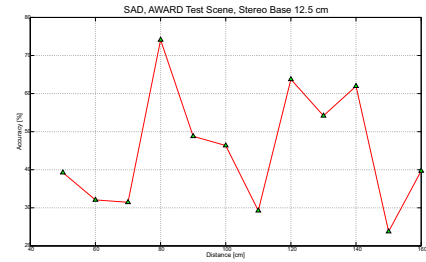


(d) SHD

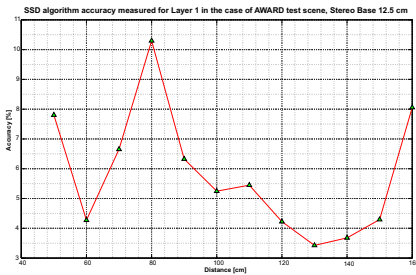
Figure B.9: Stereo base length of 10 cm



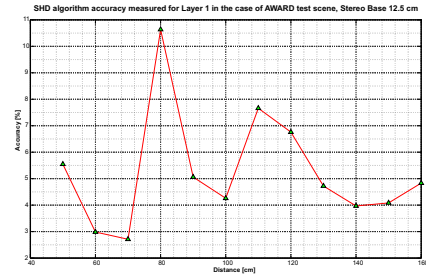
(a) NCC



(b) SAD



(c) SSD



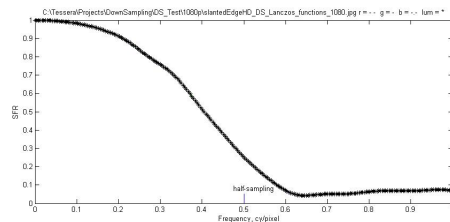
(d) SHD

Figure B.10: Stereo base length of 12.5 cm

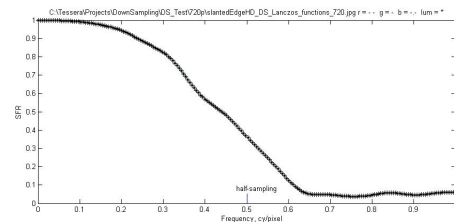
Appendix C

Choice of Down Sampling algorithms measurements

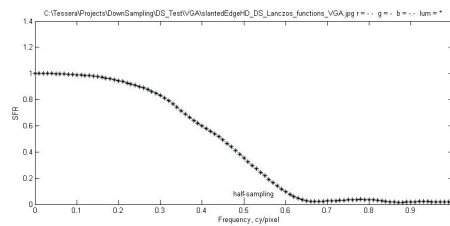
C.1 MTF results for different down sampling algorithms



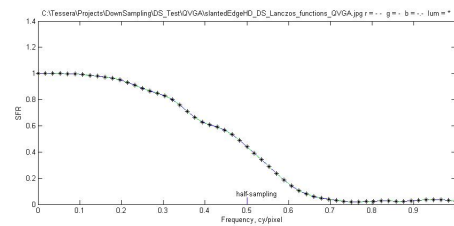
(a) 12 Mpix to Full HD



(b) 12 Mpix to HD



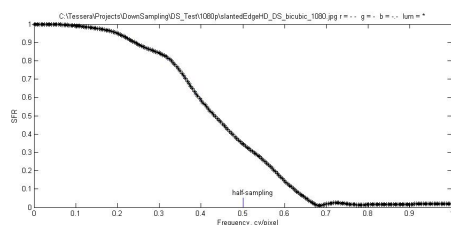
(c) 12 Mpix to VGA



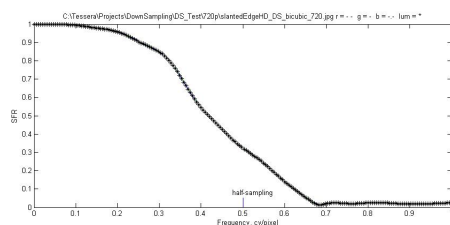
(d) 12 Mpix to QVGA

Figure C.1: Lanczos Algorithm

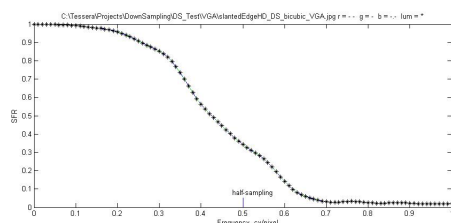
Choice of Down Sampling algorithms measurements



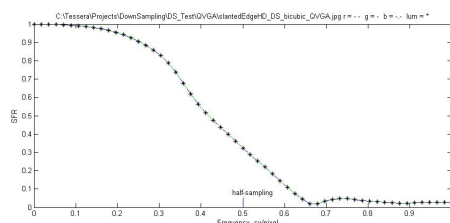
(a) 12 Mpix to Full HD



(b) 12 Mpix to HD

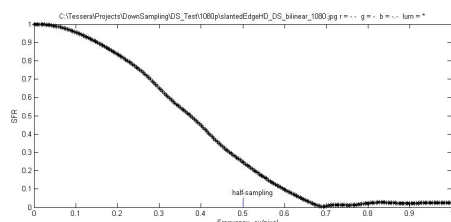


(c) 12 Mpix to VGA

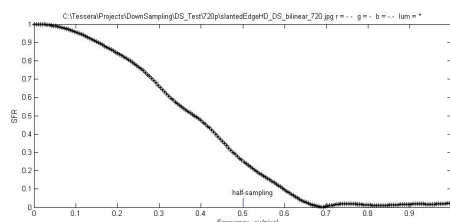


(d) 12 Mpix to QVGA

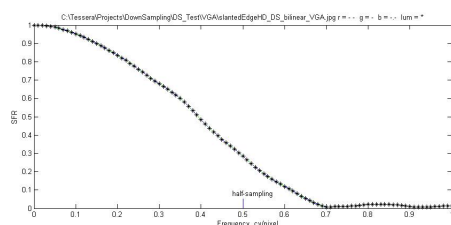
Figure C.2: Bicubic Algorithm



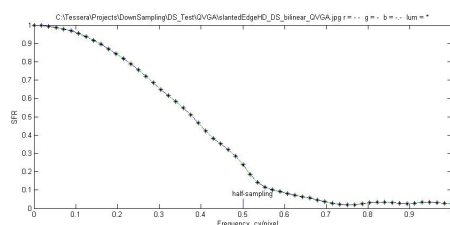
(a) 12 Mpix to Full HD



(b) 12 Mpix to HD



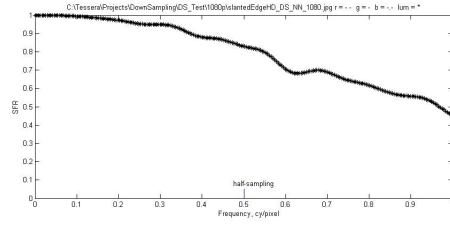
(c) 12 Mpix to VGA



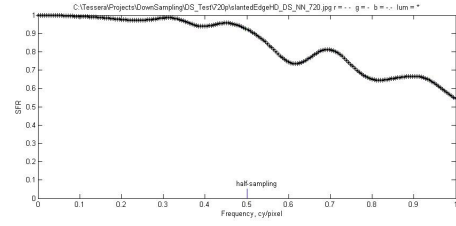
(d) 12 Mpix to QVGA

Figure C.3: Bilinear Algorithm

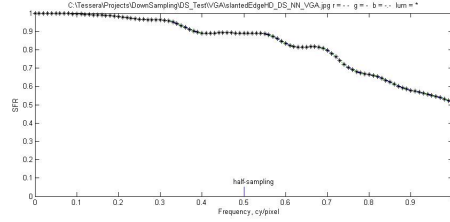
C.1 MTF results for different down sampling algorithms



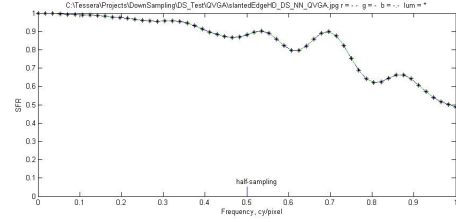
(a) 12 Mpix to Full HD



(b) 12 Mpix to HD

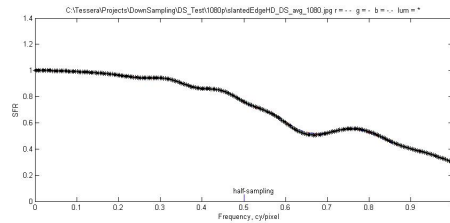


(c) 12 Mpix to VGA

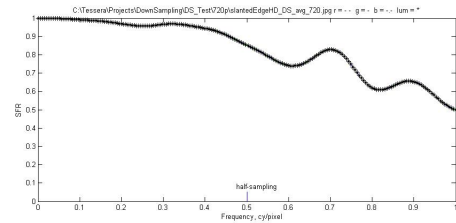


(d) 12 Mpix to QVGA

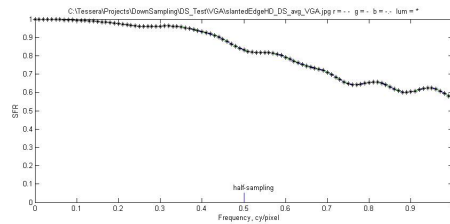
Figure C.4: Nearest Neighbor Algorithm



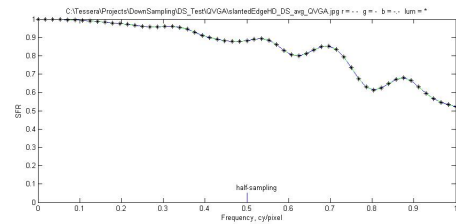
(a) 12 Mpix to Full HD



(b) 12 Mpix to HD



(c) 12 Mpix to VGA



(d) 12 Mpix to QVGA

Figure C.5: Averaging Algorithm

C.2 Downsampling of Zone Plate type of images

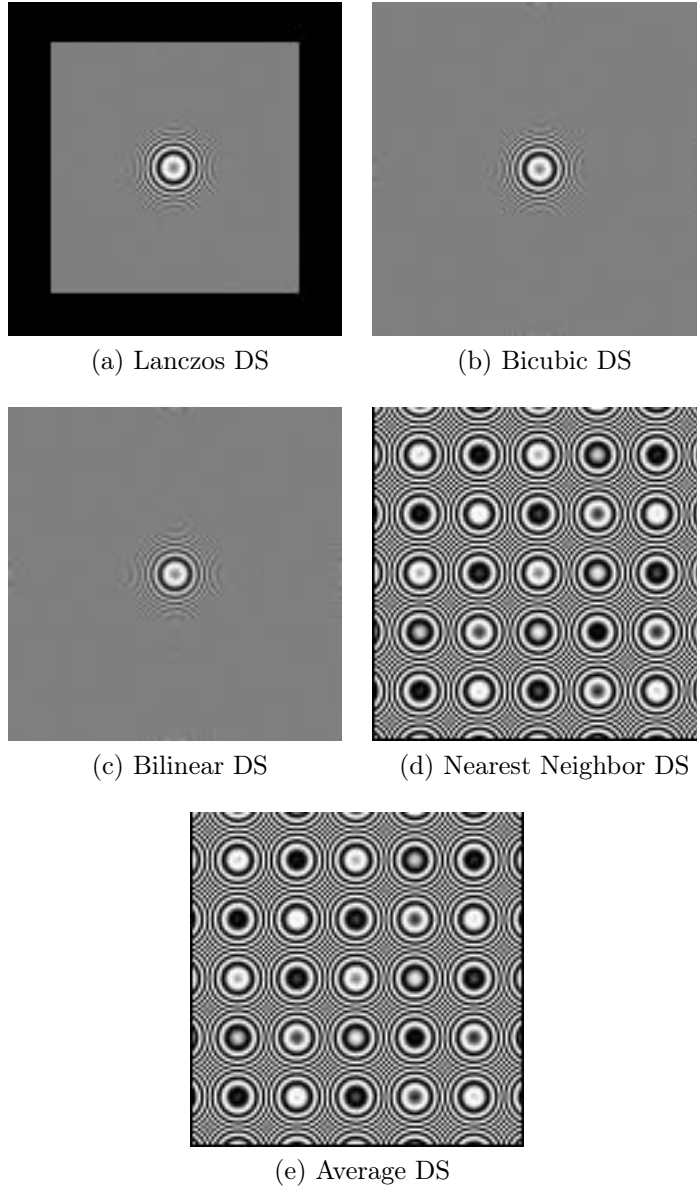


Figure C.6: Zone plate type of images

Appendix D

Details about cameras used for image acquisition

D.1 Fujifilm W3 3D

Specifications of the Fujifilm FinePix REAL 3D W3 camera

- Model Name: FinePix REAL 3D W3
- Number of effective pixels: 10.0 megapixels
- CCD Sensor: 1/2.3 inch CCD x 2
- Storage Media:
 - Internal Memory (Approx. 34 MB)
 - SD memory card
 - SDHC memory card
- File Format:
 - 3D Still image: MPO+JPEG, MPO
 - 2D Still image: JPEG
 - 3D Movie: 3D-AVI
 - 2D Movie: AVI format

- Lens:
 - Name: Fujinon 3x optical zoom lens, F3.7(Wide) - F4.2(Telephoto)
 - Focal Length: $f = 6.3 - 18.9$ mm, equivalent to 35 - 105 mm on a 35 mm camera
- Aperture: Wide: F3.7 - F8.0, Telephoto: F4.2 - F9.0
- Focus Distance:
 - Normal: Approx. 60 cm to infinity
 - Macro(3D):
 - * Wide: Approx. 38 cm - 70 cm
 - * Telephoto: Approx. 1.1 m - 2.3 m
 - Macro(2D):
 - * Wide: Approx. 8 cm - 80 cm
 - * Telephoto: Approx. 60 cm - 3 m
- Exposure mode: Programmed AE, Aperture Priority AE, Manual
- Focus:
 - Mode: Single AF
 - AF mode:
 - * 3D: Center
 - * 2D: Center, Multi

D.2 Tessera WLC Cameras

- CMOS Sensor: 2.5 mm
- Number of effective pixels: 2.0 megapixels
- Focus Distance: Fixed (infinity)

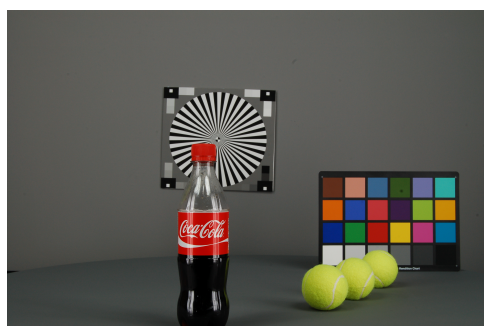
D.3 Aptina MT9P031

- CMOS Sensor, 5 megapixels
- Optical format: 1/2.5 inch (4:3)
- Active pixels: 2592H x 1944V
- Pixel size: 2.2 μm x 2.2 μm
- Color filter array: RGB Bayer pattern
- Shutter type:
 - Global reset release (GRR)
 - Snapshot only
 - Electronic rolling shutter (ERS)
- Maximum data rate/master clock:
 - 96 Mp/s at 96 MHz (2.8 V I/O)
 - 48 Mp/s at 48 MHz (1.8 V I/O)
- Frame rate:
 - Full resolution: Programmable up to 14 fps
 - VGA: Programmable up to 53 fps
- ADC resolution: 12-bit, on-chip
- Responsivity: 1.4 V/lx-sec (550 nm)

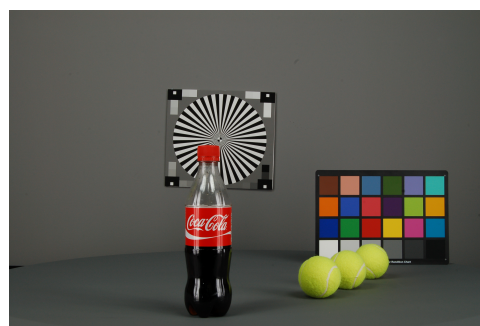
Appendix E

Proposed Test Scene

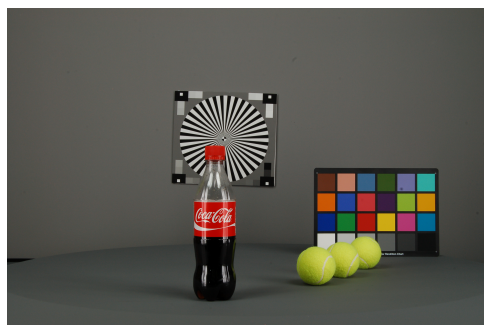
E.1 Different camera-object distances



(a) 97 cm



(b) 107 cm



(c) 117 cm

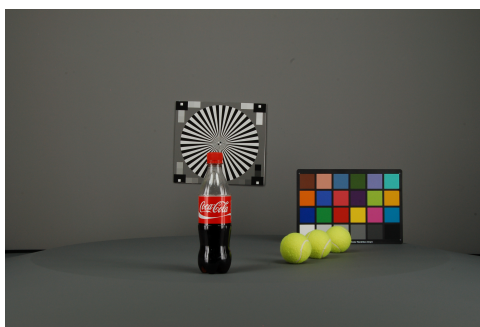


(d) 127 cm

Figure E.1



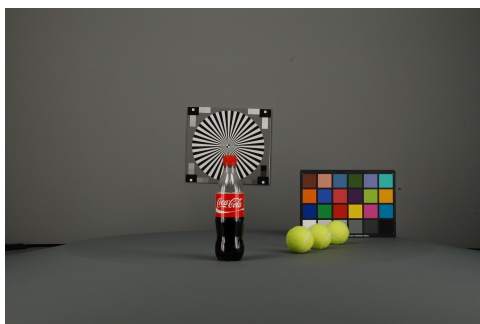
(a) 137 cm



(b) 147 cm



(c) 157 cm



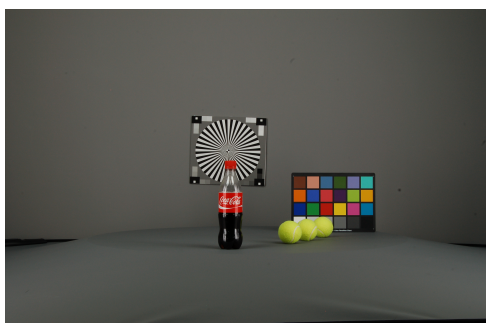
(d) 167 cm



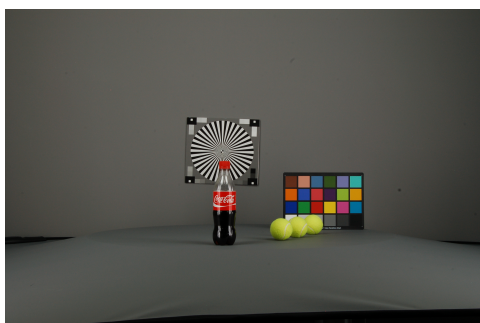
(e) 177 cm



(f) 187 cm



(g) 197 cm



(h) 207 cm

Figure E.2

E.2 Scene setup procedure



Figure E.3: Studio lights 1



Figure E.4: Studio lights 2

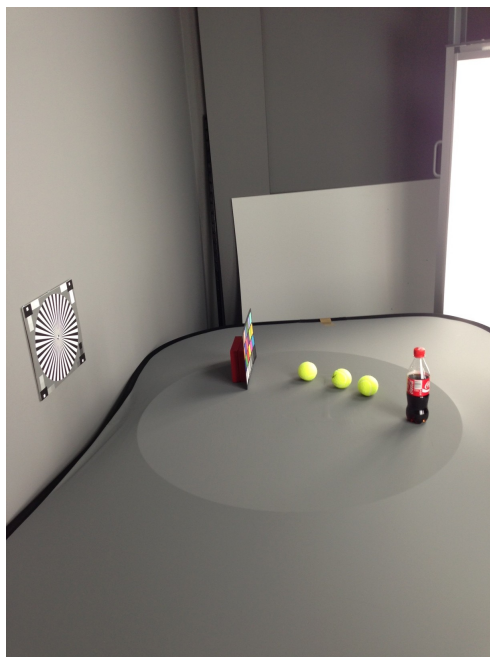


Figure E.5: Side-view of the scene



Figure E.6: Camera setup

E.2 Scene setup procedure



Figure E.7: Tools used

Appendix F

Data variation test results

F.1 Test results Nikon D50

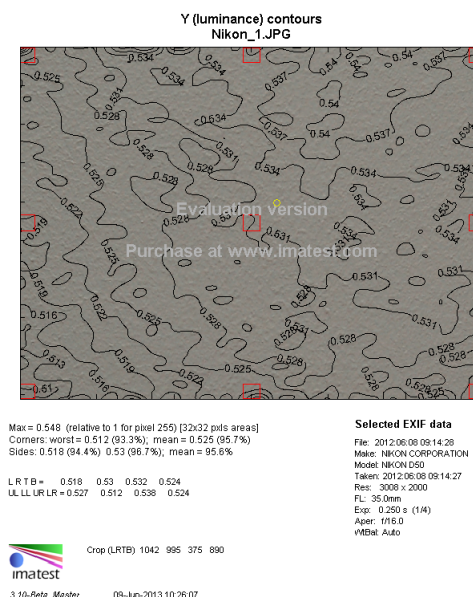


Figure F.1: Imatest Nikon 1

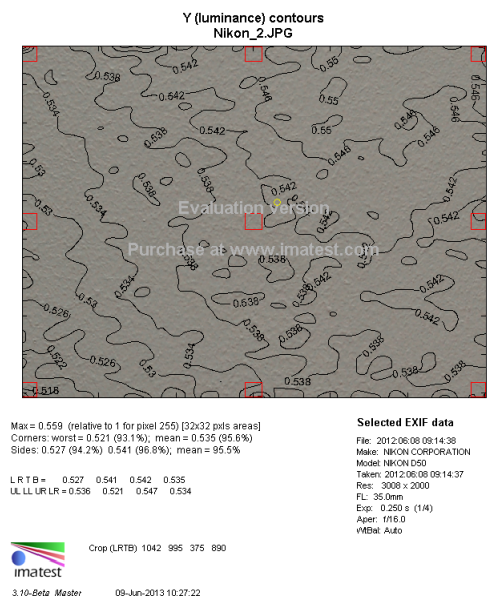


Figure F.2: Imatest Nikon 2

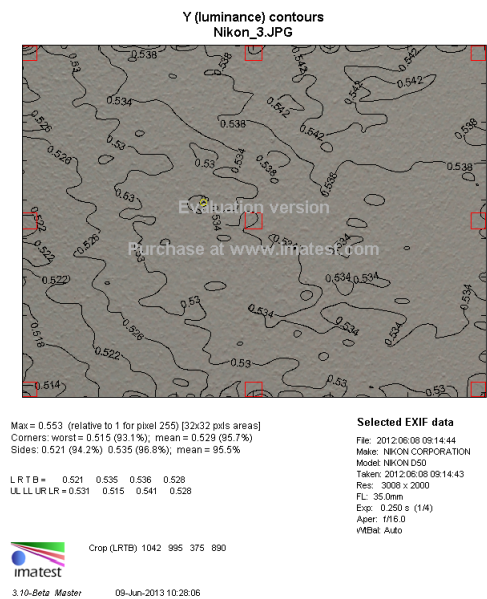


Figure F.3: Imatest Nikon 3

F.1 Test results Nikon D50

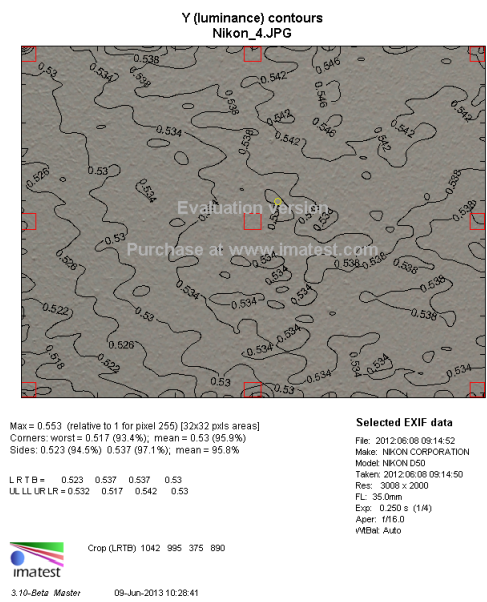


Figure F.4: Imatest Nikon 4

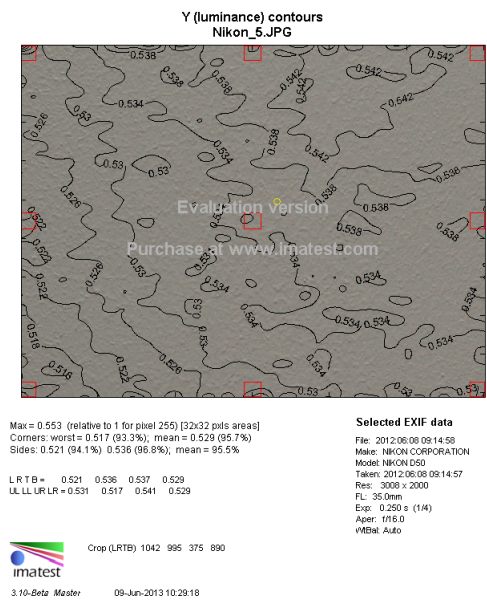


Figure F.5: Imatest Nikon 5

F.2 Test results Canon EOS 600 D

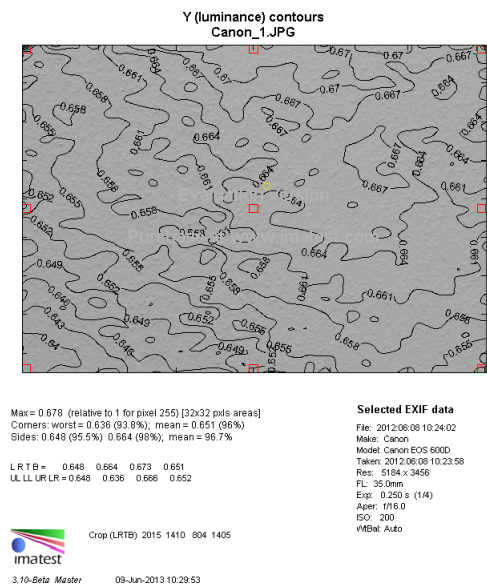


Figure F.6: Imatest Canon 1

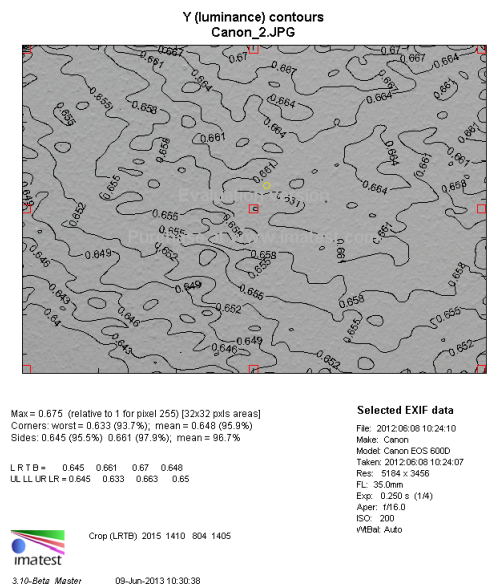


Figure F.7: Imatest Canon 2

F.2 Test results Canon EOS 600 D

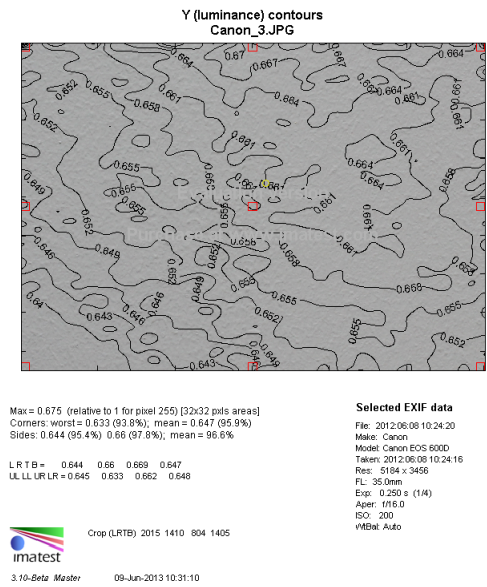


Figure F.8: Imatest Canon 3

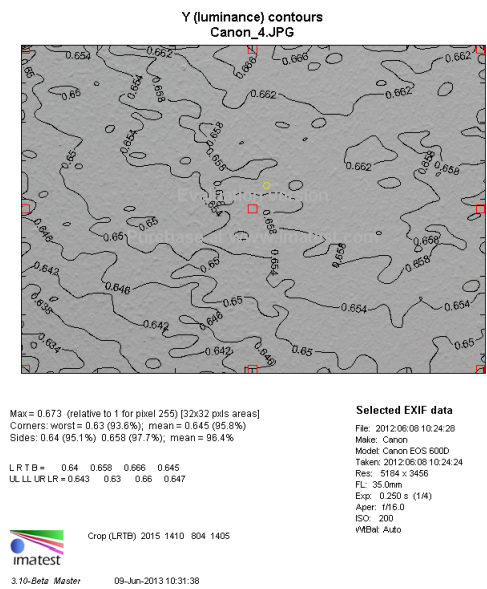


Figure F.9: Imatest Canon 4

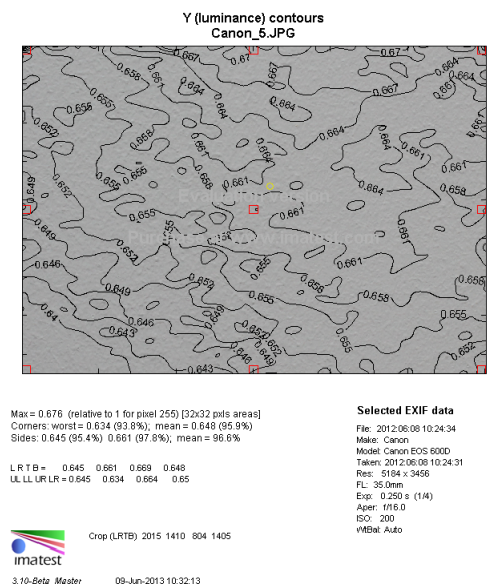


Figure F.10: Imatest Canon 5

Bibliography

- [1] Felix Albu, Eran Steinberg, Alexandru Drimbarean, Corneliu Florea, Adrian Zamfir, Peter Corcoran, and Vlad Poenaru. Image processing method and apparatus, 2008.
- [2] I Andorko, P Corcoran, and P Bigioi. Depth map generation and depth layer separation for information customization in computer gaming. *2010 2nd International IEEE Consumer Electronics Society's Games Innovations Conference*, (i):1–4, December 2010.
- [3] Istvan Andorko, Peter Corcoran, and Petronel Bigioi. FPGA based stereo imaging system with applications in computer gaming. In *IEEE Games Innovation Conference*, pages 239–245, 2009.
- [4] Istvan Andorko, Peter Corcoran, and Petronel Bigioi. Hardware implementation of a real-time 3D video acquisition system. In *International Conference on Optimization of Electrical and Electronic Equipment*, page 920, 2010.
- [5] Istvan Andorko, Peter Corcoran, and Petronel Bigioi. A dual image processing pipeline camera with CE applications. *2011 IEEE International Conference on Consumer Electronics (ICCE)*, pages 737–738, January 2011.
- [6] Istvan Andorko, Peter Corcoran, and Petronel Bigioi. Tools and techniques for the implementation of a FPGA-based stereoscopic camera. *UbiCC Journal*, 6(3):876 – 883, 2011.
- [7] Istvan Andorko, Peter Corcoran, Petronel Bigioi, and Senior Member. Proposal of a Universal Test Scene for Depth Map Evaluation. In *IEEE International Conference on Consumer Electronics*, pages 151–152, 2013.

- [8] Istvan Andorko, Student Member, and Peter Corcoran. Proposal of a Universal Test Scene for Depth Map Evaluation. *IEEE Transactions on Consumer Electronics*, (i):385–390, 2013.
- [9] Henry Harlyn Baker. *Depth from edge and intensity based stereo*. PhD thesis, Stanford University California, 1982.
- [10] Igor Barcowschi, Peter Corcoran, Alexandru Drimbarean, Larry Murray, and Piotr Stec. Panorama imaging using a blending map, 2012.
- [11] K Barnard, V Cardei, and B Funt. A comparison of computational color constancy algorithms-Part I: Methodology and experiments with synthesized data. *IEEE Transactions on Image Processing*, 11(9):972–983, 2002.
- [12] J L Barron, D J Fleet, S S Beauchemin, and T A Burkitt. Performance of optical flow techniques. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 236–242, 1992.
- [13] Ronen Basri, David Jacobs, and Ira Kemelmacher. Photometric Stereo with General, Unknown Lighting. *International Journal of Computer Vision*, 72(3):239–257, June 2006.
- [14] Philippe M. Beaudoin, Yves Audet, and Victor Hugo Ponce-Ponce. Dark current compensation in CMOS image sensors using a differential pixel architecture. *2009 Joint IEEE North-East Workshop on Circuits and Systems and TAISA Conference*, pages 1–4, June 2009.
- [15] Petronel Bigioi, Peter Corcoran, Yuri Prilutsky, and Eran Steinberg. Digital Image Processing Using Face Detection Information, 2012.
- [16] S. Birchfield and C. Tomasi. Multiway cut for stereo and motion with slanted surfaces. In *IEEE International Conference on Computer Vision*, number 1, pages 489–495. Ieee, 1999.
- [17] M. Bleyer, M. Gelautz, C. Rother, and C. Rhemann. A stereo approach that handles the matting problem via image warping. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 501–508, June 2009.

BIBLIOGRAPHY

- [18] Michael Bleyer, Carsten Rother, and Pushmeet Kohli. Surface stereo with soft segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1570–1577, 2010.
- [19] Michael Bleyer, Carsten Rother, Pushmeet Kohli, Daniel Scharstein, and Sudipta Sinha. Object stereo - joint stereo matching and object segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, number 1, pages 3081 – 3088, 2011.
- [20] Aaron F Bobick and Stephen S Intille. Large Occlusion Stereo. *International Journal of Computer Vision*, 33(3):181–200, 1999.
- [21] Glenn D. Boreman. *Modulation Transfer Function in Optical and Electro-Optical Systems*. SPIE - The International Society for Optical Engineering, 2001.
- [22] J. Y. Bouguet. Camera calibration toolbox for matlab.
- [23] Michael V. Bove. Discrete fourier transform based depth-from-focus. *Image Understanding and Machine Vision*, 14:118–121, 1989.
- [24] Y Boykov, O Veksler, and R Zabih. Fast Approximate Energy Minimization via Graph Cuts. In *International Conference on Computer Vision*, pages 377–384, 1999.
- [25] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.
- [26] P. D. Burns. Sfr evaluation for digital cameras and scanners.
- [27] Claudio Caraffi, Stefano Cattani, and Paolo Grisleri. Off-Road Path and Obstacle Detection Using Decision Networks and Stereo Vision. *IEEE Transactions on Intelligent Transportation Systems*, 8(4):607–618, December 2007.
- [28] Dalit Caspi, Nahum Kiryati, Senior Member, and Joseph Shamir. Range Imaging With Adaptive Color Structured Light. 20(5):470–480, 1998.
- [29] M.Emre. Celebi, H.a. Kingravi, and F. Celiker. Fast colour space transformations using minimax approximations. *IET Image Processing*, 4(2):70, 2010.

- [30] T. Chen, H. P. A. Lensch, C. Fuchs, and H.-P. Seidel. Polarization and phase-shifting for 3d scanning of translucent objects. *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2007.
- [31] Jung-Bum Chun, Hunjoon Jung, and Chong-Min Kyung. Dynamic-Range Widening in a CMOS Image Sensor Through Exposure Control Over a Dual-Photodiode Pixel. *IEEE Transactions on Electron Devices*, 56(12):3000–3008, December 2009.
- [32] Mihai Ciuc, Adrian Capata, Valentin Mocanu, Corneliu Florea, Alexei Pososin, and Peter Corcoran. Automatic Face and Skin Beautification using Face Detection, 2010.
- [33] S. D. Cochran and G. Medioni. 3-d surface description from binocular stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:981–994, 1992.
- [34] C Connolly and T Fleiss. A study of efficiency and accuracy in the transformation from RGB to CIELAB color space. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 6(7):1046–8, January 1997.
- [35] Peter Corcoran, Petronel Bigioi, and Mircea Ionita. Enhanced real-time face models from stereo imaging, 2011.
- [36] Peter Corcoran and Gabriel Costache. Face Recognition with Combined PCA-based Datasets, 2009.
- [37] S. T. Worrall D. V. S. X. De Silva, W. A. C. Fernando and A. M. Kondo. A novel depth map quality metric and its usage in depth map coding. *3DTV conference: The true vision - capture, transmission and display of 3D video*, pages 1 – 4, 2011.
- [38] Thao Dang, Christian Hoffmann, and Christoph Stiller. Continuous stereo self-calibration by camera parameter tracking. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 18(7):1536–50, July 2009.

BIBLIOGRAPHY

- [39] R L de Queiroz. On independent color space transformations for the compression of CMYK images. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 8(10):1446–51, January 1999.
- [40] L Di Stefano, M Marchionni, and S Mattoccia. A fast area-based stereo matching algorithm. *Image and Vision Computing*, 22:983–1005, 2004.
- [41] Piotr Didyk, Tobias Ritschel, Elmar Eisemann, Karol Myszkowski, and Hans-Peter Seidel. A perceptual model for disparity. *ACM SIGGRAPH 2011 papers on - SIGGRAPH '11*, page 1, 2011.
- [42] J Ens and P Lawrence. An Investigation of Methods for Determining Depth from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(2), 1993.
- [43] T. A. Fischer and J. Holm. Electronic still picture camera spatial frequency response measurement. In *IS&T Annual Conference*, pages 626 – 630, 1994.
- [44] D A Forsyth. A Novel Algorithm for Color Constancy. *International Journal of Computer Vision*, 5(1):5 – 24, 1990.
- [45] S Fuchs and G. Hirzinger. Extrinsic and depth calibration of tof-cameras. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 464–468, 2008.
- [46] D. Geiger, B. Ladendorf, and A. Yuille. Occlusions and binocular stereo. *ECCV*, pages 425–433, 1992.
- [47] A Gentile, S Vitabile, L Verdoscia, and F Sorbello. Image Processing Chain for Digital Still Cameras based on the SIMPil Architecture. In *INternational Conference Workshops on Parallel Processing*, pages 215–222, 2005.
- [48] C Georgoulas, L Kotoulas, G.Ch. Sirakoulis, I Andreadis, and A Gasteratos. Real-time disparity map computation module. *Microprocessors and Microsystems*, 32:159–170, 2008.
- [49] K.T. Gribbon and D.G. Bailey. A Novel Approach to Real-time Bilinear Interpolation. *Second IEEE International Workshop on Electronic Design, Test and Applications*, 1:126–126, 2004.

- [50] W. E. L. Grimson. Computational experiments with a feature based stereo algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7:17–34, 1985.
- [51] M. J. Hannah. Computer matching of areas in stereo imaging. *Stanford University California, Dept. of Computer Science*, 1974.
- [52] Anwar Hasni Abu Hasan, Rostam Affendi Hamzah, and Mohd Haffiz Johar. Region of Interest in Disparity Mapping for Navigation of Stereo Vision Autonomous Guided Vehicle. *2009 International Conference on Computer Technology and Development*, pages 98–102, 2009.
- [53] C. T. E. R. Hewage and M. G. Martini. Reduced-reference quality evaluation for compressed depth maps associated with color plus depth 3D video. *IEEE International Conference on image processing*, pages 4017–4020, 2010.
- [54] Asmaa Hosni, Michael Bleier, Margrit Gelautz, and Christoph Rhemann. Local stereo matching using geodesic support weights. In *IEEE International Conference On Image Processing*, number 1, pages 2093–2096, 2009.
- [55] Martin Humenberger, Christian Zinner, Michael Weber, Wilfried Kubinger, and Markus Vincze. A fast stereo matching algorithm suitable for embedded real-time systems. *Computer Vision and Image Understanding*, 114(11):1180–1202, 2010.
- [56] Ten-lee Hwang, J J Clark, and A L Yuille. Depth Recovery Algorithm Using Defocus Information. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 476–482, 1989.
- [57] Claudia Iancu and Peter M Corcoran. A Review of Hidden Markov Models in Face Recognition. 2005.
- [58] Mircea Ionita, Istvan Andorko, and Peter Corcoran. Enhanced real-time face models from stereo imaging for gaming applications. In *IEEE Games Innovation Conference*, pages 17–24, 2009.
- [59] M. Irani and P. Anandan. *Lecture Notes in Computer Science*, volume 1064, pages 17–30. Computer Vision - ECCV '96, 1996.

BIBLIOGRAPHY

- [60] R. Janer. Four image interpolation techniques for ultrasound breast phantom data acquired using Fischer's full field digital mammography and ultrasound system (FFDMUS): a comparative approach. *IEEE International Conference on Image Processing 2005*, pages II-1238, 2005.
- [61] N. Jojic, B. Brumitt, B. Meyers, S. Harris, and T. Huang. Detection and estimation of pointing gestures in dense disparity maps. *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, pages 468-475.
- [62] T. Kahlmann, F. Remondino, and H. Ingensand. Calibration of the fast range imaging camera swissranger for use in the surveillance of the environment. *Proc. SPIE Conference on Electro-Optical Remote Sensing II*, 2006.
- [63] J Kalomiros and J A Lygouras. Hardware implementation of a stereo co-processor in a medium-scale field programmable gate array. *IET Computers & Digital Techniques*, 2(5):336 - 346, 2008.
- [64] T. Kanade, H. Kano, S. Kimura, a. Yoshida, and K. Oda. Development of a video-rate stereo machine. *Proceedings 1995 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human Robot Interaction and Cooperative Robots*, 3:95-100, 1995.
- [65] T. Kanade and M. Okutomi. A stereo matching algorithm with an adaptive window: theory and experiment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(9):920-932, 1994.
- [66] W C Kao, S H Wang, L Y Chen, and S Y Lin. Design Considerations of Color Image Processing Pipeline for Digital Cameras. *IEEE Transactions on Consumer Electronics*, 52(4):1144-1152, 2006.
- [67] Wen-chung Kao, Senior Member, Li-wei Cheng, Chen-yu Chien, and Wen-kuo Lin. Robust Brightness Measurement and Exposure Control in Real-Time Video Recording. *IEEE Transactions on Instrumentation and Measurement*, 60(4):1206-1216, 2011.
- [68] Wen-chung Kao, Sheng-hong Wang, Lien-yang Chen, and Sheng-yuan Lin. Design considerations of color image processing pipeline for digital cameras. *IEEE Transactions on Consumer Electronics*, 52(4):1144-1152, November 2006.

- [69] Dong-Sun Kim, Sang-Seol Lee, and Byeong-Ho Choi. A real-time stereo depth extraction hardware for intelligent home assistant robot. *IEEE Transactions on Consumer Electronics*, 56(3):1782–1788, August 2010.
- [70] Sung-yeol Kim, Student Member, Eun-kyung Lee, Yo-sung Ho, and Senior Member. Generation of ROI Enhanced Depth Maps Using Stereoscopic Cameras and a Depth Camera. *IEEE Transactions on Broadcasting*, 54(4):732–740, 2008.
- [71] W. N. Klarquist and A. C. Bovik. A foveated vergent active stereo system for dynamic three-dimensional scene recovery. *International Conference on Robotics and Automation*, pages 3259–3266, 1998.
- [72] Andreas Klaus, Mario Sormann, and Konrad Karner. Segment-Based Stereo Matching Using Belief Propagation and a Self-Adapting Dissimilarity Measure. In *International Conference on Pattern Recognition*, pages 18–21, 2006.
- [73] Henry Kong, Qin Sun, William Bauson, Stephen Kiselewich, Paul Ainslie, and Riad Hammoud. Disparity Based Image Segmentation For Occupant Classification. In *Computer Vision and Pattern Recognition Workshop*, number 1, pages 126 – 134, 2004.
- [74] Tetsuya Kuno and Hiroaki Sugiura. Practical color filter array interpolation part 2 with non-linear filter. *IEEE Transactions on Consumer Electronics*, 52(4):1409–1417, November 2006.
- [75] S H Lai, C W Fu, and S Chang. A Generalized Depth Estimation Algorithm with a Single Image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(4):405–411, 1992.
- [76] Pongsak Lasang, Chin Phek Ong, and Sheng Mei Shen. CFA-based Motion Blur Removal using Long / Short Exposure Pairs. *IEEE Transactions on Consumer Electronics*, 56(2):332–338, 2010.
- [77] D. Lee and T. Pavlidis. One-dimensional regularization with discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(6):822–829, 1988.

BIBLIOGRAPHY

- [78] S H Lee, N I Cho, and J Park. Simultaneous disparity estimation and object segmentation using stochastic models. *Electronics Letters*, 40(9):2–3, 2004.
- [79] Sang-Beom Lee, Kwan-Jung Oh, and Yo-Sung Ho. Segment-Based Multi-View Depth Map Estimation Using Belief Propagation from Dense Multi-View Video. *2008 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, pages 193–196, May 2008.
- [80] T M Lehmann, C Gönner, and K Spitzer. Survey: interpolation methods in medical image processing. *IEEE transactions on medical imaging*, 18(11):1049–75, November 1999.
- [81] Jim S Jimmy Li and Sharmil Randhawa. Color filter array demosaicking using high-order interpolation techniques with a weighted median filter for sharp color edge preservation. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 18(9):1946–57, September 2009.
- [82] R Lukac and K N Plataniotis. Universal demosaicking for imaging pipelines with an RGB color filter array. *The Journal of The Pattern Recognition Society*, 38:2208–2212, 2005.
- [83] Muhammad Tariq Mahmood and Tae-Sun Choi. Nonlinear approach for enhancement of image focus volume in shape from focus. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 21(5):2866–73, May 2012.
- [84] D. Marr and T. Poggio. Cooperative computation of stereo disparity. *Science*, 194(4262):283 – 287, 1976.
- [85] Rafael Mayoral and Gabriel Lera. Evaluation of correspondence errors for stereo. *Image and Vision Computing*, 24:1288–1300, 2006.
- [86] a. Miron, a. Bensrhair, a. Rogozan, and S. Ainouz. Cross-comparison census for colour stereo matching applied to intelligent vehicle. *Electronics Letters*, 48(24):1530–1532, November 2012.
- [87] Jung-Ho Moon and Soo Yul Jung. Implementation of an image stabilization system for a small digital camera. *IEEE Transactions on Consumer Electronics*, 54(2):206 – 212, May 2008.

- [88] Michele Moscaritolo, Henry Jampel, Frederick Knezevich, and Ran Zeimer. An image based auto-focusing algorithm for digital fundus photography. *IEEE transactions on medical imaging*, 28(11):1703–7, November 2009.
- [89] M Mrak, Zgaljic T., and E. Izquierdo. Influence of downsampling filter characteristics on compression performance in wavelet-based scalable video coding. *IET Image Processing*, 2(3):116–129, 2008.
- [90] Rafael Muñoz Salinas, R. Medina-Carnicer, F.J. Madrid-Cuevas, and a. Carmona-Poyato. Depth silhouettes for gesture recognition. *Pattern Recognition Letters*, 29(3):319–329, February 2008.
- [91] K Muhlmann, D Maier, J Hesser, and R Manner. Calculating Dense Disparity Maps from Color Stereo Images , an Efficient Implementation. In *IEEE Workshop on Stereo and Multi-Baseline Vision*, pages 30–36, 2001.
- [92] Karsten Muhlmann, Dennis Maier, Jurgen Hesser, and Reinhard Manner. Calculating Dense Disparity Maps from Color Stereo Images , an Efficient Implementation. In *IEEE Workshop on Stereo and Multi-Baseline Vision*, pages 30 – 36, 2001.
- [93] Yuichi Nakamura, Tomohiko Matsuura, Kiyohide Satoh, and Yuichi Ohta. Occlusion Detectable Stereo - Occlusion Patterns in Camera Matrix -. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 371–378, 1996.
- [94] Florin Nanu, Peter Corcoran, and Cosmin Stan. Methods and Apparatuses for Eye Gaze Measurement, 2011.
- [95] P. M. Narendra. Scene-Based Nonuniformity Compensation for Imaging Sensors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (1):57–61, 1982.
- [96] Catalina Neghina, Mihnea Gangea, Stefan Petrescu, Emilian David, Petronel Bigioi, Eric Zarakov, and Eran Steinberg. Detecting Facial Expressions in Digital Images, 2009.
- [97] M.A. Nuno-Maganda and M.O. Arias-Estrada. Real-Time FPGA-Based Architecture for Bicubic Interpolation : An Application for Digital Image Scaling. In

BIBLIOGRAPHY

- International Conference on Reconfigurable Computing and FPGAs*, number 1, pages 1–8, 2005.
- [98] Yuichi Ohta and Takeo Kanade. Stereo by intra- and inter-scanline search using dynamic programming. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2):139 – 154, 1985.
- [99] M. Okutomi and T. Kanade. A multiple-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4):353–363, April 1993.
- [100] Masatoshi Okutomi and Takeo Kanade. A locally adaptive window for signal matching. *International Journal of Computer Vision*, 7(2):143 – 162, 1992.
- [101] Manu Parmar and Stanley J Reeves. Selection of optimal spectral sensitivity functions for color filter arrays. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 19(12):3190–203, December 2010.
- [102] Soo-chang Pei and Yu-ying Wang. Census-based Vision for Auditory Depth Images and Speech Navigation of Visually Impaired Users. *IEEE Transactions on Consumer Electronics*, 57(4):1883–1890, 2011.
- [103] Alex P Pentland. A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (4):523 – 531, 1987.
- [104] Georg Petschnigg, Richard Szeliski, Maneesh Agrawala, Michael Cohen, Hugues Hoppe, and Kentaro Toyama. Digital photography with flash and no-flash image pairs. *ACM Transactions on Graphics*, 23(3):664, August 2004.
- [105] M T Pourazad, P Nasiopoulos, and R K Ward. An H . 264-based Scheme for 2D to 3D Video Conversion. *IEEE Transactions on Consumer Electronics*, 55(2):742–748, 2009.
- [106] M T Pourazad, P Nasiopoulos, and R K Ward. An H.264-based scheme for 2D to 3D video conversion. *IEEE Transactions on Consumer Electronics*, 55:742–748, 2009.
- [107] L. H. Quam. Hierarchical warp stereo. *Image Understanding Workshop*, 14:149–155, 1984.

- [108] A N Rajagopalan and S Chaudhuri. A Variational Approach to Recovering Depth From Defocused Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(10):1158–1164, 1997.
- [109] A N Rajagopalan, S Chaudhuri, and U Mudenagudi. Depth Estimation and Image Restoration Using Defocused Stereo Pairs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11):1521–1525, 2004.
- [110] R Ramanath, W E Snyder, G L Bilbro, and W A Sander. Demosaicking methods for Bayer color arrays. *Journal of Electronic Imaging*, 11(3):306–315, 2002.
- [111] Rajeev Ramanath, Wesley E. Snyder, Youngjun Yoo, and Mark S Drew. Color Image Processing Pipeline [. *IEEE Signal Processing Magazine*, pages 34–43, 2005.
- [112] Jiann-Yeou Rau and Po-Chia Yeh. A semi-automatic image-based close range 3D modeling pipeline using a multi-camera configuration. *Sensors (Basel, Switzerland)*, 12(8):11271–93, January 2012.
- [113] S Roy. Stereo Without Epipolar Lines : A Maximum-Flow Formulation. *International Journal of Computer Vision*, 34(2/3):147–161, 1999.
- [114] Sebastien Roy. Stereo Without Epipolar Lines : A Maximum-Flow Formulation. *International Journal of Computer Vision*, 34:147–161, 1999.
- [115] D. Scharstein. Matching images by comparing their gradient fields. In *International Conference on Pattern Recognition*, volume 1, pages 572–575. IEEE Comput. Soc. Press, 1994.
- [116] D. Scharstein and R. Szeliski. High-accuracy stereo depth maps using structured light. *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, pages I–195–I–202.
- [117] D Scharstein and R Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(13):7–42, 2002.

BIBLIOGRAPHY

- [118] Daniel Scharstein and Richard Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*, 47(1):7–42, 2002.
- [119] Daniel Scharstein and Richard Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *Journal of Computer Vision*, 2002.
- [120] Korbinian Schmid, Heiko Hirschmüller, Andreas Dömel, Iris Grix, Michael Suppa, and Gerd Hirzinger. View Planning for Multi-View Stereo 3D Reconstruction Using an Autonomous Multicopter. *Journal of Intelligent & Robotic Systems*, 65(1-4):309–323, August 2011.
- [121] J. Shah. A nonlinear diffusion model for discontinuous disparity and half-occlusions in stereo. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 34–40. IEEE Comput. Soc. Press, 1993.
- [122] S. Shimizu, T. Kondo, T. Kohashi, M. Tsurata, and T. Komuro. A new algorithm for exposure control based on fuzzy logic for video cameras. *IEEE Transactions on Consumer Electronics*, 38(3):617 – 623, 1992.
- [123] J Paul Siebert. *AN INTRODUCTION TO 3D COMPUTER VISION TECHNIQUES*.
- [124] Aljoscha Smolic, Karsten Mueller, Philipp Merkle, Christoph Fehn, Peter Kauff, Peter Eisert, and Thomas Wiegand. 3D Video and Free Viewpoint Video - Technologies , Applications and MPEG Standards. In *IEEE International Conference on Multimedia and Expo*, pages 2161–2164, 2006.
- [125] Eran Steinberg, Peter Corcoran, and Petronel Bigioi. Perfecting the effect of flash within an image acquisition device using face detection, 2010.
- [126] Eran Steinberg, Yuri Prilutsky, Peter Corcoran, and Petronel Bigioi. Method of Improving orientation and color balance of digital images using face detection algorithm, 2008.
- [127] Eran Steinberg, Yuri Prilutsky, Peter Corcoran, and Petronel Bigioi. Perfecting of digital image capture parameters within acquisition devices using face detection, 2009.

- [128] M. Subbarao and N. Gurumoorthy. Depth recovery from blurred edges. In *Computer Society Conference on Computer Vision and Pattern Recognition*, number 1987, pages 498–503, 1988.
- [129] J Sun, S B Kang, Z B Xu, X Tang, and H Y Shum. Flash Cut : Foreground Extraction with Flash and No-flash Image Pairs. In *IEEE Conferene onComouter Vision and Pattern Recognition*, pages 1–8, 2007.
- [130] Jian Sun, Nan-ning Zheng, and Heung-Yeung Shum. Stereo Matching Using Belief Propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7):787–800, 2003.
- [131] Legong Sun and Zheng Mao. An improved Normalized Cross Correlation algorithm for object tracking. *IEEE 10th INTERNATIONAL CONFERENCE ON SIGNAL PROCESSING PROCEEDINGS*, 2:1267–1270, October 2010.
- [132] Richard Szeliski. Solving random-dot stereograms using the heat equation. In *Computer Vision and Pattern Recognition*, pages 284 – 288, 1984.
- [133] Yuichi Taguchi, Bennett Wilburn, and C. Lawrence Zitnick. Stereo reconstruction with mixed pixels using adaptive over-segmentation. *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2008.
- [134] S Takezawa and G Dissanayake. Simultaneous Localisation and Mapping Problems in Indoor Environments with StereoVision. *Annual Conference of IEEE Industrial Electronics Society*, pages 1896–1901, 2005.
- [135] Hai Tao, Harpreet Sawhney, and Rakesh Kumur. A Global Matching Framework for Stereo Computation. In *Eight IEEE International Conference on Computer Vision*, number 2, pages 532 – 539, 2001.
- [136] H Truong, S. Abdallah, S. Rogeaux, and A. Zelinsky. A novel mechanism for stereo active vision. *Australian Conference on Robotics and Automation*, 2000.
- [137] Filareti Tsalakanidou, Frank Forster, Sotiris Malassiotis, and Michael G. Strintzis. Real-time acquisition of depth and color images using structured light and its application to 3D face recognition. *Real-Time Imaging*, 11(5-6):358–369, October 2005.

BIBLIOGRAPHY

- [138] O. Veksler. Stereo matching by compact windows via minimum ratio cycle. In *IEEE International Conference on Computer Vision*, volume 1, pages 540–547. IEEE Comput. Soc, 2001.
- [139] Dingrui Wan and Jie Zhou. Multiresolution and wide-scope depth estimation using a dual-PTZ-camera system. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 18(3):677–82, March 2009.
- [140] Liang Wang and Ruigang Yang. Global stereo matching leveraged by sparse ground control points. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3033 – 3040, 2011.
- [141] Shou-Der Wei and Shang-Hong Lai. Fast template matching based on normalized cross correlation with adaptive multilevel winner update. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 17(11):2227–35, November 2008.
- [142] J. Weng, N. Ahuja, and T S Huang. Matching two perspective views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(8):806 – 825, 1992.
- [143] Li Xu and Jiaya Jia. Stereo matching: an outlier confidence approach. *Computer Vision*, 5305:775–787, 2008.
- [144] Qingxiong Yang, Liang Wang, Ruigang Yang, Henrik Stewénus, and David Nistér. Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling. *IEEE transactions on pattern analysis and machine intelligence*, 31(3):492–504, March 2009.
- [145] Yibing Yangt, Alan Yuillet, and Jie Lu. Local, global and multilevel stereo matching. In *Computer Vision and Pattern Recognition*, pages 274–279, 1993.
- [146] P T Yap and P Raveendran. Image focus measure based on Chebyshev moments. *IEE Proceedings - Vision, Image and Signal Processing*, 151(2):128 – 136, 2004.
- [147] Kuk-jin Yoon and In-so Kweon. Locally adaptive support-weight approach for visual correspondence search. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, 2:924–931.

- [148] Ramin Zabih and John Wood. Non-parametric Local Transforms for Computing Visual Correspondence. In *Proceedings of European Conference on Computer Vision*, number May, pages 151–158, 1994.
- [149] Hidemori Zen, T Koizumi, H Yamamoto, and I Kimura. A new digital signal processor for progressive scan CCD. *IEEE Transactions on Consumer Electronics*, 44(2):289 – 296, 1998.
- [150] Guofeng Zhang, Jiaya Jia, Tien-Tsin Wong, and Hujun Bao. Consistent depth maps recovery from a video sequence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(6):974–988, 2009.
- [151] Jie Zhao, Jiali Chai, and Guozun Men. A Fast Quasi-Dense Matching Method. *2009 International Asia Symposium on Intelligent Interaction and Affective Computing*, (1):100–103, December 2009.
- [152] Jiejie Zhu, Liang Wang, Ruigang Yang, James E Davis, and Zhigeng Pan. Reliability Fusion of Time-of-Flight Depth and Stereo for High Quality Depth Maps. *IEEE transactions on pattern analysis and machine intelligence*, 33(7):1400–1414, August 2011.
- [153] Barbara Zitová and Jan Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, October 2003.