

# **Free-formed Surface Mirrors in Computer Vision Systems**

## **Dissertation**

zur Erlangung des Grades eines  
Doktors der Naturwissenschaften  
der Fakultät für Physik  
der Eberhard-Karls-Universität zu Tübingen

vorgelegt von

**Alexander Würz-Wessel**

aus Albstadt

2003

Tag der mündlichen Prüfung:

Dekan:

1. Berichterstatter:

2. Berichterstatter:

20. Juni 2003

Prof. Dr. Herbert Müther

Prof. Dr. Klaus Werner

Prof. Dr. David Wharam

# Acknowledgments

---

I am very thankful to

- my advisor Dr. Fridtjof Stein at the DaimlerChrysler Research Center, Esslingen, who took the time for extensive discussions, gave valuable advice, and offered ample guidance. Beyond that, he provided an optimal work environment for this thesis with regard to all aspects, and the necessary freedom and support for this and many other ideas.
- my PhD advisor Prof. Dr. Klaus Werner at the Institute for Astronomy and Astrophysics of the University of Tübingen. He entered into productive discussions, provided lots of hints and help with the physical details and supported me managing the administrative issues.
- my supervisors Stefan Hahn and Hansgeorg Metzler at the DaimlerChrysler Research Center, Esslingen, for initiating and maintaining the project in which this thesis was developed.
- Hansgeorg Leis, Thomas Winkler, and Dr. Jörg Frauendiener for the readily offered support with the ASAP simulations, the air-pressure deformation simulations, and the differential geometry calculations, respectively.
- my students Pascal Paysan, Stefan Moosbrugger, and Daniel Schowalter for the collaboration and the implementation of pieces of this thesis in their internships and master's thesis.
- my colleagues in the environment perception team, headed by Dr. Uwe Franke, and in the RIC/AP and RIC/AA departments, for intensive discussions, an open-minded and friendly atmosphere, and the self-evident team-work.
- Dr. Stefan Gehrig for proof-reading this thesis.
- Dr. Jochen L. Deetjen and Dr. Jörn Wilms for providing  $\LaTeX$  styles and support for the layout of this thesis.
- my parents for supporting me and my academic education.
- at last, but not least, my wife Anja for the lasting support and the motivation to work purposively on this thesis.



# Zusammenfassung

---

Die Teilnahme am heutigen Straßenverkehr ist anstrengend und gefährlich. Obwohl die Fahrzeuge vermehrt passive Sicherheit bieten, werden jeden Tag Menschen getötet und verletzt. Die Möglichkeiten der passiven Sicherheitssysteme sind nahezu ausgeschöpft, so daß aktive Systeme notwendig geworden sind, um allen Verkehrsteilnehmern eine höhere Sicherheit zu bieten. Aktive Assistenzsysteme benötigen eine ins Fahrzeug eingebaute Sensorik, um den aktuellen Zustand der Fahrzeugumgebung zu erfassen. Aufgrund der Echtzeitauswertung der Daten wird entweder der Fahrer vom System gewarnt, oder das System greift z.B. durch Steuern oder Bremsen in das Fahrzeug ein.

Die Erfassung der Umgebung für eine Maschine bedingt die dreidimensionale Vermessung des Fahrzeugaußenraums. Sensoren, die eine direkte Vermessung bieten, sind z.B. Radar- oder Lidar-Systeme. Kameras, als passive Sensoren, müssen mindestens zu einem Stereosystem kombiniert werden, um die Ableitung der Entfernung eines Objektes zu ermöglichen. Bilder sind jedoch eine Quelle vielfältiger Informationen, und deshalb werden Kameras für die maschinelle Umgebungserfassung eingesetzt. Ein Beispiel für ein System ist das automatisierte Staufahren auf der Basis der Stereobildverarbeitung. Dies ist die Rahmenanwendung dieser Dissertation.

Menschen mit ihrem Augenpaar verwenden ein hohes Maß an modellbasiertem Wissen für die Wahrnehmung einer Situation. Die heutigen computerbasierten Anwendungen können eine solche Herangehensweise nicht realisieren. Das Fehlen eines Umgebungsmodells kann zu einer fehlerhaften Darstellung des Fahrzeugaußenraums führen, besonders dann, wenn sich wiederholende Muster im Bild vorhanden sind. Ein zeitweiliger Verlust eines Bildes, z.B. durch den Scheibenwischer, oder der komplette Ausfall einer Kamera beeinträchtigen die Verlässlichkeit des Systems. All diese Probleme könnten durch das Hinzufügen einer weiteren Kamera gelöst werden. Kameras sind jedoch teuer und der mögliche Bauraum im Fahrzeug ist begrenzt.

Das Stereokamerasystem für das Fahrzeugfolgefahren ist hinter der Windschutzscheibe angebracht, so daß die Kameras nach vorne schauen. Von dieser Position aus bilden die Kameras auch die Motorhaube des Fahrzeugs ab. Die Haube ist eine glatte, glänzend lackierte Freifläche, die die Umgebung reflektiert, die auch direkt von der Kamera gesehen wird. Diese Reflexionen beinhalten im wesentlichen die gleichen Informationen, die auch eine zusätzliche Kamera liefern würde. Durch die Krümmung der Fläche ist das Bild jedoch verzerrt, und die Lackierung führt zu einer Verminderung des Kontrasts.

In dieser Dissertation wird gezeigt, daß Reflexionen auf Freiflächen in der Bildverarbeitung genutzt werden können. Außerdem werden die dazu notwendigen Methoden entwickelt.

Die Dissertation besteht aus drei Teilen:

Der erste Teil beginnt mit einer detaillierten Analyse der Probleme, die sich durch die Verwendung einer Freifläche in einem optischen System ergeben. Dann wird die relevante Literatur vorgestellt. Darin zeigt sich, daß bisher Reflexionen auf Freiflächen nur als Fehlerquelle für Bildverarbeitungsalgorithmen betrachtet wurden. Zur Verwendung der in den Reflexionen beinhalteten Information gibt es dementsprechend keine Methoden. Die Entwicklung möglicher Methoden beschließt den ersten Teil.

Die Kalibrierung der Kamera und die Vermessung der Oberflächen sind fundamentale Anforderungen für jede der Rekonstruktionsmethoden und bilden den zweiten Teil dieser Arbeit. Es wird ein neues Verfahren zur bildbasierten Datengewinnung in der Kalibrierung präsentiert, das die existierende Kalibriertechnik in mehreren Stufen verbessert. Es ermöglicht zum einen die genauere Positionsbestimmung der Bildmerkmale und zum anderen die Verwendung von nur teilweise sichtbaren Kalibrierobjekten. Dies dient auch vielen anderen Anwendungen. Auf der Basis der Kalibriermethode wird eine neue Technik zur Vermessung der Oberflächentopographie eingeführt. Diese arbeitet mit einer Bilderserie und ist auf reflektierende Flächen anwendbar.

Der letzte Teil der Dissertation ist zwei Methoden zur Informationsextraktion aus den Reflexionsbildern gewidmet. Das erste Verfahren ist in sofern nicht klassisch als keine Rekonstruktion des Bildes zu einer Lochkameraperspektive stattfindet. Es basiert auf der Erweiterung einer linearen Beschränkung der Stereobildverarbeitung zu einem generelleren nicht-linearen Verständnis der Abbildungsgeometrie. Die zweite Methode ist eine klassische Rekonstruktion des Bildes durch den Vergleich der Reflexionsrichtungen der Freifläche mit denen eines virtuellen ebenen Spiegels. Dieses Verfahren wurde auf zwei Arten realisiert. Einmal wurde ein tabellarischer bzw. ein polynomialer Ansatz gewählt, der auf der zentralen Recheneinheit (CPU) des Computers verarbeitet wird. Die zweite Implementierung benutzt den Hauptprozessor der Grafikkarte (GPU) zur Bildrekonstruktion. Die Verwendung der GPU bietet viele neue Möglichkeiten zur Bildverarbeitung im allgemeinen. Beide Verfahren wurden mit Bildern aus dem Versuchsfahrzeug getestet. Für eine Methode wurde eine manuelle Abstandsmessung durchgeführt.

In dieser Dissertation wird die übliche Annahme, daß Reflexionen auf Freiflächen nur Fehlerquellen in der Bildverarbeitung darstellen, widerlegt. Reflexionen auf Freiflächen sind Informationsquellen, selbst wenn die Fläche nicht als Element eines optischen Systems konzipiert wurde.

# Abstract

---

Participating in today's road traffic is tiring and dangerous. Even though vehicles provide more and more passive safety, fatalities and injuries occur every day. The passive safety systems have reached a level of sophistication that active systems become necessary to provide additional safety to all traffic participants. An active assistance system requires on-board sensors to perceive the actual state of the vehicle's environment. Upon the real-time evaluation of the data, the system either warns the driver about a danger ahead or intervenes on its own, e.g. by steering or braking.

The perception of the environment for a machine implies the three-dimensional reconstruction of the surrounding. Sensors that directly provide the Euclidean measurement are e.g. radar or lidar. Cameras, as passive sensors, must be combined at least to a stereo system in order to enable the derivation of the distance of a certain object. However, images are a source of manifold information and hence camera systems are used for machine perception. One example of such a system is the autonomous stop-and-go driving based on stereo-image processing. This is the framework application of this thesis.

Humans with their pair of eyes use an extensive model-based knowledge to support their perception. Today's stereo-vision systems cannot depend on such an approach. The lack of an environment model may lead to an erroneous scene representation, e.g. in the presence of repetitive patterns. The temporary loss of an image, e.g. caused by a passing windshield wiper, or the long-termed loss of a camera reduce the reliability of the application. All of these problems could be solved by adding a third camera to the system. However, cameras are expensive and the space in the vehicle is limited.

The stereo-camera system for the vehicle-following application is mounted behind the windshield of the car, looking ahead. At this position the cameras also take an image of the engine hood of the car. This is a free-formed, shiny varnished, and smooth car-body part that reflects the scene which is observed directly by the cameras. This reflection essentially contains the same information that would be provided by an additional camera. However, the reflected image is distorted, due to the curvature of the surface, and contrast-diminished by the varnishing.

To prove the feasibility of using reflections on free-formed surfaces in image processing is the topic of this thesis. The development of methods to extract the information is presented as well.

The thesis is divided into three parts:

The first part starts with a detailed analysis of the problems arising from the use of free-formed surfaces in an optical system. Subsequently, a review of the related work is presented. It shows that so far reflections on free-formed surfaces are only considered to be a source of error for image-processing algorithms. Accordingly, methods to incorporate the contained information into an application are not available. The development of feasible approaches to reconstruct the distorted image-information is the conclusion of the first part.

The camera calibration and the surface reconstruction, the basic requirements for each of the reconstruction methods, form the second part of the thesis. A new method for the imaged-based data-acquisition in the calibration effort is presented. It improves the existing calibration technique in multiple steps. Not only the determination of the image features is enhanced but a method to recover only partially visible calibration targets is introduced. This serves for many other applications as well. On the basis of the camera calibration technique, a new method to recover the surface's topography is elaborated. It operates on a series of images and is suitable for reflective surfaces.

The last part of the thesis is devoted to the detailed presentation of two methods for the extraction of the contained information. The first method is a non-classical approach in the sense that the distorted image is not reconstructed to a pinhole-camera view. It depends on an extension of a usually applied linear constraint in stereo-image processing to a more general non-linear understanding of the restriction. This forms the basis of the direct information extraction. The second method is a classical reconstruction of the image, based on the comparison of the reflection directions of the free-formed surface and a virtual planar mirror. This method is implemented in two ways, both obeying the real-time requirement of the application. The first implementation is a table-based or polynomial-based reconstruction performed by the central processing unit (CPU) of the computer. The second implementation uses the graphics processing unit (GPU), i.e. the main computational unit on the computer's graphics-card, to reconstruct the image. The usage of the GPU offers manifold new possibilities for image processing in general. Both implementations are tested on images acquired in the experimental vehicle. For one method a coarse distance determination is conducted.

In this thesis the common assumption that reflections on a free-formed surface are only a source of error for image-processing algorithms is refuted. Reflections on a free-formed surface, even if it is not designed to serve as a part in an optical system, provide useful information.



# Contents

---

|          |   |           |
|----------|---|-----------|
| <b>1</b> | <b>Introduction</b>                           | <b>1</b>  |
| 1.1      | Motivation . . . . .                          | 1         |
| 1.2      | Objective of this Thesis . . . . .            | 3         |
| 1.3      | Contribution of this Thesis . . . . .         | 5         |
| 1.4      | Thesis Overview . . . . .                     | 6         |
| <b>2</b> | <b>Statement and Analysis of the Problem</b>  | <b>8</b>  |
| 2.1      | Principles of Stereo Vision . . . . .         | 8         |
| 2.2      | System Environment . . . . .                  | 11        |
| 2.2.1    | Hardware . . . . .                            | 11        |
| 2.2.2    | Software . . . . .                            | 12        |
| 2.2.3    | Image Processing . . . . .                    | 13        |
| 2.3      | Statement of the Problem . . . . .            | 16        |
| 2.4      | Processes and Models . . . . .                | 20        |
| 2.4.1    | Physical Processes . . . . .                  | 20        |
| 2.4.2    | Optical Models . . . . .                      | 21        |
| 2.4.3    | Perspective Projection . . . . .              | 22        |
| 2.5      | Conclusions . . . . .                         | 23        |
| <b>3</b> | <b>Related Work</b>                           | <b>24</b> |
| 3.1      | n-ocular Systems . . . . .                    | 24        |
| 3.2      | Catadioptric Systems . . . . .                | 26        |
| 3.2.1    | Stereo using one Camera . . . . .             | 26        |
| 3.2.2    | Image Formation . . . . .                     | 28        |
| 3.2.3    | Omnidirectional Vision Systems . . . . .      | 31        |
| 3.2.4    | Calibration of Catadioptric Systems . . . . . | 33        |
| 3.3      | Reflections . . . . .                         | 35        |
| 3.4      | Image Display . . . . .                       | 38        |
| 3.5      | Conclusions . . . . .                         | 39        |

---

|          |  |           |
|----------|--|-----------|
| <b>4</b> | <b>Image and Information Reconstruction</b>                          | <b>41</b> |
| 4.1      | Model Assumptions . . . . .  | 41        |
| 4.2      | Image Reconstruction Strategies . . . . .                            | 43        |
| 4.2.1    | Reconstruction within the Image Plane . . . . .                      | 43        |
| 4.2.2    | Ray-tracing . . . . .  | 44        |
| 4.2.3    | Comparison of Lines . . . . .  | 51        |
| 4.3      | Information Reconstruction Strategy . . . . .                        | 53        |
| 4.4      | Conclusions . . . . .  | 55        |
| <b>5</b> | <b>Camera Calibration</b>  | <b>56</b> |
| 5.1      | Introduction . . . . .   | 56        |
| 5.2      | Related Work . . . . .   | 57        |
| 5.3      | Mathematical Methods . . . . .                                       | 62        |
| 5.3.1    | Affine Transformation . . . . .                                      | 62        |
| 5.3.2    | Quaternions . . . . .  | 63        |
| 5.3.3    | Homogenous Coordinate Representation . . . . .                       | 63        |
| 5.4      | Basic Camera Models . . . . .  | 64        |
| 5.5      | Calibration Algorithms . . . . .                                     | 67        |
| 5.5.1    | Tsai/Lenz Calibration . . . . .                                      | 67        |
| 5.5.2    | Bouguet/Heikkilä Calibration . . . . .                               | 70        |
| 5.6      | Enhancement of Bouguet's Calibration Approach . . . . .              | 76        |
| 5.6.1    | Feature Detection . . . . .  | 77        |
| 5.7      | Results . . . . .  | 84        |
| 5.8      | Discussion . . . . .   | 88        |
| <b>6</b> | <b>Reconstruction of the Reflective Surface</b>                      | <b>90</b> |
| 6.1      | Introduction . . . . .   | 90        |
| 6.2      | Related Work . . . . .   | 91        |
| 6.3      | Mathematical Geometry Representation . . . . .                       | 94        |
| 6.4      | Geometry Reconstruction from a Sequence of Images . . . . .          | 98        |
| 6.4.1    | Expansion of the Checkerboard Camera-Calibration Technique . . . . . | 99        |
| 6.4.2    | Surface Reconstruction by Distance Minimization . . . . .            | 102       |
| 6.5      | Results . . . . .  | 106       |
| 6.6      | Discussion . . . . .   | 107       |

|          |   |            |
|----------|---|------------|
| <b>7</b> | <b>Non-classical Information Reconstruction</b>           | <b>112</b> |
| 7.1      | Introduction . . . . .                                    | 112        |
| 7.2      | Epipolar Geometry on Free-Formed Surfaces . . . . .       | 113        |
| 7.2.1    | Geometric Setup . . . . .                                 | 113        |
| 7.2.2    | The Law of Reflection . . . . .                           | 115        |
| 7.2.3    | Index Notation and Projection . . . . .                   | 115        |
| 7.2.4    | Linearized Law of Reflection . . . . .                    | 116        |
| 7.2.5    | Results . . . . .   | 119        |
| 7.3      | Simulations . . . . .                                     | 120        |
| 7.3.1    | Simulation setup . . . . .                                | 120        |
| 7.3.2    | Results . . . . .   | 120        |
| 7.4      | Discussion . . . . .                                      | 124        |
| <b>8</b> | <b>Comparison of Lines and Reconstruction Results</b>     | <b>125</b> |
| 8.1      | Comparison of Lines . . . . .                             | 125        |
| 8.2      | Real-time Implementation . . . . .                        | 127        |
| 8.2.1    | Implementation on Common Computer Hardware . . . . .      | 128        |
| 8.2.2    | Implementation on Specialized Graphics Hardware . . . . . | 129        |
| 8.3      | Experimental Restrictions . . . . .                       | 132        |
| 8.3.1    | Deformation of the Hood . . . . .                         | 133        |
| 8.3.2    | Vibrations . . . . .                                      | 133        |
| 8.3.3    | Soiled Surfaces . . . . .                                 | 135        |
| 8.3.4    | Varnished Surfaces . . . . .                              | 138        |
| 8.4      | Results . . . . .   | 140        |
| 8.5      | Discussion . . . . .                                      | 146        |
| <b>9</b> | <b>Conclusions and Outlook</b>                            | <b>148</b> |
| 9.1      | Summary . . . . .   | 148        |
| 9.2      | Contributions of this Thesis . . . . .                    | 150        |
| 9.3      | Conclusions . . . . .                                     | 152        |
| 9.4      | Future Work . . . . .                                     | 152        |
| 9.5      | Outlook . . . . .   | 153        |
| <b>A</b> | <b>CAD Data Representation of the Engine Hood</b>         | <b>162</b> |
| <b>B</b> | <b>Air Pressure Deformation of the Hood</b>               | <b>167</b> |

# List of Figures

---

|      |  |    |
|------|--|----|
| 1.1  | A Stereo-Camera Image-Pair with Repetitive Patterns . . . . .    | 3  |
| 1.2  | A Traffic Scene Reflected in the Hood of a Car . . . . .         | 5  |
| 1.3  | A Flipped Presentation of the Reflected Scene . . . . .          | 6  |
| 2.1  | Epipolar Geometry in Stereo Vision . . . . .                     | 9  |
| 2.2  | The Notion of Disparity in Image Processing . . . . .            | 10 |
| 2.3  | The Experimental Car with its Sensors, Actuators and Computers   | 12 |
| 2.4  | A Multiple Matching Hypotheses Situation . . . . .               | 17 |
| 2.5  | Distances in a Multiple Matching Hypotheses Situation . . . . .  | 18 |
| 2.6  | Temporary Loss of an Image . . . . .                             | 18 |
| 2.7  | Formation of a Virtual Camera by a Planar Mirror . . . . .       | 21 |
| 2.8  | A Curved Surface in General Does Not Create a Virtual Camera     | 22 |
| 3.1  | Formation of a Virtual Viewpoint . . . . .                       | 29 |
| 3.2  | Automatic Rectification of a Plane by a Mirror . . . . .         | 31 |
| 3.3  | Calibration Image of a Parabolic Omnivision Camera . . . . .     | 33 |
| 4.1  | Restricted Object Location Space . . . . .                       | 42 |
| 4.2  | Direct, Reflected, and Virtual View on a Checkerboard Wall . . . | 44 |
| 4.3  | A Simple Ray-Tracing Approach . . . . .                          | 45 |
| 4.4  | Distribution of the Rays in the ASAP Simulation . . . . .        | 46 |
| 4.5  | Topview of the Rays Traced in ASAP . . . . .                     | 47 |
| 4.6  | A Sideview of the Rays Traced in ASAP . . . . .                  | 48 |
| 4.7  | The Rays on the Destination Wall in the Simulation A . . . . .   | 49 |
| 4.8  | The Bundle Focal Points in Simulation A . . . . .                | 49 |
| 4.9  | The Rays on the Destination Wall in Simulation B . . . . .       | 50 |
| 4.10 | The Bundle Focal Points in Simulation B . . . . .                | 50 |
| 4.11 | Reconstruction by the Comparison of Lines . . . . .              | 52 |
| 5.1  | The Pinhole and the Focusing Screen Camera. . . . .              | 65 |
| 5.2  | The Tsai/Lenz-Calibration . . . . .                              | 67 |
| 5.3  | Radial-asymmetric and Tangential Distortions . . . . .           | 70 |
| 5.4  | Principle Point Location and Affinity Distortion . . . . .       | 71 |
| 5.5  | The Complete Distortion Model . . . . .                          | 72 |
| 5.6  | Calibration Matching Templates . . . . .                         | 78 |

|      |  |     |
|------|--|-----|
| 5.7  | Image of the Correlation Coefficients . . . . .                    | 79  |
| 5.8  | Hough Line Representation in Two Dimensions . . . . .              | 80  |
| 5.9  | Extracted Checkerboard Feature Points . . . . .                    | 81  |
| 5.10 | Original and Reconstructed Image of the Bouguet Calibration Rig    | 82  |
| 5.11 | Dynamically Generated Matching Templates for Corners . . . . .     | 83  |
| 5.12 | Distance Measurement Depending on the Scaling Factor . . . . .     | 87  |
| 6.1  | Acquisition Scheme for the Hood Surface Data . . . . .             | 100 |
| 6.2  | Sum of squared Euclidean distance from the line . . . . .          | 104 |
| 6.3  | Recovery of the Surface Constraining the Point to the Ray of Sight | 105 |
| 6.4  | Distribution of the Control Points for the Surface Reconstruction  | 108 |
| 6.5  | Scalar Representation of the Surface . . . . .                     | 108 |
| 6.6  | Three-dimensional Surface Point Distribution . . . . .             | 109 |
| 6.7  | Sideview and Topview of the Surface . . . . .                      | 109 |
| 6.8  | Surface Fit to the Data Points . . . . .                           | 110 |
| 6.9  | Recovered Surface Patch . . . . .                                  | 110 |
| 6.10 | Distribution of the Recovered Surface Points within the Image .    | 111 |
| 7.1  | Basic Geometric Setup of the Reflection Case . . . . .             | 114 |
| 7.2  | Setup of the Epipolar Geometry Simulation . . . . .                | 121 |
| 7.3  | Results of the Epipolar Geometry Simulation . . . . .              | 122 |
| 7.4  | Distance Relationship along the Epipoles . . . . .                 | 123 |
| 8.1  | Errors in the Reconstructed Object Location . . . . .              | 126 |
| 8.2  | Pipeline Concept of a Graphics Processing Unit (GPU) . . . . .     | 130 |
| 8.3  | Setup to Determine the Influence of Vibrations . . . . .           | 134 |
| 8.4  | Vibrations of the Camera System to a Fixed Point on a Wall . . .   | 136 |
| 8.5  | Vibrations of the Left Camera to an External and to a Hood Point   | 137 |
| 8.6  | Direct View and Intensity Histogram . . . . .                      | 139 |
| 8.7  | Direct View Region and its Histogram . . . . .                     | 139 |
| 8.8  | Reflected View Region and its Histogram . . . . .                  | 139 |
| 8.9  | A Typical Traffic Scene in a Vehicle Following Application . . .   | 140 |
| 8.10 | Error Distribution of the Comparison-of-Lines Method . . . . .     | 142 |
| 8.11 | A CPU Image Reconstruction . . . . .                               | 143 |
| 8.12 | The Surface as Viewed by the Camera . . . . .                      | 144 |
| 8.13 | A GPU Image Reconstruction . . . . .                               | 144 |
| 8.14 | Distance Evaluation with the Stereo-Vision System . . . . .        | 145 |
| A.1  | A CAD Data Representation of the Engine Hood from the Front        | 163 |
| A.2  | A CAD Data Representation of the Engine Hood from the Side .       | 164 |
| A.3  | A CAD Data Representation of the Engine Hood from the Top .        | 165 |
| A.4  | A CAD Data Representation of the Central Engine Hood . . . . .     | 166 |
| B.1  | Air Pressure Deformation of the Hood: Top View . . . . .           | 168 |
| B.2  | Air Pressure Deformation of the Hood: Side View . . . . .          | 169 |

## List of Tables

---

|     |   |     |
|-----|---|-----|
| 5.1 | Calibration Test Sets . . . . .                             | 84  |
| 5.2 | IC2 (1) Calibration Results . . . . .                       | 85  |
| 5.3 | IC2 (2) Calibration Results . . . . .                       | 85  |
| 5.4 | IC2 (3) Calibration Results . . . . .                       | 86  |
| 5.5 | IC2 (4) Calibration Results . . . . .                       | 86  |
| 5.6 | UTA Calibration Results . . . . .                           | 86  |
| 6.1 | Surface Data Point Statistics . . . . .                     | 106 |
| 8.1 | Intrinsic Camera Parameters: Original and Virtual . . . . . | 141 |

## Abbreviations and Symbols

---

The following abbreviations, symbols and conventions will be used throughout the thesis.

Scalars are represented by a minor letter, e.g.  $a$ , vectors either by a bold-faced one, e.g.  $\mathbf{a}$ , or by an index notation, e.g.  $a_i$ , a matrix by a capital bold-faced one, e.g.  $\mathbf{A}$ , and a transposed matrix by  $\mathbf{A}^T$ .  $\mathbf{a}'$  is a point in homogenous coordinates.

Pixel coordinates are defined such that  $(0,0)$  is the center of the upper left pixel of the image. Therefore,  $(n_u - 1, 0)$  is the upper right corner and  $(n_u - 1, n_v - 1)$  the lower right corner of the image.

|      |  |
|------|--|
| ASAP | Advanced System Analysis Program                       |
| CAD  | Computer Aided Design                                  |
| CCD  | Charged Coupled Device                                 |
| CMOS | Complementary Metal Oxide Semiconductor                |
| CPU  | Central Processing Unit                                |
| CV   | Computer Vision  |
| DMA  | Direct Memeory Access                                  |
| GPU  | Graphics Processing Unit                               |
| IC2  | for 'I see too', Vision-based Vehicle Following System |
| LCD  | Liquid Crystal Display                                 |
| LED  | Light Emission Diode                                   |
| UTA  | Urban Traffic Assistant                                |

|                |  |
|----------------|--|
| $\mathbf{q}$   | quaternion                               |
| $\mathbf{T}_i$ | matrix of the internal camera parameters |
| $\mathbf{T}_e$ | matrix of the external camera parameters |
| $\mathcal{R}$  | retina                                   |
| $\mathbf{w}$   | point in the world coordinate system     |
| $\mathbf{m}$   | point in the image coordinate system     |

|                      |  |
|----------------------|--|
| $u$                  | horizontal image coordinate                                  |
| $v$                  | vertical image coordinate                                    |
| $b$                  | baseline length of stereo camera system                      |
| $d$                  | disparity  |
| $k_u, k_v$           | pixel-size on the retina in $u$ - and $v$ -direction         |
| $f$                  | focal length   |
| $\alpha_u, \alpha_v$ | focal length measured in pixels, in $u$ - and $v$ -direction |
| $c$                  | center of projection, principle point                        |
| $c_u, c_v$           | image coordinates of principle point                         |
| $n_u, n_v$           | number of pixels in the image in $u$ - and $v$ -direction    |
| $\mathbf{R}$         | rotation matrix  |
| $\mathbf{t}$         | translation vector   |
| $\mathbb{1}$         | unity matrix   |
| $\kappa$             | vector of the lens distortion parameters                     |
| $\kappa_i$           | distortion parameter   |
| $\mathbf{H}$         | homogenous matrix  |
| $\hat{\mathbf{m}}$   | back-projected image point                                   |
| $\mathbf{p}$         | data point   |
| $\mathbf{k}$         | point of correspondence                                      |
| $\mathbf{n}$         | normal vector to a plane                                     |
| $t$                  | variable in time   |
| $\sigma$             | standard deviation   |



# CHAPTER 1

---

## Introduction

### *1.1 Motivation*

Participating in today's road traffic is a tiring and dangerous task. Unfortunately, the times of joyful driving become less and less. The increasing number of vehicles on a road network with fixed length leads to more congestion and stressful situations. Considering human deficiencies, such as having a short attention span or a limited multi-tasking ability, the monotonous situations and the overload of the driver stipulate potential danger. Car manufacturers have constructed counter-measures for decades. The passive security is steadily enhanced by applications such as the supplementary restraint systems or the side-impact protection. But also active components were introduced. The anti-lock brake-system (ABS) keeps a car steerable while braking. The systems detects a blocked wheel and immediately releases the brake for a short period of time. The revolving wheel is able to redirect the vehicle. Another system is the electronic stability program (ESP). It individually attaches or releases the brake on single wheels to avoid skidding in curves. According to the German Statistical Office, ESP in Mercedes passenger cars caused a drop of 15% in fatal accidents (Becker, 2002). However, all these systems take only the ego-vehicle into account. The adaptive cruise control (ACC) measures the distance to a leading car and adjusts the speed accordingly. ABS and ESP are within the category of security systems whereas the ACC increases the convenience of the driver. But all of the them are so called *driver assistance systems*, assisting the driver while driving. They acquire data from on-board sensors which perceive either the state of the ego-vehicle or its surrounding environment.

But even with these systems still thousands of people die in traffic accidents every year. Many more are injured. With the undamped need for mobility in all aspects of life, one cannot expect the traffic load to diminish. Accordingly, the security provided by the car must be increased, for the driver as well as for the other road users. Hence, the vehicle has to become more and more aware of its environment and intelligent in its reaction to it. Intelligence in this context is defined as the capability to perceive, to recollect, to adapt and to intervene.

The *intelligent vehicle* depends on the data about the current state of the ego-vehicle and the environment. Data provided by external sources such as a Global Positioning System (GPS), traffic jam warnings, or digital maps only assist in the coarse positioning of the vehicle and route guidance. Dynamically evolving situations are only perceivable with on-board sensors, fast data evaluation, and decision making.

Available sensors for intelligent vehicles are millimeter-wave radar, ultrasonic sensors, laser radar, infrared-pattern systems, and cameras in different wavelength bands. Except for the camera, the sensors are active. Active sensors send out a signal and the echo is evaluated in order to directly retrieve the depth information. However, these sensors usually lack from sufficient angular resolution or have a limited opening angle. In addition, the signal quality is subject to the environment. In the case of an ultrasonic sensor the wind literally blows away the signal, or spray in rain scatters the laser beams. Cameras, exploiting the vision cues, produce gray-scale or color images that do not provide direct Euclidean measurements. The depth has to be recovered using multiple cameras providing supplementary information.

Based on the acquired data the following, potentially dangerous situations are handled by driver assistance systems which are either today's off-the-shelf systems, or which are planned for the future:

- **Intelligent Cruise Control (ICC):** On a congested highway the task is to stay in the driving lane, follow the preceding vehicle at a safe distance, and obey the speed limit. ICC systems are usually based on radar or lidar sensors. The speed is kept by the cruise control. Hence, a system keeping a safe distance to a leading car is called intelligent cruise control.
- **Lane Keeping Assistant:** Most roads are marked with lanes. The system detects the driving lane assigned to the ego-vehicle. This is done either by an optical recognition of the lane or with the help of magnetic markers attached to the road. The system either warns the driver if the vehicle is about to leave the lane, or automatically keeps the vehicle in it.
- **Lane Changing Assistant:** Changing a lane on a busy highway is dangerous. The supporting system must be able to reliably detect the lanes, a vehicle on the side, and a vehicle coming from behind. To solve this task, either multiple cameras or a combination of cameras and other sensors are required. The joint evaluation of the data enables the system either to warn the driver or prevent the vehicle from changing lanes.
- **Intersection Assistant:** On intersections, especially in cities, most of the accidents occur. An intersection assistant system requires a combination of vision and radar or lidar systems. Intersections, with vehicles coming from the sides, pedestrians crossing the road, and cyclists, sharing the own lane, are complex.



Figure 1.1: A Stereo-Camera Image-Pair with Repetitive Patterns: The pillars of the guardrail are repetitive patterns.

- **Traffic Sign Assistant:** This system informs the driver about current speed limits and other regulations. It is vision-based and assisted by external information such as digital maps and GPS. In situations with multiple lanes and multiple signs, the applicable traffic-sign to the ego-vehicle must be determined.
- **Traffic Jam Pilot:** Automated stop-and-go driving is the task of the traffic jam pilot. This system operates in the low velocity range. The data is acquired with vision systems and radar or lidar. The objective is to reduce the stress of the stop and go driving. In contrast to the cruise controls this systems extends the velocity range down to the standing situation.
- **Platooning System:** Operating trucks on highways in platoons is appealing to companies running a large fleet. In an automated platoon only the leading vehicle is driven manually. The platooning system is an adaption of the traffic jam pilot to the truck requirements.
- **Auto Pilot:** This systems completely takes over the control of the ego-vehicle. It comprehends the functionality of the systems above. The implementation of such a system on the basis of off-the-shelf hardware components is the major goal of the assistance systems research.

With an increasing number of such intelligent vehicles on the road, traffic becomes safer. However, these applications provide security for human beings. Therefore, the systems must be safe and reliable.

## 1.2 Objective of this Thesis

This thesis is developed in a working group that focuses on vision-based vehicle-following systems. The principle of following a leading vehicle is the basis of several applications mentioned above: the intelligent cruise control (ICC), the automated

stop-and-go driving, and the platooning. The framework application for this thesis performs automated stop-and-go driving on the basis of stereo-image processing. Fig. 1.1 shows a stereo image-pair of a typical vehicle-following scene.

The distance is retrieved by a feature detection in the images, a subsequent matching process in the image pair, and a triangulation on the basis of the precedent data. The critical step in the process is the matching of the features in the left and the right image. If patterns occur several times, so called repetitive patterns, multiple solutions to the matching effort exist. The pillars of the guardrail in Fig. 1.1 are an example. Instead of an unequivocal result, multiple matching hypotheses are obtained. Choosing the wrong match leads to a false depth measurement and subsequently to a wrong environment representation. Decisions of the system on this data basis are potentially inappropriate. Hence, the data basis must be improved. This is achieved either by applying additional algorithms to the same images or acquiring an additional view onto the scene, e.g. by a third camera.

The stereo-vision-based perception of a leading vehicle depends on the reliable functionality of the camera systems. A temporary occlusion of the leading vehicle in one of the images, e.g. caused by a passing windshield wiper, leads to temporary loss of environment data. Dirt on the windshield which is not removed or an electronic failure of a camera cause a long-termed loss of one image. Both cases are potentially dangerous for the assistance system and require a fall-back layer in the application.

Repetitive patterns and the necessity of providing a fall-back layer in the case of a loss of an image triggered the work that is presented in this thesis. Considering the application, the maximum distance which must be recovered is small. On the other hand, a large field of view is desirable. Vehicles from other lanes cutting in between the ego-vehicle and the leading vehicle ought to be detected as soon as possible. These reasons opt for lenses with a small focal length. Due to the large field of view, the hood of the car is visible in the lower image region. Fig. 1.2 shows an example of this setup. The hood of the car is a smooth, shiny varnished part. Hence, it reflects an additional view onto the scene, that is directly observed by the camera. In principle, this is the same information that an additional camera provides. However, the hood of the car is not designed to serve as a part in an optical system. It is a free-formed surface without any special optical properties. The curvature distorts the image, and the varnish diminishes the contrast. Nevertheless, the information contained in this catadioptric<sup>1</sup> system should be sufficient to provide additional data to a computer vision application. Fig. 1.3 shows the lower image region of Fig. 1.2 cropped, flipped, and enlarged. The objective of this thesis is to prove the usability of the information contained in these free-formed surface reflections in computer vision algorithms.

---

<sup>1</sup>Dioptic systems consists of lenses, catoptric systems of mirrors. For the combination of mirrors and lenses the name *catadioptric* has been established in literature.



Figure 1.2: A Traffic Scene Reflected in the Hood of a Car: The task of this thesis is to make use of the information contained in the lower image region, where the scene is reflected in the hood.

### ***1.3 Contribution of this Thesis***

In this section, the main contributions of this thesis are described. First, an investigation of the possibilities of reconstructing an image from free-formed-surface reflections is conducted. It results in two novel solutions and two fundamental requirements. The fundamental requirements are a precise camera calibration and the determination of the camera position relative to the reflective surface. Therefore, a camera-calibration technique is introduced and major improvements to it are presented. On this basis, a new algorithm for the recovery of the reflective surface is described and discussed. After elaborating the fundamental principles, the two solutions are detailed. The first one requires an extension of the linear epipolar constraint to a general notion of epipolar curves. This is accomplished by a theoretical examination of the reflection process with differential geometry methods. The second method is an in-image based reconstruction that takes the geometry of the surface and the setup into account. It compares the reflective surface to a tangential planar mirror and results in an assignment of image locations in the reconstructed image and the original image. For this approach two implementation methods are presented. One of them is a new one, using the graphics processing unit (GPU) for image processing. This achieves unprecedented speed and image quality in the reconstruction.



Figure 1.3: A Flipped Presentation of the Reflected Scene: The lower part with the reflection of Fig. 1.2 is cropped, flipped and enlarged.

In this thesis the usual perception that reflections on free-formed surfaces only introduce errors in computer vision algorithms is refuted. The algorithms developed on this assumption are just not able to cope with the information contained in the reflection. With appropriate consideration of the geometric setup and adapted algorithms the reflections provide useful information to computer vision applications.

## 1.4 Thesis Overview

The remainder of this thesis is devoted to the development of reconstruction methods of free-formed surface reflections. The focus of the methods is their applicability in computer vision systems with regard to vehicle applications.

Chapter 2 analyzes the problem situation and defines the task of the thesis. To explain the circumstances in which the methods are developed and applied, a short introduction to the principles of stereo-image processing is given. It is fundamental for some of the terms used throughout the thesis, and the ideas that are extended by this work. After discussing the system environment, the task of the thesis is explicitly stated. It is followed by a short analysis of the physical and optical processes involved.

Chapter 3 presents a review of the existing literature with regard to the posed problem and to catadioptric systems. The focus of this thesis is the evaluation of the reflection on the free-formed surface. Therefore, other solutions to problem of robust object recognition in the presence of repetitive patterns are not discussed. The camera calibration and the surface recovery is a major part of this thesis. Accordingly, the related work with respect to these topics is presented in the respective chapters.

New ideas for the retrieval of the information contained in free-formed-surface reflections are developed in Chapter 4.

Two fundamental requirements for the use of free-formed-surface reflections in computer-vision systems are discussed in Chapter 5 and Chapter 6. The first is concerned with camera calibration. There, calibration techniques and respectively introduced major improvements are presented. The camera-calibration method is not only a requirement for this application but it serves and enhances other image-processing

as well. The second one, which depends on a precise camera calibration, is the recovery of the surface. There, a new approach to retrieve the surface data is presented.

Chapter 7 extends the linear epipolar constraint of stereo vision to a more general notion of epipolar curves on warped images. This is the basis of the extraction of the contained information without a classical image reconstruction.

The thesis concludes with the elaboration of a novel method for a classical image reconstruction in Chapter 8. The Comparison-of-Lines algorithm is developed, implemented, and tested on images acquired in the experimental vehicle. To prove the applicability of free-formed surface reflections in computer-vision applications a coarse, manual distance determination is conducted on the basis of a reconstructed image. In addition, this method is implemented on the graphics processing unit (GPU) of the computer graphics-card. The implementation on the GPU offers new possibilities for image processing in general.

Chapter 9 summarizes the presented thesis, states the work that remains to be done and gives a short outlook into the future.

The thesis contains two appendices. Appendix A presents views of the front of the car based on CAD data. Appendix B shows the result of an air pressure deformation of the hood at 250 km/h.

Why is this thesis structured in this way? For one, it represents the logical steps in developing a field of research. For the other, it reflects the planning and the course of work which led to this thesis. The dissertation is composed of three major parts. The first part comprehends the Chapters 2 through 4. This was also the first step in the course of work. The problem analysis, the review of the related work, and the determination of possible solutions were comprised in a feasibility study. On that basis, it was decided to continue the development. The second phase started with the camera calibration, due to its very fundamental nature of the subsequent steps. The following surface recovery uses the calibrated stereo-camera system itself to determine the surface within the camera coordinate frame. This concluded the second phase of the work and is presented in Chapter 5 and Chapter 6. The last part is the elaboration of new reconstruction methods and their implementation and testing in the given environment. This is comprised in Chapter 7 and Chapter 8.

According to the environment and the intended application, the focus is always on computer vision. More fundamental investigations, such as the reflectivity of varnishes, the magnitude of the specular and diffuse reflection components, or their angular distribution, are not conducted here. This decision is based on two reasons. One is, that such an examination is closer related to topics such as structured-lighting-techniques than to machine perception. The other is, that this kind of information is available in literature.

## CHAPTER 2

---

### **Statement and Analysis of the Problem**

The presence of repetitive patterns decreases the matching reliability of stereo vision. Using a local matching approach these patterns result in multiple matching hypotheses. Subsequently, it is necessary to decide the correct match by some means. Choosing the wrong match will lead to an incorrect representation of the surrounding environment. For applications, such as autonomously driving a car, that imply the security of a human being, wrong decisions based on the acquired potentially erroneous data are not acceptable.

To elaborate this multiple-matching-hypotheses problem in more detail, some basics about stereo vision are given in the next sections. Following that, the hard- and software environment, with special focus on the image processing, is analyzed. With these prerequisites the problem statement of this thesis is detailed. Subsequently, a look is taken at the physical processes involved in using a reflective surface in the optical system. Finally, the required solution properties are stated.

#### ***2.1 Principles of Stereo Vision***

To reconstruct a three-dimensional scene, different sensor approaches are taken. One of them is computer-based stereo-vision. Due to the projection of the three-dimensional scene onto a two-dimensional plane, depth information is lost. In order to reconstruct this lost information a second camera, imaging the same scene from a different position, is necessary. The correlation of the two information sets then yields the three-dimensional reconstruction.

The most important step in this procedure is to determine the corresponding points in the respective images. Based on this correspondence, using triangulation and the relative camera positions, the depth information is derived. However, the correspondence problem is not simple. First, in one image, significant structures are located by some kind of interest operator, e.g. an edge or corner detector. Around an interesting location an image region is extracted. This extracted image region is called a matching primitive. Searching for the location of the primitive in the second image the problems start, because there might be multiple solutions, if any at all. This



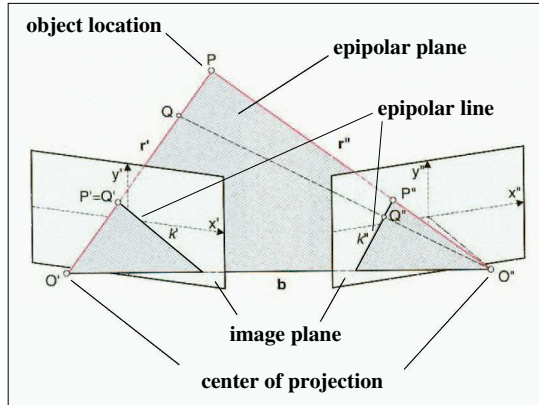


Figure 2.1: Epipolar Geometry in Stereo Vision. Image from Luhmann (2000).

multiplicity or failure might be due to one of the following reasons:

1. An object, visible in one camera, might be out of the field of view or obstructed for the second camera.
2. Repetitive patterns create multiple matching hypotheses.
3. Reflections create ambiguous object structures.
4. In image regions with low structure the solutions may be instable and sensitive to noise.
5. Depth discontinuities are difficult to retrieve. Sometimes the foreground, sometimes the background is measured.

The first two points are interconnected and their influence depends on the so called globality or locality of the primitive. A global primitive, larger in size, might occur only once in the other image. Therefore, the risk of encountering multiple solutions is low. On the one hand, larger image structures tend to be obstructed more often in the second image than smaller ones. Differences in the visible background in the global primitives account for lower correlation results. Local primitives, i.e. smaller extracted image regions, often generate multiple correlations, because the small patch from one image fits quite well to several structures in the other one. On the other hand, local primitives are less often obstructed and less sensitive to the visible background.

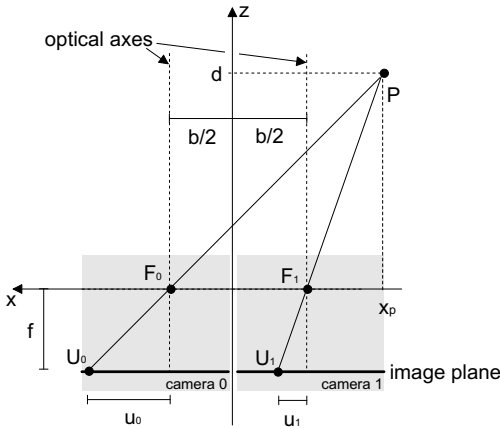


Figure 2.2: The Notion of Disparity in Image Processing: The disparity  $d$  is the difference in distance of the image locations from the respective principle point:  $d = u_0 - u_1$ . Image from Gehrig (2000).

Searching for the corresponding object all over the second image is a very time-consuming effort. With the knowledge of the camera's relative position to each other, the search is confined to a limited region, the so called epipolar line. Fig. 2.1 shows how this line is obtained from the geometric setup. The mapping rays  $r'$  and  $r''$  together with the baseline  $b$  of the cameras span the epipolar plane. This plane intersects the image plane of each camera. The intersection lines  $k'$  and  $k''$  are called epipolar lines. For the detected image location  $p'$  of an object location  $P$ , the corresponding point  $p''$  is found along the epipolar line  $k''$  in the second image. With this knowledge a search throughout the whole second image becomes obsolete. The search must only be performed along the epipolar line. The start of an epipolar line is the image of the camera point by the other camera and the end is given by the image of the direction of infinity, the epipole.

If the stereo-camera system is set up in the so called standard geometry, i.e. the optical axes of both cameras are parallel and the rotational difference in the image planes is zero, the epipolar line of a given row in one electronic image is the equivalent row in the other image. Along these epipolar lines the object's distance is recovered by the disparity. The disparity  $d$  denotes the difference in distance of the image locations from the respective principle point (see Fig. 2.2):

$$d = u_0 - u_1. \quad (2.1)$$

The disparity is directly correlated to the object distance  $z$ . Considering the sign, the disparity is zero for an infinite distance. For a finite distance, the distance-disparity relationship is (Luhmann, 2000):

$$z = b \frac{f}{d} \quad (2.2)$$

with  $f$  being the focal length of the cameras. By defining a coordinate system in between the cameras (see Fig. 2.2) the  $x$ - and  $y$  coordinates of the object are given by:

$$x = \frac{b}{2z}(u_0 + u_1) \quad (2.3)$$

$$y = \frac{b}{z}v_0 = \frac{b}{z}v_1 \quad (2.4)$$

with  $(u_0, v_0)$  and  $(u_1, v_1)$  being the coordinates of the object in the respective images.

However, if the system is esotropic either the standard geometry has to be reconstructed by means of rectification algorithms (see e.g. Kaempchen (2001)) or the epipolar notion has to be extended to a more general geometric understanding, allowing e.g. negative disparities and inclined lines.

In the correspondence analysis of real camera setups and images, one has to consider that the external and internal parameter sets are only known to a certain precision. Accordingly, the image distortions are only correctable to a certain degree and the epipolar lines are determined with some uncertainty. Subsequently, in the correlation process the epipolar line has to be extended to an epipolar band. The width of the band depends on the uncertainties in the parameter estimation. The necessary length of the epipolar band in the implementation is linearly dependent on the distance range where the object is located.

Applying the epipolar geometry to the correspondence problem enables the algorithms to run in video-real time at 25 Hz, even for dense depth maps.

## 2.2 System Environment

This section addresses the setup of the experimental car, with its hardware components and the software implemented on them.<sup>1</sup> With regard to the focus of this thesis the image processing algorithms are discussed in more detail.

### 2.2.1 Hardware

The vehicle-following system was first implemented in a Mercedes E-class 420. After about one third of the development of this thesis the research car changed to a Mer-

<sup>1</sup>Some of the system details are derived in their content from a previous dissertation application implemented on the same experimental car (Gehrig, 2000).

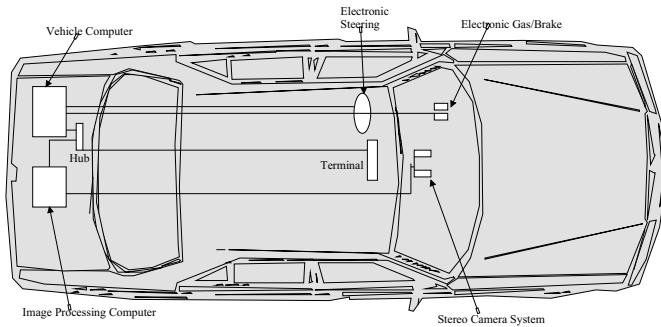


Figure 2.3: The Experimental Car with its Sensors, Actuators and Computers. Image from Gehrig (2000).

cedes S-class 500. However, the internal setup of the relevant hardware is equivalent and is shown in Fig. 2.3. The vehicle is equipped with electronically controlled throttle, brake and steering wheel. The sensor, i.e. the stereo-camera system, is mounted behind the windshield, close to the rear-view mirror, looking across the hood of the car onto the scene ahead. This position provides a wipeable region for the cameras guaranteeing adequate viewing conditions. The image processing algorithms run on an Intel-based computer in the trunk. A Motorola-based computer hosts the vehicle control algorithms and interfaces all the actuators (throttle, brake and steering wheel). The Intel machine contains a Stemmer IMPCI frame-grabber which acquires the images from the cameras. As shown in Fig. 2.3, the trunk of the car hosts the processing equipment, i.e. the Motorola 604 with Lynx OS and the image processing hardware, an Intel Pentium II with a Linux operating system. The computers are linked by an Ethernet over a communication hub. The control is provided via an X-terminal attached to the dashboard. The video output of the frame-grabber is displayed on a monitor.

### 2.2.2 Software

As mentioned above, the hardware environment in the experimental vehicle consists of two separate computers for controlling and image processing. The communication process in between the hosts is conducted via CAN (controller area network). The essential vehicle data (e.g. velocity and steering angle) is also available on that bus. The image processing algorithms are implemented in ANSI-C and compiled with a GNU-C compiler under Linux.

The off-line development environment in the laboratory enables the usage of the same image processing algorithms with additional debugging information. The system is based on Vista (Pope & Lowe, 1994) for the visualization of the images and the overlay of the detection results. Instead of using the images grabbed directly from the cameras, images from the hard-disk, recorded while driving, are taken as input. Upon the encounter of a failure or inexplicable detection results while driving, an image sequence is recorded by pressing a button on the system. Subsequently, the sequence and the corresponding vehicle data is analyzed and the image processing algorithms are debugged and improved.

### 2.2.3 Image Processing

In the following paragraphs the basic feature and subsequent object detection algorithms of the vehicle following application are detailed. The detection consists of multiple steps. First is the image acquisition, followed by the feature detection. The third is the feature matching. Finally, a cluster analysis of the three-dimensional points is performed. A confirmation of the clusters is achieved by a temporal analysis on consecutive frames. The desired result of the image processing is a consistent and reliable situation assessment which yields a list of detected objects with their metric information.

#### Image Acquisition

To acquire the images, a COBRA gray-scale CCD camera with  $768 \times 568$  pixels is used. The camera operates in an interlaced mode with a frame rate of 12 Hz. For the subsequently described algorithms only half images are used. The images of the cameras are grabbed synchronously with two out of the three RGB frame-grabber channels. The shutter time of the cameras is set automatically on the basis of the brightness of the previously acquired image. Hence the control of the shutter time needs not to be performed by the image processing. The focal setting of the lenses is infinity. Examining the intended application and the mounting of the cameras behind the windshield, relevant objects are only located at a distance larger than 2 m. Accordingly, the aperture of the camera is set to map objects of larger distance sharply to the retina. Closer objects, especially the windshield itself or raindrops on it, only appear blurry. The used lens systems have a focal length in between 7.5 and 12 mm. This corresponds to a field of view of  $48^\circ$  and  $28^\circ$ , respectively.

The baseline length of the stereo-camera setup is approximately 25 cm. In general, there is a trade-off in accuracy in the triangulation, which increases with larger baseline distances, and the matching reliability. If the baseline is too large, objects are viewed from increasingly different angles and the correlation robustness decreases, e.g. one camera looks from behind onto a truck while the other images the truck

from the side. Humans with their model based correlation, knowing that it is a truck and what the geometry of a truck looks like, are still able to correlate the views and estimate the distance. A computer based system is not able to do so yet.

The cameras in the stereo setup are allowed to be esotropic. As mentioned above, this not only requires an extended consideration of the epipolar geometry, but internal and external camera calibration are basic necessities. Rather than rectifying the images and applying a standard epipolar geometry model, a more time-efficient method is used. The lens-distortion correction is stored in an approximated lookup-table such that the operations are applied to a corrected image and epipolar lines are really lines and not curves. Subsequently, the epipolar geometry is extended such that negative disparities are allowed.

The sensor coordinate system is such, that the  $y$ -axis points up, the  $x$ -axis to the left and  $z$ -axis ahead, all with respect to the mounting in the car. The image coordinate system originates in the upper left corner of the image, with  $u$  being the horizontal coordinate pointing to the right and  $v$  being the vertical coordinate pointing downwards. The sensor coordinate system is located with its origin at the focal point of the camera. Further camera model details and calibration requirements are elaborated in Chapter 5.

### **Feature Extraction**

To initialize the feature extraction in the image pairs, a test needs to be performed whether there is enough structure in a certain image region. Regions without or with a low structure, e.g. uniformly colored walls, cannot be matched. An interest operator is applied to determine the regions of sufficient structure for subsequent feature extraction. The chosen interest determination method is a simple edge detector, the horizontal Prewitt operator. The desired result of the edge operator is not the exact edge location but the mere existence. This is sufficient to trigger an area-based correspondence search where the exact matches and their location are determined.

### **Feature Matching**

As mentioned above, the stereo-correspondence problem is significantly simplified by applying the epipolar constraint. In the described epipolar band, two types of correspondences are used: feature-based and area-based. In the feature-based approach all the features are extracted in both images. Subsequently, the best correspondence in between the extracted features is determined. The area-based method depends on the direct correlation of the gray-scale values. For previously discussed reasons, e.g. the occlusion problem, a local feature-based approach is used. The primitives are  $16 \times 3$  pixels in size.

In order to determine the quality of a match, usually a correlation function along the epipolar line is evaluated. Typical functions are the Sum-of-the-Absolute-Difference (SAD) in gray-scale values or the Sum-of-Squared-Differences (SSD). One of the more time-consuming, but most suitable correlation measure for image processing is the normalized mean-free cross-correlation function  $ccfmf$  (Aschwanden, 1993). This function is applied here. Due to the relevance for the camera calibration approach, it is introduced and discussed in more detail in Section 5.6.1.

The analysis of the stereo features is performed on all pixels within the region of interest. This region is the whole image in the horizontal direction, and the central part in the vertical direction, skipping the region where the sky and the hood of the car are imaged. To reduce the computational load only disparities corresponding to 2 m or more are considered. Additionally, once a correspondence has been established, the disparities to be searched in the consecutive image, are limited to a range around the previously retrieved object distance.

### Cluster Analysis and Object Formation

On the basis of a list of three-dimensional point locations, preliminary objects are formed. For the clustering, all points are considered except those on or below the ground. Here, a weak model of a flat road is used. All three-dimensional points below a certain height are disregarded. Note here, that beam reflections, e.g. on a wet road, are reconstructed below the road surface, and therefore omitted. Parsing the list of three-dimensional points, those with an Euclidean distance below a certain threshold are connected. The obtained cluster connectivity is protocolled using a coloring scheme known from Graph Theory. Investigation of the computational load, reveals a complexity of  $O(n^2)$  with  $n$  features in the list. This is not acceptable for real-time applications. Applying some heuristics, e.g. that a feature at the far left part of one image cannot correspond to one at the far right of the other, the complexity of the algorithms remains. However, the number  $n$  reduces to an acceptable level.

Constitutive for clusters are its dimensions (length, width, and height), represented as a bounding box, and a center location. The center location is the result of averaging all contributing three-dimensional point locations. Retrieving the object's distance, the contributing points far away are weighted much less than the close ones, because in a vehicle following application the rear end of an object ahead is the focus of interest.

So far only methods operating on one image pair were discussed. However, in the course of driving, series of images are acquired. In general, the scene imaged in a consecutive frame is quite similar to the previous one. The own vehicle is usually displaced by some distance and surrounding vehicles moved according to their relative speed. A leading car is still close to the previous relative position with respect

to the ego-vehicle. The difficulty of combining the data from consecutive frames is a matching task and called the correspondence problem. It is closely related to the problem of tracking objects in time. Here, the object matching from different frames is achieved by demanding them to be close together in space. This is equivalent to assuming only a slowly dynamic scene in the vehicle's surrounding.

Clusters determined in a single image pair constitute a raw object. If such a raw object is geometrically confirmed by objects located nearby in subsequent frames, it becomes an object. This avoids actions in the control of the vehicle on the basis of suddenly appearing objects. To increase the precision of the measured object center, the symmetry of the object in the image is evaluated. The combination of the bounding box and the cluster center is then tracked further with a Kalman filter (Kalman, 1960). However, the object dimensions are not subject to filtering. The usage of the Kalman filter allows for predictions into the future. This extrapolation is very important in situations of a temporary loss of an image.

Applying this rather simple object formation approach, leads to merging of objects, e.g. if they drive close to each other. Later on, these objects may be split again. This merging and splitting is detected by comparing the data of the previous frame and detecting overlaps. This all happens again on the basis of small interframe changes. In addition to that, a very simple object model restraining the maximum width is used. The bounding box of an extra large object is split upon exceeding the threshold.

The methods introduced above proved to be reliable under all reasonable conditions. Experiments included fog, rain, and snow, in day- and nighttime. The obstacle detection quality was evaluated empirically. No situations were encountered where the systems missed a nearby located object. However, few objects were detected where there were not any. The emission of a car ahead on a cold day forms a white cloud in front of the ego-vehicle. Before evaporating, the systems detects the cloud and confirms an object. Since it is hardly distinguishable from a human, wearing a white shirt standing in front of the car, there are almost no measures to discard such object detections.

### **2.3 *Statement of the Problem***

Computer based applications in cars, used in everyday traffic, need to be as failsafe and reliable as possible. Stereo-camera systems provide within their images a manifold source of information about an observed scene. However, the reconstruction methods must be sophisticated and operate on data with the least possible errors in extraction. If the data for the subsequent processing on a higher, more abstract level is ambiguous the decisions based upon it might be faulty as well. This is not acceptable.

Even though a very high level of reliability has been reached by the application of elaborate methods, two cases remain where progress is achievable:





Figure 2.4: A Multiple Matching Hypotheses Situation: An image pair exhibiting a multiple matching hypotheses situation. The features of the pillar marked in the left image are found in the right image several times. Choosing the wrong match leads to a wrong environment representation. A feature matching with distance information are shown in Fig. 2.5

- In general, more local matching approaches are faster and a result is more likely to be obtained. Due to the inherent susceptibility of the local approach to repetitive patterns, the result needs to be verified. Fig. 2.4 exhibits such a case. A feature located on a pillar of the guardrail in the left image is matched to each of the pillars in the right image. Due to the similarity of the features, the correlation results are close to equal in height and the choice is almost random. The result of the correlation is shown in Fig. 2.5. The distribution of the points and distances is such that an object is formed at 2.6 m. Considering that the hood of the car ends about 2 m in front of the cameras, the object is right ahead of the car, and the brake will be applied.
- Due to external circumstances, the image of one camera might be lost temporarily or even long-termed. A temporary loss is shown in Fig. 2.6. The windshield wiper passes the camera and obstructs most of the view onto the scene. The Kalman filter approach with its predictive capability provides already some stability in such a case, but with additional supporting data it is enhanced further. In the case of a long-termed loss of an image, e.g. due to an electronic failure of a camera, a reliable system should provide a fall-back layer to cope with the occurrence. The fall-back layer inhibits a sudden loss of functionality and provides additional time for the driver to take over the control again.

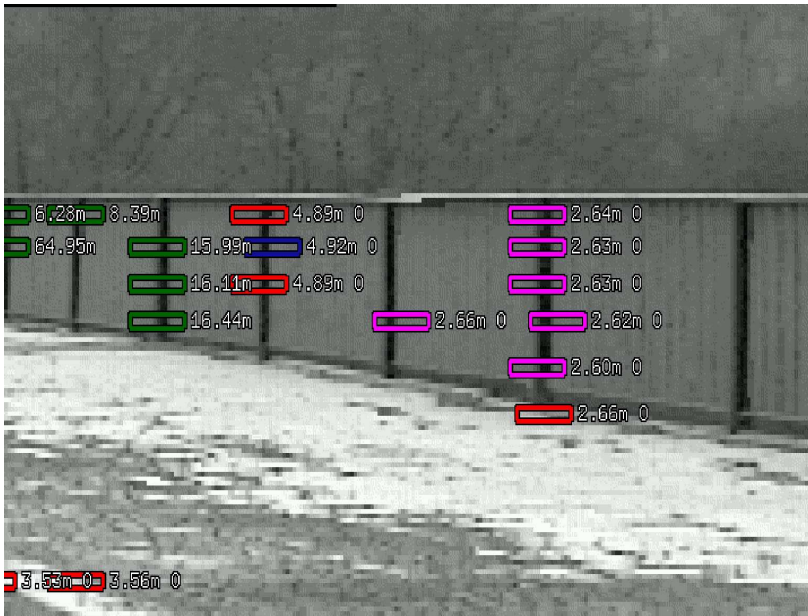


Figure 2.5: Distances in a Multiple Matching Hypotheses Situation: A feature of a pillar in the left image was correlated to the right image (see Fig. 2.4). The result is shown in this image. For the extracted primitive multiple matches are obtained. The distances are such that the points are combined to an object with a distance of 2.6 m. The comparison of this distance with the distance a human estimates from the original images, already reveals the error. Since the hood of the car is about 2 m long, the retrieved distance constitutes an object right in front of the car.



Figure 2.6: Temporary Loss of an Image: Due to the windshield wiper passing a camera, an image frame is lost for processing.

In the introduction to the basics of stereo vision (see Section 2.1), the two situations mentioned above are already named as problematic. The third source of error in that listing is that reflections create ambiguous object structures. In detailing the image processing procedure (see Section 2.2.3), two common solution strategies for errors due to reflections are mentioned. The first is to exclude known reflective image regions from processing. The second one is, that objects created by a previously unknown reflective surface are discarded, usually by the application of some heuristic.<sup>2</sup> Still, reflections on the back window of a leading car might interfere with the scene recovery and create false measurements, that are very hard to distinguish.

For humans, with their model-supported environment perception, it is easy to detect the reflections. Humans judging the smoothness of a surface use the reflections in it to do so. But why are reflections so demanding on image processing algorithms? The simple answer is: The reflections do contain information of the same nature! The usual perception that this information is erroneous is rather related to an inappropriate underlying modeling in the image processing than to a true faulty nature of the reflective information.

To obtain a sufficient field of view right in front of the car lenses with a rather short focal length are used. A side-effect of that is, that in the lower part of the image the hood of the car is visible (see Fig. 1.2). The hood of the car is a free-formed surface and was not designed with the intent to use it in an optical system. Still, the reflections form an image of the scene as if recorded by a virtual camera. Already implied in this virtual camera notion is that this virtual view and the direct view satisfy the very basic necessity of stereo-image processing, i.e. the views of the cameras onto the same scene must be from different angles.

After these preliminary remarks, the problem statement of this thesis is as follows:

The fundamental task of this thesis is to retrieve the information contained in free-formed surface reflections such that an image processing application is able to evaluate them. To do so, a general investigation into the optical properties of free-formed surfaces is necessary. An image processing application is only able to incorporate the information if the general geometry of the mapping is retrieved. Based on these results, a specific algorithmic solution is obtained to incorporate the information into the vehicle-following application.

---

<sup>2</sup>The two methods mentioned refer to the exclusion of the image region of the hood of the car from processing and discarding recovered objects below the assumed flat road, which are due reflections on a wet road.

## 2.4 Processes and Models

In order to achieve a better understanding of the posed problem, the involved basic physical processes and optical models are briefly discussed. To conclude this section, some general remarks about the perspective projection and reconstruction are made.

### 2.4.1 Physical Processes

Two kinds of reflections are distinguishable from a physical point of view:

- **Specular reflection:** this component is a surface phenomenon. It is described by Huygens' principle (1690). A wavefront is a plane of equal phase and any point on a primary wavefront is the origin of a spherical secondary wavefront. Their superposition, i.e. constructing the envelope of the elementary wavefronts, describes the primary wavefront at a later point in time. If one of these wavefronts intersects with a surface, a part of it, depending on the reflection coefficient, is reflected. This is represented by the amplitude (the radius) of the developing secondary spherical wavefront. Regarding the time-wise evolution of a wavefront, incident by an angle onto a surface, and constructing the evolving secondary wavefront the superposition of the waves leads to the law of reflection: The angle of incident equals the angle of reflection. Moving from wavefront optics to geometric optics of rays the normal vector of the plane of equal phase is taken as the direction of the ray and bundles of parallel rays are examined.
- **Diffuse reflection:** this component is due to the fact that part of the radiation that is not reflected enters into the body. There, multiple reflections and scattering occur, before part of that radiation leaves the body again in a wide range of angles around the normal vector to the surface. This results in a dull surface. Most of the times it is assumed that the diffuse reflection component obeys Lambert's Law (1760), i.e. the diffuse reflected intensity is equal under all angles of emission.

These mechanisms apply under the assumption of an idealized bordering surface between the medias. Under real conditions the surfaces are rough. However, the induced changes in the diffuse component are smaller than in the specular one. On a rough surface, the specular component is not reflected into the direction of the idealized angle of reflection but into a range of angles around it. However, except for very rough surfaces, this range is very small (Torrance & Sparrow, 1967). In addition to that, the diffuse reflection component reduces the specular intensity.

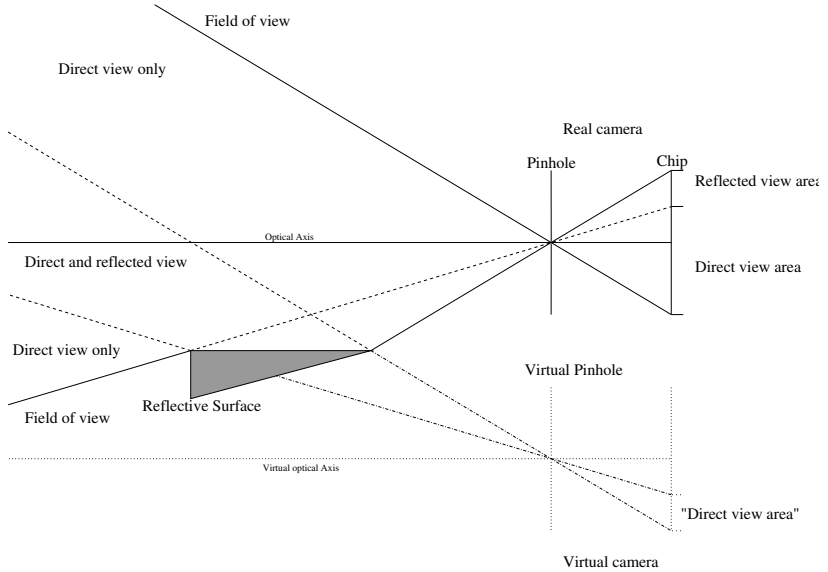


Figure 2.7: Formation of a Virtual Camera by a Planar Mirror.

### 2.4.2 Optical Models

The optical models are separable into two general cases as well:

- **Perspective projection:** this is the classical pinhole model. In it, two points on a line parallel to the optical axis are projected onto two points in the image. This model applies to most cameras and is implemented by different combinations of lenses and mirrors.
- **Orthographic projection:** this is the projection of object locations parallel to the optical axis onto the image plane. The orthographic camera is implemented either by a tele-centric lens, a pinhole aperture at the focal point between the lens and the image plane, an adjustable zoom-lens, or a relay-lens in addition to the others lenses involved. Above that, the usage of a relay-lens reduces coma and astigmatism. The field of view of an orthographic camera is very small and its irradiance too. However, one of its application will be discussed in Chapter 3: a camera with an omnidirectional field of view.

These are the two rudimentary models describing a camera. Usually, cameras composed solely of lenses are referred to as *dioptric* systems, while those only containing

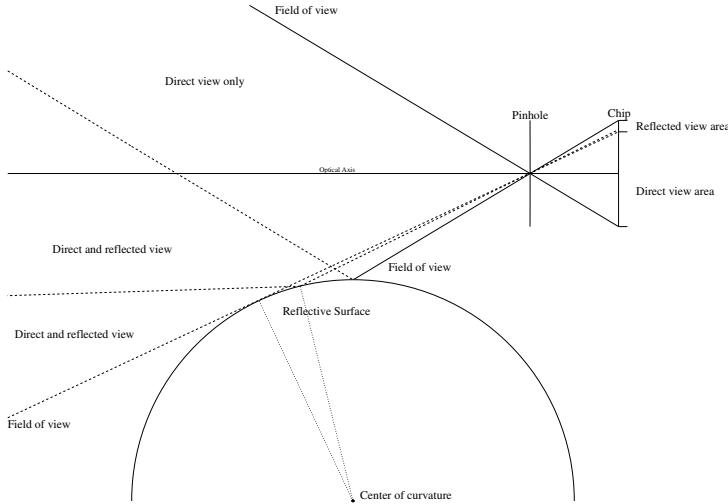


Figure 2.8: A Curved Surface in General Does Not Create a Virtual Camera.

mirrors are called *catoptric*. In the approach of this thesis, optical systems with lenses and mirrors are examined. For those, the name *catadioptric* systems is used in literature. Fig. 2.7 and 2.8 show the schematics of such systems. In Fig. 2.7 the ray propagation of a specular reflection on a planar mirror is shown. Obviously, three regions form. In two of them, the objects are only imaged in direct view, whereas in the third one a direct and a reflected view of the object is obtained. Similarly to other two-dimensional mappings of the three-dimensional space, geometric information is lost which must be recovered by triangulation. In this case, all the rays, originating from the camera, reflected on the planar mirror intersect at a singular point. This point is the location of a virtual camera sharing the same view. Fig. 2.8 shows the ray propagation in the case of a reflection on a surface with constant curvature. In contrast to the planar case, only two regions form. One of direct sight and an extended region of reflective and direct sight. However, the reflected rays do not intersect anymore. This implies, that in general an optical system containing curved mirrors does not retain a singular viewpoint. Hence, it is not a classical virtual camera.

### 2.4.3 Perspective Projection

To display an image, different back-projections of the acquired data are applicable. The one related to the human eye, and therefore often applied in computer-vision

algorithms, is the so called perspective projection. In it, the data is back-projected to a plane. If the ray propagation is undistorted a so called perspective correct projection is performed. It retains lines, sections of lines and angles, e.g. equal sections of lines at different distances are reduced in length according to the distance. Undistorted ray propagation refers in this case to a back-projection with a mathematical point projection center, the so called singular viewpoint of the system.

If the data is projected onto a cylindrical surface, a panoramic view with angles larger than  $180^\circ$  is produced. However, for the human eye the image exhibits distortions, because lines are not mapped to lines and angles are changed. Like a planar surface, a cylinder has no intrinsic curvature. However, the three-dimensional projection on an extrinsically curved surface distorts the image.

Projections onto other geometric shapes, e.g. spheres, with an intrinsic curvature are not suitable because the image is acquired with a planar retina.

## 2.5 Conclusions

To conclude this chapter, the required and some desired solution properties are stated. The application to develop serves as a verification stage in multiple matching hypotheses situations or as a fall-back layer in the case of a camera failure. The current system runs in video real-time, i.e. a cycle time of 80 ms per frame. Considering the large amount of information to be processed, the algorithmic approach for the verification step needs to be very fast. In the fall-back case, the information acquired in the reflection replaces the second image and its processing. Apart from deciding that the second image is not available, the respective time slot is fully available for the reconstruction and the evaluation of the reflection image.

Another requirement is that the reconstructed image must be dense. This is due to the fact that the sparse data together with the image background color introduces additional structure to the image. In the case of a sparse distribution, either an unequivocal resolution is obtained using inverse ray-tracing methods, or interpolation schemes must be applied. However, interpolation schemes are very time consuming and not the way to go on ordinary processing units.

In order to apply the already highly sophisticated and tuned image processing algorithms of the existing system, it is desirable to reconstruct a perspective correct image. This pinhole projection is the underlying model of the existing application. However, as indicated above, a reflection image on a curved surface must not retain a singular viewpoint. Does the specific combination of the hood as a reflective surface and a perspective camera yield a singular viewpoint? Or is there at least a distribution of viewpoints allowing a pinhole reconstruction with reasonable errors? These are questions to answer in the course of this thesis.

## CHAPTER 3

---

### Related Work

In this chapter related work in literature with respect to the given task is reviewed. The objective of this thesis, detailed in Section 2.3, is to increase the robustness of a stereo-vision system. This is intended to be done with the information contained in the reflections in a rigidly mounted free-formed specular surface. Therefore, other means of improving the reliability, such as the development and implementation of feature matching algorithms, are not considered here. An extensive review of existing image-processing algorithms is found in Gehrig (2000). Due to the relevancy of the tasks, related work with regard to camera calibration and surface geometry reconstruction is introduced in the respective chapters (see Section 5.2 and Section 6.2).

The exploitation of the reflection images provides a third respectively fourth image for processing. Hence, the given review starts with publications on  $n$ -ocular systems. In the second section, the state of the art with respect to lens-mirror systems is presented. After that, the basics of reflections and their incorporation into computer-vision systems is examined. Finally, image display approaches are regarded to see whether a suitable technique of treating warped images is available.

#### 3.1 *n*-ocular Systems

As discussed in Section 2.1, a problem of stereo vision is the robust local feature matching in the presence of repetitive patterns. One solution to this problem is the usage of additional images, acquired from different points of view. To use the information the usual feature matching algorithms must be extended to cope with three or more images. A publication relating to this is Okutomi & Kanade (1993). There, an algorithm for the processing of  $n$  images with different camera distances along the same baseline is developed. It is based on the Sum-of-Squared-Difference (SSD) matching algorithm. This operator is applied to each image pair. The results are represented in inverse distance rather than in disparity. This new sum operator SSSD-in-inverse-distance exhibits, even for repetitive patterns, a unique solution at the correct distance. This is related to the representation in inverse distance. Comparing the matching results of different pairs, the wrong matches move inverse to the baseline



distance while the correct matching location is invariant. Adding the results yields a peak at the object location while the other incorrect matches level out.

Kanade et al. (1995) built on the basis of the SSSD-in-inverse-distance operator a systems which is able to process six images in real-time. The computational load for this is high: the system evaluates 30 images per second, i.e. 30 million pixels  $\times$  disparity measurements per second. The disparity search range in the application is 60 pixels and the result is a  $256 \times 240$  pixel depth map.

If the cameras are not distributed along a line, another property can be invoked. Image processing, based on two cameras arranged along a horizontal line, is not able to determine the distance of horizontal edge features. The same holds true for vertical lines in a vertical camera arrangement. However, if the cameras are e.g. arranged at the corner points of an isosceles triangle the horizontal as well as the vertical features are exploitable (Stewart, 1989). The TLMA (Trinocular Local Matching Algorithm) considers the three images simultaneously and it does not require a match to be present in all of them. The final processing stage is an area-based consistency check to verify the matches. The computational time is only increased insignificantly compared to the two-image matching but the results are more reliable. The depth map is denser and the susceptibility to repetitive patterns is decreased but not eliminated.

Hanson et al. (1988) present a similar approach. However, the cameras are arranged freely. In order to speed up the computation the number of epipolar intersections, retrievable in the three images, is reduced. This is done first by rectifying the images. It simplifies the hypotheses formation in the matching stage. After validating the acquired triplets on the basis of a disparity-gradient method a robust result is obtained.

A trinocular vision application with freely arranged cameras is presented by Williamson & Thorpe (1998). They use an 'L'-shaped arrangement of the cameras, similar to Stewart (1989). However, the camera position are not restricted to an isosceles triangle. The free 'L'-arrangement still gives different epipolar directions in the camera pairs. The horizontal stereo camera system allows the detection of vertical edges while the vertical system provides the horizontal ones. The comparison of the extremum location in the disparity - matching-error graphs and the respective confidence level of the matches determine whether the pixel belongs to a vertical or a horizontal surface. False positives are rejected by a simple confidence measure. With this approach, they are able to reliably detect small object at a long range in a vehicle-specific environment.

Another method to evaluate the information contained in three images, is to transform the images into a parameter space. Shen & Paillou (1995a) and Shen & Paillou (1995b) use the Hough transformation (see Section 5.6.1) to do so. Apart from the reduction in the computational load, this allows a completely parallel matching process in all three images. In addition, it does not require an a-priori-similarity in the images, i.e. very different angles of view are allowed. The results with regard to false

matches and even in the case of partially obstructed objects the recovery process is satisfying.

To conclude the review of multi-baseline image processing the publication of Bortolozzi & Dubuisson (1991) is discussed. The depth recovery there, is based on a segment extraction and matching. After determining points of interest by a Deriche filter, the points are approximated with polygons. Due to noise and errors in segment extraction a verification step is necessary. In it, collinear and almost collinear segments are merged, and parallel segments with a distance lower than a set threshold are discarded except for the longest segment. Segments sharing a common region are transformed to a single segment. After this preprocessing, segments are matched in two images and a projection of the reconstructed segment into the third image is compared with the respective, extracted segment there. If the correlation is sufficient the segment is accepted. The computation time is satisfying. This is due to the possible parallel processing of the segment extraction and verification stage for each image. The subsequent processing is on a higher, more abstract, and less time-consuming level. However, the approach is susceptible to obstructed views which result in very different segment lengths in the images and a subsequent discarding of the match.

## 3.2 *Catadioptric Systems*

In this section lens-mirror systems in different distinctions are the focus of interest. An application of a single camera with a mirror system, which suggests itself, is the formation of a stereo-vision system. More general setups are discussed in the second passage. Catadioptric systems with a very large field of view are the topic in the third paragraph. Remarks about the calibration of catadioptric systems conclude this section.

### 3.2.1 *Stereo using one Camera*

There are different approaches to implement a stereo-vision system. The most obvious is to use two physical cameras. As already discussed in Chapter 2, these either require an adjustment effort to ensure a standard stereo geometry or a calibration method to determine the system parameters. Even if the cameras are set up in the standard stereo geometry the calibration might be necessary due to different intrinsic parameter sets of the two cameras.

To reduce these difficulties different measures are taken. Most of them aim at using only one physical camera. This implies only one set of intrinsic parameters and therefore truly equivalent cameras in the stereo system produced by mirrors. However, in these systems the mirrors require the adjustment to guarantee the intended setup. One of these approaches is described by Cafforio & Rocca (1986). In it, both images are

reflected on a mirror, i.e. the system is based on two virtual cameras. Another system, requiring four mirrors, is presented in Inaba et al. (1993). In this distinction, the mirrors are movable. This provides a zooming, a yawing and a focusing ability. In addition, the image processing allows motion-tracking and stereo-matching. With this system a three-dimensional trajectory of an object is tracked in real-time.

Goshtasby & Gruver (1993) use only two mirrors and a single camera, but the field of view of the mirrors and therefore the virtual camera locations are behind the physical camera. The positioning of the mirrors in this application is not as critical as in those mentioned above. This is due to the description of a supplementary rectification algorithm that restores the parallelity of the optical axes. Still a specific geometric setup is required.

This special geometry is not necessary anymore in Gluckman & Nayar (1998b), respectively Gluckman & Nayar (1999). The only restriction there is, that the respective orientations of the two mirrors are within a plane. Subsequently, the location of the virtual camera is determined by the reflection of the aperture and the knowledge of the cameras' principle points. The epipolar lines are confined to planes which reduces the traditional seven stereo parameters to six.

In Gluckman & Nayar (2000) the problem is addressed from a different side. The task is a rectified stereo image-pair in a single camera with a predefined virtual baseline. On this basis the number of planar mirrors, their location and the size of the sensor are determined. It is obvious, that an odd number of mirrors will result in a mirrored image of the scene. Gluckman derives this approach mathematically, but disregards the treatment of rotational differences in the images.

Other stereo-vision approaches use curved mirror surfaces to increase the field of view. In order to simplify the mathematical description, orthographic cameras are used. One of the first applications is Nayar et al. (1998). Nayar uses two specular spheres, i.e. surfaces with a constant non-zero curvature. The calibration is a very challenging task and a rigid camera mounting relative to the spheres is necessary. The aim of the publication is to show, that it is possible to perform a vision-based object recognition with curved mirrors. With the knowledge of the radii and the relative position of the camera the normal vector on the spheres and subsequently the directions of reflection are determined. This is the basis for retrieving a direction in space of the object. Due to the quantization of the image an uncertainty is introduced to the determination of the direction. However, this is not a restriction in this special case. It holds true not only for reflective images but for computer vision with quantized retinas in general. The spherical approach reveals other problems. The rays reflected on the sphere cannot be traced back to a point. This prohibits a perspective reconstruction. However, due to the rigid mounting and the calibration, the geometric parameters can be determined and that accounts for the feasibility.

Nene & Nayar (1998) examine four different systems, with planar, parabolic, elliptical, and hyperbolic mirrors. The task is to calculate a correct depth map by using a system containing mirrors with variable non-zero curvature. The derived classes of mirrors and cameras are discussed in more detail in Section 3.2.2.

To round up the possibilities for stereo image processing with a single camera, a bi-prism approach from Lee et al. (1999) must be mentioned. The system enables a simple calibration and results in two coplanar images with parallel baselines, i.e. the images are rectified to the standard stereo geometry. Due to the usage of the same camera twice there are no differences in intrinsic parameters and object intensities. The angle of view of the system depends on the prism angle  $\alpha$ . The problem in the approach however is to obtain a sufficient baseline distance of the virtual cameras. For a prism angle of  $\alpha = 12.4^\circ$  the baseline distance is only 38.8 mm in that specific setup.

A very different kind of stereo system using one camera is presented in Pajdla (2002). It deals with stereo vision on the basis of oblique cameras. An oblique camera is a generalized model of a camera. The rays that map the observed scene to the image plane must not all intersect at one point in this camera model. This makes it a non-central camera model. The only restriction in order to perform a stereo-vision matching is that each object point is only imaged once. Considering the non-classical and non-central property of the free-formed surface imaging system this publication is of interest. Pajdla (2002) derives in the mathematical model the feasibility of such a stereo-vision approach. The central projection assumption is not constitutive for a scene reconstruction. However, the implementation of his non-classical and non-central camera is the rotation of a tele-centric lens and conic mirror on a ring. This is still a closed-form geometry.

### 3.2.2 Image Formation

Catadioptric systems have long been used in astronomical instrumentation. They are still developed further there. Beach (1999) introduced a new optical system with almost zero-valued spherical mirrors for classical Seidel-aberrations<sup>1</sup>. This property is implemented for a relatively small aperture ( $< 1/f$ ), an opening angle of more than three radians, a retina with about  $10^6$  pixels, 600 nm bandwidth (visible light to the near infrared), and without vignetting. The shutter speed is increased such that opening diameters of more than 1 m are achievable. However, only one spherical mirror is of that size while the other optical elements are smaller.

Another enhancement of an astronomical instrumentation is described by Rogers

---

<sup>1</sup>The classical Seidel-aberrations are the five mapping imperfections of the third order: spherical aberration, coma, astigmatism, warped image field and distortion. These phenomena were first examined in detail by Ludwig von Seidel (1821-1896) in the mid 19th century.

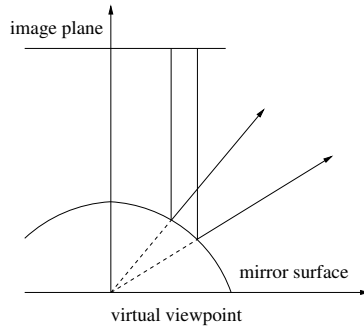


Figure 3.1: Formation of a Virtual Viewpoint: For a perspective reconstruction a singular real or virtual viewpoint is necessary. Here, an orthographic projection with a parabolic mirror surface was assumed, i.e. the surface is rotational paraboloid.

(1999). It is a coaxial optical system, consisting of a Cassegrain telescope with a movable primary mirror and a catadioptric connection. The system operates in the visible, the near infrared, and the millimeter waveband. The task is to implement a system that exhibits all over the stated wavebands two field of views: a wide angle and a tele-lens. The geometric setup is such that the movable optical parts are not used in the tele-lens application, in order to avoid dynamic misadjustments.

Meanwhile catadioptric systems are used for manifold applications in computer vision. Amongst them are area surveillance and robot guidance. An overview to this aspect is given in Nayar et al. (1998). Most of the systems aim to provide a singular viewpoint of the entire system. This is known as the *Single-Viewpoint-Constraint*. Neglecting this requirement implies the impossibility of a correct perspective image reconstruction. Even parts of the image are only reconstructible to a pinhole-camera view if all the rays originate from a single point or seem to do so (see Fig. 3.1). It is important to note that for a singular viewpoint, the position as well as the focal length of the virtual replacement camera is arbitrarily chosen. Otherwise, the knowledge of depth information is necessary to achieve a correct reconstruction.

Nayar & Baker (1997), Baker & Nayar (1998), and Baker & Nayar (1999) extensively deal with such catadioptric systems. They conclude that only rotational conic sections and a plane retain a single viewpoint. In the course of the investigation certain assignments of projective methods to the conic sections are developed. The simpler configuration, with regard to calibration and mathematical description, is the orthographic projection with the paraboloid surface. As long as the image plane is perpendicular to the axis of symmetry of the paraboloid, the system is invariant to translations of the image plane to the mirror surface. The hyperboloidal mirror is

assigned to the perspective projection. However, the aperture of the camera must be located at the second focal point of the hyperboloid. In addition to that, the optical axis of the camera and the axis of symmetry must be superimposed. This causes a very demanding adjustment effort. For a planar mirror all projections are applicable. The demand for adjustment is low, but the field of view too. The equations for construction of the systems are derived and blur regions are numerically determined. The resolution of the entire system and the defocusing, due to the finite size of the lens aperture and the mirror curvature, are examined.

In a further development of the systems reviewed above, Nayar & Peri (1999) consider the minimizations of mirror parts by folding the system. Folding refers to the combination of several planar or warped mirrors, such that the single viewpoint is retained but the geometric size of the system is reduced. The authors restrict their examination to such systems where the axis of the respective components are superimposed. All of the systems in general image a hemisphere with an exclusion cone in which the physical camera is located. The mapping of the hemisphere onto a plane always introduces distortions and the describing transformation is always non-affine.

To eliminate the exclusion cone, Chahl & Srinivasan (1997) develop systems containing multiple curved mirrors. However, the acquired data can only be reconstructed on a cylindrical surface, i.e. as a panoramic image. This is due to the distribution of the viewpoints along an axis. In a previous discussion Chahl & Srinivasan (1997) derive the mirror classes retaining a singular viewpoint by analyzing a differential equation. The results are again conic sections and the plane. Subsequently, the conic sections are compared to polynomial surfaces of higher order. The results are equivalent to Nayar & Baker (1997). An interesting idea for the back-projection procedure is introduced in this publication. The back-projection for rotational surfaces is implemented on a hardware-basis by a retina exhibiting a concentric circular quantization instead of a linear one. Then, the successive read-out of the circularly distributed pixels performs the reconstruction task.

Another idea for the automated reconstruction is introduced by Hicks & Bajcsy (1999). Instead of designing a new retina, a mirror is designed such that the view onto it retains the geometry of a plane perpendicular to its symmetry axis. In addition, a large field of view is demanded. If these mirrors are used with conventional imaging systems, the surfaces act as an image forming sensor, providing undistorted images which do not require further processing. The surfaces are derived by solving a differential equation of the surface geometry function. Using an orthographic projection not only simplifies the differential equation but it provides, as mentioned above, a translational invariance of the mirror to the camera. This is important for the adjustment of the system. Here, the orthographic projection is approximated by a tele-centric lens. To prove the functionality a chess-grid is imaged (see Fig. 3.2). In principle, it is possible to construct a mirror that images the chess-grid, except for

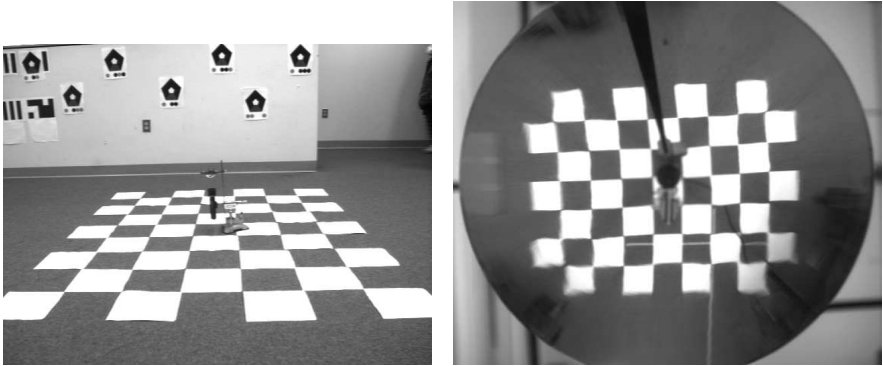


Figure 3.2: Automatic Rectification of a Plane by a Mirror: In the left image the setup is shown. The ground plane is imaged by a camera, placed in the middle of the chess-grid, looking upwards into a specially formed mirror. The right image shows the view of the camera into the mirror. The images are taken from Hicks & Bajcsy (1999).

scaling, without distortions. However, the rectification only works for a single plane. All other planes, even parallel ones, are imaged with distortions.

### 3.2.3 Omnidirectional Vision Systems

Omnidirectional systems are special cameras or setups that allow views into all directions. Among these, special panoramic sensors with a field of view larger than  $180^\circ$  are distinguished. Considering the geometry, it is obvious that fields of view larger than  $180^\circ$  cannot be reconstructed on a plane, due to its required infinite size. To completely display such an image, only panoramic projections are feasible. However, these do not conserve lines and angles. Usually, the projection is applied to a cylindrical surface, virtually placed such that the axis of the cylinder and the symmetry axis of the mirror or the optical axis of the fish-eye lens coincide. The cylindrical surface has the advantage that it does not exhibit an intrinsic curvature. Therefore, it can be cut and rolled out on a plane, e.g. displayed as an image on a screen, without additional distortions. With respect to the Single-Viewpoint-Constraint the remarks made in Section 3.2.2 apply here as well.

The very first approaches to extreme wide-angle cameras were so called fish-eye lenses. Miyamoto (1964) constructed a lens with a field of view of  $180^\circ$ . However, the image exhibits large distortions, due to the inability of projecting a hemisphere onto a plane without error. The distortions are larger at the border of the lens. Ideas to correct these imperfections are given. The fish-eye lenses known before exhibited large lateral chromatic aberrations to such a degree that images can only be taken

with a color filter. By introducing a doublet, i.e. a positive flint glass and a negative crown glass<sup>2</sup>, in front of the pupil the color error is corrected. Fish-eye lenses retain a physical relationship between the object and the image location (Oh & Hall, 1987). The previous studies with regard to the linear relationship of the zenith angle and the radial location of the object show two general properties of omnidirectional sensors:

1. Invariance of the azimuth angle: The object and the image are to be found at the same azimuth angle.
2. The relationship of the zenith angle and the radial distance of the image location from the principle point is linear.

Both of them are very helpful for the image reconstruction.

The first developments of wide-angle mirror systems, e.g. described by Yamazawa et al. (1993) in the COPIS (COnic Projection Image Sensor) project, use a conic mirror geometry. However, cones exhibit very large errors due to the location of the retained viewpoint on the tip of the cone. Subsequently, Yamazawa et al. (1993) examines hyperbolic mirrors in the HyperOmniVision-System, but notes the unisotropic properties. The spherical but correctable and the astigmatic errors are investigated as well.

Further developments were published by the working group of Shree K. Nayar at the University of Columbia. Based on the image formation publications (see Section 3.2.2) Nayar (1997a), Nayar (1997b), and Nayar (1997c) develop different kinds of omnidirectional mirror-based cameras. Subsequently, Peri & Nayar (1997) implement the reconstruction software for these systems on a personal computer. With it, an omnidirectional, panoramic, and perspective reconstruction at 30 Hz is achievable.

Gluckman et al. (1998) used the Nayar systems for image-processing tasks, such as a stereo application. The stereo system consists of two orthographic cameras and two rotational paraboloids, mounted vertically above each other, with coinciding optical axes. In this arrangement the upper camera is located in the exclusion cone of the lower one, to minimize the obstructed space. This ensures parallel epipolar lines, allowing for real-time implementation. Panoramic images are created on cylindrical surfaces. To enhance the images a bi-linear gray-value interpolation is applied. The baseline length of the two virtual cameras is the distance of the singular viewpoints. To derive the baseline length which optimizes the depth resolution for a given application, is a remaining task for research. Another application is the representation of optical flow in omnidirectional systems (Gluckman & Nayar, 1998a). Traditional cameras suffer from a noise sensitivity and a subsequent inaccurate flow determination if the direction of ego-motion is not visible in the image. In an omnidirectional

---

<sup>2</sup>Flint and crown are two sorts of glass, different in the diffraction index and the dispersion, which are combined to correct for chromatic aberrations.



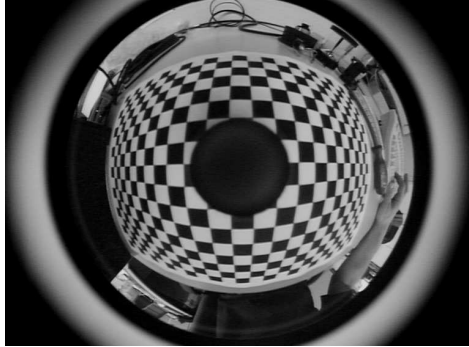


Figure 3.3: Calibration Image of a Parabolic Omnivision Camera: According to Geyer & Daniilidis (1999) a parabolic omnivision camera is calibrated by sets of parallels lines. In contrast to Geyer & Daniilidis (1999) the calibration targets are not dots but the intersections of a checkerboard pattern, a superior method (see Section 5.2). The image was taken with a proprietary omnivision-camera set.

image, the center of motion is always visible. If the system retains a singular view-point, the flow vectors are projected onto a sphere. On the sphere the rotational and translational motion components are easily distinguished, in contrast to a planar flow representation.

Takeya et al. (1998) develop a folded system, consisting of two axial-symmetric mirrors and a CCD-camera. In addition, a theoretical model, estimating the aberrations of a two-mirror-system with arbitrary curvatures, is derived. The design of the system is such that the aberrations after two reflections are minimized. The introduced system is compared to a conventional omnidirectional system based on a hyperboloidal mirror. For the hyperboloidal system the sagittal and meridional images are further apart, leading to larger astigmatisms. The mirror design in Takeya et al. (1998) however minimizes the distance of these images, allowing a focused mapping.

### 3.2.4 Calibration of Catadioptric Systems

The calibration of catadioptric systems is the topic of Geyer & Daniilidis (1999). The main interest is on the paraboloid-orthographic camera combination. The approach to determine the intrinsic parameters of the system shows that the projection of sets of parallel lines is sufficient to determine the parameters set (see Fig. 3.3). For the method to work, the sets of lines must not be parallel nor the outer circle of the parabolic image must be fully visible. Prior to that publication the intrinsic calibration

was performed manually under the condition that the whole parabolic image is visible in the image. The user had to determine the principle point of the system, i.e. the intersection point of the axis of the rotational paraboloid with the image plane, and the radius of the outer circle just based on the appearance.

The approach is based on the fact that lines are mapped to sections of circles in the parabolic image. By determining the vanishing points of the parallel line set, i.e. fitting circles to the images of the lines, two intersection points are obtained. Those points represent the image of the vanishing point of the respective line set. The principle point of the system is located somewhere along the line connecting the two points. This is a property of the system's geometry. Evaluating a second set of parallel lines yields another line connecting the two vanishing points of this set. The intersection of the two connecting lines is the principle point of the system. With this knowledge and the vanishing point locations the focal length of the parabola is determined. These three parameters fully describe the mapping of the parabolic camera.

Geyer & Daniilidis (2001) provide an enhancement of the method reviewed above for the parabolic camera case. The necessity of imaging at least two sets of parallel lines is reduced to the requirement of imaging just two single lines. This is achieved by extending the notion of the stereographic projection to the central catadioptric projection. This creates a virtual projection sphere. Its equator is a line image which contains the point  $(c_u, c_v, 2f)$ , with  $(c_u, c_v)$  being the principle point location in the image, and  $f$  being the focal length of the system. The three-dimensional intersection of two such virtual spheres determines the intrinsic parameter set. Geyer & Daniilidis (2001) give a simple algorithm how this intersection point is obtained in the case of a parabolic mirror system. This publication not only introduces a calibration method but it explicitly states the analogies between the stereographic projection, i.e. a central projection such as performed by an ordinary perspective camera, and the catadioptric projection retaining a singular viewpoint.

Barreto & Araujo (2001) discuss a three-step calibration method for central catadioptric camera. The first function is a linear mapping of the world onto an oriented projective plane. The second one describes a non-linear transformation between two such projective planes. The final analysis is a collineation in the plane. The whole method is used to study the projection of lines by a catadioptric system. The results are the descriptive intrinsic parameters of the system. However, the approach is not restricted to a parabolic mirror but other systems retaining a singular viewpoint are discussed as well.

Especially in catadioptric system, the mounting of the mirror is essential for the precision of the whole system. In optical systems, the alignment should be a design property rather than a topic of a subsequent system analysis. Tingstad (1991) describes a procedure with an aspheric mirror to align a two-mirror-subsystem relative to a lens unit. This is being done to improve a weak design with a superior adjust-

ment. To satisfy strict geometric requirements a catadioptric system has been chosen. For the alignment a Zygo interferometer with a working point at the wavelength  $\lambda = 632.8$  nm of the He-Na laser is used. The measure for the system's proficiency is that the aligned system should exhibit similar optical paths through the entrance pupil. This results in a null interference pattern. The application of interference pattern in laser tracking systems is used for high-precision coordinate determination (Zhuang & Roth, 1995). They provide the possibility of continuous, non-invasive, and highly precise measurements in a large working space. In addition, the publication deals with the modeling of errors, which is essential for the enhancement of such high precision applications. This is the topic of Tamura et al. (1994). There, a model describing the systematic errors, based on the geometry of the system, is developed. Its parameter values are estimated iteratively, to minimize a cost-function that describes the deviation of the measurement error from the model.

Complex optical systems exhibit assembly and production errors. In a real system, single components cannot be separated and calibrated individually. Therefore, an indirect and adaptive method for a simultaneous estimation of the variable parameters of the system must be derived (Kim & Cho, 2000). This publication not only presents a generalized projection model of a  $n$ -mirror system, but it introduces a learning-based approach of recursively determining the parameter set. After building the model, the constitutive parameters are extracted. These are the mirror positions, their orientations, i.e. the normal vectors, the distortion parameters, the scaling factors, the effective focal lengths, and the rotational and translational vectors. Depending on the mounting some or all of them are variable. The dependencies of the parameters on each other is usually unknown and non-linear. The solution to this complex problem is achieved by the minimization of a criterion-function (energy-function) with a steepest-gradient method. To do so, a Delta-learning function is used. The starting point is the parameter vector  $\Theta_0$ , containing the designed parameters. Subsequently, the gradient  $\nabla E(\Theta)$  of the actual error function is determined. The consecutive value of  $\Theta$  is retrieved by a motion along the negative gradient of the error-surface. With this learning procedure the positioning errors of the detected points are reduced by 89% with respect to the use of the design parameters. However, only mirrors of known, closed-form geometry are used: cones and planes.

### 3.3 Reflections

Reflections are examined under different points of view. One of the first publications is Longuet-Higgins (1960). There, the reflection of light incident on a rough surface is investigated. The generated patterns move across the surface and the image points correspond to extrema and saddle-points of certain functions. An observer looking

onto a rough surface (e.g. the surface of a lake, roughened by the wind) sees images of the light source at different locations. These observations are called *Specular Points*. The number of specular points changes: points are created and vanish again. It is shown that the intensity in this blinking is very high, due to a partial focusing of the light onto the observer. Water waves are irregular and random. The inclinations of the wave tangents, if the waves are created by the wind, have an approximately Gaussian distribution. The water surface is an infinite sum of extended waves of different wavelengths and directions. The number of observable specular reflections on an isotropic Gaussian surface depends quadratically on the distance and linearly on the mean of the squared curvature. The number of light flashes in a time interval is a function of the frequency-spectrum of the surface and the distance of the source to the observer. It is shown that the blinking frequency depends, omitting a scaling factor of the surface, only on two parameters: one is the source-observer distance and the other is the wave type - static or dynamic.

Longuet-Higgins restricts himself to specular reflections. However, Torrance & Sparrow (1967) analyze the directional distribution of the radiation flow of reflected light on a rough surface. That model, based on geometric ray optics, describes the surface as a sum of small arbitrarily distributed facet-like mirrors. It includes the effects of obstruction of the facettes by each other. The basic mechanism is the specular reflection on those facettes and an additional diffuse component created by multiple reflections and/or intrinsic scattering. The resulting angular distribution is in good accordance with the experimental measurements on metal and non-metal surfaces. In addition to that, it represents the non-specular maxima of the intensity distribution if the angle of incident is increased. Based on this reflection model Oren & Nayar (1994a) and Oren & Nayar (1994b) develop a more comprehensive model. It is shown that the surface roughness is constitutive for the deviation from Lambert's Law (see Chapter 2). The deviation increases with significant and rising roughness and the angle of incidence. The model is applicable to isotropic and anisotropic rough surfaces with an arbitrary observer and light source position. Approaching the limit of a smooth surface, it reduces itself again to Lambert's Law.

All the publications reviewed so far dealt with the basic processes invoked in reflections. The following are more concerned with reflections in image processing. To guarantee a robust localization of objects in image processing, reflective surfaces need to be recognized as such. Otherwise, in the context of this thesis wrong locations in space are determined. In single images the extraction of reflective surfaces is almost impossible. However, in image sequences it is feasible. There are two approaches to do so. The first one is associated with a moving camera and introduces the notion of a *caustic* (Oren & Nayar, 1996). This one- to three-dimensional surface is the envelope of the reflected rays created by the camera movement. Oren & Nayar (1996) use image sequences recorded during a smooth, i.e. continuously derivable, camera move-

ment. In addition, it is assumed that the surfaces have a Gaussian curvature, i.e. they are neither planes nor parabolas. Evaluating the caustics of objects within the image yields a differentiating property. The caustic of a reflected object exhibits a different geometry than an object in direct view. Directly viewed objects usually have a point caustic, whereas reflected objects have elongated curve segments. The algorithm is able to do this for unknown objects and without the knowledge of supporting points. The second approach, so far rather neglected in literature, takes the object's physical properties, such as reflectivity, roughness, and material type into account. The projection of a three-dimensional scene onto a two-dimensional image plane results in the loss of a fundamental part of the information about the scene geometry. Geometric properties are therefore not always useful for a robust object recognition. Determining objects on the basis of the physical properties is only applicable if the properties are derivable from the image. The radiation density of a surface, i.e. the observable intensity, is a product of the reflectivity and the illumination. Hence, these components are not separable regarding a single point. Neighboring points on a curved surface share the same illumination condition and usually possess similar normal vectors. Therefore, the intensity distribution allows the calculation of the reflection coefficient. Based on these facts, an algorithm is developed which is able to estimate the reflectivity parameters of an image region relative to the image background. This is done in only two processing runs across the image data. In Nayar & Bolle (1996) this image-invariant object parameter is used to locate an object in an intensity image.

As mentioned before, reflections are usually treated as sources of error. Bhat & Nayar (1995), respectively Bhat & Nayar (1998), treat the problem of a correct depth recovery in the presence of specular reflections on rough surfaces. With a given lower limit in surface roughness an optimal stereo configuration is derived that maximizes the depth recovery precision. Matching operators, such as the Sum-of-Squared-Differences (SSD), assume a correspondence, if the intensities in both images are identical for an object. Due to the possibly large difference in viewing angles, this assumption cannot be upheld anymore. Even using three cameras, verifying a stereo image pair with a third image, it happens that, due to reflections, a match is found in the stereo image pair which does not correspond to a real match in all three images. In addition, a real match determined in the stereo pair is removed by the failure of the verification in the third image. The algorithm introduced is able to cope with these situations. Even in the presence of peak reflections (highlights) correct matches are obtained, without any preprocessing of the three images. The approach is generally applicable due to the lack of requiring a specific reflection model. However, the publications use the reflection model introduced by Torrance & Sparrow (1967) which is reviewed above.

### 3.4 Image Display

These publications are related to the topic of this thesis because there might be methods in the display of warped images relevant to the solution of the posed problem.

One such method where image distortions are used intentionally is the anamorphic technique in the corrective optics and in the arts. All optical systems with two different values in the main optical sections are called anamorphic. This is used to correct astigmatic distortions if the error occurs only in one meridian. In projective optics, e.g. in the movies, it is used to achieve a different image size on the screen than available on the film. However, the reconstruction is only achieved if the film is recorded using the respective lens counterpart. Anamorphic displays have a long history in the arts. There, methods were developed to paint an image on a warped surface, e.g. the ceiling of a church, such that the observer sees an undistorted scene.

The step into the computer-vision age is performed by Lippmann (1980). The introduction of the *Optical Videodisk* enlarged the possibilities in television. Image manipulations and image processing merged (overlays etc.).

Virtual reality systems use three-dimensional computer graphics to model and render virtual environments in real-time. This usually requires extensive modeling and specialized and expensive computer hardware. The reproduction quality and the image detailing is limited by the real-time requirement. Chen (1995) presents a new approach, using 360° panoramas to create a virtual environment representation. The displayed image is distorted in real-time to simulate pans and tilts of the camera or the observer. The approach is commercialized in *QuickTime VR*, an extension of Apple Computers digital multimedia framework *QuickTime*. To create the panoramic views first a grid is determined where images with an ordinary camera are recorded. These are stitched together in the application and existing gaps of information are filled using interpolation routines. If a movement requires more than one panorama the viewing directions are adapted manually.

Another system which does not depend on the explicit knowledge of the scene geometry is presented by McMillan & Bishop (1995). This image-based rendering is able to produce highly sophisticated animations. It uses the *Plenoptic Function*  $\mathcal{P}$ , introduced by Adelson & Bergen (1992), to deliver a precise problem definition for image-based display paradigms, such as morphing or interpolation of viewing directions. The Plenoptic Function  $\mathcal{P}$  is a parametric function describing everything visible, in all desired wavelength bands, from a point of view in space. Subsequently, it is the

**Problem definition of image-based rendering systems:** Given are discrete, complete or fragmented samples of the Plenoptic Function  $\mathcal{P}$ . The task of image-based rendering is to derive a continuous representation of the Plenoptic Function  $\mathcal{P}$  from that information.

The presented system is based on sampling, reconstruction, and resampling of the Plenoptic Function  $\mathcal{P}$ . In addition, a new algorithm for visible surfaces is introduced. It contains a geometric invariant for the cylindrical projection, equivalent to the epipolar constraint in planar projections. Morphing is a common image manipulation method. Two images are the starting and the ending point in space and time. Hence, the images in between are a path through space and time. Considering this, morphing becomes a partial reconstruction of the Plenoptic Function  $\mathcal{P}$ . Morphing usually uses additional information, such as a predefined optical flow. To represent the plenoptic samples different methods are applicable. The spherical projection is difficult to use, due to its intrinsic curvature. Greene (1986) proposed a storage on the six surfaces of a cube. However, this is not suitable with regard to the alignment and the registration of real images. In addition, the information is not evenly distributed. The density along the edges and in the corners is much higher. But the biggest problem of Greene's proposition is the representation of the optical flow across the edges and corners. This becomes even worse, if the images are not only displayed but used for processing. Considering all this, cylindrical projections are used. The advantages are, that they are displayable on a plane and there are no borders in the azimuthal direction. This simplifies the correspondence searches necessary for the determination of the optical flow.

Catadioptric systems are used for further enhancements of projection optics. This is mainly due to the lack of chromatic aberrations in mirror systems. One of these enhancements is the i-line projection technique for scanning systems (Otha, 1994). However, due to its irrelevancy to the posed problem in this thesis, this shall not be detailed any further.

### 3.5 Conclusions

In this chapter various publications are reviewed to see whether solutions for the problem given in this thesis are available. According to the desired strategies, i.e. exploiting the information in a free-formed surface reflection, other methods to solve multiple matching hypothesis situations, such as a new matching algorithms, are disregarded.

In the context of stereoscopic image processing the reflections provide a third respectively fourth image. Therefore, algorithms based on multi-image processing are examined. Usually, these depend on special geometries, such as a predefined relative camera position or on rectifications of the images to the standard stereo geometry.

In general, the desire to rectify the image reflected on a free-formed surface to a pinhole view cannot be satisfied. To reconstruct a correct perspective image the combination of the mirror and camera must retain a single viewpoint. Baker & Nayar

(1999) derive the classes of mirrors and the necessary types of cameras that obey the constraint. The result is that only conic sections and planar mirrors in combination with perspective and orthographic cameras have a single viewpoint.

One major field of mirror applications in computer vision is omnivision. The mirrors, usually members of the single-viewpoint classes, are designed such that the field of view of the camera looking into it is enlarged up to  $360^\circ$ . With the single viewpoint, perspective and panoramic images are reconstructed and displayed or used for further processing.

A review of the publications about the calibration of catadioptric systems shows that even for the closed-formed geometries calibration is a necessity. A lot of effort is put either into the adjustment and its measurement or in the determination of the descriptive parameter set of the system. One must expect, that in the case of a free-formed surface an application without a profound system calibration is not feasible.

Concerning the basics of reflections mostly surface roughness and its treatment are of interest. In extension of that, regarding free-formed surfaces, the determination of specular and diffuse surfaces as a source of error in image processing is a topic of research. The aim is clearly the exclusion of the reflections.

Image display methods are concerned with mirrors in the projection systems and warped surfaces as a display screen. However, with respect to the intended application here, the results are not readily applicable.

Only two publications are closer related to the topic of this thesis: Pajdla (2002) and Hicks & Bajcsy (1999). Pajdla (2002) introduces a non-central and non-classical camera. The basis of it is a rotating combination of a tele-centric lens and a conic mirror. He proves the feasibility of stereo vision with a generalized camera model. The only application using a free-formed surface as an image-formation sensor is Hicks & Bajcsy (1999). There, a mirror automatically rectifies the view of a predefined plane. In order to do so, the mirror is especially designed.

The view onto a free-formed surface with an independently designed geometry, and obviously without the intent of being used in an optical system, was so far never considered. Hence, methods to extract the contained information are not existent either. The examination of the free-formed mirror properties and the new methods to use them appropriately are the contents of the consecutive chapters.



# Image and Information Reconstruction

In this chapter different solution approaches, considering the reviewed literature and the demands stated in Section 2.5, are investigated. First, some model assumptions, relevant for the most of the thesis, are stated. In the second part, classical reconstruction efforts are examined, amongst them ray-tracing methods and other image operations using different geometric foundations. Subsequently, thoughts about the retrieval of the contained information without an image reconstruction are presented. Finally, common dependencies for most of the efforts are elaborated.

In order to develop the solution strategies, it is assumed that the geometric setup is known, i.e. the intrinsic and extrinsic camera parameters, the relative position of the camera to the surface, and the surface geometry itself with its normal vectors and directions of reflection for each surface point. This data might either be retrieved through calibration methods (discussed in Chapter 5 and 6) or known from other sources, such as the CAD design (illustrated in Appendix A).

### *4.1 Model Assumptions*

In the course of this thesis, the theoretical suitability as well as the practical applicability of free-formed surface mirrors are investigated. In order to do so, some assumptions are made. Most of them hold true throughout the whole thesis while others only apply to the theoretical part. In the experimental discussion in Chapter 8 the focus is on practicability.

#### **Assumptions**

1. *The free-formed surface mirror is a specular reflector.*

One distinguishes specular and diffuse components in reflection processes. In computer graphics or shape-from-shading techniques, the directional distribution of the reflected radiation is dependent on the relative position of the camera, the surface, and the light-source. This distribution is described by a BRDF model (Bi-directional Reflectance Distribution Function). These functions are

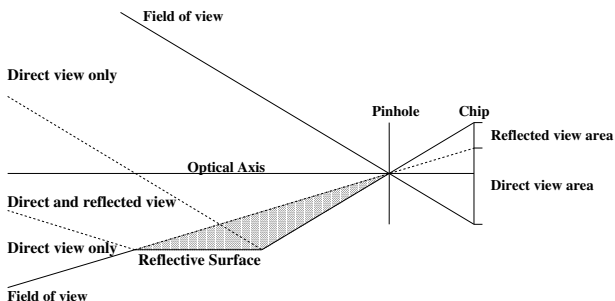


Figure 4.1: Restricted Object Location Space: Objects located in the hatched area cannot be correctly retrieved. Therefore, it is assumed that no objects are found in this space.

more or less sophisticated and physically correct (see Torrance & Sparrow (1967) or Oren & Nayar (1994b)), depending on the desired visual result. However, in this context the task is not modeling as such. Therefore, BRDFs are not relevant. The smooth and shiny varnished hood of the car, one of the best varnished parts on the whole car (according to the varnishing laboratory), exhibits a surface roughness at the scale of micrometers. Subsequently, the hood will be treated as specular reflector. This is the basis for further investigations using ray optics rather than wave optics.

This assumption will be reconsidered in the experimental section.

## 2. The camera is modeled as a perspective pinhole camera.

Due to the technical setup of the experimental car and the applied cameras, the perspective mapping is given. In reality, based on the low irradiance of the pinhole camera, lenses are used. However, these lenses introduce distortion in the image which have to be corrected, based on a calibration. Orthographic cameras are not in use in the intended application and therefore they are not considered here.

## 3. There are no objects located between the camera and the reflective surface.

This problem is depicted in Fig. 4.1. If there are objects in this restricted space, an approach such as Oren & Nayar (1996) must be applied. Oren & Nayar (1996) conduct an analysis of the object's caustic<sup>1</sup> and distinguish direct view objects and reflected view objects on that data basis. With regard to the intended application, i.e. cameras mounted in a car, such an analysis is obsolete.

<sup>1</sup>The envelope of reflected rays produced by a relative sensor motion.

4. *The reconstruction effort is focused on the central, concave part of the hood.*

Fig. 1.2 shows a reflected road scene. Obviously, at the convex side-parts the distortion is at its maximum and the reconstruction is complex. With regard to the application, the direction of view in those parts is of no interest. Fig. A.1 through A.4 show CAD data representation of the front of the car and the central, concave part of the hood. There, the most interesting information is visible from the camera position in the traffic application. In addition to that, the central part, shown in Fig. A.4, is a single CAD data patch and not segmented.

5. *The configuration is static.*

A static configuration refers to a constant geometric setup of the camera position relative to the surface and of the surface itself. With the application of an on-line camera calibration and an adaptive surface recovery, this restriction is abandoned.

This assumption will be reconsidered in the experimental section.

## **4.2 Image Reconstruction Strategies**

It is desirable to reconstruct a pinhole image from the reflections in the hood of the car. Apart from obtaining an image which is pleasant for the human eye, all the ordinary, well-known algorithms of stereo computer-vision are applicable to evaluate the scene. Recalling the reviewed literature with regard to this desire, only certain classes of mirrors allow such a reconstruction. Still, in-image operations, based on different geometric considerations, are examined.

### **4.2.1 Reconstruction within the Image Plane**

A possibility to reconstruct an image, is to record a known pattern and compare the reflection to the original. In Fig. 4.2 the simulated reflection of an upright checkerboard wall in the hood is displayed. The distortion of the quadratic structure in the lower part of the left image is plainly visible. A conceivable approach is a direct search in the image for corresponding points in the direct and the reflected image. The result is a two-dimensional field of translational vectors. This type of approach is similar to a calibration. For stereo computer-vision it is essential to obtain views onto the same scene from different angles. This is necessary to supplement the lost information in the two-dimensional mapping of a three-dimensional scene of each camera. However, if the reflected image is reconstructed with this concept the necessary additional information is lost. The right image in Fig. 4.2 shows the perspective of a camera placed underneath the hood. To the perspective of that camera, the reflection image needs to be reconstructed. However, neither the extrinsic nor the intrinsic

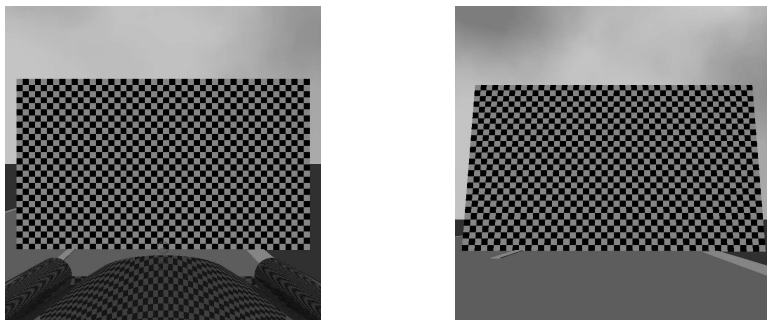


Figure 4.2: Direct, Reflected, and Virtual View on a Checkerboard Wall: (Left) The view of a camera onto a checkerboard wall, directly, and reflected in the hood. (Right) View of the same checkerboard wall from a camera located 'inside' the hood of the car.

parameters of this virtual camera are known. The application of other methods to determine these parameters would solve the task and this two-dimensional image-based procedure is obsolete. A merely image-based approach, without consideration of the three-dimensional geometric setup in some way, cannot be appropriate.

### 4.2.2 Ray-tracing

Several methods in geometric optics depend on ray-tracing. Rays, originating from a source in a certain direction, are traced. Along their given path they intersect with objects in general and their surfaces in particular. Considering their structure, especially with regard to the optical properties such as the refraction and transmission indices, and their geometry on a large and on a small scale, i.e. the relative position of the surface to the ray and the surface roughness, the altered direction of the ray is obtained. Usually, in ray optics, surface roughness and the related diffuse reflection component, are not considered. In more sophisticated approaches the reduction of intensity is taken into account. To obtain the altered ray direction methods such as the law of reflection or Fermat's principle are applied. This procedure is continued until the desired destination is reached, e.g. the intersection with an image plane of a camera.

In the first of the following paragraphs, a more general ray-tracing method is presented, according to the historical conduct and development of this thesis. In the second paragraph an advanced examination with a professional optical design tool is introduced.

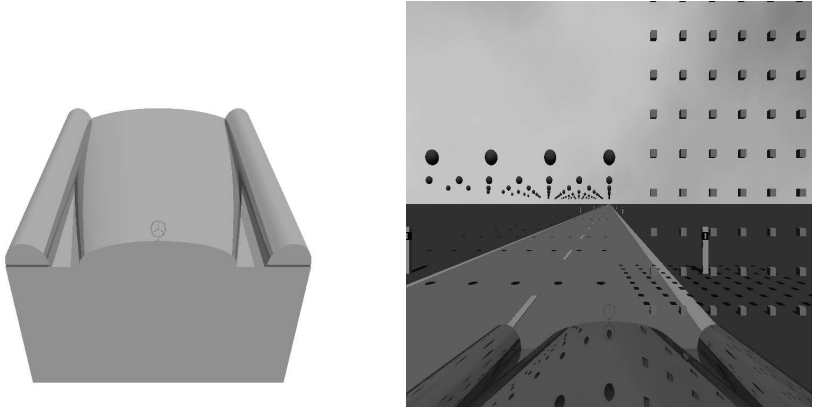


Figure 4.3: A Simple Ray-Tracing Approach: a generic car front, consisting of boxes, cylinders and a stretched cone, is modeled with POV-Ray (left). A simulation of the reflection of test objects with the generic model (right).

### General Approach

In this paragraph the first models on the basis of POV-Ray<sup>2</sup> are presented. To examine the rudimentary properties of the reflection of objects in the hood of the car, a simple car front model was generated and tested (see Fig. 4.3). The dimensions of the model and the road, and the relative position of the camera to the reflective surface are realistic. The test objects have a diameter of 10 cm. However, POV-Ray only approximates reflections on free-formed surfaces. Due to the necessity of a physically correct solution, further POV-Ray simulations were not implemented.

### Advanced Ray-Tracing Methods

A much more sophisticated ray-tracer for the professional simulation and design of optical systems is the Advanced System Analysis Program (ASAP) by Breault Research Organization Inc. This software offers the possibility to directly use CAD data to define free-formed surfaces. It traces the rays in a physically correct way, including the reflections on free-formed surfaces.

The reviewed literature alleges that only certain geometrically close-formed surfaces in combination with specific cameras retain a singular viewpoint. The singular viewpoint enables a correct perspective reconstruction of the image. The existing image-processing algorithms are dependent on a singular viewpoint. In order to apply

<sup>2</sup>POV-Ray (Persistence of Vision - Ray Tracer), is a registered trademark of POV-Ray Team, Australia.

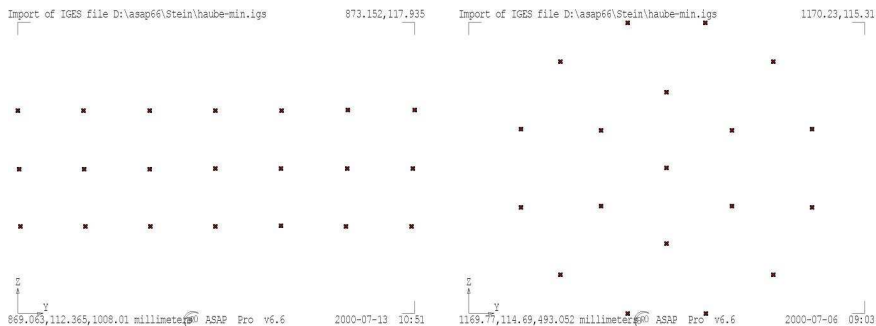


Figure 4.4: Distribution of the Rays in the ASAP Simulation: In the left graph, the distribution of the 21 central rays in the image plane is shown. The right graph depicts the distribution of the 18 additional rays around each central ray.

them, a singular viewpoint of the reconstruction image is desirable. To substantiate the claim of the reviewed literature, an extensive simulation was conducted.

To perform the simulation an inverse ray-tracing approach is chosen. An inverse ray-tracing process defines the destination and traces the origin of the rays. To calculate this, the following setup is used. The central part of the hood, shown in Fig. A.4, is introduced as a reflective surface, represented by its CAD data. The CAD coordinate system is the center of the front axis of the car. Accordingly, this is the origin of coordinate system in the simulation. All of the evaluation graphs use the same origin. The SI-Standard is valid on the coordinate axes, if not stated otherwise. The camera is a pinhole model with a focal length of 7.3 mm, similar to the cameras used in the car. The camera is placed at the relative position of the real cameras, off the symmetry plane of the hood. For the simulation only the right camera of the stereo system is considered. In the image plane of the camera 21 center rays are evenly distributed. The distribution is shown in the left graph of Fig. 4.4. To form a bundle, around each of these central rays 18 additional ray origins are placed. Six in a inner circle, and 12 in an outer one, such that the spanned cone fills the aperture. This distribution is depicted in the right graph of Fig. 4.4. Starting from these 399 points, the rays are traced through the focal point of the camera. Their intersection points with the surface are determined.

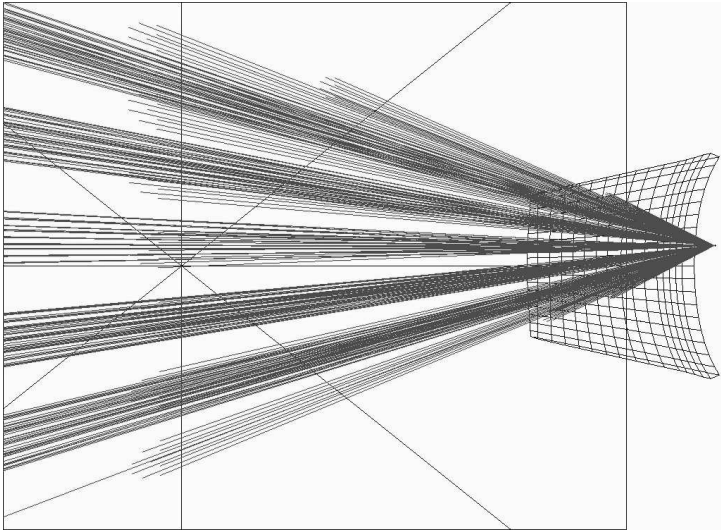


Figure 4.5: Topview on the Rays Traced in ASAP: The camera is on the right side, where the rays originate. The rays are traced to their destination on the left side, a wall at a distance of 5 m.

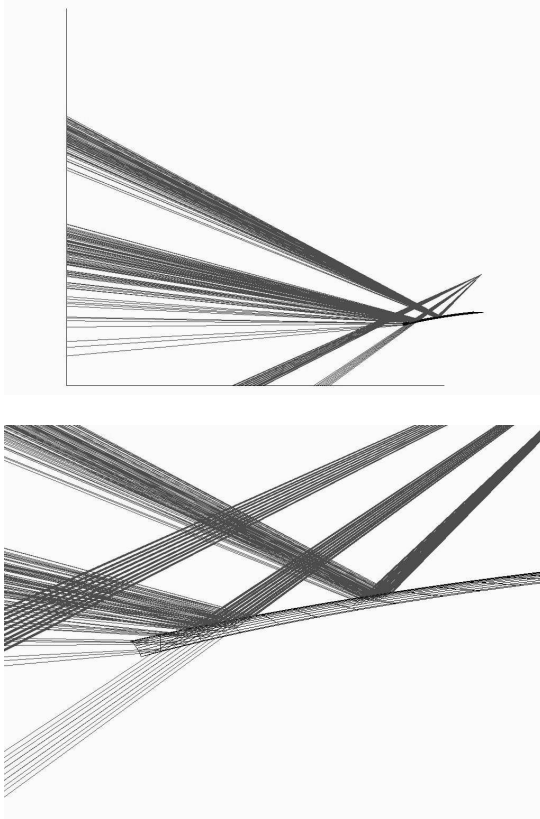


Figure 4.6: A Sideview of the Rays Traced in ASAP: In the upper graph, the whole ray propagation is shown. On the right side of the that graph is the camera location, where the rays originate, on the left the evaluation wall at a distance of 5 m. The lower graph shows an enlargement of the reflection on the hood. Not all of the rays intersect with the reflective surface.



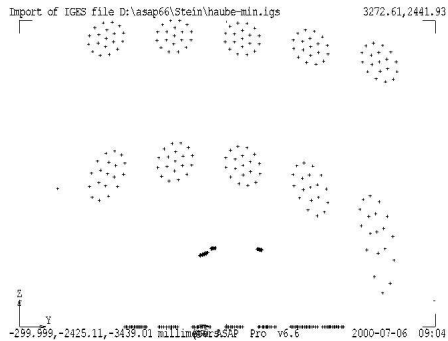


Figure 4.7: The Rays on the Destination Wall in the Simulation A: 399 rays are traced from the image plane of the camera through its focal point. The rays that intersect with the surface are reflected on it and traced to a wall at a distance of 5 m. The result is shown in this graph. Compare it with the original pattern shown in Fig. 4.4.

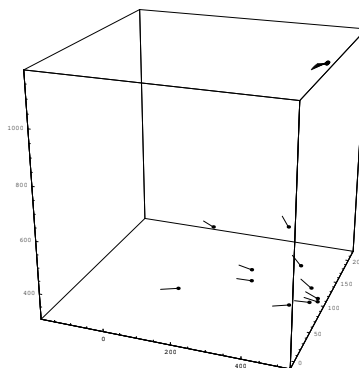


Figure 4.8: The Bundle Focal Points in Simulation A: In the upper right part of the graph, the original camera location with its bundles is shown. In the lower part the foci and mean directions of the reflected bundles, shown in Fig. 4.7, are depicted. Obviously, they do not form a singular viewpoint.

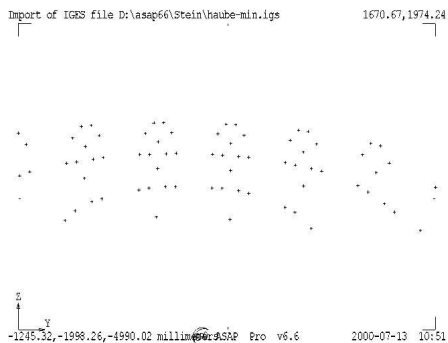


Figure 4.9: The Rays on the Destination Wall in Simulation B: Compare it with the original pattern shown in Fig. 4.4 and the results of the simulation A in Fig. 4.7.

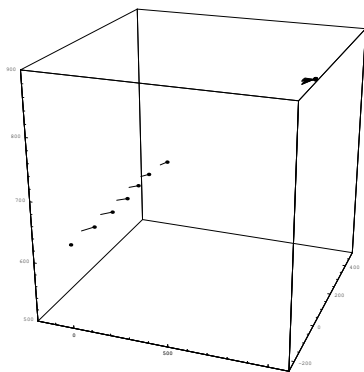


Figure 4.10: The Bundle Focal Points in Simulation B: Compare the distribution to the results of simulation A in Fig. 4.8. The sensitivity of the system to the camera position relative to the surface and the angle of view onto the surface is obvious.

Notice that in Fig. 4.5 and 4.6 not all of the rays intersect with the surface. Those intersecting with it are specularly reflected and traced to an upright wall at a distance of 5 m from the coordinate origin. This wall is the destination for the rays in the inverse tracing.

Since the rays incident on the wall are reflected, they seem to originate from a point below the surface. The ASAP software determines for each reflected bundle its focal point and mean direction of propagation. If the foci of all the bundles coincide at a single point the catadioptric system retains a singular viewpoint. To evaluate the simulation, the focal points and the mean directions are depicted.

Two simulations, referred to as A and B, with slightly different relative camera positions were conducted. In simulation A the camera exhibits a pitch angle of  $-35^\circ$  and is positioned at (0.493, 0.115, 0.870) m. The yaw and roll angle to the world coordinates are zero. The resulting distribution of the rays on the destination wall is shown in Fig. 4.7. Obviously, the circular aperture is imaged as an elliptical shape. This astigmatic behavior is expected from a curved mirror. The distribution of focal points and mean directions of the reflected bundles is presented in Fig. 4.8. The focal point and the directions shown in the upper right part of the image correspond to the location of the original camera. The foci and mean bundle directions in the lower part are those corresponding to the reflected ray bundles (see Fig. 4.7). The mean directions sway the desired field of view quite satisfactory. However, the focal points do not coincide. On the contrary, they are distributed in a rather large space.

One might expect that not only the surface geometry but also the relative camera position influences the result of the ray-tracing. To investigate this, a second simulation, called B, is performed. In this simulation the camera is positioned at (1.000, 0.115, 0.870) m, i.e. slightly higher than in simulation A, but with a smaller pitch angle of only  $-15^\circ$ . The yaw and roll angle are again zero. The ray bundle and foci distributions are shown in Fig. 4.9 and 4.10 respectively.

The comparison of the results clearly exhibits that the relative camera position with respect to the surface and the angle of view onto the surface are of essential influence on the imaging properties of the system.

### 4.2.3 Comparison of Lines

Another possibility to solve the reconstruction, reducing it to a search problem, is shown in Fig. 4.11. The idea behind the procedure is as follows: At a certain point a tangential plane to the surface is used as a virtual planar mirror. At the tangential point, the direction of both, the planar and the free-formed mirror, is identical. For the planar mirror the directions of reflection associated with each image location are determined. The intersection point of the ray of sight with the planar mirror and the determined direction of reflection define a line. In the following step the point on the warped surface and its associated direction of reflection are determined such that this

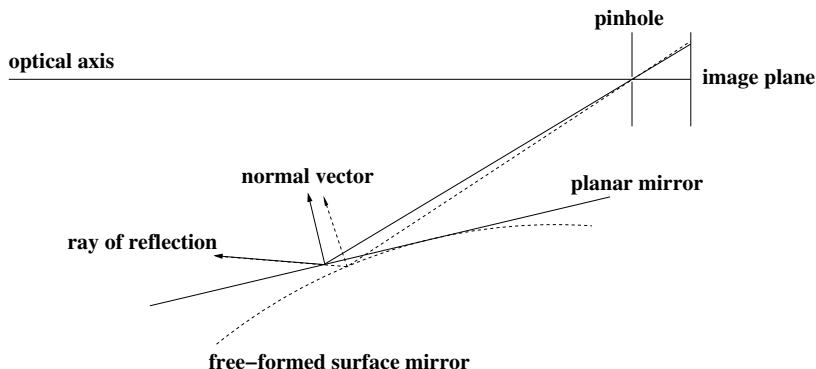


Figure 4.11: Reconstruction by the Comparison of Lines: The warped surface is approximated by a planar mirror. The image is reconstructed such that the ray of reflection of the planar mirror approximates the ray of reflection of the curved mirror to the achievable maximum degree.

line and the line of the planar mirror are the best approximations. Considering the advanced ray-tracing results, identity cannot be expected. The point on the free-formed surface corresponds to a point in the image plane. The line on the planar mirror corresponds to an image location. By solving the line-comparison search-problem a correlation of the two image locations is established. Repeating this procedure for all the image location of the planar mirror results in a two-dimensional vector field, expressing the image location associations. In contrast to the method discussed in Section 4.2.1, this is not a mere in-image operation, although in results in one. It incorporates the geometry in the vector field. The virtual camera associated with the planar mirror is the reflection of the original camera location at the mirror plane. The optical axis of the virtual camera is defined by reflecting the real optical axis at the planar mirror.

One problem with this method is that it might result in a sparse distribution of associated image locations in the reconstructed image. This requires either a gray-value interpolation to densify the information grid or a sub-sampling of the planar mirror. Another problem is, that by using just one planar mirror for the whole reflection image errors are introduced. Around the point of the tangential plane the best results are to be expected. Away from that point the quality of the reconstruction will diminish. This is due to the increasing distances of the warped surface from the planar one.

### 4.3 Information Reconstruction Strategy

Reconstructing the image by the Comparison of Lines (see Section 4.2.3) uses one planar mirror for the whole reflection image. As mentioned above, this results in errors and a sub-optimal image reconstruction. To overcome some of these flaws an approach based on active and adaptive optics, used in astronomy, is taken. In order to present the idea for the posed problem of this thesis, first the applied methods in active and adaptive optics are introduced, following the publication of Hubin & Noethe (1993).

Ever since the introduction of Galileo's telescope two parameters determine the development of instrumentation in astronomy: the diameter of the telescope, i.e. its ability to collect light emitted by objects far away, and the angular resolution, i.e. the depth of focus. The resolution of a perfect telescope in a vacuum is only enhanced by increasing its diameter. A planar wavefront of a star at infinite distance is transformed to a perfectly spherical wavefront. The maximum angular resolution achievable that way is called the *diffraction limit*. The reality is different. For one, the planar wavefront is altered by the atmosphere, for the other, imperfections of the telescope account for a non-spherical wavefront. This is equivalent to errors in the propagation of rays in the projection. Even perfect earthbound telescopes with an arbitrary diameter are not able to achieve a better angular resolution than a telescope with 10 to 20 cm in diameter. This is caused by the turbulence in the earth's atmosphere. Due to the atmosphere, the point resolution of a 4 m telescope is diminished by a magnitude compared to the diffraction limit, and the maximum intensity of an observed star by a factor of 100. The reasons are found in random spatial and time-wise disturbances of the wavefronts, caused by turbulences of the different atmospheric layers. These considerations are decisive for the development of satellite telescopes. In addition to that, the image quality is reduced by long-term effects. These are thermal and mechanical changes, optical effects in the telescope, such as defocusing, alteration of the centering, and deformation of the mirrors by the supporting mechanics. The correction of such errors are summarized as *active optics*.

So far, telescopes were understood as passive instrumentation. However, technical advances allow the construction of active telescopes. The idea of a real-time compensation of atmospheric influences was introduced by Babcock in 1953. He demanded active optical elements, working at higher frequency, to compensate continually for wavefront deformations. The usage of adjustable optical elements rather for the compensation of atmospheric distortions than for flaws in the telescope, became known as *adaptive optics*. It took about 20 years, advances in computer technology, and a better understanding of atmospheric processes before his demands became reality. After the development was secretly driven by the military it became available for astronomy during the last decade.

The elementary definition of adaptive optics is, that this technology eliminates in real-time distortions of the wavefronts and therefore almost achieves the diffraction limit. In reality, this is an optimization process on a multitude of parameters in a closed loop.

One problem implied, is the measurement of the atmosphere. The object to be observed cannot serve as a control and others, with a sufficient brightness, might not be within the field of view. To solve this, artificial reference stars, so called laser guide stars, are generated. In the upper mesosphere (90 km height) of the earth a layer of sodium, and in the lower stratosphere (10-20 km height) sodium molecules are found. Excited with a sodium laser, their light emission is recorded by a wavefront sensor, focused at the respective height. This data is the basis of the correction of the telescope mirror.

On the technical part there are many implementations. Most of them use piezoelectric regulators to deform the mirror. If the mirrors, due to their thickness, are not flexible enough, segmented mirrors are used. A well known example is the 10 m Keck telescope on Mauna Kea, Hawaii, USA. In it, 36 hexagonal elements form a mirror. Each of the segments has a diameter of 1.8 m. The four new ESO telescopes of the Very Large Telescope (VLT) have monolithic mirrors of 8 m diameter, supported by 150 active regulators.

Considering the adaptive optics approach, an image acquired by an electronic retina with its pixel raster is understood as a scanning of the reflective surface. Subsequently, each pixel is associated with a patch of the surface. To a certain degree, this is the extension of the method presented in Section 4.2.3. However, in this case not just a single plane is used but rather one plane for each pixel. With the knowledge of the mapping of the camera, i.e. the intrinsic parameter set, each of the finite image locations is transformed into a ray of light propagation, i.e. a line in three-dimensional space. On this line the observed object is located. In adaptive optics a known object, the laser guide star, is used to determine the geometry of the mirror. Here, this is inverted, because the geometry is known but not the object.

This approach does not reconstruct the image in a classical sense. Each single pixel forms an independent camera. As a consequence, the procedure of extracting the information is different from the usual stereoscopic approach. If the image is reconstructed to a perspective one with a different view onto the same scene, the reconstruction comes first, followed by a correspondence search along an epipolar line. The recovered disparity, i.e. the displacement of the object along the epipolar line within the image, results in the objects distance. In this case, the image is not reconstructed and the back-projection of the ray, which generally defines an epipolar line, is not a line anymore but rather a set of single pixel cameras forming a curve which is associated with the surface geometry. Along this epipolar curve a correspondence

search is performed, in search windows adapted to the warped geometry. If the image location in direct sight and the image location in reflected sight, together with the direction of reflection, is known, the three-dimensional object location is retrieved by intersecting two three-dimensional lines.

Extending the notion of an epipolar line to a curve and extracting the information contained in the image without transforming it to a classical perspective one suffices to call this approach an information rather than an image reconstruction strategy.

#### **4.4 Conclusions**

In this chapter solution strategies for a given problem are elaborated. Among the classical reconstructions methods the conditions for a pinhole perspective are verified. As expected, the free-formed surface combined with a perspective camera does not retain a singular viewpoint. Therefore, other methods need to be developed. However, mere in-image operations without the consideration of the geometric setup, cannot fulfill the task.

In the course of the investigation two methods emerged as the most promising. One is the Comparison-of-Lines method (see Section 4.2.3), a classical image reconstruction approach. It is chosen for experimental implementation and is described in more detail in Chapter 8. The other one is also a new approach, derived from ideas of adaptive and active optics in astronomy. It is non-classical in the sense that it does not reconstruct the image as such. It rather extends the notion of an epipolar line to a general curve, associated with the surface geometry. This implies to understand each pixel as an individual camera. Accordingly, the information is retrieved directly from the reflection region of the original image. How the surface geometry influences the epipolar geometry is regarded in depth in Chapter 7.

Implied in all the approaches, and stated in the introduction, the knowledge of the camera parameter set and the surface geometry is absolutely essential. But not only the two of them separately but also combined, i.e. the relative camera position to surface must be known. This is verified with the help of advanced ray-tracing methods in Section 4.2.2. According to their importance the task of camera calibration and surface recovery are treated in their own chapters. Respecting their natural sequence, the camera calibration is elaborated first in the following Chapter 5. Subsequently, a new surface recovery method on the basis of the camera calibration method is introduced in Chapter 6.

## CHAPTER 5

---

# Camera Calibration

In this chapter, a look is taken at camera calibration techniques for computer-vision systems. After a short introduction into the matter a review of related work and mathematical methods that apply to the field is given. Discussing some calibration algorithms with their respective camera models in more detail leads to the presentation of self-developed enhancements to our preferred method. The chapter is concluded by a discussion of the results of the presented improvements.

### *5.1 Introduction*

In principle, there are two possibilities for cameras intended to serve as a measurement system: adjustment and calibration. Setting up a camera in its world to a pre-defined position and assuming that the camera maps the scene according to a certain precept is considered adjustment. However, as soon as the accuracy requirement increases especially the mapping precept is not sufficient anymore. Accuracy is crucial for applications that depend on quantitative measurements, such as depth-from-stereoscopy, dimensional measurements, or motion analysis from images. At this point methods are required that provide the general camera parameters to the desired accuracy level. Camera calibration has a long history, originating in astronomy and photogrammetry, e.g. Conrady (1919).

The basis of the calibration approach is a mathematical model describing the camera. Depending on the required accuracy this model is more or less sophisticated. But not only a model for the camera itself is necessary. For monocular camera systems with a measurement task, such as structure-from-motion, or multiple camera systems a description of the positioning of the camera within its 'world' is required. Accordingly, the calibration is divided in determining the intrinsic and the extrinsic parameters.

A camera consists of an image plane and a lens which provides a transformation between the object space and the image space. This transformation cannot be described by a perspective mapping alone because of distortions which occur in between the object points and the location of those points in the image. These distortions are de-



scribed by the intrinsic model. The first step in the calibration task is to determine these parameters.

A camera occupies a location in three-dimensional space described by certain angles and a translational motion relative to its surrounding. This placement is described by a rigid body transformation. Therefore, the second calibration task is to determine this extrinsic parameter set.

In the following section on related work several approaches are reviewed which fulfill these tasks.

## **5.2 *Related Work***

Publications dealing with the matter of calibration go back to the beginning of the last century. During the period of 1950 to 1970 calibration was an important matter. That was driven by the necessity of understanding aerial lenses for photogrammetry. After achieving a maturity in the required models the interest in the following two decades diminished. However, with computer-vision and the readily available and affordable hardware components and the related applications the necessity for suitable calibration algorithms and methods returned.

Rather than giving a historical overview of the camera calibration development a classification on the basis of the respective mathematical solution strategy is used. Historical overviews are found in two remarkable papers. The first is Roelofs (1951) that lists 91 articles from 1889 until 1951. The second one with an exhaustive literature review and an emphasis on the development of models for lens distortions and the evolution of methods to measure and account for that distortion is Clarke & Fryer (1998).

Calibration algorithms may be divided in two basic classes: linear and non-linear approaches. Linear approaches are usually set up in a closed form which is solved either by a direct linear transformation, a least-squares method, an eigenvector analysis, or a singular value decomposition. Non-linear methods often depend on iterative solutions. Both approaches have certain advantages and disadvantages:

- A linear procedure is stable and there is an unequivocal solution. Iterative solutions of non-linear methods may end up in local extrema or even diverge.
- A linear optimization is fast due to the necessity to solve an equation system only once.
- A disadvantage of linear methods is its susceptibility to noise. Therefore, calibration results on real images or image sequences are rather poor.

- The complexity of applicable camera models to linear methods is restrained and subsequently the achievable accuracy as well. Only non-linear optimizations account for radial-symmetric distortions of higher order and tangential-asymmetric imperfections. This is especially important for wide angle, poor quality lens systems or high precision requirements.

In the following paragraphs closed-formed solutions, direct non-linear methods, and the combination of both are introduced. Some remarks are made to the camera model and the solution strategy. The conclusion of the section is a discussion on calibration targets.

### **Closed-Formed Solution**

The following list describes some linear calibration methods:

1. Abdel-Aziz & Karara (1971): This paper introduces the direct linear transformation (DLT) algorithm. A set of intermediate parameters is defined in terms of the original ones. A linear equation system is set up and solved. The final parameter set is determined on the basis of the intermediately defined. There is an extension to the usual method to incorporate distortion coefficients. However, the corresponding formulation is not exact: The depth components of the control points, in a camera-centered coordinate system, are assumed to be constant. In addition, the constraints on the intermediate parameters are not considered in order to construct a linear algorithm. As a result, on real images in the presence of noise the calibration accuracy is rather poor.
2. Lenz (1987): This method is based on a seminal paper in the two-step method class (Tsai, 1986). It provides a linearization of the second step. It has been implemented and is described in detail in Section 5.5.1.
3. Hengst (2001): This implementation is derived from Lenz (1987). It uses a least-squares solution rather than the direct linear transformation to solve the calibration tasks. Further remarks are found in Section 5.5.1.
4. Bouguet & Perona (1998): This algorithm uses a dual-space formalism common in tensor analysis on manifolds to solve a closed-formed problem setup. A dual space of a vector space is defined by a set of linear forms on the vector space. This approach takes advantage of simple relationships between plane coordinates in projective geometry and dual-space geometry and solves the geometric projection problems with it. However, there are restrictions to the complexity of the camera model. In addition, the requirements on the calibration

object increases with the degrees of freedom in the camera model, e.g. to determine just the focal length of a system a two-dimensional rectangular shape suffices while four degrees of freedom (focal lengths in  $u$ - and  $v$ -direction and coordinates of the principle point  $(c_u, c_v)$ ) demand a three-dimensional parallelepiped with one square face.

5. Longuet-Higgins (1981): This article introduces a widely used algorithm for the determination of the camera matrix, the so called fundamental matrix. The procedure is called the 8-point-algorithm because it relies on eight point correspondences. It is fast and easily implemented. However, the algorithm has often been criticized for being excessively sensitive to noise in the specification of the matched points. Hartley (1995) shows that a simple transformation, consisting of a translation and scaling, of the points in the image before the formulation of the linear equation system brings about an enormous improvement in the conditioning of the problem and hence in stability of the result.

### Direct Non-linear Minimization

The following lists describes some non-linear calibration methods:

1. Brown (1966): This is one of the first introductions to photogrammetry of a camera model containing radial and tangential distortions. The tangential distortion is due to a slightly decentered lens and is modeled as a thin prism in front of a perfectly centered lens. The model is compared to an alternative one, the truly first introduction of the decentered lens model by Conrady (1919). The models are consistent regarding tangential distortion but deviate in the radial component. Conrady's model is the more general one but still the Brown model became seminal for sophisticated camera description in photogrammetry and computer vision.
2. Brown (1971): This publication introduces the plumb line calibration method. The calibration is based on the image of a test field consisting of a set of plumb lines. Their lack of straightness in the image is used to determine the distortion parameters. In addition to the usual concepts, the variation of the distortion with the object distance is taken into account. Brown (1971) uses his own camera model introduced in Brown (1966). In the solution process the plumb lines are represented in a Hough notation (see Section 5.6.1).
3. Stein (1997): This method uses only point correspondences without the knowledge of the three-dimensional point location of the observed objects. As a camera model a pinhole projection with a radial distortion to the second order and tangential distortion is used. The algorithm is an iterative one, beginning with

finding three point correspondences. After an initial guess of the distortion parameters a trilinear tensor constraint is applied to calculate the back-projection and the error depending on the actual parameter set. Minimizing this error leads to the calibration result. However, it is to note that this is a non-metric calibration for the internal parameters. Without the knowledge of at least some three-dimensional point locations a metric calibration suitable for measurement tasks cannot be obtained.

4. Swaminathan & Nayar (2000): This non-metric calibration assumes a camera model that takes radially symmetric and decentering distortions into account. In contrast to Stein (1997) the calibration is based on straight lines that are distorted by the imaging process. In some sense this idea has been picked up from the plumb line approach of Brown (1971). The solution is then obtained by formulating an objective function that is minimized with a simplex search algorithm. The disadvantages formulated for Stein (1997) do apply here as well.

## Two-Step Methods

These methods combine linear and non-linear steps to obtain better results. Usually, the result of the linear stage serves as an initial guess for the non-linear optimization. Publications, such as Bouguet (2000), Zhang (1999) and Heikkilä & Silvén (1997) are omitted here. They are discussed in depth in Section 5.5.2.

1. Tsai (1986): This paper introduces a model which is widely used in computer vision in general. The camera model is a pinhole mapping extended by radial distortion of the first order. Tangential distortions are considered negligible in industrial camera applications. The first step consists of the linear estimation of the external parameters and some of the internal ones. The second step is a non-linear optimization of the distortion coefficients. However, the non-linear stage does not optimize all parameters. This and the rather simple camera model result in a mediocre calibration.
2. Bacakoglu & Kamel (1997): This calibration uses a camera model based on the pinhole projection with one radial and two tangential terms describing the lens distortions. The first part is divided in two stages: the first consists of a least-squares method for the parameter approximation and the second one is an optimization of the determination of the rotational transformation component. The rotation is alternatively computed using quaternions or the equivalent-angle-axis method. The second part is a non-linear optimization of the distortion angles.

3. Chatterjee et al. (1997): The camera model of Chatterjee contains the usual extrinsic parameters and the focal length, the image center displacement, the skew angle, and radial lens, decentering, and tangential distortions. The model is derived from Brown (1966). The algorithm initializes with a linear estimation of the parameter vector for a fast and accurate convergence of the main estimation method. The final optimization imposes all orthonormality constraints on all the parameters to guarantee accurate convergence.

### Calibration Targets

As a conclusion to the related work some remarks on calibration targets are made. Apart from certain non-metric calibrations, such as Stein (1997) or Swaminathan & Nayar (2000), the calibrations depend on an image, or an image sequence, taken from a calibration target built to predefined specifications. The control points on the target are chosen with special regard to their detectability within the image. An additional objective is the easy correspondence analysis between the three-dimensional object points and the two-dimensional image points. The camera model and the algorithms applied as well as the calibration target have an influence on the quality of the calibration.

Usually control points of some sort are arranged in a two- or three-dimensional setup. Taking the fact into account that the focal length and the distance of the calibration target are correlated, a planar equidistant calibration target cannot be used, i.e. a flat-on image of a planar control point arrangement. If a planar arrangement is used it has to be viewed at a certain inclination angle (see Section 5.5.1).

Three-dimensional objects do not carry this disadvantage. However, the precise construction is more challenging. In addition to that, the points must be unique in some way such that the algorithm is able to identify them unequivocally. One example is to use circular shapes. These mark the control point itself. They are surrounded by circular segments coding the control point. Methods such as this control-point encoding are used for planar calibration rigs. Another method for a planar rig is shown in the left image of Fig. 5.2 in Section 5.5.1. The three crosses identify certain points among the control points and therefore allow a determination of the rig orientation.

Approaches operating on image series of planar calibration rigs use the sequentially obtained information to build a highly contoured three-dimensional object. In the case of Bouguet (2000) this is done with a calibration rig consisting of identical points. However, the number of points and their distribution within the image provides the identification.

A final remark must be made on the nature of the control points. Methods, such as Lenz (1987) and others, use circular shapes to mark the control point. Within the image, these elliptical shapes are easy to detect. Usually, the center of mass of the

image contour is taken as the sub-pixel precise image location of the control point. However, Heikkilä & Silvén (1997) showed that this is wrong and a source of error for the extracted data. Due to the lens distortions, the circular or elliptical centers are not projected to the center of the image contour. Therefore, only real points account for a precise detectability within the image. One good solution is a chessboard type of calibration rig with the calibration targets represented by the intersection points (see Fig. 5.10).

### 5.3 Mathematical Methods

The following paragraphs introduce notation and necessary methods for a convenient description of the camera models and the calibrations algorithms.

#### 5.3.1 Affine Transformation

A form-invariant transformation of one Cartesian coordinate system, denoted by  $\mathbf{x}_0 = \{x_0, y_0, z_0\}$  into another one,  $\mathbf{x}_1 = \{x_1, y_1, z_1\}$ , is e.g. given by (Luhmann, 2000):

$$\mathbf{x}_1 = m\mathbf{R}\mathbf{x}_0 + \mathbf{t} \quad (5.1)$$

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = m \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \\ z_0 \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}. \quad (5.2)$$

$\mathbf{R}$  represents an orthonormalized rotation matrix and  $\mathbf{t}$  a translation vector.  $m$  is a scaling factor which is set to 1 in order to obtain a length-invariant transformation.

The orthonormality of the rotation matrix  $\mathbf{R}$  implies that its row and column vectors are linearly independent. Therefore:

$$\mathbf{R}^{-1} = \mathbf{R}^T \quad \text{and} \quad \mathbf{R}\mathbf{R}^T = \mathbb{I}. \quad (5.3)$$

The rotation depends on the pitch angle  $\mathbf{R}_\alpha$  (around the  $x$ -axis), the yaw angle  $\mathbf{R}_\beta$  (around the  $y$ -axis), and the roll angle  $\mathbf{R}_\gamma$  (around the  $z$ -axis); but not only on themselves but also on their order. Therefore, the order is of concern for the transformation as well as for the inverse transformation. A common definition for the order is:

$$\mathbf{R} = \mathbf{R}_\alpha \mathbf{R}_\beta \mathbf{R}_\gamma. \quad (5.4)$$

A sequence-free representation of a rotation is obtained using quaternions. This will be introduced in the following section.

### 5.3.2 Quaternions

In the 1840es Sir William Rowan Hamilton, Astronomer Royal of Ireland, and Olinde Rodrigues, directory of the Caisse Hypothécaire in Paris, France, introduced the calculus of quaternions (Altmann, 1989). The perception of their usefulness experienced first an upraise and then a decline. Meanwhile however, they not only successfully entered physics in quantum field theory but their properties in representing rotations charm elsewhere too.

The quaternion notation surmounts the problems of the order of rotational angles. It introduces an axis of rotation and a direction-bound angle of rotation around it. This is sufficient for all rotations in three-dimensional space. The direction of the axis is a unit vector  $\mathbf{a}$  and the rotational direction of the angle  $\theta$  is given by the right hand rule.

In a general form a quaternion  $\hat{\mathbf{q}}$  is given by a real part  $q$  and vector  $\mathbf{q}$ . The algebra of the vector  $\mathbf{q}$  is similar to the rules that govern the imaginary part of a complex number:

$$\hat{\mathbf{q}} = q + \mathbf{q} = q + q_x i + q_y j + q_z k. \quad (5.5)$$

For a given rotational axis  $\mathbf{a}$  and angle  $\theta$  the quaternion is constructed as:

$$\hat{\mathbf{q}} = \cos\left(\frac{\theta}{2}\right) + \sin\left(\frac{\theta}{2}\right)(a_x \mathbf{e}_x + a_y \mathbf{e}_y + a_z \mathbf{e}_z). \quad (5.6)$$

According to Horn (1987) the rotation  $\mathbf{R}$  related to a quaternion  $\hat{\mathbf{q}}$  is:

$$\mathbf{R} = \begin{pmatrix} q^2 + q_x^2 - q_y^2 - q_z^2 & 2(q_x q_y - q q_z) & 2(q_x q_z + q q_y) \\ 2(q_x q_y + q q_z) & q^2 - q_x^2 + q_y^2 - q_z^2 & 2(q_y q_z - q q_x) \\ 2(q_x q_z - q q_y) & 2(q_y q_z + q q_x) & q^2 - q_x^2 - q_y^2 + q_z^2 \end{pmatrix}. \quad (5.7)$$

Apart from the advantage of a simpler and unambiguous representation, quaternions reduce the calculation load for a sequence of rotations. Two subsequent rotations in quaternion notation require 16 multiplications while a matrix rotation needs 27. This should not be disregarded with respect to computational load.

### 5.3.3 Homogenous Coordinate Representation

A more convenient formulation of transformations is achieved by the homogenous coordinate representation. It is widely used in computer graphics applications and it is advantageous in camera calibration as well. The homogenous coordinate system is introduced by the following notion:

$$\mathbf{x}' = \begin{pmatrix} x' \\ y' \\ z' \\ s \end{pmatrix} \quad (5.8)$$

$$\text{with } x = \frac{x'}{s}, \quad y = \frac{y'}{s} \quad \text{and} \quad z = \frac{z'}{s}. \quad (5.9)$$

For Cartesian coordinates the scaling factor is  $s = 1$ . The mapping matrix for a such a vector is obviously a  $4 \times 4$  matrix. In it not only the rotation and translation are unified but also scaling, reflection, homogenous scaling, and perspective are incorporated. However, for camera calibration tasks only the rotation  $\mathbf{H}_R$ , the translation  $\mathbf{H}_T$ , and the homogenous scaling  $\mathbf{H}_S$  are of interest. Combining these, the following general homogenous representation is obtained for the affine transformation (see Eq. 5.1):

$$\mathbf{x}_1 = \mathbf{H}_T \mathbf{H}_S \mathbf{H}_R \mathbf{x}_0 \quad (5.10)$$

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} m & 0 & 0 & 0 \\ 0 & m & 0 & 0 \\ 0 & 0 & m & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \\ z_0 \\ 1 \end{pmatrix} \quad (5.11)$$

$$= \begin{pmatrix} mr_{11} & mr_{12} & mr_{13} & t_x \\ mr_{21} & mr_{22} & mr_{23} & t_y \\ mr_{31} & mr_{32} & mr_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \\ z_0 \\ 1 \end{pmatrix}. \quad (5.12)$$

## 5.4 Basic Camera Models

To model the light ray propagation through a camera system a mathematical representation of the camera with its components is required. Depending on the desired accuracy the modeling must be more or less detailed. This complexity distinguishes the sophistication of the calibration method itself. In this section a short introduction to the two most basic and equivalent camera models is given.

The pinhole and the focusing screen camera are the basic models for all the extended descriptions. The more sophisticated models recreate a pinhole image from the distorted one. The pinhole model assumes that all incident rays intersect at a projection center. This center is an infinitesimal aperture stop (see Fig. 5.1). Objects are imaged upside down on the projection plane which is located at the distance of the focal length  $f$  from the projection center. Geometric evaluation yields the following relationship between a point  $\mathbf{w}(x, y, z)$  in world coordinates and the point  $\mathbf{m}(u_p, v_p)$  in image coordinates:

$$u_p = -f \frac{x}{z} \quad \text{and} \quad v_p = -f \frac{y}{z}. \quad (5.13)$$



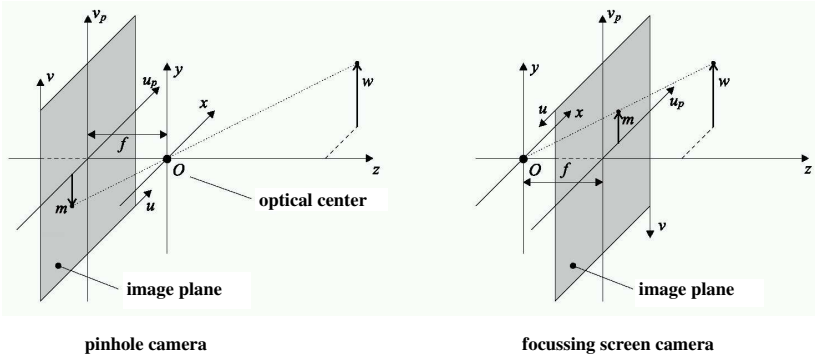


Figure 5.1: The Pinhole and the Focusing Screen Camera.

In principle, the focusing screen camera is equivalent to the pinhole camera. However, due to the arrangement of the image plane in between the focal point and the object location, a change in sign occurs:

$$u_p = f \frac{x}{z} \quad \text{and} \quad v_p = f \frac{y}{z}. \quad (5.14)$$

Calculating these locations in sensor coordinates implies not only an affine transformation but also a quantization. The quantization is due to the pixel nature, or more physically speaking, the array arrangement of finite photon receptors that integrate the received intensity. The pixel size is given by  $k_u$  and  $k_v$  in  $u$ - and  $v$ -direction of the image, respectively. The focal lengths  $\alpha$  along the image-coordinate directions are defined as:

$$\alpha_u = \frac{f}{k_u} \quad (5.15)$$

$$\alpha_v = \frac{f}{k_v}. \quad (5.16)$$

A pinhole camera with principle point coordinates  $(c_u, c_v)$ , given by the intersection of the optical axis with the image plane, is described by:

$$u = c_u - \frac{1}{k_u} u_p = c_u - \alpha_u \frac{x}{z} \quad (5.17)$$

$$v = c_v - \frac{1}{k_v} v_p = c_v - \alpha_v \frac{y}{z}. \quad (5.18)$$

A similar expression is obtained for the focusing screen camera:

$$u = c_u + \frac{1}{k_u} u_p = c_u - \alpha_u \frac{x}{z} \quad (5.19)$$

$$v = c_v + \frac{1}{k_v} v_p = c_v - \alpha_v \frac{y}{z}. \quad (5.20)$$

At this point the equivalence of the two models is obvious: they share the projection equations.

A formulation of the pinhole projection Eq. 5.17 in homogenous coordinates is desirable. The matrix of internal parameters  $\mathbf{T}_i$  transforms a point from the world coordinate system, with its origin at the projection center, to its image coordinates. For the equivalent models the same relations hold, with  $s$  being a scaling factor that takes into account that a three-dimensional world is mapped to a two-dimensional plane:

$$\begin{pmatrix} us \\ vs \\ s \\ 1 \end{pmatrix} = \mathbf{T}_i \mathbf{w} = \begin{pmatrix} \alpha_u & 0 & c_u & 0 \\ 0 & \alpha_v & c_v & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}. \quad (5.21)$$

In order to obtain this equation the assumption of a world coordinate system located at the projection center is made. If more than one camera is used in a system a joint coordinate frame has to be established. The transformation of this world coordinate frame into the respective camera coordinate system is then given by  $\mathbf{T}_e$ , the matrix of the external camera parameters. In homogenous coordinates this matrix consists of a rotation  $\mathbf{R}$  and a translation  $\mathbf{t}$ :

$$\mathbf{T}_e = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (5.22)$$

Therefore, the complete pinhole – focusing screen mapping is

$$\begin{pmatrix} us \\ vs \\ s \\ 1 \end{pmatrix} = \mathbf{T}_i \mathbf{T}_e \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}. \quad (5.23)$$

This model is, as mentioned above, certainly the most basic one. It is applicable for systems that do not require a high precision. In reality, the infinitesimal aperture stop cannot be provided and in order to increase the intensity of light incident on the retina, lenses have to be incorporated into the system. However, lens systems distort images. In the following section models are discussed that describe such distortions.

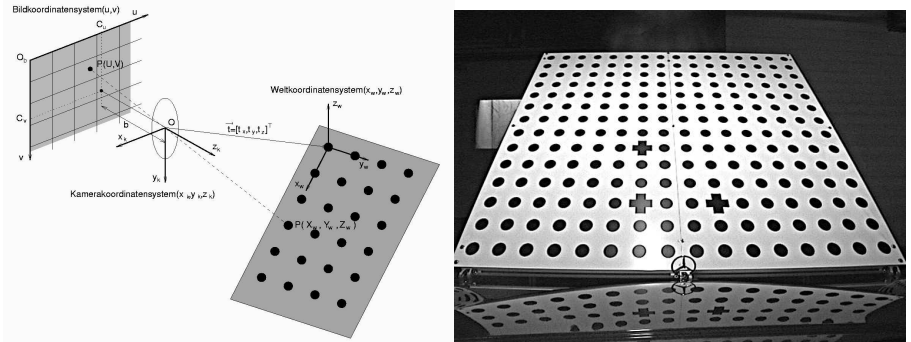


Figure 5.2: The Tsai/Lenz-Calibration: Coordinate system relationship (left), image of the self-built calibration wall (right).

## 5.5 Calibration Algorithms

In this section two calibration methods are discussed in more detail. For each method the respective camera model is introduced and the algorithmic parameter determination is presented.

### 5.5.1 Tsai/Lenz Calibration

This calibration algorithm was originally presented by Tsai (1986). It is a two-step method, consisting of a linear first and a non-linear second stage. With the knowledge of the principle point a determination of other intrinsic and extrinsic parameters is achievable. However, Tsai needs two or more planes with control points. Lenz (1987) simplified this approach and introduced an algorithm that estimates the parameters on a single image of a control point arrangement on a single planar object. Lenz & Tsai (1987) develop a method to determine the scaling factor and the principle point of a machine-vision system. However, it has not been incorporated into this approach.

The Lenz version of this method was implemented during the first single image calibration efforts in our research group (Hengst, 2001). In Fig. 5.2 a graphic of the coordinate system relationship and a real image of the self-built calibration plane is shown.

### Camera Model

The Tsai/Lenz camera model is in principle the pinhole model described in Section 5.4. A point in world coordinates  $\mathbf{w}$  is transformed by the extrinsic camera parameter matrix  $\mathbf{T}_e$  (see Eq. 5.22) to camera coordinates. Subsequently, the projection

from Eq. 5.13 onto the image plane is performed. At this stage a radial distortion of the first order ( $\kappa_1$ ) is introduced. The distorted coordinates ( $u_d, v_d$ ) on the image plane are given by:

$$u_d = \frac{2u_p}{1 + \sqrt{1 - 4\kappa_1 R_p^2}} \quad (5.24)$$

$$v_d = \frac{2v_p}{1 + \sqrt{1 - 4\kappa_1 R_p^2}} \quad (5.25)$$

with  $R_p^2 = u_p^2 + v_p^2$ . According to Eq. 5.17 these are transformed to sensor coordinates:

$$u = c_u + \frac{1}{k_u} u_d \quad (5.26)$$

$$v = c_v + \frac{1}{k_v} v_d \quad (5.27)$$

or in two-dimensional homogenous coordinates:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \mathbf{T}_i \mathbf{m}_d = \begin{pmatrix} \frac{1}{k_u} & 0 & c_u \\ 0 & \frac{1}{k_v} & c_v \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_d \\ v_d \\ 1 \end{pmatrix}. \quad (5.28)$$

The Tsai/Lenz model incorporates six external parameters (three for the rotation, three for the translation) and six internal parameters (focal length, radial distortion of the first order, two for the principle point and two for the pixel size). However, the internal calibration only determines the focal length and the radial distortion. Therefore, there are eight parameters to optimize.

### Algorithm

The algorithms in these approaches are two-step estimations. In the original Tsai (1986) algorithm the first step is a linear estimation of the external parameters (rotation and two translational components) and the effective focal length of the camera. To do so, a radial alignment constraint is assumed. Subsequently, an iterative scheme is applied to derive a closed-form solution for the estimation of three more parameters: the depth component of the translation vector, the effective focal length, and a radial distortion coefficient. In Lenz & Tsai (1987) two more parameters are added to the iterative step: the coordinates of the principle point which are assumed to be known in Tsai (1986). Lenz (1987) altered the algorithm, such that the second-step estimation of the radial distortion coefficient is performed by a non-iterative method.

## Discussion

The main reason to implement the Lenz (1987) algorithm is that it requires only a single image of the calibration target. This is considered useful for the industrial application. A planar arrangement of control points formed by elliptical shapes, as proposed by Lenz (1987), was built to serve as calibration target (see the right side of Fig. 5.2). The elliptical shape is chosen because it is imaged close to a circular shape if the plane is tilted relative to the image plane. Subsequently, the locations of the control points within the image are detected using a center-of-mass approach. Other advantages of the method are taken into account: The process does not apply any non-linear optimizations. Therefore, it is very stable and fast. The simplicity and the single image do not require any prior user knowledge.

On the other hand there are disadvantages to the Tsai/Lenz calibration as well: in the Lenz (1987) version the principle point coordinates and other intrinsic parameters, such as the affinity<sup>1</sup> are not estimated. The application of the algorithm to a single image increases the influence of noise. This is compensated to a certain degree by the introduction of additional control points. In addition to that, the control points have to be exactly known in the world coordinate system and their distribution must be planar. In order for the algorithm to work properly at least five control point must be used. Due to the correlation of the focal length and the distance of the control point plane from the image plane the control point plane must not be parallel to the image plane. To increase the accuracy the control point plane should be imaged as large as possible to scan the image to the outmost points. A final disadvantage that needs to be mentioned is that the second step only computes a few parameters that cannot be provided by the first step. There is no optimization of all parameters in the second step.

To enhance the method ourselves the estimation of the principle point was tried. In several attempts  $i$  different points are chosen as the principle point and the calibration is performed. For each calibration the standard deviation  $\sigma_i$  of the back-projection error is computed. To the  $\sigma_i$  a parabola is approximated using a least-squares method. However, it is not possible to robustly compute a minimal value of the parabola and therefore the true location of the principle point.

The conclusion is, that the Tsai/Lenz calibration provides a simple method which is easy to use. However, the results are only sufficient for lower accuracy applications. For the intended application of this thesis, i.e. the determination of the surface geometry and reconstruction of the information contained in the reflection, a more sophisticated approach is necessary.

---

<sup>1</sup>Affinity: The ratio of the pixel length in  $u$ - and  $v$ -direction

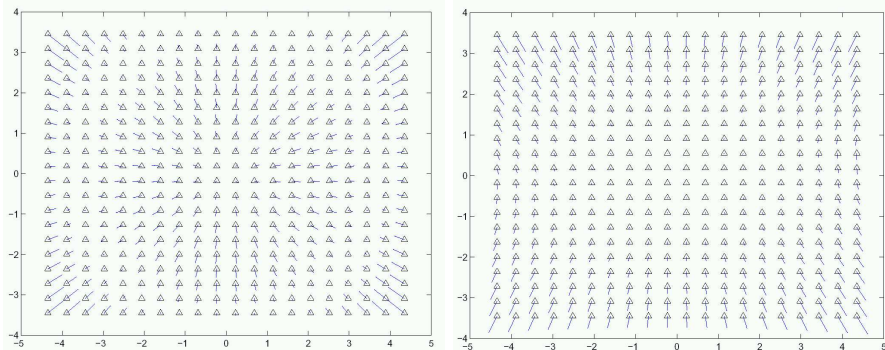


Figure 5.3: Radial-asymmetric and Tangential Distortions: A qualitative depiction of the radial-asymmetric (left) and the tangential distortion (right). Graphics from Kaempchen (2001).

### 5.5.2 Bouguet/Heikkilä Calibration

Bouguet (2000) develops a camera calibration toolbox for MATLAB. It is mainly based on publications by Zhang (1999) and Heikkilä & Silvén (1997). The camera model is more sophisticated and contains the most parameters of the examined approaches. The method is very suitable for stereo-camera systems. As the Bouguet implementation is freely available on the internet it was tested for the suitability of our needs. It operates on images of multiple views of a planar calibration rig which is deemed rather unsuitable for industrial applications. However, due to the superiority of the results the method is considered favorable. Kaempchen (2001) conducted the first tests and comparisons with other calibration methods. F.J. Stein implemented a stereo-calibration routine based on the Bouguet code, consisting of a C++ mono camera calibration for each of the cameras in the stereo system and a following stereo optimization in MATLAB. After that, enhancements of the feature detection stage are introduced (see Section 5.6). In order to increase the usability of the C++/MATLAB implementation Schowalter (2002) recoded the MATLAB part in C++. The changes in the feature point evaluation that resulted from Moosbrugger (2002) are incorporated as well. A. Moosbrugger developed code for a robust recognition of partially visible calibrations rigs. A self-built calibration rig is shown in Fig. 5.10.

### Camera Model

The camera model of Bouguet is inspired by Brown (1966), Fryer & Brown (1986), and Heikkilä & Silvén (1997). It not only considers radial distortions to the third

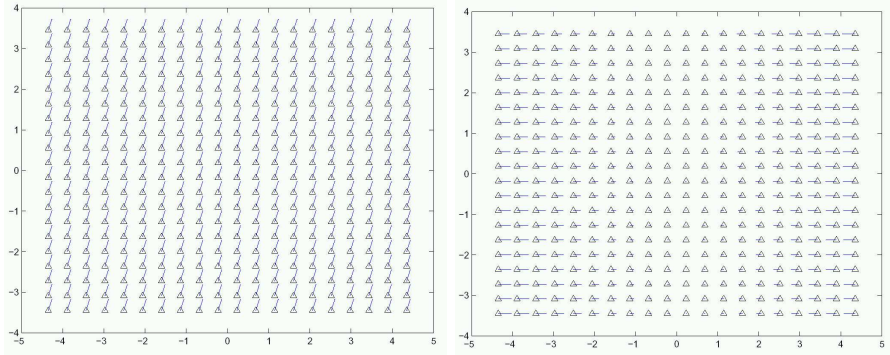


Figure 5.4: Principle Point Location and Affinity Distortion: A qualitative depiction of the movement of the principle point (left) and the influence of the affinity and the skew (right). Graphics from Kaempchen (2001).

order but also introduces tangentially-asymmetric components, affinity and skew (see Fig. 5.3 and 5.4). The tangential distortion is due to defects within the lens systems or oblique mounting of a lens. Affinity and skew are related to inconsistencies and skewed mounting of the retina in the system (Brown, 1966).

The coordinates of the camera system  $(x_c, y_c, z_c)$  are projected onto the image plane and normalized:

$$\begin{pmatrix} u_n \\ v_n \end{pmatrix} = \begin{pmatrix} \frac{x_c}{z_c} \\ \frac{y_c}{z_c} \end{pmatrix}. \quad (5.29)$$

These pinhole mapping coordinates  $(u_n, v_n)$  are radially distorted to the third order to obtain:

$$u_{\text{rad}} = u_n \left( 1 + \kappa_1 R_p^2 + \kappa_2 R_p^4 + \kappa_3 R_p^6 \right) \quad (5.30)$$

$$v_{\text{rad}} = v_n \left( 1 + \kappa_1 R_p^2 + \kappa_2 R_p^4 + \kappa_3 R_p^6 \right) \quad (5.31)$$

with

$$R_p^2 = u_n^2 + v_n^2. \quad (5.32)$$

To that a tangential distortion component is added:

$$\Delta u_{\text{tan}} = 2\kappa_4 u_n v_n + \kappa_5 (R_p^2 + 2u_n^2) \quad (5.33)$$

$$\Delta v_{\text{tan}} = 2\kappa_4 u_n v_n + \kappa_5 (R_p^2 + 2v_n^2). \quad (5.34)$$

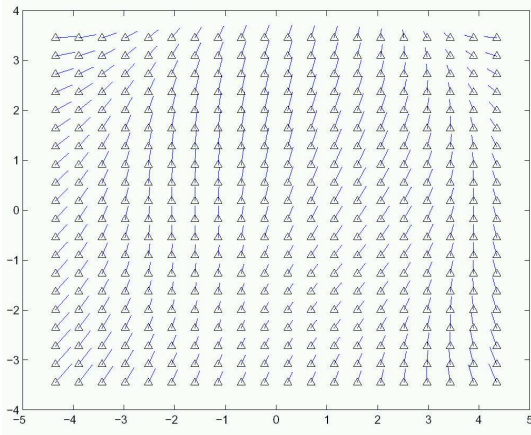


Figure 5.5: The Complete Distortion Model: A qualitative depiction of the complete Bouguet/Zhang distortion model. Graphic from Kaempchen (2001).

The sum of the radial and the tangential distortion components yields the distortion model:

$$u_d = u_{\text{rad}} + u_{\text{tan}} \quad (5.35)$$

$$v_d = v_{\text{rad}} + v_{\text{tan}}. \quad (5.36)$$

These coordinates are still expressed in the principle point coordinate system. Image coordinates are obtained by adding the principle point coordinates  $(c_u, c_v)$ . However, for decentered sensors they must not be equivalent to half the image size respectively (see Fig. 5.4). At this stage the skew  $\alpha_c$  parameter is introduced. This parameter accounts for non-rectangular pixels on the retina. A homogenous two-dimensional representation of the internal parameter matrix and mapping is:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \mathbf{T}_i \mathbf{m}_d = \begin{pmatrix} \alpha_u & \alpha_c & c_u \\ 0 & \alpha_v & c_v \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u_d \\ v_d \\ 1 \end{pmatrix}. \quad (5.37)$$

All these parameters are optimized in the calibration algorithm. There are 16 parameters altogether: six external and ten internal (focal length, two for the pixel size, three for radial distortion, two for the radial-asymmetric and the tangential distortion, one for the skew, and two for the principle point coordinates).



### Algorithm

The Bouguet (2000) method is based on the work by Zhang (1999) and Heikkilä & Silvén (1997). The non-linear transformation from world- to image-coordinates does not allow a direct inversion of the transformation. Additionally, there is no closed-formed solution of it in this calibration. Therefore, there must be an iterative solution to the minimization problem of the non-linear error-function. However, iterative solutions always carry two disadvantages: they are slow and the obtained solution is not guaranteed to be the global minimum. The iteration might come up with a local minimum solution. To minimize this risk Zhang (1999) proposed an optimization in several steps. Speed is not critical: camera calibration is an off-line algorithm and is usually only performed once for a given setup of a stereo-camera system.

Now Zhang's approach and subsequently the implementation of it in Bouguet's calibration is introduced. Zhang (1999) assumed that the two-dimensional control point array is located in the  $x$ - $y$ -plane. Therefore, the homogenous transformation of a control point  $\mathbf{w}' = (X, Y, 1)^T$  to point on the retina  $\mathbf{m}' = (u, v, s)^T$  is described by the matrix  $\mathbf{H}$ , consisting of  $\mathbf{T}_i$  (see Eq. 5.21) and the first two rows of  $\mathbf{T}_e$  (see Eq. 5.22):

$$\mathbf{m}' = \mathbf{H}\mathbf{w}' \quad (5.38)$$

$$\mathbf{H} = \mathbf{T}_i \begin{pmatrix} r_1 & r_2 & t \end{pmatrix}. \quad (5.39)$$

The image coordinates  $\mathbf{m} = (u, v)^T$  of control point  $\mathbf{w}$  is:

$$\mathbf{m} = \frac{1}{m'_3} (m'_1, m'_2)^T = \frac{1}{\bar{h}_3^T \mathbf{w}} \begin{pmatrix} \bar{h}_1^T \mathbf{w} \\ \bar{h}_2^T \mathbf{w} \end{pmatrix} \quad (5.40)$$

with  $\bar{h}_i^T$  being the  $i$ -th row of  $\mathbf{H}$ .

To calculate the transformation, it is assumed that the detected image point location  $\mathbf{m}_i$  is distorted by Gaussian noise with a covariance matrix  $\Lambda_{m_i}$ . However, in the practical application  $\Lambda_{m_i} = \sigma^2 \mathbb{I}$ .  $\mathbf{H}$  is then recovered from the image point locations  $\mathbf{m}_i$  and the model control point locations  $\hat{\mathbf{m}}_i$  by a Maximum-Likelihood estimation:

$$\sum_i (\mathbf{m}_i - \hat{\mathbf{m}}_i)^T \Lambda_{m_i} (\mathbf{m}_i - \hat{\mathbf{m}}_i) = \sigma^2 \sum_i \|\mathbf{m}_i - \hat{\mathbf{m}}_i\|^2 \rightarrow \min. \quad (5.41)$$

This non-linear minimization is iteratively performed by the Levenberg-Marquardt algorithm (Moré, 1977). The necessary initial guess is derived by Zhang (1999).

The internal parameters  $\alpha_u, \alpha_v, \alpha_c, u_0$  and  $v_0$  of the homogenous transformation  $\mathbf{H} = (\mathbf{h}_1 \mathbf{h}_2 \mathbf{h}_3)$  are formulated with the orthogonality condition of  $\mathbf{r}_1$  and  $\mathbf{r}_2$ :

$$\mathbf{h}_1^T (\mathbf{T}_i^{-1})^T \mathbf{T}_i^{-1} \mathbf{h}_2 = 0 \quad (5.42)$$

$$\mathbf{h}_1^T (\mathbf{T}_i^{-1})^T \mathbf{T}_i^{-1} \mathbf{h}_1 = \mathbf{h}_2^T (\mathbf{T}_i^{-1})^T \mathbf{T}_i^{-1} \mathbf{h}_2. \quad (5.43)$$

With the introduction of a matrix

$$\mathbf{B} = (\mathbf{T}_i^{-1})^T \mathbf{T}_i^{-1} \quad (5.44)$$

and a vector

$$\mathbf{b} = (B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, B_{33})^T \quad (5.45)$$

an equation system of the form  $\mathbf{V}\mathbf{b} = 0$  is formulated. At this point the transformation estimations  $\mathbf{H}_i$  of  $n$  images are taken into account. The solution is either the eigenvector of the smallest eigenvalue of  $\mathbf{V}^T \mathbf{V}$  or the rightmost singular vector to the smallest singular value of  $\mathbf{V}$ . From the solution vector  $\mathbf{b}$  the intrinsic parameter matrix  $\mathbf{T}_i$  and subsequently the rotation and the translation are obtained (Zhang, 1999).

Due to fact that here only matrix entries rather than physical parameters are optimized an enhancement of the intrinsic parameter matrix  $\mathbf{T}_i$  is necessary. Zhang (1999) proposes a Maximum-Likelihood estimation similar to Eq. 5.41. This time however, the parameters  $\mathbf{T}_i$ ,  $\mathbf{R}$ , and  $\mathbf{t}_i$  are optimized over  $n$  images:

$$\sum_{i=1}^n \sum_{j=1}^m \| \mathbf{m}_i - \hat{\mathbf{m}}_i(\mathbf{T}_i, \mathbf{R}_i, \mathbf{t}_i, \mathbf{M}_{ij}) \|^2 \rightarrow \min \quad (5.46)$$

with  $\hat{\mathbf{m}}_i(\mathbf{T}_i, \mathbf{R}_i, \mathbf{t}_i, \mathbf{M}_{ij})$  being the projection of the image point  $\mathbf{M}_j$  of image  $i$ . Similarly, the solution is obtained by the Levenberg-Marquardt algorithm.  $\mathbf{R}$  is represented by a three-dimensional vector  $\mathbf{r}$ , given by the Rodriguez formula (Faugeras, 1993), a quaternion approach (see Section 5.3.2).

The additional lens distortion equations are depending on the measured image point location  $(u, v)$  of the distortion-free coordinates  $(u_n, v_n)$  and the vector  $\boldsymbol{\kappa}$  containing the distortion parameters. In Zhang (1999) an equation system of the form  $\mathbf{D}\boldsymbol{\kappa} = \mathbf{d}$  is set up and solved by a least-squares estimation. Finally, all parameters are again processed by Levenberg-Marquardt:

$$\sum_{i=1}^n \sum_{j=1}^m \| \mathbf{m}_i - \hat{\mathbf{m}}_i(\mathbf{T}_i, \boldsymbol{\kappa}, \mathbf{R}_i, \mathbf{t}_i, \mathbf{M}_{ij}) \|^2 \rightarrow \min. \quad (5.47)$$

In principle this Zhang (1999) approach has been adopted by Bouguet (2000). However, some minor changes have been made. The algorithm of Bouguet (2000) is as follows:

1. Maximum-Likelihood estimation of the transformation  $\mathbf{H}$  with the Levenberg-Marquardt algorithm.
2. Closed-formed solution of the intrinsic parameters  $\mathbf{T}_i$ ,  $\mathbf{R}$ , and  $\mathbf{t}$ . However, to obtain the homography the orthogonality of vanishing points is used explicitly, as published by Caprile & Torre (1990).

3. The enhancement of the intrinsic parameters is performed by a gradient-based method.
4. The initial guess of the distortion parameter vector  $\kappa$  is skipped as it has been considered unnecessary in practical applications.
5. Finally, the iterative optimization of all parameters is performed with the Levenberg-Marquardt minimization. However, instead of Zhang's approach the distortion model of Heikkilä & Silvén (1997) is used, because it contains the tangential distortion coefficients (see above).

### Discussion

Bouguet (2000) provided in his Camera Calibration Toolbox for MATLAB an approach to the calibration task which is far superior to many other methods and readily available for usage. The multiple step approach, although consisting of non-linear parts, combined with a very comprehensive camera model is very suitable for all kinds of applications, especially stereo-vision systems. The quality of this calibration is tested and compared to others in the diploma thesis of N. Kaempchen (2001). To compare the calibrations he considered the following values amongst others:

- The residual error: the distance between a measured image location  $\mathbf{m}_i$  and the back projected location of the model control point based on the calibration parameters.
- The baseline length of the stereo-vision system: a measurement of the recovered external camera parameters.
- The absolute error in distance: this measure is very important for the quality of the stereo-camera calibration and the subsequent distance estimation with it.
- Inspection: the visualization of the three-dimensional points shows errors and inconsistencies of the stereo algorithm and the calibration which are not obtainable by the above values.

To summarize, the Bouguet (2000) calibration provides results with only 20% of the error burden compared to the Tsai/Lenz algorithm. The Bouguet solution is superior to the discussed approaches. For more information consult Kaempchen (2001).

To support this analysis a few remarks are added to some of the details:

- The multiple stage method with an initial guess assures that the iterative solution approach starts close to an optimum. On the one hand, there are only a few iterations necessary to obtain the optimal solution. On the other hand,

the danger of ending up in a local extremum far from the global one or even performing a divergent algorithm are minimal.

- Working on an image series with different views of a calibration rig contains the risk of obtaining views which lead to singularities and numerical instabilities in the calculation of the homographies. Sturm & Maybank (1999) describes a method to obtain an optimal set of calibration parameters on a series of images. The calibration target is a planar control point arrangement, imaged from different views. Their publication lists singularities which might occur depending on the position of the image plane and the calibration rig relative to each other.
- The Bouguet (2000) algorithm is especially suitable for stereo-vision system. In the MATLAB implementation the constraints of a rigidly mounted stereo-camera system are taken into account. Regarding the external parameters of the cameras, a change in the baseline is not allowed to occur from one image to the next. This consideration reduces the parameter set by six and makes the system more stable and reliable.

To conclude the discussion, a general point to the camera calibration has to be stated: each of the approaches is only as good as the determination of the original data, i.e. the location of the control point features within the image. Bouguet uses a planar checkerboard calibration rig which is in itself far superior to a dotted calibration plane (see Section 5.2). However, the determination of the features is not optimal. At this point further enhancements to the Bouguet calibration are made.

## 5.6 *Enhancement of Bouguet's Calibration Approach*

Each system providing results which are derived from measurement data are only as good as the original data itself. Evaluating the results of the Bouguet calibration, especially comparing the procedure to others proposed, it is obvious that significant progress potential lies in the improvement of basic data acquisition step.

As mentioned above, the original implementation of the stereo-camera system is divided in two parts. The first part is a C++ implemented procedure to calibrate each camera as a monocular system. The code is based on the freely available OpenCV<sup>2</sup> library. The result of this calibration is the input for the MATLAB implementation by Bouguet (2000).

It is obvious that such a two system implementation is in no way desirable. Before discussing some implementation steps, an enhancement of the original OpenCV detection of the chessboard corners is introduced.

---

<sup>2</sup>Open Source Computer Vision Library: A collection of computer-vision algorithms, implemented in C++, published by the Intel Cooperation, <http://developer.intel.com>

### 5.6.1 Feature Detection

#### The OpenCV Corner Detection

The OpenCV corner detection is a multiple step procedure based on an initial scanning process with a subsequent sub-pixel precision search of the checkerboard crossing:

1. The initial scanning process: The function tries to determine whether a given feature is a chessboard corner. It locates and sorts the chessboard corner guesses. To do so, it initializes a contour scanning process. The scanning process operates on binary images and extracts the contours. A polygon approximation is then applied to the contours with rejection of non-quadrangle structures. In a loop the structures are collected and sorted until either all desired corners are found and sorted or no more suitable contours are available. A positive result is piped into the sub-pixel precise search.
2. Sub-pixel precision search: The idea of this function is based on the observation that every vector from a center  $q$  to a point  $p$  located within the neighborhood of  $q$  is orthogonal to the image gradient at  $p$ , subject to image and measurement noise. Therefore, it is a gradient minimum search with a relocation of a neighborhood window until the center keeps within a set threshold. That is the sub-pixel precise location of the feature.

In practice the contour analysis proved to be rather instable. Following the maxim that the space of measurement should be equal to the space of calibration it is necessary, especially for automotive applications, to record the calibration images at larger distances. This leads to rather small features with close distances within the image. The OpenCV algorithm is often not able to properly detect and sort them. In order to avoid singularities due to parallel image and calibration planes (see Section 5.5.1 and 5.5.2) the calibration rig must be imaged at a certain angle. The control point analysis of the OpenCV implementation often fails if the angle is too large, mainly due to an improper polygon approximation. Finally, the gradient based determination of the sub-pixel precise location of the checkerboard crossing is not satisfactory. Even operating on a rather large image basis some of the parameters are only determined with a rather large error burden, e.g. the focal length up to an uncertainty of 2 pixels.

To improve the quality of the data basis a template matching approach is developed.

#### Template Matching Corner Detection

The developed template matching procedure is based on an algorithm usually applied in stereo-vision matching. According to Aschwandten (1993) the most suitable for image processing is the normalized mean-free cross-correlation function ( $ccfmf$ ) defined

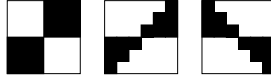


Figure 5.6: Three calibration matching templates used in the first approach to locate the chessboard corners with sub-pixel precision. The depiction is enlarged. The original templates are  $8 \times 8$  pixels.

as:

$$ccfmf = \frac{\sum_{j=0}^{n-1} \sum_{i=0}^{m-1} \left( \overline{P^l(u_p^l + i, v_p^l + j)} \cdot \overline{P^r(u_p^r + i, v_p^r + j)} \right)}{\sqrt{\sum_{j=0}^{n-1} \sum_{i=0}^{m-1} \overline{P^l(u_p^l + i, v_p^l + j)}^2 \cdot \sum_{j=0}^{n-1} \sum_{i=0}^{m-1} \overline{P^r(u_p^r + i, v_p^r + j)}^2}}. \quad (5.48)$$

Here  $P^l(u_p^l, v_p^l)$  denotes an intensity value of a window in the left camera image and  $P^r(u_p^r, v_p^r)$  a value of a window in the right camera image of size  $n \times m$  along the epipolar line. The values  $\overline{P^l(u_p^l + i, v_p^l + j)}$  and  $\overline{P^r(u_p^r + i, v_p^r + j)}$  represent intensities of mean-free windows

$$\begin{aligned} \overline{P^l(u_p^l + i, v_p^l + j)} &= P^l(u_p^l + i, v_p^l + j) - \mu^l \\ \overline{P^r(u_p^r + i, v_p^r + j)} &= P^r(u_p^r + i, v_p^r + j) - \mu^r \end{aligned} \quad (5.49)$$

with  $\mu$  representing the mean value of the window,

$$\begin{aligned} \mu^l &= \frac{1}{nm} \sum_{j=0}^{n-1} \sum_{i=0}^{m-1} P^l(u_p^l + i, v_p^l + j), \\ \mu^r &= \frac{1}{nm} \sum_{j=0}^{n-1} \sum_{i=0}^{m-1} P^r(u_p^r + i, v_p^r + j). \end{aligned} \quad (5.50)$$

Aschwanen (1993) shows that the  $ccfmf$  is optimal with respect to Gaussian image noise. The result of the  $ccfmf$  is a value between -1 and 1. Higher values represent higher similarity. The window pair with the highest correlation value is the determined correspondence point.

Instead of using this to match extracted features from one image in the other, a set of  $8 \times 8$  pixels templates representing the chessboard crossings is used to perform a coarse search for the chessboard corners. The templates represent the chessboard pattern at different tilt angles (see Fig. 5.6), in order to detect tilted calibration rigs reliably. The chessboard crossing carries the advantage of being the only size-invariant

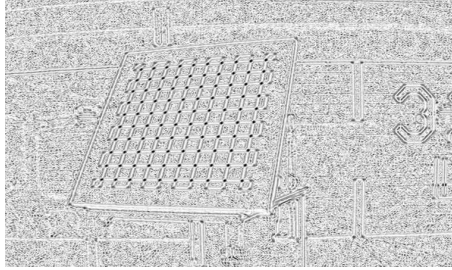


Figure 5.7: Image of the Correlation Coefficients: Black represents a high correlation value, white a low one. See left image in Fig. 5.10 for the original image and Fig. 5.6 for the matching templates.

primitive. A correlation window of  $8 \times 8$  pixels is moved across the original image, pixel by pixel, and the *ccmf* correlation coefficient is calculated for each template.

In contrast to the original Bouguet (2000) implementation the step of creating a binary image from the original one becomes obsolete. The contour based localization of the feature points works only properly on a binary image. The template matching operates on the original image. It fails on a binary one, because almost perfect matches are detected throughout the whole image.

In order to speed up the calculation, matching the inverted template is obsolete. This information is already included in the first computation: The correlation coefficient for a binary color inverted template is the negated correlation coefficient of the original template. Considering this, only the absolute value of each obtained correlation coefficient is taken into account.

The calibration rig contains 100 line intersections. Therefore, the 100 most significant points at pixel precision are extracted. However, in order not to obtain pixel locations that belong to the same crossing an adjustable minimum distance in between the maxima has to be kept.

To check a correct extraction of the points and sort them according to the Bouguet requirements first a Hough line detection is performed. Usually, a line is given by an equation such as

$$ax + by + c = 0.$$

The Hough representation, which is very suitable to determine lines in a given set of points, describes a line by the rectangular distance  $r$  from the origin and the angle  $\varphi$  of the distance line to the  $x$ -axis (see Fig. 5.8). With this representation one builds a so called Hough accumulator, i.e. for each data point all lines in a  $r$  over  $\varphi$  coordinate system are calculated. For those points which belong to a line their Hough accumula-

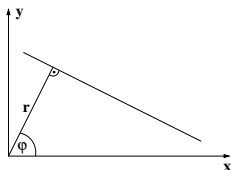


Figure 5.8: Hough Line Representation in Two Dimensions: A line is described by its rectangular distance  $r$  from the coordinate origin and the angle  $\varphi$  of the distance line to the  $x$ -axis.

tor curves intersect at one point, representing the  $r$  and  $\varphi$  of the point approximation line.

Even in images of highly tilted calibration rigs the horizontal lines are imaged as such. Therefore, the Hough accumulator is built for the horizontal lines, i.e.  $\varphi$  is confined to  $0^\circ \pm 30^\circ$ . From this accumulator the ten lines supported by the most points are extracted. The points are sorted with respect to the lines. In order to disregard wrong feature points, e.g. located at the image background, a mean distance from point to point along all the lines is calculated and lines with a point distribution far from this mean are omitted for the vertical line determination.

Using this first sorting, vertical lines  $i$  are fitted to the respective set of point  $i$  of each horizontal line by the means of a least-squares method. To robustly do so, only horizontal lines with ten points to them are considered. These ten vertical lines are assumed to be correct and the feature points are again sorted with respect to the lines. Again lines with a distribution departing from a newly calculated mean are disregarded. Subsequently, the same procedure is repeated for the horizontal lines using a least-squares fit.

This procedure uses a Hough determination for the horizontal lines and two least-squares fits first for vertical lines and then again for horizontal lines. In between the fits point distribution checks are performed. It proves to be very robust even for very slanted calibration rigs.

In the next step the 100 intersection points of the horizontal and vertical lines are calculated and represent the pixel precision feature point locations. Due to the calculation of the points by sorted sets of lines the sorting order demanded by Bouguet is obviously always obeyed.

To increase the accuracy a sub-pixel precision search for the maximum location around the pixel precision location of each feature point is performed. The used method is based of the one-dimensional sub-pixel search for correspondence maximum. To increase the accuracy of the correspondence, a widely used sub-pixel interpolation is applied (see e.g. Brandes (1995)). The correlation function is ap-



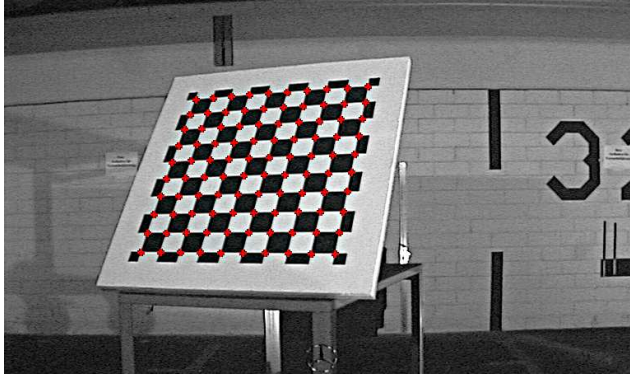


Figure 5.9: Extracted checkerboard feature points by the template correlation approach.

proximated around the maximum by a parabola. The exact location of the correlation maximum is the maximum of the fitted parabola:

$$\bar{u}_p^r = u_p^r + \frac{\frac{1}{2}(corr_1 - corr_{-1})}{2 \cdot corr_0 - corr_1 - corr_{-1}}. \quad (5.51)$$

This has been extended to two dimensions, e.g. in a  $5 \times 5$  pixel area around a feature point location the correlation values are collected a rotational paraboloid is fitted to the point distribution. The maximum of the paraboloid represents the sub-pixel precise location of the chessboard crossing. The performance is discussed in Section 5.7.

As mentioned above, Bouguet (2000) operates on an image series which is deemed rather unsuitable for an automated industrial application. However, a car driving by a series of identical checkerboards surmounts this disadvantage. In Bouguet's implementation an image is only valid if all corner points of the calibration rig are visible. This assumption is hard-coded in the MATLAB section. This has two effects: (1) If not all the control points are properly detected the image is discarded and (2) if the calibration rig is not fully visible the image must be omitted. Due to the lack of an automated error recovery in the Bouguet implementation this is a major disadvantage. To overcome this, algorithms which are able to robustly detect partially visible checkerboards were developed.

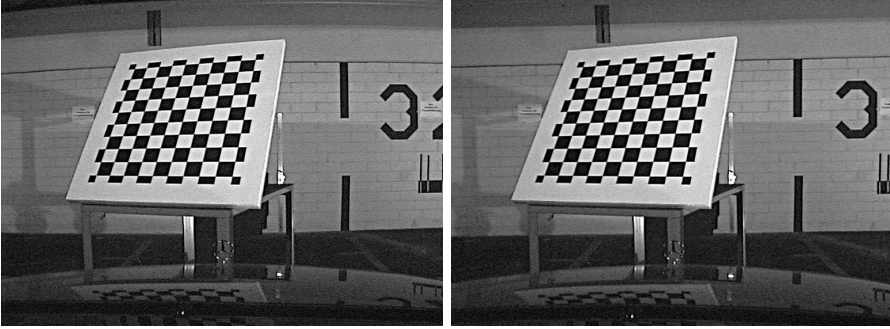


Figure 5.10: Original and Reconstructed Image of the Bouquet Calibration Rig. Left: Instead of dots (see left image in Fig. 5.2) a checkerboard is used. In order to make the outer corners available smaller rectangles and squares are added to the rim. Now there are 100 checkerboard crossings visible and detectable. Right: A lens distortion free image. The reconstruction is especially visible by comparing the pattern on the background wall in the left and the right image.

### Partial Visibility Detection (PVD)

Especially in stereo-camera systems the full visibility of a calibration rig in both images is not guaranteed. Additionally, if a car is passing a series of calibration rigs, other objects may occlude certain areas. These images still contain information which is useful in the calibration process. However, the control points must be robustly detected and the sorting order determined. This is especially true if the view is obstructed and different parts of the rig are imaged by the cameras.

The primary feature extraction of the partial visibility detection (PVD) is the template-based correlation approach described above. The determination of the initial horizontal lines is done by the evaluation of a Hough accumulator. With the determination of the vertical lines the difference in the approach starts. The vertical lines are extracted with a Hough accumulator. With these two sets of lines the feature points are sorted. Using the determined mean distance of the feature points along the horizontal and vertical lines respectively, correlation templates for each of the corners are generated dynamically (see Fig. 5.11). To care for skewed corner images, i.e. tilted calibration rigs, the templates are created with different rotational angles. One might think about using the line inclination to create the templates. However, this has not proved useful in the application.

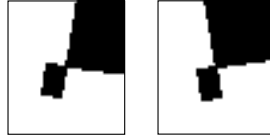


Figure 5.11: Dynamically Generated Matching Templates for Corners: Depending on the feature point distance along the horizontal and vertical lines dynamic masks for the corner detection are used. Here, a lower left corner is depicted at different rotational angles, but templates are generated for all corners.

The extracted feature points are now checked with the dynamic templates on their corner characteristics. To increase the robustness of the detection an outlier<sup>3</sup> detection is performed. This is done by examining the point distribution based on the mean distance calculation. If only one corner is visible the sorting starts there. Otherwise the corner with the highest correlation coefficient is taken as the starting point. To sort the points into the required absolute raster, first the line inclination and the point distances are used. Later on, the search is performed with pieces of lines that are supported by multiple points.

To conclude the pixel precision detection, lines are determined by least-squares methods based on the point-to-lines sorting obtained before. If there are no inconsistencies the pixel-precise location-phase is terminated. The sub-pixel precision detection is the same as described above.

This approach, implemented during an internship by A. Moosbrugger, proved to be very efficient and robust. For more details consult the internship report Moosbrugger (2002).

While the first correlation based algorithm is used as a replacement of the OpenCV approach, the PVD has implications on the subsequent procedure. The division of the application in a C++-based monocular calibration and a subsequent MATLAB-based determination of the external camera parameters is impractical. In addition to this more convenience-driven point, the results of the partial visibility detection (PVD) require alterations in the implementation. The Bouguet implementation discards the whole image if points are not properly detected. In the PVD approach an images must only be discarded if no points are detected. This change and the monolithic implementation is accomplished by the master's thesis of D. Schowalter (2002).

<sup>3</sup>Outliers are points which are detected as corners but they are not located on the calibration rig but in the image background.

## 5.7 Results

In this section a short presentation of a performance comparison of the different Bouguet calibrations is given. As mentioned above, tests between different calibration approaches are presented in the diploma thesis of N. Kaempchen (2001). In addition, during the tests numerical problems in the reconstruction based on the homogenous coordinate representation were encountered and are presented here.

### Comparison of the Bouguet implementations

To compare the performance of the different Bouguet implementations, tests on different sets of images were conducted. Four of the images sets were taken with one stereo-camera pair (named IC2 (1) through (4)), a fifth with a different one (named UTA). The four IC2 sets mainly differ in the number of images and in their composition, i.e. certain 'difficult' images are added intentionally to produce situations that are hard to detect or which cause numerical instability (flat-on images). All the images contain 100 control points on the checkerboard calibration rig. The image size for all test sets is  $768 \times 568$  pixels. Even though the cameras operate in interlaced mode, full images are grabbed. This is unproblematic due to the static setup of the calibration environment. The IC2 cameras are equipped with 7.5 mm and the UTA cameras with 12 mm lenses.

|                    | IC2 (1) | IC2 (2) | IC2 (3) | IC2 (4) | UTA |
|--------------------|---------|---------|---------|---------|-----|
| # images           | 25      | 40      | 40      | 80      | 15  |
| # difficult images | 0       | 5       | 10      | 15      | 0   |

Table 5.1: Calibration Test Sets

As measurements the following values of the left camera in each system are examined:

- The standard deviation of the focal lengths  $\alpha_u$  in  $u$ - and  $\alpha_v$  in  $v$ -direction,
- the standard deviation of the principle point location  $(c_u, c_v)$ ,
- $\overline{\text{res}}$  the mean residual error,
- $\text{res}_m$  the maximum residual error,
- and the standard deviation deviation  $\sigma_{\text{res}}$  of the residual error.

Note that the residual errors are calculated on the basis of the monocular calibration. All values in the tables are given in pixels. The conversion factor is 120.00 pixel/mm.

In the set IC2 (2) only the stereo calibration in MATLAB failed whereas in the sets IC2 (3) and IC2 (4) the monocular stage failed, both due to a numerical instability error message.

|                         | Original Bouguet    | Correlation Based   |
|-------------------------|---------------------|---------------------|
| $\alpha_u$              | $897.974 \pm 1.223$ | $897.522 \pm 0.578$ |
| $\alpha_v$              | $890.694 \pm 1.434$ | $890.282 \pm 0.680$ |
| $c_u$                   | $385.527 \pm 2.261$ | $386.355 \pm 1.079$ |
| $c_v$                   | $296.008 \pm 1.857$ | $296.529 \pm 0.878$ |
| $\overline{\text{res}}$ | 0.901               | 0.088               |
| $\text{res}_m$          | 2.853               | 0.264               |
| $\sigma_{\text{res}}$   | 0.322               | 0.045               |

Table 5.2: IC2 (1) Calibration Results

|                         | Original Bouguet | Correlation Based   |
|-------------------------|------------------|---------------------|
| $\alpha_u$              | no result        | $898.812 \pm 0.807$ |
| $\alpha_v$              | no result        | $891.095 \pm 0.857$ |
| $c_u$                   | no result        | $385.312 \pm 1.178$ |
| $c_v$                   | no result        | $298.268 \pm 1.163$ |
| $\overline{\text{res}}$ | 1.150            | 0.623               |
| $\text{res}_m$          | 8.950            | 1.981               |
| $\sigma_{\text{res}}$   | 0.899            | 0.572               |

Table 5.3: IC2 (2) Calibration Results

|                       | Original Bouguet | Correlation Based   |
|-----------------------|------------------|---------------------|
| $\alpha_u$            | no result        | $898.419 \pm 0.797$ |
| $\alpha_v$            | no result        | $890.710 \pm 0.914$ |
| $c_u$                 | no result        | $386.946 \pm 1.260$ |
| $c_v$                 | no result        | $294.557 \pm 1.056$ |
| $\bar{\text{res}}$    | no result        | 0.685               |
| $\text{res}_m$        | no result        | 2.010               |
| $\sigma_{\text{res}}$ | no result        | 0.601               |

Table 5.4: IC2 (3) Calibration Results

|                       | Original Bouguet | Correlation Based   |
|-----------------------|------------------|---------------------|
| $\alpha_u$            | no result        | $898.383 \pm 2.371$ |
| $\alpha_v$            | no result        | $891.935 \pm 2.462$ |
| $c_u$                 | no result        | $380.980 \pm 3.526$ |
| $c_v$                 | no result        | $297.065 \pm 3.074$ |
| $\bar{\text{res}}$    | no result        | 1.538               |
| $\text{res}_m$        | no result        | 2.321               |
| $\sigma_{\text{res}}$ | no result        | 0.891               |

Table 5.5: IC2 (4) Calibration Results

|                       | Original Bouguet     | Correlation Based    |
|-----------------------|----------------------|----------------------|
| $\alpha_u$            | $1527.032 \pm 2.520$ | $1525.792 \pm 0.837$ |
| $\alpha_v$            | $1520.201 \pm 2.586$ | $1518.838 \pm 0.846$ |
| $c_u$                 | $392.823 \pm 4.166$  | $387.410 \pm 2.074$  |
| $c_v$                 | $270.060 \pm 3.531$  | $268.221 \pm 1.739$  |
| $\bar{\text{res}}$    | 2.481                | 0.567                |
| $\text{res}_m$        | 6.331                | 1.251                |
| $\sigma_{\text{res}}$ | 1.839                | 0.295                |

Table 5.6: UTA Calibration Results

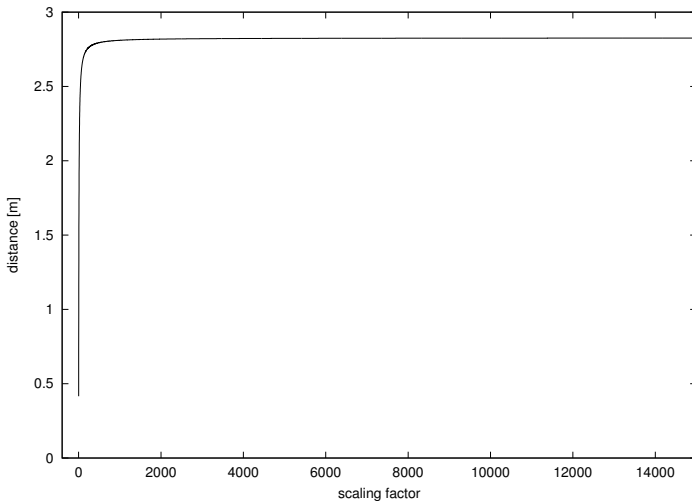


Figure 5.12: Distance Measurement Depending on the Scaling Factor: The object distance measured by a laser unit is 2.88 m.

### Numerical instabilities in the three-dimensional reconstruction

Another point which has been detected during the investigation of the Bouguet approach is related to numerical instabilities of the three-dimensional point reconstruction in the homogenous coordinate representation.

Usually, the following procedure is used to reconstruct the three-dimensional object location. First the image locations  $(u, v)$  of an object are determined in both images. In the case of the Bouguet (2000) camera calibration the object is a chessboard crossing and the correspondence is given by the sorting order of the feature points. Then, a camera coordinate system is chosen as the reference frame, e.g. the left camera coordinate system. For the left camera system the viewing ray is a line through the origin with the direction  $(x, y, z)$  given by Eq. 5.21. For the right camera the viewing ray is determined equivalently. The representation of the right ray of sight in the left coordinate system is obtained by transforming the right camera coordinate system origin as well as the direction vector with the extrinsic parameter matrix  $\mathbf{T}_e$  (see Eq. 5.21). The theoretical intersection<sup>4</sup> of these two lines in the left camera coordinate system represents the three-dimensional object location. The intersection is described by a set of linear equations.

<sup>4</sup>In a system containing errors the lines are warped.

The remark that must be made refers to the scaling factor in Eq. 5.21. Since the longitudinal distance  $z$  of the object is unknown, the scaling factor is freely chosen. It is straightforward to set the scaling factor  $s = 1$  to ease the calculational load. However, it has been discovered that the computation is numerically sensitive to the scaling factor. Following that convenient way by setting  $s = 1$ , wrong distances of the calibration rig were obtained. For further investigation a test setup was made. A flat-on image of the calibration rig is taken. The distance from the camera location to the rig is measured using a laser distance measurement unit to be  $d = 2.88$  m. The obtained distances are shown in Fig. 5.12.

This problem becomes obvious by examining the minimal distance between two lines. The distance is (Bronstein & Semendjaev, 1985):

$$d = \frac{\begin{vmatrix} x_1 - x_2 & y_1 - y_2 & z_1 - z_2 \\ a & b & c \\ a' & b' & c' \end{vmatrix}}{\sqrt{\begin{vmatrix} a & b \\ a' & b' \end{vmatrix}^2 + \begin{vmatrix} b & c \\ b' & c' \end{vmatrix}^2 + \begin{vmatrix} c & a \\ c' & a' \end{vmatrix}^2}} \quad (5.52)$$

for two lines given in coordinate representation:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} + \lambda \begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} + \lambda' \begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} \quad (5.53)$$

Comparing this to the setup of the direction vectors and the influence of the scaling factor  $s$  on them, it is obvious that a small scaling factor  $s$  accounts for denominator instabilities upon numerical calculation.

## 5.8 Discussion

The conclusion of Section 4.4 is that camera calibration is a very basic requirement for the dependable extraction of information from an image. With the objective to build an easy to use but yet reliable and accurate calibration tool which suits industrial needs an investigation into calibrations methods started. Conducting comparative tests it became obvious that the Bouguet method is very suitable and provides results superior to the other approaches. However, the accuracy is still not sufficient and a new method to reliably detect the control points, fully visible or not, was developed. Tests show that the new method improves the accuracy of the calibration by 15 – 45%. The reliability is increased as well. The new method still provides acceptable results



in image sequences where the old one fails completely due to numerical instabilities or the inability of a robust control point detection.

As mentioned above, speed is not critical for a calibration implementation, because the calibration is usually done only once and off-line. Still, by operating on the original image and taking advantage of the algorithmic properties, such as the symmetry in the *ccfmf* for color inverted binary templates, the processing speed is increased.

Altogether, a basic requirement for the extraction of image information on free-formed surface mirrors has been fulfilled. In addition, a tool has been developed that provides necessary data to many other applications as well.

# Reconstruction of the Reflective Surface

In this chapter the second basic requirement determined in Section 4.4 is fulfilled: the determination of the relative position of each camera in the stereo system to the surface geometry. To do so, the geometry itself is reconstructed on the basis of an acquired image sequence. After giving a short introduction and a review of the related work, the applicable mathematical representations are discussed. Subsequently, a newly developed method on the basis of the camera-calibration technique is introduced and its results are discussed.

### 6.1 Introduction

The image of a scene reflected on a free-formed surface is distorted. The distortion depends on the surface geometry as well as on the camera position. Therefore, not only the determination of the surface geometry but also of the relative camera position to the surface is necessary. Bearing the application in mind, the obtained surface representation should be smooth and the curvature information must be retrievable, i.e. the surface has two continuous derivatives ( $C^2$ ) (see Chapter 7).

In the special case of a car body part serving as a free-formed surface mirror, the surface geometry is available in CAD formats. Nevertheless, rather than using this surface data and determine only the relative camera position respectively, the decision was made to develop an approach on the basis of the camera-calibration technique as described in Chapter 5. The decision is based on the consideration of the following points:

- An assembled car body might deviate from the designed form represented by the CAD data. The deformation might even occur after the assembly itself.
- A CAD data representation with its control net, supporting planes, and so on is a rather bulky geometry representation. The data formats are not designed with regard to storage efficiency.

- A method to reliably determine the relative position has to be developed. This includes the case of CAD data usage.
- The decision not to use existing methods, such as structured-lighting techniques, is driven by the demand not to introduce additional hardware. Usually, these methods require a calibrated setup of an back-illuminated screen, displaying the structures for the geometry determination, and the camera. In addition to that, some of these techniques are either not suitable for reflective surfaces or they are not applicable to a large object such as the hood of a car.
- And last but not least, developing a method to recover the geometry of a reflective surface is an interesting task.

## **6.2 *Related Work***

In this section techniques for the geometry reconstruction are introduced and a few remarks are made to each approach.

The focus of this overview is mostly on structured-lighting methods because these dominate the field. The principle technique is to set up a controlled environment. It consists of a screen for the structure display and a calibrated camera to image the projection of the structures on the object. The camera is calibrated with respect to the internal parameters as well as the relative position to the screen. The displayed structures are usually gray-scale line patterns with varying spatial frequency. By exploiting the correspondence between the emitted and the observed spatial frequencies and the distortion of the parallel line pattern the object's topography is obtained. However, the usually applied structured-lighting techniques are not suitable for reflective surfaces and therefore they are omitted here. The reasons for the failure of the ordinary structured-lighting techniques are stated in the remarks about Kammel (2001).

Other techniques, such as shape-from-shading or triangulation methods, are not adequate for the inspection or reconstruction of specularly reflective surfaces. They depend on the diffuse reflection component. Stylus instruments and optical autofocus scanners provide highly accurate results but the data acquisition is very slow. Methods related to interferometry are too sensitive and too expensive for the industrial application.

1. Beyerer & Pérard (1997): In this approach, a series of CCD-camera images of the reflections of a freely programmable back-illuminated LCD are evaluated. Displaying a known pattern sequence and applying an inverse ray-tracing method the surface relief is reconstructed. In contrast to the methods of line projection, where the measurement is locally affected mainly by the topography, this method is more susceptible to the inclination of the surface, i.e. the

measured effect is dependent on the surface gradient. Usually, the obtained linear equation system contains more unknowns than equations. This is especially true for a local examination of the reflection points. In this approach the solution is obtained by a global approach, demanding smoothness of the surface and exploiting the thereto related correlation of neighboring points.

2. Pérard (2000): The task of that thesis is to develop an automated inspection system for curvature defects on free-formed surfaces. The methodical approach is similar to Beyerer & Pérard (1997). The algorithm is extended to cope with varying curvatures within a surface. To estimate the local enlargement a special pattern must be used. Pérard considers the linear polarization of light emitted by an LED. However, the sensitivity of the reconstruction is not constant and depends on the geometric setup of the measurement system as well as on the surface curvature. For automated inspection of predefined reflective objects, such as a car body part at the end of a production line, the method is inverted. The display pattern is altered, such that the distorted image, reflected on the surface, produces a parallel fringe pattern. Subsequently, a curvature defect, e.g. due to a faulty varnishing of the part, is easily detectable as a deviation from the parallelity.
3. Kammel (2001): The introduced method is a reflection technique. In contrast to the usual projection method in structured-lighting techniques, it is applicable to reflective surfaces.

The ordinary projection techniques are based on triangulation. The camera is focused on the surface to be measured and the structured light is projected onto it. The obtained image is distorted compared to the original structure because the camera exhibits a parallax.

In the case of a reflection technique the camera is focused on a screen, displaying the light pattern, that is reflected by the surface. This requires the surface to be at least partially specular. The surface becomes a part of the optical system, and therefore distorts the observed structural pattern. If the surface is warped at an observed point, in the case of a projection method still the same part of the pattern is imaged. This is due to the focusing on the surface. However, if a reflection method is applied, another part of the pattern is observed, again due to the focusing, in this case on the screen as seen in the reflection. This constitutes a sensitivity to the surface curvature. Kammel (2001) applies the same smoothness constraint to solve the sub-determined linear equation system explained in Beyerer & Pérard (1997).

The paper investigates the special properties of convex and concave surfaces. Concave surfaces are much harder to recover due to their light collecting prop-

erties. In the structured-lighting techniques it is assumed that the fringes are distinguishable, i.e. the image-function is coded. Observing concave or partially concave surfaces, parts of the pattern might overlap or points in the image cannot be assigned to points on the screen explicitly. Considering this, the fringe pattern resolution must be lower for concave regions compared to those for convex ones.

As mentioned above, the technique is directly sensitive to the surface curvature. However, the recovery of absolute values is only possible if the measurement of the warped surface is compared to the measurement of a planar surface. The planar surface must replace the warped surface in the experimental setup. A requirement, that is only fulfilled in a controlled laboratory setup and for objects of a rather small size.<sup>1</sup> For the surface reconstruction of the hood of a car, especially in an assembled state, the technique is not applicable.

4. Andresen & Morche (1983): This method evaluates an image series of projections of cross-raster structures on a dynamically deformed surface. The cross-raster is a pattern of horizontal and vertical parallel lines. This pattern is used because the intersection points of the cross-raster are easy to detect with image-processing algorithms and therefore accounts for automation. The evaluation provides curvature and inclination information along the coordinate lines which correspond to the cross-raster directions. The evaluation method is based on the integration of the cross-raster point movements and its derivatives in locally skewed coordinate systems. However, for the intended application, a controlled deformation of the object, the hood, is not feasible.
5. Ritter & Hahn (1984): The method is based on an extended analysis of the reflection Moiré effect (Ligtenberg, 1954). The interpretation of Moiré lines is used to examine the deformation of objects. To a certain degree the approach is similar to the structured-lighting techniques. However, here the information is not obtained by projecting patterns with varying spatial frequency onto a static object but by deforming the object and observing the change in the static pattern. The focus of the experiments is stress analysis of the objects. Assuming small changes in the topography the reflection Moiré effect is used to directly measure the curvature of an object. Ritter & Hahn (1984) enhanced the Ligtenberg (1954) approach with regard to the interpretation of object points with the same inclination. The closing remarks for Andresen & Morche (1983) apply here as well.

---

<sup>1</sup>During the course of the investigation of this paper, a meeting with S. Kammel was arranged at the University of Karlsruhe. The experiment of curvature recovery was personally observed. The used object is about 10 cm<sup>2</sup> and is replaced in the optical bench by a planar mirror of equivalent size.

6. Savarese & Perona (2001): In this publication, the geometry linking the shape of a curved mirror-surface to the distortions it incorporates in a reflected scene is explored. The analysis is local and differential and operates on a calibrated scene composed of lines passing through a point. The study assumes perfectly specular mirrors and shows that local information about the geometry of the surface is recovered up to the second order from either the orientation and curvature of the images of two intersecting lines, or from the orientation of the images of three or more intersecting lines. Other proposed patterns in the article are circles or non-planar patterns. However, circles will not qualify for a proper recovery of the surface (see Section 5.2). It is necessary to image the calibrated pattern at an angle of inclination, otherwise the correlation between the line in the image and the line in the scene does not carry meaningful information. The topology of the surface is locally approximated by a paraboloid. An explicit calculation is given for a spherical mirror surface.

### 6.3 Mathematical Geometry Representation

The shape of manufactured specular surfaces is generally known, and the mathematical description, used in the CAD systems, should be taken into account. In the case of developing a measurement system, a mathematical model has to be derived from the scattered data points.

In order to achieve flexibility, the model must be sufficient to represent arbitrary topologies. In addition to this, the model should meet several requirements:

- To obtain normal vector and curvature information, the surface must be at least  $C^2$  continuous, as used by most CAD applications. This is due to the fact, that these surfaces are more pleasing to the human eye. Another reason is, that in most cases,  $C^2$  continuous patches are better approximated by their control net than patches with  $C^0$  or  $C^1$  continuity, e.g. bicubic compared to biquadratic B-spline patches.
- The surface model that interpolates data points from a set of points should converge to a certain shape if data points are added.

In general, there are two basic types of functions:

1. Given data points  $(x_i, y_i, z_i) \in \mathbb{R}^3$ , find a scalar function  $F : \mathbb{R}^2 \rightarrow \mathbb{R}$  that approximates or interpolates the value  $z_i$  at  $(x_i, y_i)$ , i.e.  $F(x_i, y_i) \approx z_i$ . This problem is known as *Scattered Data Interpolation*. There are many solutions to this problem, among them finite element methods.

In this case, the approach is applied to a height field representation of the surface. In the height field representation, the distance  $s_i$  of the surface point along the ray of sight, associated with an image location  $(u_i, v_i)$ , is recorded.

2. In contrast to the scalar problem, there is the *Parametric Problem*. The task here, is to find a parameterized surface  $F : \mathbb{R}^2 \rightarrow \mathbb{R}^3$  that approximates or interpolates the data points. This is usually done by specifying additional parameter values  $(u_i, v_i) \in \mathbb{R}^2$  and by determining  $F$  such that  $F(u_i, v_i) \approx (x_i, y_i, z_i)$ . This theory of parametric surfaces is well understood in differential geometry and dates back to the time of Gauss.

In this case, the additional parameter space is introduced such that it coincides with the image locations.

In the case of using a parameterized surface representation, which suggests itself, additional requirements need to be met:

- Fast estimation of intersection points, which is a property of the *convex hull*. The convex hull refers to the property that the surface is described by a control net and that the surface points always lie inside this net. Subsequently, the control net serves as a frame for an interference check of light rays and the surface to decide whether it is necessary to calculate the intersection point. This speeds up a ray-tracing process significantly.
- To calculate intersections of viewing rays with the surface, the surface has to be subdivided in smaller patches iteratively. This is the most time-consuming calculation step in a ray-tracing method and should therefore be implemented in a time-efficient way.

Data interpolation and the convex hull property are contradictory. A control net that is the convex hull of a given set of data points cannot interpolate all of them unless a reduced continuity ( $C^0$ ) is accepted. According to this, requiring at least  $C^2$ -continuity demands a data approximation process.

One model fulfilling the stated demands are the non-uniform bicubic tensor product B-spline patches. In this model, the coordinates of a surface point

$$\mathbf{p}(u, v) = \begin{pmatrix} x(u, v) \\ y(u, v) \\ z(u, v) \end{pmatrix} = \sum_{i=0}^3 \sum_{j=0}^3 \mathbf{d}_{ij} N_i^3(u) N_j^3(v) \quad (6.1)$$

are expressed in terms of the points of the control net  $\mathbf{d}_{ij}$  and the bicubic blending functions  $N_i^3(u)$  and  $N_j^3(v)$ . Tensor product patches are a simple extension of curves to two dimensions and are separable.

Another basis for the representation of the surface topography and the slope is the mathematical representation of the design. These are usually Non-Uniform Rational B-Splines (NURBS). A NURBS of the fourth order with  $n \times m$  patches is described by:

$$\mathbf{r}(u, v) = \frac{\sum_{i=0}^m \sum_{j=0}^n \mathbf{C}_{ij} N_{i,4}(u) N_{j,4}(v) w_{ij}}{\sum_{k=0}^m \sum_{l=0}^n N_{k,4}(u) N_{l,4}(v) w_{kl}} \quad (6.2)$$

with  $N_{i,4}(u)$  and  $N_{j,4}(v)$  being the two basis functions,  $\mathbf{C}_{ij}$  being the bi-directional control net and  $w_{ij}$  being weights. The basis functions  $N_{i,p}$  are the normalized B-spline basis functions of degree  $p$ , defined recursively as (Piegl, 1991):

$$N_{i,0}(u) = \begin{cases} 0 & \text{if } u_i \leq u < u_{i+1} \\ 1 & \text{otherwise} \end{cases} \quad (6.3)$$

$$N_{i,p}(u) = \frac{u - u_i}{u_{i+p} - u_i} N_{i,p-1}(u) + \frac{u_{i+p+1} - u}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(u), \quad (6.4)$$

where  $u_i$  are the so called knots forming a vector

$$\mathbf{U} = \{u_0, u_1, \dots, u_m\}. \quad (6.5)$$

The degree, number of knots and number of control points are related by the formula  $m = n + p + 1$ . For non-uniform and non-periodic B-splines, the knot vector takes the form

$$\mathbf{U} = \{a, a, \dots, a, u_{p+1}, \dots, u_{m-p-1}, b, b, \dots, b\}, \quad (6.6)$$

where the end knots  $a$  and  $b$  are repeated with multiplicity  $p + 1$ . In most practical applications  $a = 0$  and  $b = 1$ .

NURBS possess several advantages:

- They offer a common mathematical form for representing and designing free-formed surfaces and analytical shapes, e.g. conics, quadrics, and surfaces of revolution.
- In the NURBS representation the control net as well as the weights are manipulated. The weighted consideration implies additional geometric advantages:
  - Generalization: If all the weights are set to 1, the NURBS are equal to non-rational B-spline forms.
  - If a particular weight is zero, the respective control point does not have an influence on the surface.



- The surface is only affected locally by the change of a weight.
- The evaluation is reasonably fast and computationally stable.
- NURBS have a clear geometric interpretation.
- NURBS have a powerful geometric tool-kit, such as insertion, refinement, and removal of knots, degree elevation, and splitting.
- Invariance to scaling, rotation, translation, and shear as well as parallel and perspective projection. This allows for cross checks of the separately determined surface geometries relative to the camera, provided that the camera to camera transformation is known.
- Locality: Outside the respective interval the rational basis function equals zero.
- Partition of unity: The sum over all rational basis functions equals 1.
- Differentiability: In the interior of a knot span, the rational basis functions are infinitely, continuously differentiable if the denominator is bounded away from zero. At a knot they are  $p - k$  times continuously differentiable where  $k$  is the multiplicity of the knot.

But NURBS carry disadvantages as well:

- To define traditional curves and surfaces extra storage is needed. A circle in space is represented by a circumscribing square requiring seven control points and ten knots. Compared to a traditional representation (the center, the radius and the normal vector of the containing plane) requiring seven numbers in three-dimensional space, the NURBS representation needs 38.
- Improper application of the weights may result in a very bad parameterization, subsequently destroying the surface representation.
- Interrogation techniques, such as the intersection curve of two surfaces that are just touching, are more easily represented in traditional forms.
- Fundamental algorithms, e.g. inverse point mapping, depend on iterative approaches and are therefore subject to numerical instability. However, this property not only applies to NURBS. Other free-formed schemes exhibit this property as well.

## 6.4 Geometry Reconstruction from a Sequence of Images

After analyzing the demands and the approaches reviewed in Section 6.2, it was decided to develop a surface recovery method based on the following considerations:

- Most ready-to-use techniques require additional hardware and the data representations cannot be chosen closely related to the specific task. In addition, some of the methods are only applicable to small objects.
- Andresen & Morche (1983) introduces a method of surface recovery based on the projection of cross-raster patterns. The checkerboard pattern used in Chapter 5 (see Fig. 5.10 and Fig. 6.1) for the camera calibration constitutes a cross-raster. The necessary algorithms are developed and coded for the camera calibration. Therefore, only an extension and accommodation is necessary.
- The system should be self-contained, i.e. it depends only on data acquired by the camera system itself.
- In analogy to the structure-lighting techniques, the surface is scanned with varying spatial frequency. But rather than to alter the projected pattern in a fixed setup, the pattern remains the same and the longitudinal distance is altered, i.e. an image series of the calibration pattern is recorded with a varying camera to rig distance.

Accordingly, an approach consisting of the following steps is developed:

1. The original image data basis is obtained by recording an image series of the calibration checkerboard with a calibrated camera system, i.e. all the camera parameters are known (see Chapter 5). To obtain the data across the whole object the lateral starting point of the image series is altered as well. Having a mobile camera system, mounted in a car, the distance variation is easy to obtain. The image series is recorded with available software while slowly driving relative to the calibration rig, starting at different lateral offsets.
2. With the knowledge of the internal camera parameters, the obtained images are corrected for lens distortions (see Fig. 5.10). Subsequently, all algorithms, references and methods are applied to reconstructed images.
3. The correlation algorithm, developed with regard to the calibration (see Section 5.6), is applied as stand-alone process. In it, according to the calibration procedure, the chessboard features in direct sight are detected. The correspondence of the feature points in the left and the right camera is established by means of the sorting order of control points.

4. In this step the point correspondences of the features in direct sight and in reflected sight are extracted. To do so, the feature detection algorithm has been extended to work on the reflected image as well. The extensions are described in the following Section 6.4.1.
5. With the correspondence of the feature points, the internal camera parameters, i.e. the mapping of the image points to lines in the three-dimensional space (see Eq. 5.21), and the knowledge of the external camera parameters (see Eq. 5.22) the three-dimensional point locations of the control points are obtained. This is done for each camera separately, as each of the following steps.
6. Using the correspondence between the features in direct sight and reflected sight, the obtained three-dimensional point locations are assigned to image locations in the reflection area. This data is collected while processing the whole image series. It provides a data set that contains single as well as multiple three-dimensional point locations for as many image locations in the reflected sight as possible.
7. Image locations with multiple corresponding three-dimensional object locations are used for further processing. Image locations with only a single point assigned to them are disregarded. By the means of a least-squares method, lines are fitted to the three-dimensional object locations. These lines represent rays of reflection.
8. With the knowledge of the direction of reflection for a given image location and the associated ray of sight, determined by the internal camera parameters, the surface point is obtained. The method of surface point determination is described in more detail in Section 6.4.2.
9. For the acquired data points either a height field or a parameterized representation is obtained (see Section 6.3). With this knowledge a dense surface point net is calculated.

#### ***6.4.1 Expansion of the Checkerboard Camera-Calibration Technique***

A very important step in this whole surface recovery approach, is a robust establishment of the feature correlation in the direct view and in the reflected view. The foundation of the procedure is the feature detection of the camera calibration as described in Section 5.6.1. However, to achieve the desired accuracy, some modifications are necessary.

In contrast to the feature extraction in direct sight, the feature location in the reflective area is a little more challenging. Instead of extracting the 100 most prominent features and trying to approximate sets of horizontal or vertical lines to them, a partial

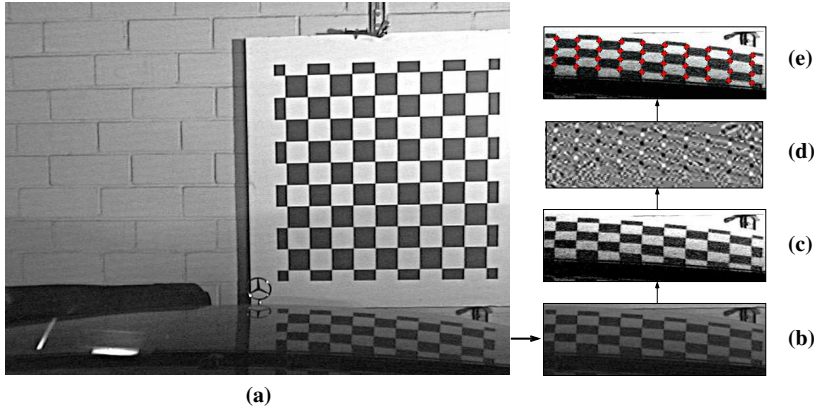


Figure 6.1: Acquisition Scheme for the Hood Surface Data: (a) lens-distortion-free original image, (b) extracted reflection image region, (c) contrast enhanced image, (d) image of correlation coefficients and (e) extracted sub-pixel precise feature locations.

visibility approach, introduced in its principle in Section 5.6.1, is necessary. This is due to the fact that in most cases only a part of the calibration rig is visible. At close range, even if the control points are all detectable in direct view, the reflection will only contain some of them. At far range, due to the diminishing resolution correlated with the curvature of the surface, not all points are detectable.

As depicted in Fig. 6.1 the feature point extraction is a multiple step procedure. The whole process is directly applicable to the original image. However, in that case the recovered surface is additionally warped by the lens distortions. Hence, everything operates on lens-distortion-free images (see Fig. 6.1 (a)).

The first step is a coarse cropping of the region of interest. Based on the leftmost and rightmost corner points of the calibration object in direct view, plus a certain amount of pixels to the left and right (see Fig. 6.1 (b)).

Due to the varnishing of the hood, the contrast within this image is diminished. In principle, the mean-free cross correlation function ( $ccmf$ , see Eq. 5.48), described by Aschwanden (1993) is able to deal with this. The subtraction of the mean gray-value of the image constitutes a susceptibility to a relative change rather than an absolute one. However, to reach the desired accuracy in the presence of noise an image enhancement is performed. The observable improvement (see Fig. 6.1 (c)) is obtained by applying a histogram spreading, i.e. the gray-value distribution is stretched to maximum and minimum gray-values while preserving its shape. Due to this preservation, the precise feature location, is retained.

Subsequently, a template correlation with a single  $8 \times 8$  pixel masks is performed (see the leftmost mask in Fig. 5.6). In contrast to the feature extraction in the calibration approach, the symmetry in the *ccfmf* (see Section 5.6.1) is not disregarded but used later on. The *ccfmf* provides a correlation result between  $-1$  (anti-correlated) and  $1$  (correlated). Matching a binary template and obtaining a certain correlation coefficient implies that correlating the inverted template will yield the negative result. This is applied here to speed up the process. In the camera-calibration technique only the absolute value of the correlation is regarded (see Fig. 5.7). Here, positive and negative values are kept (see Fig. 6.1 (d)). There, white represents a match with the applied template and black an anti-correlation, that is a match with the inverted template.

In the calibration effort the 100 most prominent features are extracted, disregarding the value of the correlation coefficient or its sign. Here, in order not to obtain false matches, relative thresholds, positive and negative, are set. Typically, an absolute threshold value of 0.8 of the maximum correlation coefficient within the image, has proven to be robust.

Similar to the approach in the Partial Visibility Detection (PVD), corner masks are created on run-time, based on the mean distance of the feature points (see Fig. 5.11). These are used to test the previously extracted coarse features for being a corner. Similar to the calibration case, the dynamic masks are created at different rotational angles. This is even more important in the reflective image region, due to the distortions caused by the curvature. In the upper image parts of the hood, where the curvature is higher than in the lower ones, the feature distance might be too small for the dynamically allocated templates. To overcome this difficulty, the image is resized by a factor of two in each direction. The intermediate gray-values are obtained by applying a bilinear interpolation. Performing the corner test on these images proves to be robust. It must be noted that the enlarged image is only used to characterize a feature for being a corner. The extraction of the location is only performed in the originally-sized image. Upon the robust detection of a corner, the sorting process is started. In contrast to the PVD in direct sight, the lines or rather curves, due to surface curvature, are reconstructed by examining partial lines in between the points. The replacement of these lines by larger segments is not performed as it has proven to be rather unhelpful. To check the sorting, making sure that there are no jumps in between lines, the distribution of the correlation coefficients is used. Along a reconstructed line, positive and negative maximum values must take turns. The encounter of the consecutive equal signs indicates an error in the reconstruction or the sorting.

The second to last step, is the sub-pixel precise location of the feature points. It is performed according to the sub-pixel search in the calibration algorithm (see Section 5.6.1). At this point the features, previously disregarded due to their correlation coefficient falling below the threshold, are reconsidered. This is done on the basis of

intersecting the partial lines in between pairs of points of equally signed correlation coefficients. An extraction example is shown in Fig. 6.1 (e).

Finally, the correlation between the points is established on the basis of the determined sorting order. Notice that the lower left corner in direct sight corresponds to the upper left corner in the reflection.

#### 6.4.2 Surface Reconstruction by Distance Minimization

Given a set of three-dimensional object locations associated with an image location in the reflection area, the task is to obtain the three-dimensional surface point that fits best to the data.

Through  $\mathbf{T}_i$ , the matrix of internal camera parameters (see Eq. 5.21), the image location  $(u, v)$  is correlated to a ray of sight, an origin line in the camera coordinate system. This line intersects the reflective surface. How is that point determined?

A first approach is to solve for the ray of reflection by setting up a least-squares problem and solving for the supporting point and the directional vector. The solution has to minimize the sum of squared distances of the data points to the line. In almost any case, the obtained ray of reflection and the ray of sight will not intersect. Determining the point of closest distance of the two lines on the ray of sight yields a result. However, this result is most certainly not the desired one. For one, the determined point is not part of the ray of reflection. Another ray of sight, associated with a different image location, does intersect this ray of reflection, but most certainly not at a true surface point.

In the description above the point of closest distance on the ray of sight is chosen as the resulting point. Picking some other point is equally meaningless. The point of closest distance on the ray of reflection, carries the problem of being part of the ray of sight. This implies that a projection of that surface point results in a different image location than the one originally determined.

Considering these misfits, some other solution has to be obtained. One solution, that does not carry the problems mentioned above, is calculated by constraining the supporting point of the ray of reflection to ray of sight. However, this approach requires a different determination of the directional vector. This is elaborated in the subsequent presentation.

In general, a line in three-dimensional space is approximated by minimizing the Euclidean distance of the data points to the potential line. The Euclidean distance  $d$  of point  $p = (x_3, y_3, z_3)$  from a line, given by  $\mathbf{x}_1 = (a, b, c) = (x_2 - x_1, y_2 - y_1, z_2 - z_1)$ ,

is (Bronstein & Semendjaev, 1985):

$$d^2 = \frac{1}{a^2 + b^2 + c^2} \left[ ((x_3 - x_1)b - (y_3 - y_1)a)^2 + ((y_3 - y_1)c - (z_3 - z_1)b)^2 + ((z_3 - z_1)a - (x_3 - x_1)c)^2 \right]. \quad (6.7)$$

Setting up a least-squares method based on this bulky expression is rather inconvenient. However, using an analog problem situation in physics and its solution provides the better way. The analog problem is the determination of the rotational ellipsoid and its main axes for a given distribution of point masses. In this case the free axis with the lowest rotational energy is the focus of interest. If the point mass distribution is given by  $P = \{\mathbf{p}_1, \dots, \mathbf{p}_n\}$ , the center of mass  $\mathbf{c}_m$  is obtained by

$$\mathbf{c}_m = \frac{1}{n} \sum_{i=0}^n \mathbf{p}_i. \quad (6.8)$$

The center of mass is the supporting point of the main axes. Transforming the point mass locations to the center of mass coordinate system reduces the problem to the estimation of the directional vector of an origin line. This vector is calculated by an eigenvector and eigenvalue analysis of a covariance matrix. Putting the transformed point locations columnwise in a matrix  $\mathbf{A}$ , the covariance matrix is  $\mathbf{C} = \mathbf{A}\mathbf{A}^T$ . Then, the directional vector is the eigenvector corresponding to the largest eigenvalue. It is

$$\mathbf{C} = \mathbf{A}\mathbf{A}^T \quad (6.9)$$

$$\mathbf{q}_i = (x_i^q, y_i^q, z_i^q)^T = \mathbf{p}_i - \mathbf{c}_m \quad (6.10)$$

$$\mathbf{A} = \begin{pmatrix} x_1^q & \dots & x_n^q \\ y_1^q & \dots & y_n^q \\ z_1^q & \dots & z_n^q \end{pmatrix}. \quad (6.11)$$

The advantage of this approach is, that is extendable to arbitrary dimensions. In the three-dimensional case, the covariance matrix is given by

$$\mathbf{C} = \sum_{i=0}^n \begin{pmatrix} (x_i^q)^2 & x_i^q \cdot y_i^q & x_i^q \cdot z_i^q \\ y_i^q \cdot x_i^q & (y_i^q)^2 & y_i^q \cdot z_i^q \\ z_i^q \cdot x_i^q & z_i^q \cdot y_i^q & (z_i^q)^2 \end{pmatrix}. \quad (6.12)$$

According to Fischer & Kaul (1990), the eigenvector-eigenvalue equation to be solved is:

$$\mathbf{C}\mathbf{x} = \lambda\mathbf{x}. \quad (6.13)$$

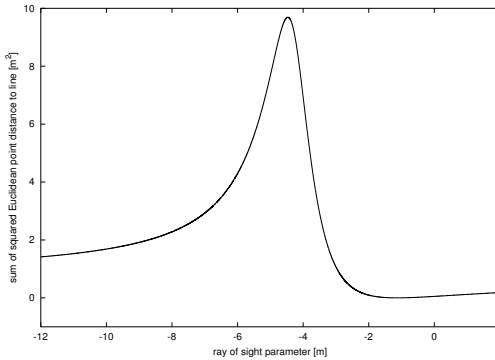


Figure 6.2: Sum of squared Euclidean distance from the line: the supporting point of the ray of reflection is constrained to the ray of sight. The directional vector is calculated using the covariance method. Subsequently, the sum of squared Euclidean distance is recorded for each line point. The line clearly exhibits a minimum, the location of the surface point and a maximum, the location where the retrieved directional vector is perpendicular to the actual one.

$\lambda \in \mathbb{R}$  is an eigenvalue of  $\mathbf{C}$ , if there is a vector  $\mathbf{x} \in \mathbb{R}^3$  unequal to zero, that solves the equation. Hence,  $\mathbf{x}$  is called an eigenvector. To obtain the eigenvalue of a matrix, the determinant equation needs to be solved:

$$\det(\mathbf{C} - \lambda \mathbf{I}) = 0. \quad (6.14)$$

As mentioned above, this general approach results in warped lines with stated misfits. Instead of determining a supporting point, the supporting point is set to be a point on the ray of sight along which the surface points must be located.

Analyzing Eq. 6.14 for the given case, yields an equation of the sixth order which has no closed-form solution. Therefore, a numerical approach is taken. The supporting point is moved along the ray of sight, the directional vector problem is solved for that respective point, and the sum of squared Euclidean distance is calculated for the data points to the obtained line, i.e. the reflected ray. To increase the robustness of the recovery, only those image locations with more than 15 three-dimensional data points assigned to them, are used.

Due to noise in the data and the possibility of erroneous three-dimensional point assignment to an image location, an outlier detection is necessary. To remove outliers from the data a  $3\sigma$ -test is performed, i.e. for the given line the distribution of the Euclidean distances is calculated and points with a distance larger than  $3\sigma$  of the distribution are removed. This is performed for each potential line recovery.



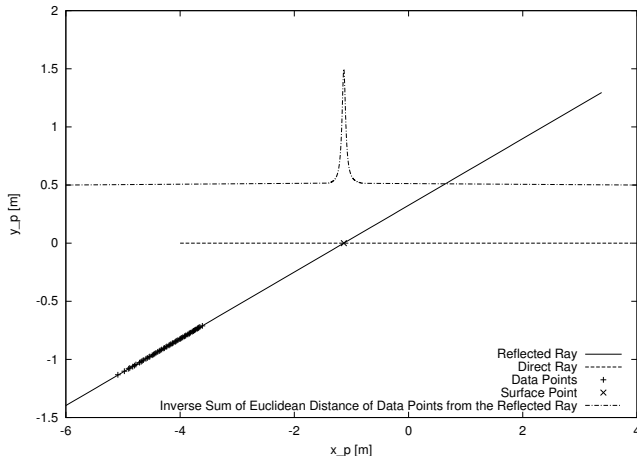


Figure 6.3: Recovery of the Surface Constraining the Point to the Ray of Sight: The line of sight (direct ray) and the recovered ray of reflection (reflected ray) determine a plane in space. This plane is shown with its coordinate system  $(x_p, y_p)$ . In it the data points, the recovered rays, and their intersection, the surface point, are displayed. In the upper part a shifted inverse sum of Euclidean distance function is depicted. It exhibits a clear maximum at the surface point location. The graph is based on real data.

The sums are recorded and are shown in Fig. 6.2. That function exhibits a minimum, the location of the surface point, and a maximum where the retrieved directional vector is perpendicular to the true one. However, the surface point determination is numerically more stable, if the inverse sum of Euclidean distance is evaluated. Fig. 6.3 presents such an evaluation. The line of sight (direct ray) and the recovered ray of reflection (reflected ray) determine a plane in space. The figure shows that plane. In it, the data points, the recovered rays, and their intersection, the surface point, are displayed. In the upper part a shifted inverse sum of Euclidean distance function is depicted. It exhibits a clear maximum at the surface point location.

With this approach, the point along the ray of sight, associated with the image location  $(u, v)$ , is extracted. For this point, the retrieved line minimizes the sum of squared Euclidean distance to the data points. It therefore yields the most likely surface points. A point which does not carry any ambiguities with respect to back-projection and the like mentioned above. By using the normalized directional vectors for the lines, the line parameter automatically yields the distance of the surface point location along the ray and therefore the height field representation (see Section 6.3).

## 6.5 Results

On the following pages several images present a surface reconstruction. To recover the surface, 41 image sequences, containing 150 images each for each camera, were recorded. In between the sequences a lateral movement of the car relative to the calibration rig is applied. During a longitudinal displacement 150 stereo images are taken. The images are processed according to the algorithm described above. The data points are collected to form a 'calibration object' for the surface recovery. The collected data points, 615000 altogether, are shown in Fig. 6.4. The evaluation statistics of the control points is as follows:

|                                     | Left Camera | Right Camera |
|-------------------------------------|-------------|--------------|
| Number of Images                    | 6150        | 6150         |
| Number of Available Control Points  | 615000      | 615000       |
| Number of Data Points               | 325409      | 308006       |
| Recovered Surface Points            | 1032        | 1075         |
| Average Number of Points per Line   | 8.27        | 8.56         |
| Removed Outliers                    | 13519       | 14281        |
| Max. Number of Points to a Line     | 57          | 41           |
| Number of Points below Minimum (15) | 95183       | 85231        |

Table 6.1: Surface Data Point Statistics

For the recovery of the surface for each camera, all 2107 determined surface points are taken into account by respectively transforming them in between the camera reference frames.

So far, the cameras are referenced as left and right. However, the experimental setup requires more flexibility. It is rather inconvenient, if the software needs to be adapted upon a change e.g. in the wiring of the cameras. Therefore, left and right are only tags for the cameras and the 'left' camera may be physically the right one and vice versa. Switched cameras lead to negative scene representation, i.e. negative distance ( $z$ ) values are obtained. Accordingly, negative line parameters are recovered in the surface estimation. This is the case here.

As mentioned above, the line determination automatically yields the height field representation of the surface, i.e. the distance of the surface along the ray of sight recorded for each image location. A display of the height field is shown in Fig. 6.5. The distance information can be converted to three-dimensional surface point locations. Fig. 6.6 displays the 2107 recovered surface points. Projective views from the side and the top onto the recovered surface points are depicted in Fig. 6.7.

To the three-dimensional surface points a parametric surface is approximated. This approximation, together with the data points, is contained in Fig. 6.8. A surface patch located in the middle of the hood, the area where the most interesting objects for the original application are located, is seen in Fig. 6.9 to left, a view from the top onto that patch in the same figure on the right.

An image of the distribution of surface points in the reflection region concludes the presentation of the results (see Fig. 6.10).

## 6.6 Discussion

In this chapter the second basic requirement, the reconstruction of the surface, is fulfilled. Rather than applying an available hardware-intensive method to recover the surface and the relative camera positions to it, a new method is developed on the basis of the camera-calibration techniques, presented in Chapter 5.

The camera calibration-technique cannot be used as developed but some adaption to the curved surfaces, and the therefore warped images, is necessary. However, the Partial Visibility Detection (PVD) proves to be flexible enough in its core, to solve the task.

In the subsequent surface point determination, a distance minimization technique with additional constraints is used. The solution to that problem is borrowed from experimental physics, i.e. the determination of the rotational ellipsoid, and its axis, of a point mass distribution. An alteration of that procedure, with respect to the supporting point of the line, provides a reliable method.

The surface geometry is retrieved for the experimental car available for this investigation. The final test for the accuracy of the recovered surface with this method is the application of it for image reconstruction and object detection (see Chapter 8).

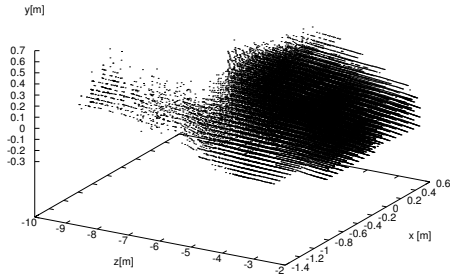


Figure 6.4: Distribution of the Control Points for the Surface Reconstruction: From the recorded images 615000 three-dimensional data points were reconstructed with the stereoscopic information.

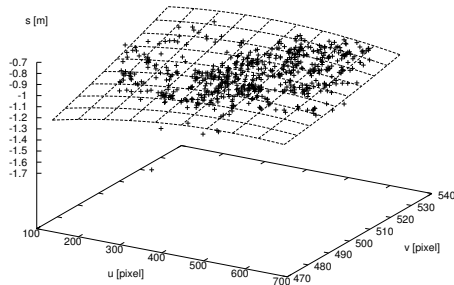


Figure 6.5: Scalar Representation of the Surface: For each image location  $(u, v)$  the distance of the surface is recorded. Here the data points for the left camera as well as an approximation of the data points by a scalar function is shown.

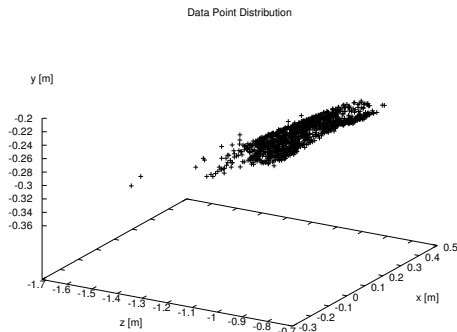


Figure 6.6: Three-dimensional Surface Point Distribution: 2107 surface points displayed in the reference frame of the left camera. The camera location relative to the surface is to the upper right.

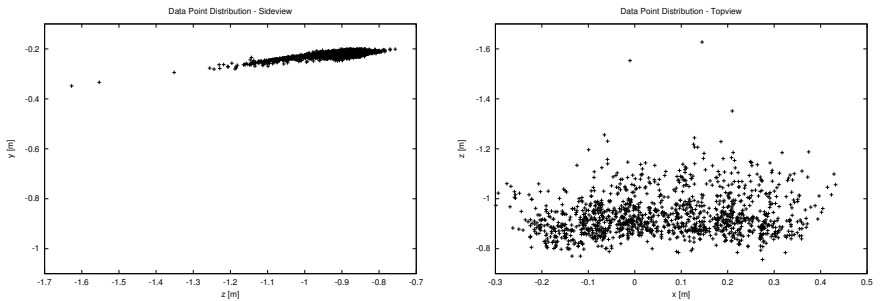


Figure 6.7: Sideview and Topview of the Surface: The three-dimensional data point distribution (see Fig. 6.6) is projected onto the  $z$ - $y$ -plane (sideview, left) and the  $x$ - $z$ -plane (topview, right).

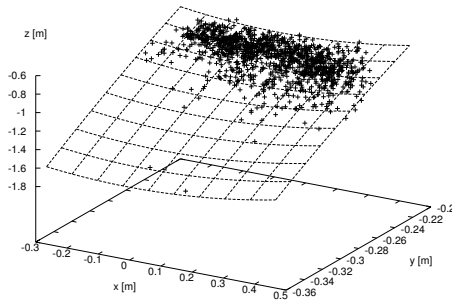


Figure 6.8: Surface Fit to the Data Points: A smooth surface is approximated to the 2107 three-dimensional surface points.

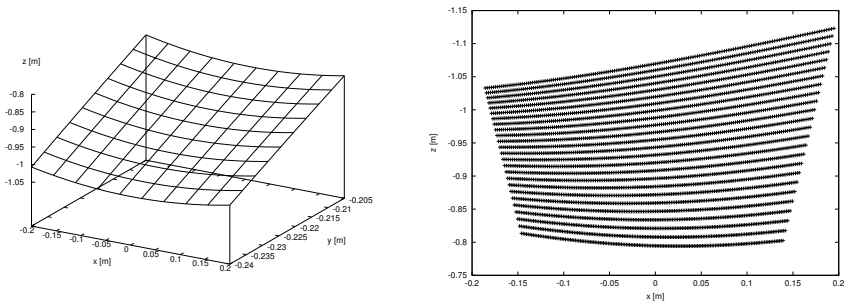


Figure 6.9: Recovered Surface Patch: A patch of the recovered surface, located in the middle of the hood, viewed from below as grid (left) and from the top (point distribution).

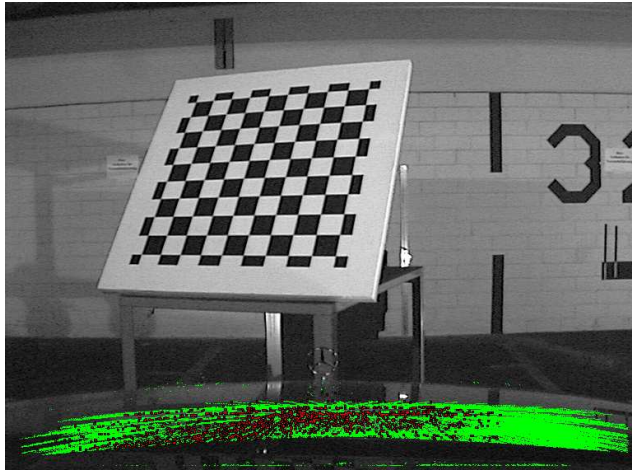


Figure 6.10: Distribution of the Recovered Surface Points within the Image: Valid locations, where a surface point is recovered, are marked green. Red locations were discarded in a multiple step process, either due to single three-dimensional object assignment or not achieving the minimum number of required object points.

# Non-classical Information Reconstruction

In this chapter, an investigation is conducted on how to retrieve the information contained in the image without reconstructing it in a classical sense, e.g. a pinhole reconstruction similar to a calibration effort. Following a short introduction to the matter a formula is derived which describes the general influence of the surface geometry on the epipolar geometry. After that, simulations are presented that verify the findings.

### 7.1 Introduction

In stereo image processing an epipolar line is the image of a ray of sight of one camera as seen by the other. The epipolar line starts with the image of the projection center of the first camera by the second one and ends by the image of the infinite direction. Along the epipolar line a feature of the first image is sought in the second one and the location, expressed by the disparity, determines the distance. In this context the distance-disparity-relationship is governed by a hyperbolic function.

Since the free-formed surface in combination with a perspective camera is not part of the class derived by Baker & Nayar (1999) it does not retain a singular viewpoint. It forms a non-classical camera. Due to this fact a somewhat different definition of an epipolar line or rather curve must be used, because one cannot expect such a system to keep a linear epipolar relationship and a hyperbolic distance law. Choosing an image location in direct view and moving an object along the ray of sight, associated with that image location, the point of reflection of the object draws a curve on the surface. The image of that surface curve is what will be called an epipolar curve. The curve will depend on the surface geometry as well as on the camera's relative view of the surface. Not only the surface geometry will decide the epipolar geometry: Another way to describe the formation of the epipolar curve is that it is the image of a direct ray of sight reflected on the surface. This image depends on the relative position.

With this curve one knows where to look for an object detected in direct view. The location along the curve and the knowledge of the surface geometry and therefore the related direction of reflection allows to retrieve the location of the object by inter-



secting lines in three-dimensional space. This is the basic idea of any stereo-vision system transformed to a free-formed catadioptric system.

In order to achieve this a formula is necessary which describes the relationship between the surface geometry and the camera's relative view of it. Another interesting information is whether there are certain cases where free-formed surfaces retain a linear epipolar relationship. This decreases the computational load tremendously.

## 7.2 Epipolar Geometry on Free-Formed Surfaces

To find a formula that describes the epipolar relationship on a reflective free-formed surface a variation approach is taken. With the definition of the epipolar curve in mind, a basic geometric setup is introduced that describes the reflective case. To initialize the variation the law of reflection is examined and linearized. Performing a variation, i.e. examining the infinitesimal change of the point of reflection on the surface upon a infinitesimal movement of the object along the ray of direct sight, will yield the result.

### 7.2.1 Geometric Setup

The basic geometric setup of a directly and reflected view of a scene is depicted Fig. 7.1. An object located at  $\mathbf{T}$  is observed by a camera at  $\mathbf{O}$  directly and via a reflection at  $\mathbf{R}$  indirectly. The reflection is governed by Fermat's principle which results in the angle of incident and reflection being equal. The angle of reflection is denoted by  $\gamma$ . The angles are measured with respect to the surface normal vector  $\mathbf{n}$  at the point of reflection  $\mathbf{R}$ . The direction of reflection, i.e. the vector from  $\mathbf{R}$  to  $\mathbf{T}$ , is called  $\mathbf{b}$  where as the direction in which the reflection is observed, i.e. the vector from  $\mathbf{O}$  to  $\mathbf{R}$ , is called  $\mathbf{e}$ . The points  $\mathbf{T}$ ,  $\mathbf{R}$ , and  $\mathbf{O}$  define a plane which is called the *optical plane*. By its definition the ray of direct sight, i.e. the line between  $\mathbf{O}$  and  $\mathbf{T}$ , is part of the optical plane.

To put this setup into more mathematical terms: The origin of the coordinate system is  $\mathbf{O}$  which is the location of the camera. The directions of view to the object location  $\mathbf{T}$  (ray of direct sight) and to the point of reflection  $\mathbf{R}$  (ray of reflected sight) are given by unit vectors. It is:

$$\mathbf{R} := \mathbf{OR} = R\mathbf{e} \quad (7.1)$$

$$\mathbf{T} := \mathbf{OT} = T\mathbf{t} \quad (7.2)$$

$$\mathbf{B} := \mathbf{T} - \mathbf{R} = B\mathbf{b} \quad (7.3)$$

where  $\mathbf{b}$  is a unit vector as well.

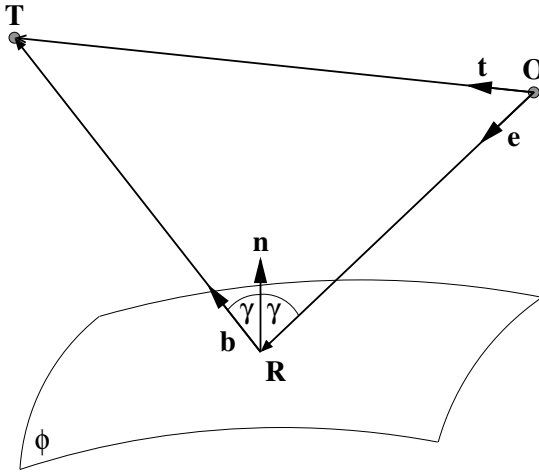


Figure 7.1: Basic Geometric Setup of the Reflection Case:  $\mathbf{T}$  denotes the object location,  $\mathbf{R}$  the point of reflection on the surface  $\phi$ ,  $\mathbf{O}$  the coordinate origin and the camera location,  $\mathbf{n}$  the normal vector at  $\phi(\mathbf{R})$ ,  $\gamma$  the angle of reflection and the angle of incident, and  $\mathbf{e}$  and  $\mathbf{b}$  the unit vectors in the respective directions.

The surface of reflection is described by an equation of the form

$$\phi(\mathbf{R}) = 0, \quad (7.4)$$

e.g.  $R = F(\mathbf{e})$ , a function which gives the distance of the surface from the camera location  $\mathbf{O}$  for a certain direction  $\mathbf{e}$ . At each point of the surface a normal vector  $\mathbf{n}$  can be defined as:

$$\mathbf{n} = \frac{\text{grad } \phi(\mathbf{R})}{\|\text{grad } \phi(\mathbf{R})\|}. \quad (7.5)$$

The optical plane together with the wanted information mentioned above is the starting point of the variation approach. By close examination of the optical plane the general nature of the epipolar curves is determined. In addition, the degenerate cases, where the curve is again a line, are revealed. Recalling the definition of an epipolar curve the object location  $\mathbf{T}$  is moved along the ray of sight. Obviously, the object location remains within the original optical plane as the ray of direct sight is always part of the optical plane. If the point of reflection  $\mathbf{R}$  remains within the original optical plane, when the object location moves along the ray of direct sight, then the epipolar curve is a line, namely the line of intersection between the image plane and the optical plane. However, if the point of reflection  $\mathbf{R}$  leaves the original

optical plane the epipolar relationship is a general curve and not a line. Whether this happens or not is an intrinsic property of the surface geometry and the relative camera position to it.

To examine the behavior of the point of reflection  $\mathbf{R}$  relative to the original optical plane depending on the surface geometry and the variation of the object location  $\mathbf{R}$  a closer look at the law of reflection is taken.

### 7.2.2 The Law of Reflection

According to Fermat's principle the provision for the reflection is that the path of light must be minimal. Additionally, the point of reflection must be within the reflective surface  $\phi$ , i.e.:

$$\delta(\|\mathbf{R}\| + \|\mathbf{B}\| - \lambda\phi(\mathbf{R})) = 0 \quad (7.6)$$

under variation of  $\mathbf{R}$  with  $\lambda$  being a Lagrange multiplier. Performing the variation yields:

$$\mathbf{e} - \mathbf{b} = \lambda \text{grad}\phi(\mathbf{R}). \quad (7.7)$$

The geometric content of this equation is the usual law of reflection, i.e. the angle of incident is equal to the angle of reflection, where the angles are measured to the normal vector of the surface. Four equations for  $\mathbf{R}$  and  $\lambda$  are set up based on the equation of the surface. Therefore, the point of reflection  $\mathbf{R}$  and the ray of sight  $\mathbf{e}$  for a given object location  $\mathbf{T}$  are calculated. The main unknown is  $\mathbf{e}$ . With it, the real image on an arbitrary surface or the virtual image observed by an eye or a camera at the origin  $\mathbf{O}$  is obtainable.

Instead of evaluating this non-linear equation system to analyze the influence of the surface geometry, an investigation is conducted how a small change, a variation, in the object location influences the location of the point of reflection. To do so, a linearization of the law of reflection around a given reflective situation is necessary. However, this linearization is described in a more convenient way if an index notation for the vectors is used.

### 7.2.3 Index Notation and Projection

The vector notation is rather bulky for the intended kind of calculation. Therefore, a vector is now represented by indices, i.e.  $\mathbf{T} = T^a$  or  $\mathbf{e} = e^a$ . Especially  $\mathbf{R} = R^a = Re^a$  is written. In addition, Einstein's sum convention is applied for scalar products. Einstein's sum convention states that within one side of an equation one summarizes quantities with an upper and a lower equal index:  $\sum_i a_i = a^i a_i$ . An upper index denotes a contravariant vector, a lower index a covariant vector. Contravariant and covariant vectors are related to each other through the metric tensor. Therefore:

$$\|\mathbf{e}\|^2 = 1 \quad \iff \quad e^a e_a = 1. \quad (7.8)$$

For the derivative the equivalence  $\partial_a = \frac{\partial}{\partial x^a}$  holds, such that

$$\text{grad } \phi = \partial_a \phi. \quad (7.9)$$

Using  $N^2 = \|\text{grad } \phi\|^2 = \partial_a \phi \partial^a \phi$ , the following representation for the law of reflection is obtained:

$$e^a - b^a = \lambda N n^a. \quad (7.10)$$

One determines  $\lambda$  by expanding the scalar product of this form of the law of reflection with itself:

$$\lambda^2 N^2 = 2 - 2e_a b^a = 2(1 + \cos(2\gamma)) = 4\cos^2(\gamma). \quad (7.11)$$

Due to orientation of the vectors and angles it is:

$$e^a - b^a = -2\cos(\gamma)n^a. \quad (7.12)$$

For later use, a projection  $p^a_b$  onto the plane perpendicular the normal vector is defined as:

$$p^a_b p^b_c = p^a_c, \quad p^a_b n^b = 0. \quad (7.13)$$

The same projection is described by:

$$p^a_b = \delta^a_b - n^a n_b. \quad (7.14)$$

By projecting the law of reflection onto the plane one sees that the projects of  $e^a$  and  $b^a$  coincide. Therefore, the projections onto the normal vector coincide as well. The angles between  $b^a$  respectively  $-e^a$  and  $n^a$  are equal. This is still the law of reflection.

#### 7.2.4 Linearized Law of Reflection

The law of reflection is now written as:

$$(e^a - b^a)p^a_b = 0. \quad (7.15)$$

Performing a variation of the object location  $\mathbf{T}$  yields as a result one of the vector  $\mathbf{B}$ . Obviously, a change in the direction of  $\mathbf{B}$  is extraneous. Therefore, only a variation of the direction of  $b^a$  yields consequences. This variation induces a change  $\delta R^a$  of the point of reflection and subsequently one in the direction of the reflection  $e^a$ , which is the desired result. The variation of the point of reflection brings about a variation in the direction of the normal vector which is dependent on the geometry of the reflective surface.

The variation of the law of reflection is:

$$(\delta e^a - \delta b^a)p^a_b + (e^a - b^a)\delta p^a_b = 0. \quad (7.16)$$

To evaluate this expression the determination of the variation of the projection  $p^a_b$  is necessary:

$$\delta p^a_b = -\delta n^b n_a - n^b \delta n_a. \quad (7.17)$$

The variation of the normal vector is according to surface theory given by the extrinsic curvature  $K_{ab}$ . It is a tensor field, defined on the surface, defining how the surface is warped in the three-dimensional space. In order to obtain the extrinsic curvature the surface needs to be twice continuously differentiable ( $C^2$ ). It is (disregarding a sign):

$$\delta n_a = K_{ab} \delta R^b \quad (7.18)$$

and subsequently (see Eq. 7.16):

$$(\delta e^a - \delta b^a) p^a_b + (e^a - b^a) n_a K^b_c \delta R^c = 0. \quad (7.19)$$

Applying the law of reflection together with the identity  $R^a = R e^a$  and its variation  $\delta R^a = R \delta e^a + \delta R e^a$ :

$$(\delta e^a - \delta b^a) p^a_b + 2 \cos(\gamma) K^b_c \delta R^c = 0 \quad (7.20)$$

is obtained. The linearization of the constraining condition  $\phi(\mathbf{R}) = 0$  gives  $\delta R^a \partial_a \phi = 0$ , respectively  $n_a \delta R^a = 0$ , i.e. the variation of the point of reflection is tangent to the surface, which is rather obvious. Therefore,  $\delta R^a$  has two independent components. Eq. 7.20 describes only two independent components because the normal component vanishes identically.

An orthonormal coordinate system to which from now on the vectors are related is introduced. The z-axis is chosen equal to the normal vector, i.e.  $z^a = n^a$ , the x-axis perpendicular to it, within the plane spanned by  $e^a$  and  $b^a$ . Finally, the y-axis is perpendicular such that  $x^a, y^a$ , and  $z^a$  build a right-handed normalized basis system. This implies:

$$x^a y_a = 0 \quad (7.21)$$

$$x^a x_a = 1 \quad (7.22)$$

$$\{x^a, y^a\} p^b_a = x^b, y^b \quad (7.23)$$

$$z^a p^b_a = 0, \quad (7.24)$$

and

$$e^a = -\cos(\gamma) z^a + \sin(\gamma) x^a, \quad (7.25)$$

$$b^a = \cos(\gamma) z^a + \sin(\gamma) x^a. \quad (7.26)$$

Accordingly, the variation of  $R^a$  is divided with respect to the same basis:

$$\delta R^a = R(p x^a + q y^2), \quad (7.27)$$

where the constraint  $\delta R^a n_a = 0$  is considered. The change of  $R^a$  consists of two parts: a change within the optical plane and a lateral one. This leads to the change in the length:

$$\frac{\delta R}{R} = \frac{1}{R^2} R_a \delta R^a = p \sin(\gamma) \quad (7.28)$$

and in the direction:

$$\begin{aligned} \delta e^a &= \frac{1}{R} (\delta R^a - \delta R e^a) \\ &= p \cos^2(\gamma) x^a + q y^a + p \cos(\gamma) \sin(\gamma). \end{aligned} \quad (7.29)$$

Similar considerations lead to a representation of the infinitesimal change of the object location. However, this is not the focus of this derivation, the concern is on the change of  $B^a$  that it induces. Again, the change is composed of two components. First, there is the possibility to change the length and keep the direction. Second, the length is kept and the direction altered. Related to the coordinate basis the following representation is obtained:

$$b^a = \sin(\gamma) x^2 + \cos(\gamma) z^a \quad (7.30)$$

and subsequently

$$\delta b^a = P (\cos(\gamma) x^a - \sin(\gamma) z^2) + Q y^a \quad (7.31)$$

and

$$\begin{aligned} \delta B^a &= (\delta B \cos(\gamma) + P \sin(\gamma)) z^a + \\ &(\delta B \sin(\gamma) - P \cos(\gamma)) x^a + Q y^a. \end{aligned} \quad (7.32)$$

$\delta B^a$  is a change in length of  $B^a$ , whereas  $P$  and  $Q$  induce changes in direction in such a way that  $P$  remains within the original optical plane and  $Q$  perpendicular to that plane. From this equation one easily calculates the infinitesimal change in the object location. As mentioned above, a variation of  $B^a$  along  $b^a$ , i.e. a mere change in length, is irrelevant.

The extrinsic curvature of the reflective surface has only a component perpendicular to  $z^a$ :

$$K_{ab} = K_{11} x_a x_b + 2K_{12} x_{(a} y_{b)} + K_{22} y_a y_b. \quad (7.33)$$

Due to the usage of a right-handed, orthogonal and normalized basis the metric tensor is represented by a unit matrix. Therefore:

$$K^b{}_c = K_{11} x^b x_c + K_{12} x^b y_c + K_{21} y^b x_c + K_{22} y^b y_c. \quad (7.34)$$

In addition, the extrinsic curvature of a surface is represented by a symmetric matrix, i.e.  $K_{12} = K_{21}$ : This is due to the fact that a continuous differentiable surface is assumed and that the curvature tensor is derived from the second fundamental form depending on the second derivatives of a function. According to the law of Schwarz (Bronstein & Semendjaev, 1985), if the derivatives do exist and if they are continuous then they are interchangeable. The interchangeability accounts for the symmetry in the matrix representation.

Incorporating these expressions in Eq. 7.20 the wanted result is obtained in a matrix representation:

$$\begin{pmatrix} P \\ Q \end{pmatrix} = \begin{pmatrix} 2RK_{11} + \cos(\gamma) & 2RK_{12} \\ -2RK_{12}\cos(\gamma) & -1 - 2RK_{22}\cos(\gamma) \end{pmatrix} \begin{pmatrix} p \\ q \end{pmatrix}. \quad (7.35)$$

It is discussed in the following section.

### 7.2.5 Results

The starting demand was to obtain a formula that governs the variation of the point of reflection with respect to the optical plane when the object location changes within the plane especially along the ray of direct sight which is always part of the optical plane. Examining Eq. 7.35, it is obvious that a movement of the object location  $\mathbf{T}$  along the reflected ray  $\mathbf{B}$  does not change the point of reflection  $\mathbf{R}$  at all. However, the matrix in Eq. 7.35 has off-diagonal elements. Due to this, a movement of the object location within the optical plane induces a lateral movement of the reflection point  $\mathbf{R}$  relative to plane. This is a consequence of the surface geometry.

Only in cases where the extrinsic curvature is described solely by diagonal elements in its matrix representation a longitudinal movement of the object location induces only a longitudinal movement of the reflection point. The extrinsic curvature has a diagonal matrix if the optical plane coincides with one of the main curvature directions of the surface at the reflection. One of these coincidences occurs looking along a symmetry-axis of the surface.

Apart from these general results a look should be taken at more specific computer-visualization related cases. Setting  $q = 0$ , i.e. a variation only within the optical plane, in Eq. 7.35 results in:

$$P = (2RK_{11} + \cos(\gamma)) \cdot p \quad (7.36)$$

$$Q = -2RK_{12}\cos(\gamma) \cdot p. \quad (7.37)$$

Eq. 7.37 again describes the lateral movement of the reflection point relative to the optical plane. To observe an epipolar line rather than a general integrated curve,  $Q$  has to be zero or at least negligible. This is the case, if  $\gamma$  is close  $90^\circ$ , i.e. a grazing angle of incident, and if  $K_{12}$ , the off-diagonal element in the curvature matrix, is

small. A small  $K_{12}$  implies an only slightly curved surface. In the special case of a reflected image in the hood of a car both provision mentioned above might apply: the hood is only slightly warped and the camera position behind the windshield provides a grazing incident. In general, the quantization of the image due to the retina, sets the mark for accepting a curve as line: If the deviation of the curve from the respective line is below the accuracy of the matching, which is somewhere between one half and one quarter of a pixel for a video-real-time system, the epipolar relationship may be regarded as a line. However, the disparity-distance-relationship along that line might not be described by a hyperbolic function, depending very much on the surface curvature as well.

### 7.3 Simulations

To verify whether the constraints discussed in Section 7.2.5 apply in the case of a hood of a car a simulation has been setup and evaluated. The simulation was performed by ray-tracing methods provided by the Advanced Systems Analysis Program (ASAP, see Section 4.2.2) software application. ASAP is a design tool for optical systems and is able to use CAD surfaces as part of an optical system and simulate light propagation in a physically correct way.

In the first section the setup is introduced. In the second one the results are presented.

#### 7.3.1 Simulation setup

The setup of the simulation is chosen according to the geometry setup of the theoretical examination (see Fig. 7.1) and is shown in Fig. 7.2. Considering the intended application, CAD data of the hood is used to represent the reflective surface. Not to restrict the generality of the simulation, a pinhole camera with a focal length of 10 mm is placed at an arbitrary point off the symmetry plane of the surface. Then, arbitrary rays of direct sight, determined by their image coordinates and the pinhole in 3D space, are chosen. A point object is moved along these rays to distances ranging from  $x = 0.6$  to 100.0 m. The points of reflection on the surface are calculated and the reflected rays are traced through the pinhole onto the image plane. The simulation results are image coordinates for different object distances related to a single image point in direct view.

#### 7.3.2 Results

Fig. 7.3 shows the result of the simulation. In the upper part of the image eleven points of direct view are marked *a* through *k*. In the lower part of the image the resulting reflection points corresponding to the object locations *a* through *k* and the fitted straight lines are shown. Considering a focal length of 10 mm and 120 pixels/mm the stan-



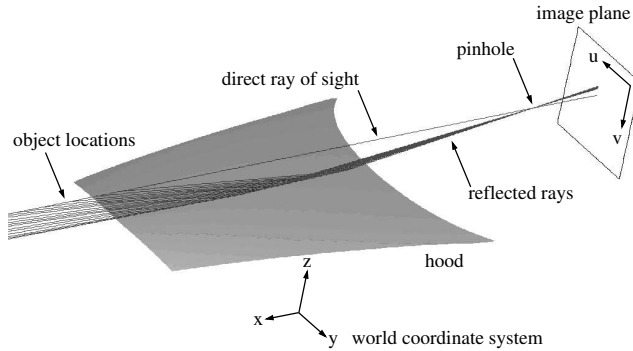


Figure 7.2: Setup of the Epipolar Geometry Simulation: A pinhole camera is placed off the symmetry plane of the hood. The hood is represented by its original CAD data. An arbitrary ray of sight off the optical axis of the camera is chosen. A point object is moved along that ray of direct sight to distances ranging from  $x = 0.6$  to  $100.0$  m. The rays from the object are reflected on the hood and traced back to the image plane where the point of incident is recorded. Results are shown in Fig. 7.3 and Fig. 7.4.

standard deviations of the points to the fitted lines range from 1.12 to 1.21 pixels. The epipoles of certain points (here arranged in a line) overlap even though the camera is not placed in the plane of symmetry of the surface but at an arbitrary position.

Fig. 7.4 shows the distance relationship along the epipole for the original points  $f$  and  $g$  (see Fig. 7.2). Since a disparity in the usual sense cannot be determined the line parameter  $t$  in pixels is used instead to display the result. For the line a geometric representation is assumed:  $\mathbf{x} = \mathbf{x}_0 + t \cdot \mathbf{r}$  with the supporting point  $\mathbf{x}_0$ , the line parameter  $t$  and the direction of the line  $\mathbf{r}$ .

In an ordinary stereo camera setup the epipolar distance relationship is a reciprocal function of the disparity along the linear epipole (see Eq. 2.2). The disparity-distance function depends on the baseline distance and the focal length of the equivalent pinhole replacements for the cameras. The relative angle of the optical axes in space, with special regard to esotropia, is a parameter of the function. However, it is possible to identify a general hyperbola that governs the distance relationship for each pixel and its respective epipole. This is based on the assumption that the obtained images are corrected to comply with a pinhole camera model. If a system contains e.g. astigmatic lenses and the images are not correct for lens distortions the determination of a single hyperbola for the disparity-distance function will not be successful.

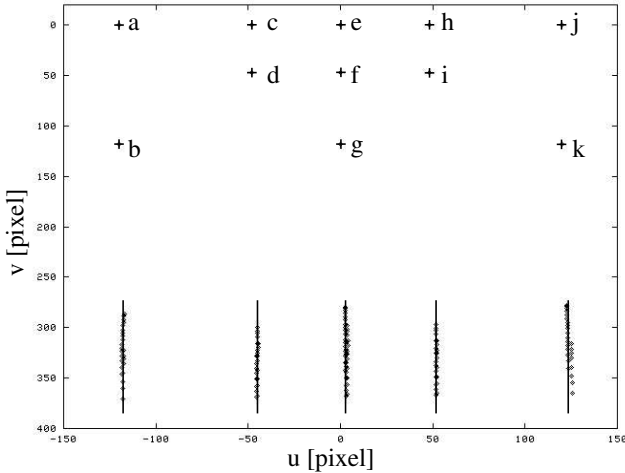


Figure 7.3: Results of the Epipolar Geometry Simulation: Eleven rays of sight, marked  $a$  through  $k$ , represented by their image coordinates in the upper part of the figure, are chosen. Reflected coordinates of the object point, that are moved along the direct ray of sight, are calculated. The lines are least-squares fits to the data points. The mean distance of the data points to the line ranges from 1.12 to 1.21 pixels with a standard deviation  $\sigma$  between 0.67 and 0.78 pixels. To calculate these values a focal length of 10 mm and 120 pixels/mm in the image plane are assumed.

The catadioptric system shares properties of the ordinary stereo setup but it also exhibits differences. Even though the rays are reflected on a free-formed surface the reciprocal relationship still yields. The resulting functions and standard deviations of the mean distance from the data points to the fitted hyperbola are:

$$\begin{aligned}
 d_f(t)[\text{m}] &= \frac{1}{-3.578 + 1.441 \cdot t} - 5.515 \cdot 10^{-2} \\
 \sigma(d_f) &= 1.86 \cdot 10^{-3} \\
 d_g(t)[\text{m}] &= \frac{1}{-6.881 + 2.954 \cdot t} + 7.743 \cdot 10^{-2} \\
 \sigma(d_g) &= 2.08 \cdot 10^{-3}
 \end{aligned}$$

Obviously, the distance functions along the epipoles do neither match nor overlap even though the points in direct view are in one column of the retina array. The functions are not shifted as their maximum intrinsic curvature differs.

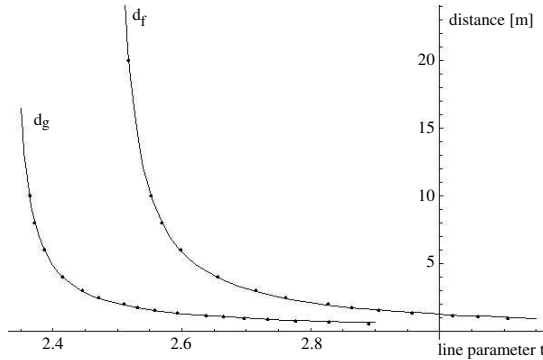


Figure 7.4: Distance Relationship along the Epipoles  $f$  and  $g$  (see Fig. 7.2). Like in an ordinary stereo camera setup the distance relationship is a reciprocal function. Instead of the disparity in a common stereo system the line parameter  $t$  in pixels along the epipole is used as ordinate. Although both direct view points share an epipole their distance functions along it do not overlap. They are neither shifted as the evaluation of the maximum intrinsic curvature shows.

This can also be understood from a more geometrical viewpoint: The retina column mentioned above and the pinhole determine a plane in space. This plane corresponds to the 'optical plane' introduced in Section 7.2. However, there the plane is defined by the object location, the camera location, and the point of reflection on the surface. The point associated with the direct view of the object location is part of the retina column related to in this argument. This plane intersects with the free-formed mirror surface. The shape of the intersection curve is a free form with varying slope. Different rays of direct sight correspond to certain heights at a given distance. These points are reflected at different locations on the free-formed surface mirror. Varying the distance of the objects alters the location of the reflection. But due to the difference in curvature around the original reflection point the variation of the reflection point itself and therefore its projection onto the image plane is different, i.e. the same variation in distance leads to different variations in the image location depending on the ray of direct sight and therefore on the curvature of the surface where the ray is being reflected.

Having intrinsically different hyperbolas for each pixel validates again the claim that a pinhole image cannot be reconstructed. Nevertheless, a distance calculation from the line parameter to sub-pixel precision is possible. On the other hand, the epipolar line and the respective distance relationship need to be determined for each pixel individually.

## **7.4 Discussion**

In the first section of this chapter differential geometry mathematics are applied to the problem of finding an epipolar curve in the case of a free-formed surface mirror serving in an optical system as source of information. A variation problem is set up and solved and the result provided a new formula which is able to govern epipolar geometry in general. In the presentation of the theoretical results constraints are elaborated which must be obeyed in order to treat the epipolar curve still as a line.

In the second part a simulation is conducted to examine whether the special case of a slightly curved hood of a car in combination with a quantized retina is within the limits.

Considering the results of the simulation the epipolar curve is treated as a line but the line and its intrinsic distance relationship need to be computed for each pixel separately. The computation is based on the calibration of the catadioptric system, i.e. the surface geometry and the relative positioning of the camera to it.

## CHAPTER 8

---

### **Comparison of Lines and Reconstruction Results**

This chapter elaborates the Comparison-of-Lines method, an in-image based reconstruction taking the geometry into account (see Section 4.2.3). It is described in some more detail in the first section. The second section is dedicated to the real-time implementation, one of the requirements stated in Section 2.5. There, two different approaches, a more conservative one and a new and progressive one are described. In the following section the deviations from the model assumptions, explained in Section 4.1, the restrictions in the real application, are examined. To conclude the chapter the results are presented and discussed.

#### ***8.1 Comparison of Lines***

In Chapter 4 two feasible reconstruction methods are developed. One is the Comparison-of-Lines method, the other the non-classical information extraction elaborated in Chapter 7. The non-classical method, comprehending each original image pixel as an individual camera, results generally in epipolar curves with their own distance-disparity function each for every direct view pixel in the original image. This is more memory consuming and slower compared to the method presented here. The Comparison-of-Lines method reconstructs an image in the classical sense. However, according to the results of the solution development (see Chapter 4) an error-free pinhole reconstruction is not feasible. Hence, the task is to keep the error as small as possible.

The idea of the Comparison-of-Lines method is presented in Section 4.2.3 and depicted in Fig. 4.11. It reduces the reconstruction to a search problem. The search problem is solved once and off-line and the result is implemented obeying the real-time constraint. In the following, the idea is shortly revisited and the solution presented.

The Comparison-of-Lines method chooses a point on the surface. Its tangential plane serves subsequently as a planar mirror. At the tangential point the direction of reflection of the free-formed surface and the tangential plane are identical. For the planar as well as the free-formed surface mirror the directions of reflection are com-

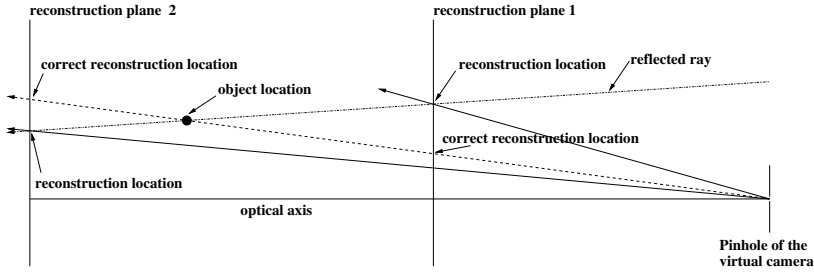


Figure 8.1: Errors in the Reconstructed Object Location: By choosing a virtual camera location in a non-singular viewpoint system some rays of reflection do not coincide with the rays of the virtual camera. Dependent on the choice of the reconstruction plane the error varies.

puted. The task in the search problem is to identify the three-dimensional reflection line off the free-formed surface that fits best to the reflection ray off a given planar mirror location. Once this problem is solved a relationship between the original image, i.e. the reflection on the hood, and the reconstructed image, i.e. the reflection on the virtual planar mirror, is given. This implies that the reconstruction problem is solved. The Comparison-of-Lines method carries two problems. The first one is that the reconstruction may lead to a sparse distribution of information in the reconstructed image. This is not acceptable if the images are to be used for further processing. The second problem is that errors are introduced into the reflection image. Around the point of the tangential plane the best results are to be expected while far away from that point the quality of the reconstruction will diminish. This is due to the increasing distances of the warped surface from the planar one. The principle error introduced by the choice of a reconstruction plane in the case of warped rays of sight is shown in Fig. 8.1.

This general error consideration leads to a choice of the tangential point location where the most interesting objects for the vehicle-following applications are expected. This is the central upper part of the hood, visible in the central lower region of the original image. In addition, in this surface region the curvature is small and the errors are reduced. In Chapter 6 the surface is reconstructed within the camera reference frame. For each image location  $(u, v)$  the parametric surface representation contains the three-dimensional surface point, its normal vector and therefore the direction of reflection. With the knowledge of the free-formed surface point and its normal vector the tangential plane is defined. The next step is a scanning of the planar mirror. For each image location in the reflected region of the original image the intersection point with the planar mirror is computed. This is done by solving a linear equation system

resulting from the comparison of the line and the plane equations. The intersection points in combination with the respective direction of reflection define a line in space. Considering the restriction of the reconstruction region to an area of small curvature, the free-formed surface point corresponding to the planar mirror reflection line cannot be located far away. Therefore, the search problem is solved by a straightforward implementation limited to a preset area around the coarse location. As measure for the line correspondence the distance of the lines (see Eq. 6.7) and the angle between the directions is taken. Using 7.5 mm lenses, the field of view is  $48^\circ$ . With a PAL-standard retina, the image is 768 pixels wide. Hence, the pixel-based angular resolution is  $0.0625^\circ$ . This is used as a threshold for the angular correspondence. If no match is obtained, the closest candidate is used to perform an interpolation with the direct neighbors to assign an original image location to the virtual mirror. Upon finishing the scanning process a relationship between the planar mirror and the original image region is established. Note that this is an inverse ray-tracing process, starting at the virtual image location. It guarantees a dense reconstructed image without a gray-value interpolation.

The calibration parameters for the virtual camera are either determined by the relative position of the planar mirror to the real camera or chosen freely. The external parameters, i.e. the rotation and the translation of the virtual camera are obtained by the reflection of the real camera point and the optical axis at the planar mirror. The quantization of the image and the focal length are free parameters. The projection of the corner points of the reconstruction region in the original image onto the planar mirror and the subsequent mapping of those points by the virtual camera with a certain focal length determines the corner points of the virtual image. The number of pixels in between is arbitrary and adapted to the resolution of the original image. The metric conversion in the camera-calibration method is defined by the number of pixels per mm on the retina. If this factor is retained for the virtual camera, its focal length and the quantization are correlated. The virtual camera performs a perfect pinhole mapping, i.e. there are no distortions parameters in the intrinsic set. Hence, with the knowledge of the optical axis, the principle point is determined as well. This sums up the constitutive parameters of the virtual camera.

## ***8.2 Real-time Implementation***

As mentioned in the solution requirements (see Section 2.5) the data contained in the reflected image must be provided and evaluated as fast as possible. The vehicle-following application is running in video real-time, i.e. 25 images per second. The grabbing of the half images requires about 12 ms. The rest is available for the evaluation of the image-based data and the derived information on an abstract level.

In this section different approaches to reconstruct the images are presented. The first subsection details a table-based implementation and an approximation solution each running on the central processing unit. The third implementation uses the graphics-card of the computer to reconstruct the image and is discussed in the second subsection. The results of both implementation schemes are presented in Section 8.4.

### ***8.2.1 Implementation on Common Computer Hardware***

In Section 8.1 the result is an assignment of image locations in the reconstructed image to locations in the original image. In order to obtain a dense reconstructed image without a time-intensive interpolation an inverse ray-tracing method is used. There are two methods to implement this mapping in a fast way. The first one is more memory consuming but faster than the second one.

In the first approach the mapping is stored according to the image locations in the reconstructed image. Pixel by pixel the image locations in the original image are expressed by its number in an integer format. On startup of the application the integer reconstruction table is read. The memory for the original image is allocated and subsequently used by the frame-grabber. The fastest way to access a memory location is to follow a pointer to it. Therefore, the integer reconstruction table is converted into a pointer array onto the original image. The storage of the pointer array is memory intensive. However, the reconstruction is performed just by stepping through the pixels of the reconstruction image. To each of these pixels the gray-value of the respective pointer is assigned. If the lens-distortion correction for the original image is represented in a similar table format, the consecutive table evaluation is performed off-line and the tables are merged into one.

The second method is to approximate the reconstruction by a closed-formed surface. This method is applied to the original image in the case of lens distortions. There, the function describing the lens distortions is approximated by a fourth-order polynomial in tiles. The tiles are computed iteratively. The criterion for the tile size is that the polynomial fits the original vector function with a sufficient precision. If not, the tile is split and new polynomial parameters are computed such that the precision is achieved. Instead of using the original image each pixel is referenced by the polynomial function in its tile. This method is much more memory efficient. Only the polynomial coefficients for the tiles are stored. However, each in-image operation requires the evaluation of a polynomial function. This is a higher computational load and hence slower. In this case, the lens-distortion polynomials with their tiles and the reconstruction polynomials with their tiles must be kept and evaluated separately.



### 8.2.2 Implementation on Specialized Graphics Hardware

The processing of digital images and the contained information is a time consuming effort. Images are memory intensive and the applicable algorithms are manifold and complex. The simultaneous application of algorithms such as image enhancement (color interpolation), image preprocessing (transformation to a birds-eye view), correction of image distortions, and information extraction (object or driving lane detection) are not feasible on computer hardware with the available clock rates. To apply redundant or supplementary methods to verify the results is prohibited by the real-time constraint.

So far the standard approach is to implement the algorithms such that the central processing unit (CPU) performs the tasks. The computational performance of the CPU is therefore often the limiting factor. Others are the available memory or the bus rate with which large amounts of data, e.g. images, are transferred. The development of the CPU during the last years provides the possibility to use ever more complex algorithms. Still, many applications are not as save as they could be due to the impossibility to apply at least several of the well understood but computationally intensive methods.

Today, computer graphics-cards serve not only for the display of image data or graphical user interfaces. Furthermore, they are hardware specialized on vector calculus and color mixing. These graphics-cards, e.g. 3D-accelerated cards with hardware-implemented OpenGL or DirectX standard, are developed mainly for entertainment (computer games). Nevertheless, the algorithms available on these cards can be used for image-processing as well.<sup>1</sup> In the following paragraphs the setup and functionality of such graphics processing units (GPU) and their application in image processing and reconstruction are described.

#### GPU - Setup and Functionality

The major difference in the setup of a GPU compared to a CPU is the pipeline-processing concept. The CPU consists of a control unit, an arithmetic-logical unit (ALU), and registers. These exist permanently and can be filled with arbitrary types of data. The prefetch queue anticipates the required data and moves it across the data bus into the cache memory. The data-bus interface provides the communication with the other hardware components such as the graphics-card, the hard-drive, and so on. The CPU is designed for sequential data processing.

The GPU components form a pipeline (see Fig. 8.2). The data is processed sequentially by the units (shader) of the GPU. CPU units, such as the prefetch queue or the ALU are not available. Especially the lack of an arithmetic unit requires the adaption of common image processing algorithms to graphics processing units.

---

<sup>1</sup>Patent pending.

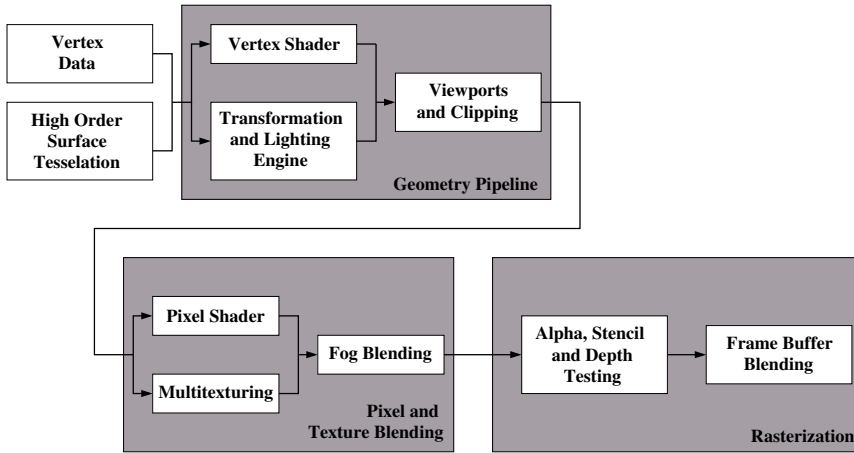


Figure 8.2: Pipeline Concept of a Graphics Processing Unit (GPU).

The vector analysis is performed by the geometry pipeline. It is capable of processing three-dimensional vectors. If the standard algorithms are not sufficient, additional methods are implemented in the vertex shader. Hence, a projection and various calculations, expressed e.g. by matrix operations, are assigned to each pixel individually. Using the geometry pipeline reconstructions and perspective views are obtainable. Almost any image processing that relies on vector calculation, such as stereo vision, is supported or replaced by a GPU implementation.

The color and texture mixing is performed in the texture- and pixel-shader units. These shaders move color values within the image according to a preset rule. They interpolate color values for regions of sparse information. Today, there are between 4 and 16 texture units, depending on the GPU model, such that multiple operations are applicable to the result of the geometry pipeline. In addition, the units may be arithmetically combined. Hence, it is possible to implement image processing that depends on the comparison of color values on a GPU.

The graphics-cards usually contain several memory units for images. In these units either several images are processed or different algorithms are applied to the same image. Within the pipeline, data elements may be abandoned, but so far new elements cannot be generated. Each data element is piped through by itself, without the consideration of neighboring pixel. This implies the following consequences:

- The lack of permanent registers prohibits the calculation of a sum of color values in an image region. Register are only allocated during the processing of

an element, e.g. a vector. Afterwards the register memory is freed. Nevertheless, color values are added by applying the interpolation routine. Minimizing the image by a factor of 2 in each dimension, summarizes and averages four adjacent pixels. With the knowledge of the factor 4 the sum is recovered.

- Due to the independency of the pixels, the sequence of processing is irrelevant. This is the fundamental prerequisite for the parallel and extremely fast computation.

The graphics-cards operate with Direct Memory Access (DMA). Therefore, images are directly moved from the frame-grabber or the hard-drive to the graphics-card memory. Usually, either the image is moved into the memory of the CPU or the algorithms operate on the frame-grabber image memory. The processed images from the graphics-card are moved to the CPU memory by DMA or displayed. If they are moved to the CPU memory, additional algorithms are applicable before the images are displayed or reprocessed by the GPU.

### **Image Processing and Reconstruction on a GPU**

This technology is very useful for a large variety of image processing methods. One of them is the reconstruction of distorted images. The lens distortion parameters, determined in the camera calibration, describe a rule of how to relocate the color values to obtain the reconstructed image. This is a task for the pixel shading unit.

The geometric reconstruction of an image is implemented in the vertex shaders of the geometry pipeline. Special camera geometries, such as a parabolic omnivision camera (see Fig. 3.3), apply a defined geometry model to the grabbed image to reconstruct the views. The vector model of the paraboloid is used in the GPU to produce an overlay of the image. In this case the intrinsic geometry of the image and the projection geometry are identical. The GPU assumes a virtual camera position to observe the three-dimensional structure. The image of that camera is what is actually displayed on the computer screen. Placing the virtual camera viewpoint at the singular viewpoint of the camera system in the model (see Section 3.2.3), the mapping performed by the GPU automatically reconstructs the image.

The vertex shaders are not restricted to a closed-formed geometry. To each pixel an arbitrary geometric point location can be assigned. That is how a free-formed surface is modeled. If the geometric surface data is not dense across the pixels different interpolation methods, such as triangulation or quadric approximation, are performed on a hardware basis. The usual approach to calculate a triangular mesh in the CPU to approximate the surface data becomes obsolete.

Due to its parallel structure, the GPU is extremely fast in image processing. The color-value interpolation which is usually only applicable to a small image region of

interest, is now computationally for free. The same is true for the correction of lens distortions. So far tables that contain the correct pixel locations or an approximated polynomial function to the distortion are used. Now, the calibration parameter set is transformed into a vector dislocation field which is stored in the GPU. The subsequent reconstruction is performed in real-time. Tests with a Nvidia GeForce3 GPU proved the applicability. The images are processed by the GPU at a rate of up to 30 frames per second, more than the camera can provide. Within the cycle the original image is corrected for lens distortions, mapped onto the geometry, is reconstructed, interpolated, and written back to the CPU memory. So far, these steps are hardly performed by the CPU in a cycle-time of the camera. Now, the methods are applied without computational load on the CPU, leaving it available for other sophisticated information evaluation.

In contrast to the implementation on the common computer hardware, the GPU method directly uses the surface data recovered in Chapter 6. The GPU projects the original image region onto the surface geometry. The virtual camera location retrieved with the Comparison-of-Lines method is used to define the viewpoint for the GPU mapping. The lens-distortion correction is implemented as two-dimensional vector field in the pixel shader unit. It is performed ahead of the projection onto the surface geometry. The final step is an interpolation of the reconstructed image.

### 8.3 *Experimental Restrictions*

In this section restrictions to the usability of car body parts as mirrors are investigated. The focus is on the relevancy to the image processing. More general and theoretical phenomena are not included.

The first effect that may influence the application is a deformation of the hood under driving conditions. The second one is caused by vibrations. The cameras are mounted to the inner side of the windshield. If the cameras are exhibited to a torsional force relative to the hood of the car, the obtained calibration as well as the surface reconstruction are insufficient for the driving application. Another limitation is the soilage of the hood. If the image is diminished by a soiled hood the processing will not be successful. The last restriction investigated is the influence of the varnish on the reflection. However, a determination of the diffuse and the specular component is not the scope here. According to Pérard (2000) the diffuse component reduces the specular intensity by about 30%. In the paragraph about the varnish of the hood, its influence on the gray-value histogram and the subsequent image processing is examined. According to Otha (1994), mirrors do not exhibit a chromatic aberration by themselves. Therefore, the image is not additionally blurred and only the varnish accounts for a change in the gray-value distribution.

### 8.3.1 Deformation of the Hood

One effect which might influence the functionality of the application is a deformation of the hood under driving conditions. In principle two deformations occur. One is due to heat, e.g. when the sun is shining onto a dark varnish of the hood. The other one is caused by air pressure variations related to the driving velocity.

A measurement for the quality of a car is the gap size of its body parts. Therefore, gap sizes are designed to be as small as possible under basically all circumstances. One of these cases is the expansion and contraction of body parts in heat and coldness. For the aluminum hood of a S-class Mercedes, the deformation due to temperature variations is so small that the design department refrained from simulations to quantify the change of the gaps relative to the fenders. Hence, the influence of the temperature deformation with respect to the vision application is considered negligible.

To quantify the air pressure deformation a simulation was set up in cooperation with T. Winkler from the car design and simulation department (EP/CSB). According to the real driving conditions, an air flow along undisturbed streamlines across the upper side of the hood is assumed. To obtain the maximum effect the velocity is set to 250 km/h. In the assembled state the distance of the hood to the engine body is less than 2 cm. In addition, the engine compartment is tightened with respect to turbulences from the air flow generated in between the lower car side and the road. Hence, the air flow velocity underneath the hood is set to zero. The design and simulation tool, built by the EP/CSB department, is usually used to simulate the air flow across a newly designed vehicle and calculate its air resistance  $c_w$ . The tool operates on the CAD design information of the vehicles and therefore the hood is represented by its construction data. Due to the required size, the results are presented in the appendix (see Fig. B.1 and Fig. B.2). The maximum deformation occurs in the central front part of the hood with an amplitude of about 2.2 mm. In the upper central part, where the interesting objects of the application are reflected, the amplitude is about 0.3 mm. According to Bernoulli's theorem the effect is influenced by the velocity to the square. The intended velocity range of the application is up to 50 km/h. Compared to the simulation velocity of 250 km/h the deformation is only about 4% of the amplitudes mentioned above. Considering the accuracy of the point location extraction in the image processing, a deformation in order of  $10^{-2}$  mm is negligible.

### 8.3.2 Vibrations

One of the results of the reconstruction method development (see Chapter 4) is that the relative position of the camera to the surface influences the reconstruction. A change in the camera position during the run-time of the application reduces the reliability of the system. One effect that causes a variation in the relative camera position



Figure 8.3: Setup to Determine the Influence of Vibrations: Patterns are attached to the wall and to the hood. Note that the pattern on the hood are not reflections but attached markers.

is a twisting force on the car body due to vibrations. To investigate the influence correlation patterns are attached to the hood of the car and to a wall (see Fig. 8.3). First a reference image sequence is taken with the car standing in front of the wall. The car engine is turned off during the recording of the reference sequence. The location of the reference points is determined by averaging the extracted locations. Subsequently a measurement sequence is recorded. This sequence consists of 400 half frame images, to exclude interlaced-frame-mode implications on the result. The video rate is 25 images per second, i.e. 40 ms in between the frames. After about 9 s the engine of the car is started and subsequently driven without a gear through the r.p.m.-range. For each of these images the pattern location is extracted and compared to the reference location. To evaluate the results, the distance of the point locations to the reference is represented in polar coordinates, i.e. a distance to the reference point and an angle to an arbitrary but predefined reference direction.

Fig. 8.4 shows the point extraction in direct view of the two cameras. During the first 9 s the graphs exhibit the noise in the point extraction. The polar distance error is less than 0.1 pixel so that the absolute maximum error is less than 0.2 pixels. The angular distribution for the cameras is random and differs from the left to the right one. The vibration caused by starting the engine is clearly visible and the movement up and down the r.p.m.-range as well. It is obvious that the whole stereo-camera system moves relative to the pattern on the wall, because the displacement of the pattern in the image relative to the reference is synchronous for both cameras. This

is true for the distance as well as the angle. Only small variations related to noise are observable. These results confirm as well that the calibration of the stereo-camera system obtained on a static scene is applicable under driving conditions.

Fig. 8.5 shows the point movement of the wall pattern and the hood pattern in the left camera. Hence, the graph titled 'Wall Reference' is identical to the graph of the left camera in Fig. 8.4. The other graph in Fig. 8.5 shows the movement of the hood pattern. It exhibits a slightly higher noise in the point extraction. This is due to the relative inclination of the pattern attached to the hood. When the engine is started a twist of the hood pattern relative to the wall pattern is observable in the distance as well as in the angular distribution. However, shortly after the engine start the twisting force fades away and the background noise prevails again. Hence, except for very high torsion forces such as the start of a large engine, the surface is rigidly mounted to the camera system.

The general resume of this investigation is that vibrational forces are negligible for camera systems mounted in a car. If the car is vibrating, it moves as a whole. Torsional changes of the body parts are well below the achievable resolution of the image processing. This is true for dioptric as well as catadioptric systems mounted into the car.

### **8.3.3 Soiled Surfaces**

A restriction to the applicability of reflected images on a car body part is soilage. The influence of it has been evaluated on an empirical base. The image shown in Fig. 1.2 is recorded with a car that had not been washed for at least four weeks. Nevertheless, the image quality in the lower region is acceptable. As mentioned in Section 2.2.3, the aperture is set such that objects at a distance larger than 2 m are imaged sharply. Accordingly, raindrops or other residues on the windshield only blur the image. The same applies to the soilage on the hood. The camera is set to infinity and the objects reflected in the hood, rather than the surface itself, are imaged.

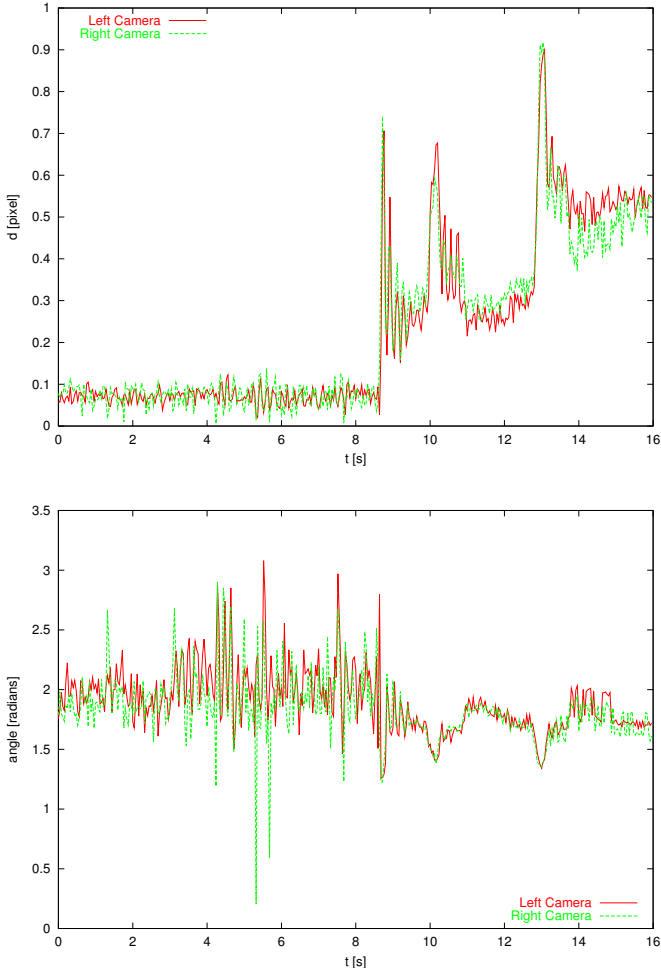


Figure 8.4: Vibrations of the Camera System to a Fixed Point on a Wall: The graphs show the distribution of the point locations with respect to a reference frame. The movement is represented in polar coordinates. The distance from the reference point is shown in the upper graph, the angle to a reference direction in the lower graph. After about 9 s the engine is started and subsequently driven through the r.p.m.-range.



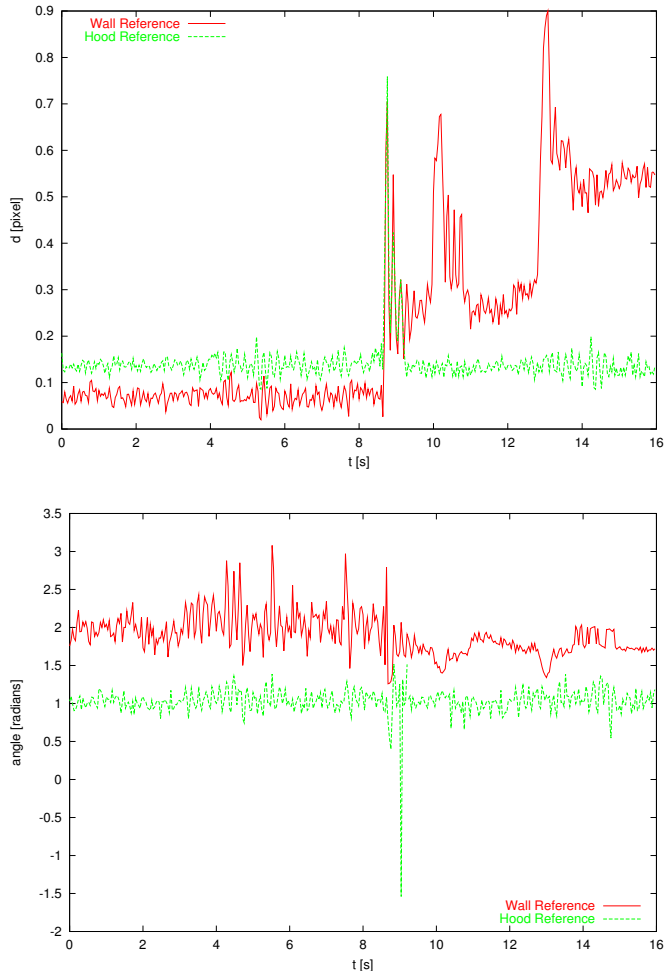


Figure 8.5: Vibrations of the Left Camera to an External and to a Hood Point: The graphs show the distribution of the point locations with respect to a reference frame. The movement is represented in polar coordinates. The distance from the reference point is shown in the upper graph, the angle to a reference direction in the lower graph. After about 9 s the engine is started and subsequently driven through the r.p.m.-range.

### 8.3.4 Varnished Surfaces

Most matching algorithms are based on the assumption that the gray-values of an object in the images are at least closely related. Especially those, that are not mean-free, achieve the best matches if the values are identical. A correlation method, such as the mean-free cross-correlation (see Section 5.6.1), that corrects the image region with its mean gray-value is less susceptible. However, the reflection on the varnished surface changes the gray-value distribution. If the two images reflected on the hood are taken for the matching process it does not matter. However, if the process is used in the fall-back mode the direct view and the reflected view from one image are matched against each other. In this case the difference in intensity might be too large for the algorithm to work properly. This has already been discovered in the surface recovery method. There, a histogram spreading is performed in order to reliably match the binary chessboard template to the reflection in the hood (see Section 6.4.1).

The left side of Fig. 8.6 shows an image of a black and white calibration wall and its reflection in the hood of the car. On the right hand side of that figure the intensity distribution is shown. In that histogram, and similarly in those shown in Fig. 8.7 and 8.8, the value range is cut off at a lower and an upper limit. This is due to the digitization of the image by the frame-grabber and without implication on the conclusions. In Fig. 8.6 there is a marked region in direct sight and its correspondence in the reflection. Fig. 8.7 shows the direct view region and its histogram, Fig. 8.8 the same for the reflection. To obtain comparable results the reflection region has been resized to match the direct sight. The comparison of the histograms shows a displacement of the maxima locations and a diminishing of the maximum height in the reflection histogram. The combination of the direct- and the reflected-sight histogram explains the double peak in the lower intensity region of the histogram in Fig. 8.6.

The basis of the intensity correction is the comparison of a direct view and a reflected view of a gray-value wedge.

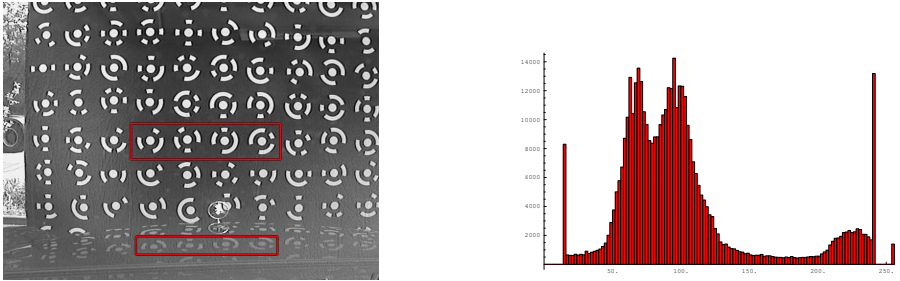


Figure 8.6: Direct View and Intensity Histogram. Left: An image of a calibration wall. In it two corresponding regions are marked. Right: The histogram of the intensities in the whole image.

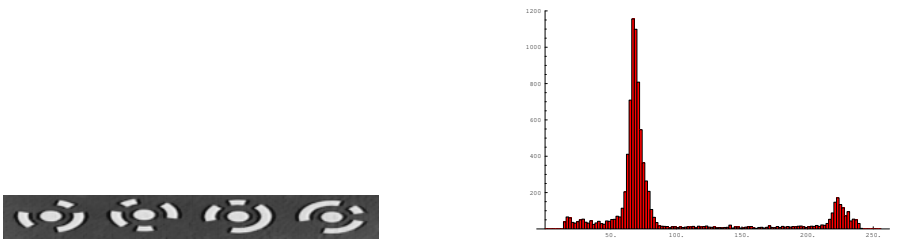


Figure 8.7: Direct View Region and its Histogram. Left: The cropped and enlarged direct view region from Fig. 8.6. Right: The histogram of the intensities in the direct view region.

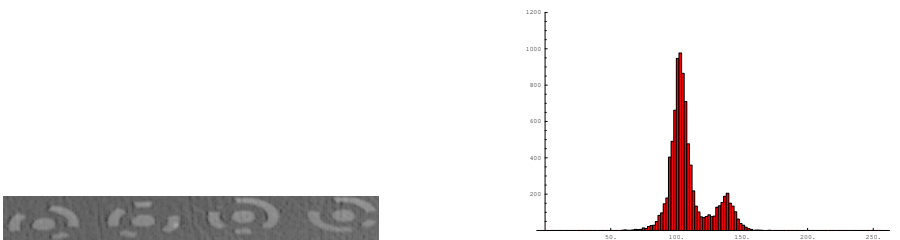


Figure 8.8: Reflected View Region and its Histogram. Left: The reflected view region from Fig. 8.6 is cropped and enlarged to the same size as the direct view region (see Fig. 8.7). This is done to achieve a normalization in the histograms of the direct and the reflected view regions. Right: The histogram of the intensities in the reflected view region.



Figure 8.9: A Typical Traffic Scene in a Vehicle Following Application. The box marks the region of interest which is reconstructed. The cross is the location of the tangential point chosen for the Comparison-of-Lines method.

## 8.4 Results

In this section the results of the reconstruction method is presented. Fig. 8.9 shows a typical traffic scene in a vehicle-following application. A leading car at a certain distance is observed directly and in the reflection on the hood. The region of reconstruction is marked as a box in the lower part of the image. This box as well as the tangential surface point are determined by visual inspection of an image in the setup phase of the reconstruction. Considering this information the Comparison-of-Lines method is applied. It results in the following transformation from the virtual camera coordinate frame to the original camera coordinate frame:

$$\mathbf{x}_{\text{orig}} = \begin{pmatrix} 1.00015947 & -0.01789754 & 0.00000000 & 0.00284484 \\ 0.00011452 & 0.99357952 & 0.06549896 & -0.31724867 \\ 0.00116478 & -0.13072821 & 1.00000000 & 0.02077946 \\ 0 & 0 & 0 & 1 \end{pmatrix} \mathbf{x}_{\text{rc}} \quad (8.1)$$

where  $\mathbf{x}_{\text{orig}}$  denotes a vector in the original camera coordinates and  $\mathbf{x}_{\text{rc}}$  one in the reconstructed coordinate frame. The representation is in homogenous coordinates (see Section 5.3.3). Therefore, the upper left  $3 \times 3$  elements of the transformation matrix contains the rotational information, whereas the last column is the translation vector. Obviously, the system is only very slightly esotropic. The virtual camera location is about 32 cm below the real camera location. This is still within passenger cabin and not well below the hood as one might expect. Accordingly, the hope to achieve

a larger baseline, which is associated with a higher precision in the stereoscopic triangulation, is not fulfilled. The location of the virtual camera implies as well that the tangential plane at the cross-marked point in Fig. 8.9 exhibits a large inclination relative to the original camera. The planar mirror surface divides the translation line between the original and the virtual camera in halves. Hence, the planar mirror is located at a distance of about 16 cm below the original camera. A reconstruction point in a region of larger curvature, i.e. further away from the original camera, diminishes the virtual baseline. The virtual baseline is almost identical to the baseline of the physical stereo-camera system. Therefore, an increased measurement accuracy in the triangulation cannot be achieved. On the other hand, the decreased matching reliability caused by a large baseline distance is not of concern.

The number of pixels per mm  $k_u$  for the virtual camera is set to the value of the physical camera. As mentioned above, this correlates the focal length of the virtual camera and the size of the reconstruction image. To reconstruct the image a subsampling rate of 2.5 in  $u$ -direction and 2.8 in  $v$ -direction is used. There are two reasons to do so. First, the subsampling in general adapts the resolution of the reconstructed image to the resolution of the reconstruction image. Second, the different subsampling rates account for the compression of the reflection region in the  $v$ -direction. Hence, the aspect ratio of an object is restored in the reconstructed image. Tab. 8.1 summarizes the intrinsic parameters:

|            |            | Original Camera | Virtual Camera |
|------------|------------|-----------------|----------------|
| $k_u$      | [pixel/mm] | 120.0           | 120.0          |
| $\alpha_u$ | [pixel]    | 897.5           | 2243.8         |
| $\alpha_v$ | [pixel]    | 890.3           | 2492.8         |
| $n_u$      | [pixel]    | 768             | 802            |
| $n_v$      | [pixel]    | 568             | 202            |
| $c_u$      | [pixel]    | 386.4           | 390.0          |
| $c_v$      | [pixel]    | 296.5           | 521.2          |

Table 8.1: Intrinsic Camera Parameters: Original and Virtual

$k_u$  is the number of pixels per mm in  $u$ -direction on the retina,  $\alpha_u$  and  $\alpha_v$  is the focal length in  $u$ - and  $v$ -direction respectively,  $n_u$  and  $n_v$  is the image size, and  $c_u$  and  $c_v$  denote the principle-point location. Note that the virtual camera's principle point is not within the image. This is due to the geometric setup and the region of interest within the reconstructed image.

The planar mirror only approximates the reflection directions of the free-formed surface mirror. Fig. 8.10 shows the projection of the deviation for the reconstruction box in Fig. 8.9. The coordinate system is centered on the tangential point, marked

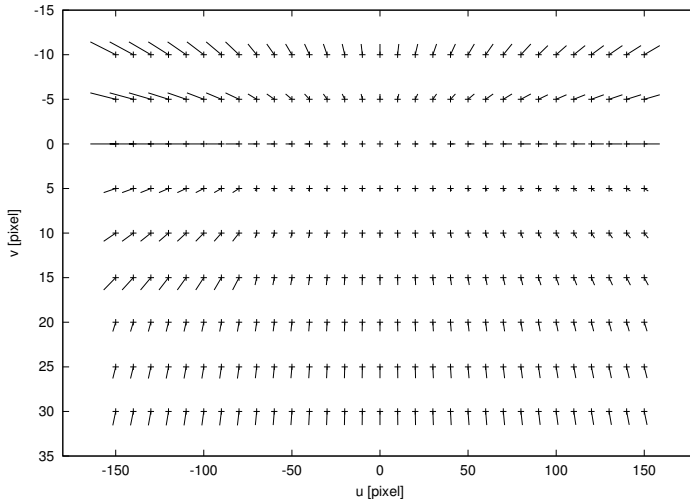


Figure 8.10: Error Distribution of the Comparison-of-Lines Method: The planar mirror only approximates the reflection directions of the free-formed surface mirror. The graph shows the projection of the deviation for the reconstruction box in Fig. 8.9. The coordinate system is centered on the tangential point (marked with a cross in Fig. 8.9). The box is depicted as viewed in Fig. 8.9. The size of the box is  $320 \times 50$  pixel. The length of the line encodes the angular difference. The maximum difference, located at the upper left side, is  $0.061^\circ$ .

with a cross in Fig. 8.9. The box is shown as viewed in Fig. 8.9. The length of the line encodes the angular difference. The maximum difference is  $0.061^\circ$ . The camera and the tangential point are located on the right side of the symmetry axis of the hood. Therefore, the errors are larger on the left side of the reconstruction. The curvature of the hood in  $v$ -direction is larger than in  $u$ -direction. Hence, the plane approximates the  $u$ -direction reflection component better than the one in  $v$ -direction. Close to the tangential point the approximation of the reflection direction is best.

With the intrinsic and extrinsic parameter sets and the surface geometry the image reconstruction is feasible. Fig. 8.11 shows the reconstruction of the reflection region by the CPU implemented method. On an Intel Pentium II with 400 MHz, the image processing host in the car, the image reconstruction takes 3.2 ms. There, the reconstruction uses a merged lens-distortion and reconstruction table. If an additional histogram spreading is performed the reconstruction time rises to 24.6 ms per image. Using a predefined histogram rather than extracting the histogram for each image, the

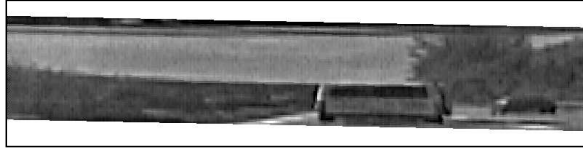


Figure 8.11: A CPU Image Reconstruction: The central part of the hood (see Fig. 8.9) is reconstructed with the CPU by the table-based implementation of the Comparison-of-Lines method. To avoid a sparse distribution of the pixel data in the image, an inverse ray-tracing method is used. The desirable interpolation to enhance the image cannot be applied due to the limitation in computation time. Compare this result with the GPU reconstruction in Fig. 8.13.

reconstruction time is 12.1 ms. Using an Intel Pentium 4 with 2 GHz, the next generation of vehicle hardware, the reconstruction times are 1.2 ms, 18.1 ms, and 8.9 ms, respectively. Considering the time limitations for the vehicle-following application, the histogram spreading is only applicable to parts of the reconstruction image. The regions are chosen depending on the object location in direct sight. The reconstruction of the region of reflection is shown in Fig. 8.11. An image interpolation is not applicable due to the time limits.

As described in Section 2.2.1, the image processing hardware consists of an Intel Pentium II with 400 MHz clock rate, 512 MB memory, and an ordinary graphics-card for the X-Window display of the Linux-OS. It does not contain an up-to-date graphics processing unit (GPU). Therefore, in the car only the common hardware solution is implemented.

The GPU reconstruction method is implemented in a laboratory setup with previously recorded images on the next generation of vehicle hardware. This is an Intel Pentium 4 processor with 2 GHz clock-rate, 1 GB memory, and a Nvidia GeForce3 GPU. Fig. 8.12 shows the GPU mesh representation of the surface data as viewed from the original camera. Fig. 8.13 depicts the reconstructed image on the basis of the Comparison-of-Lines method. For this image the original image was first corrected for lens-distortions, then projected onto the surface and subsequently mapped to the virtual image plane. Finally, an interpolation is performed. The image moved directly from the camera to the graphics-card memory. After the processing the image is moved by Direct Memory Access (DMA) to the CPU memory for further processing. This is all being done without computational burden on the central processing unit. The time consuming steps are the transfers of the images by DMA from the camera to the graphics-card and on to the CPU memory. Nevertheless, the reconstructed image is available for evaluation less than 1 ms after the grabbing.

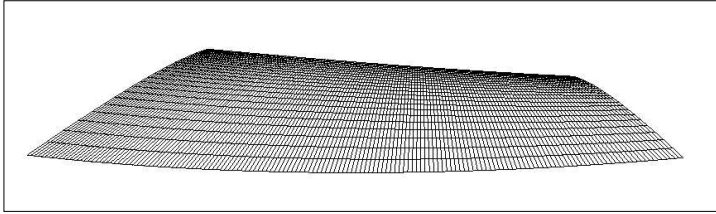


Figure 8.12: The Surface as Viewed by the Camera in Fig. 8.9. The quadric mesh is produced by the GPU on the basis of the surface reconstruction data obtained in Chapter 6

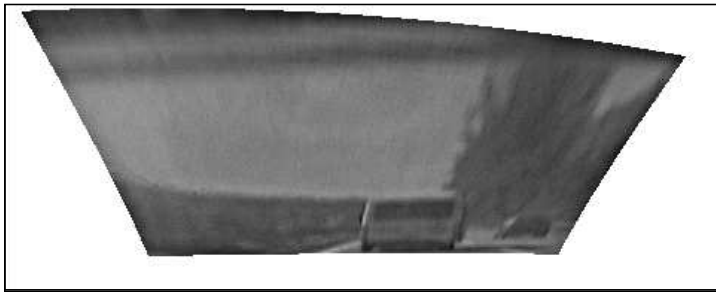


Figure 8.13: A GPU Image Reconstruction: The central part of the hood (see Fig. 8.9) is reconstructed by the GPU using the surface mesh model shown in Fig. 8.12. The camera location is determined by the Comparison-of-Lines method. The image is automatically interpolated by the GPU. Compare this result with the CPU reconstruction in Fig. 8.11.



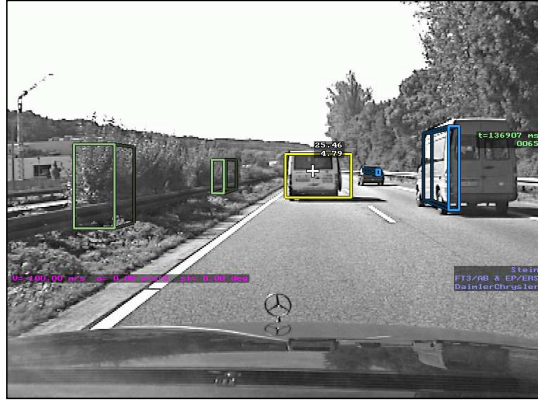


Figure 8.14: Distance Evaluation with the Stereo-Vision System: The traffic scene is evaluated by the stereo-vision application as described in Section 2.2. The leading car is observed at a distance of 25.46 m.

### Coarse Distance Determination

With the calibration information, the reflection images, and the original images form a stereo-vision system. Hence, the final test of the applicability is a distance measurement using a reconstructed image. Due to the fact that the surface recovery and the subsequent calculations are based on the stereo-camera system the result of this distance determination must be compared to the distance recovered by the stereo-image processing. The evaluation result of Fig. 8.9 is shown in Fig. 8.14. There, the leading car is observed at a distance of 25.46 m. Considering the parameters of the stereo-camera system this corresponds to a disparity of 7.98 pixels.

The basis of the distance determination are the original image, the reconstructed image, the system parameters, and the transformation matrix in between the systems (see Fig. 8.9, Fig. 8.11, Tab. 8.1, and Eq. 8.1, respectively). To calculate the distance of the leading car, the lower left corner of its back window is chosen. Its image location is (410, 239) in the original image and (427, 128) in the reconstructed image. Using the system parameters in Tab. 8.1 and the transformation matrix in Eq. 8.1, the image locations result in a disparity of 132.1 pixels. This result is very different from the result obtained in the original images. The pixel density on the retinas  $k_u$  is identical and the baseline distances of the systems are similar. However, note that the focal lengths of the original camera and the virtual camera are very different, as well as the image sizes and the principle point locations. Transforming the disparity obtained from the catadioptric stereo-vision system into a metric distance, one ob-

tains a distance of 26.28 m. This is very similar to the result of the dioptric stereo system, and proves the usability of the free-formed surface mirror as valuable source of information.

## 8.5 Discussion

In this chapter the experimental implementation is described. The chosen method is the Comparison-of-Lines algorithm. It is faster and less memory consuming than the non-classical information extraction elaborated in Chapter 7.

The implementation is performed in different ways. The first and second method are classical image processing approaches using the central processing unit. The first one uses an off-line calculated table, for the reconstruction as well as for the correction of the lens-distortions. These tables are merged to additionally speed up the process. The second method approximates the function for the reconstruction and the lens-distortions. In order to achieve a sufficient precision, the approximative polynomials are only defined in regions of the image. These regional splits may differ for the lens-distortion and the reconstruction. This method is more memory efficient because only the polynomial coefficients and the tiles are stored. On the other hand, the consecutive evaluation of the polynomials, which is performed for each pixel operation, is a higher load on the CPU. The third method performs the reconstruction, and multiple other image operations, without a computational load on the CPU. This method is implemented on the graphics processing unit (GPU) of the graphics-card. Instead of moving the grabbed image to the main memory unit of the CPU the image is transferred by Direct Memory Access (DMA) to the graphics processing unit. The GPU is specialized on image operations and vector calculus. Hence, the surface data is directly moved as vertices to the GPU. A sparse distribution of the surface data is interpolated automatically by the unit. The lens-distortion correction is implemented as a two-dimensional vector field. All these operations and an interpolation of the resulting image are performed at extremely high speed by the GPU.

The second part of the chapter is dedicated to the investigation of experimental restrictions. Considering the application of free-formed surface reflections in a vehicle environment four cases come to mind. One is the a deformation of the hood during the run-time of the application. A change due to the temperature is not of concern. The gap sizes between the body parts of the vehicle are a measurement of its quality. In the case of the experimental car the differences in the gaps due to the temperature are negligible. Hence, the influence on the image processing application is deemed to be equally minimal. A deformation is induced by air pressure variations. To examine their influence a simulation was conducted in collaboration with the design and simulation department. In the intended range of velocity the effect proved to be in the order

of  $10^{-2}$  mm. This must not be taken into consideration. One result of Chapter 4 is that the optical properties of the system depend on the relative position of the camera to the free-formed surface. Therefore, a change of this position during the application interferes with the result. Considering this, the second case is the influence of vibrations onto the system. If the vibrations induce a torsional force on the surface or the cameras the obtained calibration is not applicable anymore. However, the results showed that torsion is only influential for the maximum force induced by the start-up of the engine. Otherwise, the cameras and the surface vibrate synchronously. Another problem is caused by a soiled hood. However, the aperture is set such that object further away than 2 m are imaged sharply. Hence, the soilage of the hood blurs the image but does not render it unusable. This is supported by an empirical investigation. The last restriction is due to the varnish of the surface. The focus of the examination is the influence of the varnish onto the image processing. The physical properties of the varnished surface, such as the specular and the diffuse component, are not the scope. The gray-value histogram is indeed influenced by the varnish. However, the shape of the distribution is preserved, such that mean-free cross-correlation algorithms are capable of processing the images. In the case of comparing a reflection image with a direct view image, the differences are too large for the correlation. In this case the histogram is corrected by the reference of a gray-scale wedge.

The reconstructions and the coarse distance evaluation presented in this chapter prove that the reflections on a free-formed surface mirror provide image information which is sufficient for processing. Not only the images are reconstructed but the calibration parameters for the virtual camera are retrieved as well. They are the basis of the stereo-image processing.

## CHAPTER 9

---

# Conclusions and Outlook

### 9.1 Summary

This thesis presents a new approach to make computer-vision applications more reliable: free-formed surface reflections. The reliability increases with the evaluation of additional information. Usually, the data is provided by applying additional algorithms on the same image or by introducing additional hardware to the system, such as a third camera. Considering the requirements of the vehicle-following application, wide-angle lenses are used. They provide a view onto a reflective body part of the car, the hood. It is not designed to serve as part in an optical system. Still it serves as an image forming sensor. Nayar et al. (1998) proves that a vision-based object recognition is feasible with curved mirrors. However, that approach depends on geometrically closed-formed surface. This thesis extends it to the free form.

The stated task of this thesis is to develop and implement methods that retrieve the information in the reflection region of the image. The survey of the related work shows that reflections in general have become of interest within the last few years. The applications are focused on closed-formed geometries with special properties, such as a large field of view or a singular viewpoint. If the catadioptric systems obey the singular-viewpoint constraint, correct pinhole projections are reconstructible from the data. Surfaces that are not closed-formed are almost neglected in literature. Only Hicks & Bajcsy (1999) develop a catadioptric image formation sensor that automatically rectifies a predefined plane. If the surfaces are not designed to serve as mirror they are only treated as source of error. Accordingly, methods to process the reflection data were not available.

The first step in the development of methods and algorithms is an examination of the image and information reconstruction options. Operations that do not take the geometry into account, i.e. mere in-image methods, are not a suitable solution for the posed problem. Two methods emerge as the most promising:

- One is an approach that understands each retina element as an independent

camera. It requires an extension of the planar epipolar constraint to a more general notion of epipolar curves.

- The other one compares the directions of reflection of the surface to a planar reference mirror. It results in an in-image operation where the surface geometry is incorporated.

Apart from this, there are two more general results:

- To use the information the camera system must be calibrated. The intrinsic and extrinsic parameters have to be determined with high precision.
- The second fundament is the determination of the relative position of the camera to the free-formed surface. The distortions in the reflection image depend very much on these parameters.

Accordingly, camera calibration is the topic of Chapter 5. After reviewing different approaches, the method introduced by Bouguet (2000) is favorable as it provides the best results. However, during the course of investigation it became obvious, that the basic data acquisition in the Bouguet calibration is subject to improvement. The enhancement is implemented by a template-based correlation approach with a subsequent geometric point distribution analysis. Another property of the available Bouguet implementation is that it only operates on fully visible calibration rigs. Especially in vehicle-based applications in an industrial environment this is a disadvantage. Hence, a new approach, called the Partial Visibility Detection (PVD) is presented. It allows the reliable usage of partially visible calibration rigs. Images that were discarded before or needed to be removed manually from the image sequence now provide additional information. The freely distributed implementation consists of two parts. The initial monocular camera calibration is implemented in C++ and contained in the Intel OpenCV library. The stereo camera calibration and the non-linear optimization of the complete parameter set is implemented in MATLAB. The change in image handling, caused by the PVD, and the desire to have a monolithic application required a reimplemention of the MATLAB part. These improvements provide an easier to use, more reliable and more precise camera calibration.

In extension of the camera calibration algorithms a cross-raster-based method to recover the surface geometry is developed. The representation of the surface in the camera coordinate system automatically solves the problem of determining the relative camera position to the surface. The method does not require additional hardware but uses an image sequence of the camera calibration rig to retrieve the surface. The core of the reconstruction is an adapted version of the Partial Visibility Detection (PVD) to the curved surfaces and the therefore warped image regions. In the determination of the surface points a minimization technique constraining the surface point to the ray of sight is developed and applied.

After fulfilling the basic requirements of camera calibration and surface recovery the two promising methods, the non-classical information reconstruction with an extended epipolar-curve notion and the Comparison-of-Lines method are developed, implemented, and tested. The methods are applied to images acquired in the experimental vehicle. The results of the reconstruction and the coarse distance determination of an object prove the claim, that the reflections in a free-formed surface mirror are a source of information. The images as well as the calibration data is retrievable. These are the two components of a successful image processing.

Not only the idea of using a car body part as a reflective surface in an image-processing systems is patented but nine other patents or patent applications originated from this work as well.

## 9.2 Contributions of this Thesis

The main contributions of this thesis are:

- A general investigation of the possibilities to reconstruct an image or extract information from free-formed-surface reflections. There are two major results. First, that a pinhole image cannot be correctly reconstructed due to the lack of a singular viewpoint of the system. And second, a method that does not take the geometry into account is not able to provide the contained information.
- An improvement, the implementation, and the testing of a camera calibration method. Based on the comparison of different calibration algorithms the Bouguet (2000) method is considered superior to others. This approach is improved in several steps. First, the OpenCV implementation of a contour coarse-fine search and a gradient-based sub-pixel location of the chessboard line-intersection is replaced by a template-correlation method with a consecutive geometry analysis. In a second step, the restriction that only fully visible calibration rigs are considered is removed. To do so, the Partial Visibility Detection (PVD) of the chessboard rig is developed. It reliably detects and classifies corners of the calibration pattern. This is the basis of evaluating the geometry information of partially visible rigs. Apart from the desire to have a monolithic implementation, the PVD method requires a reimplementation of the MATLAB-based stereo-calibration algorithm in C++. The testing shows that the calibration method now is easier to handle, more reliable, and provides results with a much smaller error burden.
- A self-contained surface recovery on the basis of an acquired image-series is designed, implemented, and tested. A result of the investigations of the possible

reconstructions is that the geometry of the surface and the relative camera position to it must be determined. The Partial Visibility Detection (PVD) forms the basis of a cross-raster-based measurement of the surface geometry in the camera coordinate system. To reliably detect the reflections of the calibration pattern the PVD is adapted to curved geometries in the pattern analysis. The method is generally applicable to reflective surfaces and provides either direct measurement data or an approximated functional representation of the surface.

- The epipolar constraint of stereo vision is extended to a more general notion of epipolar curves. The evaluation of the relative camera positions in stereo vision provides a restriction to the object locations in the respective images. This is known as the epipolar constraint. This constraint is extended to free-formed surface reflections. Similar to the stereo-vision case, the geometry contains the constraining information. However, the free-formed surface is more difficult to handle. With the understanding of each retina element as an individual camera, differential-geometry calculus and variation methods are applied to evaluate the system setup. It results in a general formula that governs the epipolar curve. Approximations of the general case in the vehicle-specific setup are investigated as well.
- The Comparison-of-Lines method, an image-based reconstruction considering the surface geometry, is developed, implemented and tested. The method compares the reflection directions of a planar mirror, tangential to the surface, to the reflection directions of the surface itself. It results in an assignment of locations in the reconstructed image to locations in the original image. This is the basis of any classical reconstruction. Using an inverse ray-tracing method a densely reconstructed image is obtained. To use the image in a stereo-vision application intrinsic and extrinsic parameters of the virtual camera are necessary and provided. The method is implemented in two classical ways, using tables and polynomial approximations, on a the central processing unit. These implementations obey the video-real-time requirement demanded on the solution. A new method, using the graphics processing unit (GPU) of the graphics-card, is presented. This implementation provides an extraordinary speed and quality of the reconstruction. The idea of implementing a computer vision algorithm on the GPU is extendable to many other applications. It provides an image-processing capability which is unprecedented.
- The usability of reflections on free-formed surfaces is proved by a distance determination on the basis of a reconstructed image and an original image forming a stereo-vision system.

### 9.3 Conclusions

What has been achieved in this thesis? The common assumption that reflections on a free-formed surface are only a source of error for image-processing algorithms is refuted. Reflections on a free-formed surface, even if it is not designed to serve as a part in an optical system, provide information. The usual algorithms are just not able to cope with it. With the knowledge of the geometry of the surface and the system setup the information is retrievable. Specifically designed algorithms incorporate the information in real-time into applications. It supports image-processing algorithms in different cases, such as the decision making in multiple matching hypothesis situations or the temporary or long-termed loss of an image.

### 9.4 Future Work

A lot of work remains to be done in the future. Apart from challenging tasks that remain in computer vision, the work related to this thesis is summarized as follows:

- So far the camera-calibration method relies on the evaluation of an image sequence in a static and well-known scene. The geometric information incorporated in the calibration pattern enables the algorithm to correlate the extracted features and determine the intrinsic and extrinsic parameter set. Especially in a vehicle application an online calibration is desirable. So far, approaches that achieve a calibration in an unknown environment are non-metric. However, for the vehicle applications the metric information is indispensable.
- The surface recovery relies on an acquired image series. In order to obtain a sufficient number of data points a rather large series of images must be recorded. This is a lengthy procedure. The correction of the lens distortions, the feature extraction, the matching, the surface-point determination, and the surface approximation are optimized for speed. Nevertheless, the evaluation still takes time. Hence, an extension of the online camera calibration mentioned above is desirable to form an online surface recovery. The online recovery solves a still existing problem. Deformations of the hood as restrictions to the applicability are excluded by the investigations in Chapter 8. However, a permanent structural damage, as caused by a stone or a bump onto another object, is not considered. An online surface reconstruction is able to detect such a change in the surface. A subsequent adaption to the new circumstances is feasible.
- The non-classical information reconstruction is developed but not yet implemented in an experimental vehicle.



- As mentioned in Chapter 8, the experimental vehicle is not equipped with a graphics processing unit (GPU). After introducing the next generation of hardware in the vehicle, an experimental verification is to be performed.
- The free-formed surface mirror application is designed for two scenarios. One is supporting the stereo vision system in multiple matching hypothesis situations. The other is the provision of a fall-back layer in the case of a temporary or long-termed loss of an image. Both cases stipulate the making of a decision. The situations when to fall back to the reflection image must be determined. Subsequently, the manner in which to invoke the reflection image must be decided. These decision making algorithms are not developed yet.
- The algorithms and methods presented in this thesis prove the feasibility of the ideas. The reliability and sophistication necessary for the incorporation into a product are not achieved yet. With that level the replacement of one camera by the reflection image of the other is possible.
- Image processing on the graphics processing unit (GPU) is a technology that is just emerging. The GPU is developing rapidly and with each generation the possibilities rise. The precision of the calculation, the variety of implemented commands, and the amount of parallel shader units is increasing. With this perspective it is a very promising research area.

## 9.5 Outlook

Traffic is steadily increasing. Analogously, the overload of the driver with stressful situations is increasing as well. Systems which assist the driver in these complex situation are a necessity to at least keep the risk of participating in road traffic at the actual level. Driver assistance systems must rely on sensor technology to perceive the environment of the ego-vehicle. Computer vision and camera technology are still rapidly developing fields. But other approaches with different sensors such as radar-, laser-, or infrared-based systems are competitive, especially in vehicle applications. Nevertheless, vision with its passive nature, is a source of abundant information, more information than other sensors are able to provide. Above this, the visual information itself, i.e. without processing, is intuitive for the human user. With steadily increasing computational power in general, the evaluation of this information will be more and more feasible, reliable, and applicable. Systems exploiting the visual cues will increase the driver's convenience and the security of all traffic participants.

## Bibliography

---

- Abdel-Aziz Y.I., Karara H.M., 1971, *Direct Linear Transformation from Comparator Coordinates into Object Space Coordinates in Close-Range Photogrammetry*, In ASP Symposium on Close-Range Photogrammetry, American Society of Photogrammetry, 1-18
- Adelson E.H., Bergen J.R., 1992, *The Plenoptic Function and the Elements of Early Vision*, Ch. 1, 3-20, MIT Press, Cambridge, Mass., USA
- Altmann S.L., 1989, *Hamilton, Rodrigues, and the Quaternion Scandal*, Mathematics Magazine, 62, 291-308
- Andresen K., Morche B., 1983, *Digitale Verarbeitung von Kreuzrasterstrukturen zur Verformungsmessung von Flächen*, Technical Report 480, VDI
- Aschwanden P., 1993, *Experimenteller Vergleich von Korrelationskriterien in der Bildanalyse*, Hartung-Gorre Verlag Konstanz
- Bacakoglu H., Kamel M.S., 1997, *A Three-Step Camera Calibration Method*, IEEE Transactions on Instrumentation and Measurement, 46, 1165-1172
- Baker S., Nayar S.K., 1998, *A Theory of Catadioptric Image Formation*, In Proceedings of International Conference on Computer Vision (ICCV), IEEE Computer Society, 35-42
- Baker S., Nayar S.K., 1999, *A Theory of Single-Viewpoint Catadioptric Image Formation*, International Journal of Computer Vision, 35, 175-196
- Barreto J.P., Araujo H., 2001, *Issues on the Geometry of Central Catadioptric Image Formation*, In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, IEEE Computer Society, 422-427
- Beach D., 1999, *Novel technique for the Design of Ultra-Fast, High Resolution, Broad Spectrum, Wide Angle Catadioptric Lenses*, Optical Engineering, 38, 1627-1634
- Becker U., 2002, *Crash Rates Drop for Mercedes Passenger Cars*, DaimlerChrysler Headline, Newsletter for the Management
- Beyerer J., Pérard D., 1997, *Automatische Inspektion spiegelnder Freiformflächen anhand von Rasterreflexionen*, Technisches Messen, 64, 394-400

- Bhat D.N., Nayar S.K., 1995, *Stereo in the Presence of Specular Reflection*, In Proceedings of International Conference on Computer Vision (ICCV), IEEE Computer Society, 1086-1092
- Bhat D.N., Nayar S.K., 1998, *Stereo and Specular Reflection*, International Journal of Computer Vision, 26, 91-106
- Bortolozzi F., Dubuisson B., 1991, *Trinocular Vision: a 3-D Solution*, In SPIE Intelligent Robots and Computer Vision, Vol. 1608, 97-106
- Bouguet J.Y., 2000, *A Camera Calibration Toolbox for Matlab*, [http://www.vision.caltech.edu/bouguetj/calib\\_doc/index.html](http://www.vision.caltech.edu/bouguetj/calib_doc/index.html)
- Bouguet J.Y., Perona P., 1998, *Closed-form camera calibration in dual-space geometry*, In European Conference on Computer Vision (ECCV).
- Brandes M., 1995, Entfernungsbestimmung markanter Punkte aus einem Fahrzeug mit Hilfe eines Stereokamerasystems, *Master's thesis*, Fakultät für Informatik, Universität Stuttgart
- Bronstein I.N., Semendjajev K.A., 1985, *Taschenbuch der Mathematik*, Verlag Harri Deutsch, Thun
- Brown D.C., 1966, *Decentring Distortions of Lenses*, Photogrammetric Engineering, 32, 444-462
- Brown D.C., 1971, *Close-Range Camera Calibration*, Photogrammetric Engineering, 37, 855-866
- Cafforio C., Rocca F., 1986, *Precise Stereopsis with a Single Video Camera*, Signal Processing III: Theories and Application, 641-644
- Caprile B., Torre V., 1990, *Using Vanishing Points for Camera Calibration*, International Journal of Computer Vision, 4, 127-140
- Chahl J., Srinivasan M., 1997, *Reflective Surfaces for Panoramic Imaging*, Applied Optics: Optical Technology and Biomedical Optics, 36, 8275-8285
- Chatterjee C., Roychowdhury V.P., Chong E.K.P., 1997, *A Nonlinear Gauss-Seidel Algorithm for Noncoplanar and Coplanar Camera Calibration with Convergence Analysis*, Computer Vision and Image Understanding, 67, 58-80
- Chen S.E., 1995, *QuickTime VR - An Image Based Approach to Virtual Environment Navigation*, In Proceedings of SIGGRAPH, 29-38
- Clarke T., Fryer J., 1998, *The Development of Camera Calibration Methods and Models*, Photogrammetric Record, 16, 51-66
- Conrady A.E., 1919, *Decentered Lens-Systems*, Monthly Notices of the Royal Astronomical Society, 79, 384-390

- Faugeras O., 1993, *Three-Dimensional Computer Vision*, Ch. 5, 150, Cambridge University Press, Massachusetts, USA
- Fischer H., Kaul H., 1990, *Mathematik für Physiker*, Vol. 1 of *Mathematik/Physik*, Mathematik/Physik, Teubner Studienbücher
- Fryer J.G., Brown D.C., 1986, *Lens Distortion for Close-Range Photogrammetry*, *Photogrammetric Engineering and Remote Sensing*, 52, 51-58
- Gehrig S.K., 2000, *Design, Simulation, and Implementation of a Vision-Based Vehicle-Following System*, Ph.D. thesis, University of Tübingen
- Geyer C., Daniilidis K., 1999, *Catadioptric Camera Calibration*, In Proceedings of International Conference on Computer Vision (ICCV), Vol. 1, IEEE Computer Society, 398-404
- Geyer C., Daniilidis K., 2001, *Catadioptric Projective Geometry*, *International Journal of Computer Vision*, 45, 223-243
- Gluckman J., Nayar S.K., 1998a, *Ego-Motion and Omnidirectional Cameras*, In Proceedings of International Conference on Computer Vision (ICCV), IEEE Computer Society, 999-1005
- Gluckman J., Nayar S.K., 1998b, *A Real-Time Catadioptric System Using Planar Mirrors*, In Proceedings of the DARPA Image Understanding Workshop.
- Gluckman J., Nayar S.K., 1999, *Planar Catadioptric Stereo: Geometry and Calibration*, In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, 22-28
- Gluckman J., Nayar S.K., 2000, *Rectified Catadioptric Stereo Sensors*, In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, 380-388
- Gluckman J., Nayar S.K., Thoresz K.J., 1998, *Real-Time Omnidirectional and Panoramic Stereo*, In Proceedings of the DARPA Image Understanding Workshop, 299-303
- Goshtasby A., Gruver W.A., 1993, *Design of a Single Lens Stereo Camera System*, *Pattern Recognition*, 26, 923-937
- Greene N., 1986, *Environment Mapping and Other Applications of World Projections*, *IEEE Computer Graphics and Applications*, 21-29
- Hanson C., Ayache N., Lustman F., 1988, *Efficient Depth Estimation using Trinocular Stereo*, In *Sensor Fusion: Spatial Resolution and Scene Interpretation*, Vol. 1003, Society of Photo-Optical Engineering (SPIE), 124-131
- Hartley R.I., 1995, *In Defence of the 8-Point Algorithm*, In Proceedings of International Conference on Computer Vision (ICCV), IEEE Computer Society, 1064-1070

- Heikkilä J., Silvén O., 1997, *A Four-step Camera Calibration Procedure with Implicit Image Correction*, In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, 1106-1112
- Hengst M., 2001, *Kalibrierung eines elektronischen Stereokamerasystems*, *Master's thesis*, University of Stuttgart
- Hicks R.A., Bajcsy R., 1999, *Reflective Surface as Computational Sensors*, In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, 82-86
- Horn B.K., 1987, *Closed-form solution of absolute orientation using unit quaternions*, Journal of the Optical Society of America, 4, 629-642
- Hubin N., Noethe L., 1993, *What is Adaptive Optics*, Science, 262, 1345-1484
- Inaba M., Hara T., Inoue H., 1993, *A Stereo Viewer Based on a Single Camera with View Control Mechanisms*, In Proceedings of International Conference on Intelligent Robots and Systems, IEEE Computer Society
- Kaempchen N., 2001, *Modellbasierte Lagebestimmung von Objekten in Stereobildsequenzen*, *Master's thesis*, Institut für Physikalische Elektronik, University of Stuttgart
- Kalman R.E., 1960, *A New Approach to Linear Filtering and Prediction Problems*, Transactions ASME Journal of Basic Engineering, 82, 35-45
- Kammel S., 2001, *Topography reconstruction of specular surfaces from a series of grayscale images.*, In Harding K., Miller J., Batchelor B. (eds.) *Machine Vision and Three-Dimensional Imaging Systems for Inspection and Metrology*, Vol. 4189, Society of Photo-Optical Engineering (SPIE), 136-144
- Kanade T., Kano H., Kimura S., et al., 1995, *Development of a Video-Rate Stereo Machine*, In Proceedings of International Conference on Intelligent Robotics and Systems (IROS), 95-100
- Kim W., Cho H., 2000, *Learning-Based Constitutive Parameters Estimation in an Image Sensing System with Multiple Mirrors*, Pattern Recognition, 1199-1217
- Lee D.H., Kweon I.S., Cipolla R., 1999, *A Bisprium Stereo Camera*, In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, 82-87
- Lenz R., 1987, *Linsenfehlerkorrigierte Eichung von Halbleiterkameras mit Standardobjektiven für hochgenaue 3D-Messungen in Echtzeit*, In Paulus E. (ed.) *Informatik-Fachberichte 149 Musternererkennung. 9. DAGM-Symposium*, Braunschweig, Springer-Verlag, 212-216
- Lenz R.K., Tsai R.Y., 1987, *Techniques for calibration of the scale factor and image center for high accuracy 3D machine vision metrology*, In IEEE International Conference on Robotics and Automation, 68-75

- Ligtenberg F.K., 1954, *The moiré-method, a new experimental method for the determination of moments in small slab models*, In Proceedings of the Soc. Exp. Stress Anal., Vol. 12, 83-98
- Lippmann A., 1980, *Movie Maps: An Application of the Optical Videodisc to Computer*, In Proceedings of SIGGRAPH, 32-42
- Longuet-Higgins H.C., 1981, *A Computer Algorithm for Reconstructing a Scene from two Projections*, Nature, 293, 133-135
- Longuet-Higgins M., 1960, *Reflection and Refraction at a Random Moving Surface*, Journal of the Optical Society of America, 50, 838-856
- Luhmann T., 2000, *Nahbereichsphotogrammetrie: Grundlagen, Methoden und Anwendungen*, Heinz Wichmann Verlag, Heidelberg
- McMillan L., Bishop G., 1995, *Plenoptic Modeling: An Image-Based Rendering System*, In Proceedings of SIGGRAPH, 39-46
- Miyamoto K., 1964, *Fish Eye Lens*, Journal of the Optical Society of America, 54, 1060-1061
- Moosbrugger A., 2002, *Robuste Erfassung von partiell sichtbaren Kalibrierfeldern*, Internship Report, DaimlerChrysler AG, Berufsakademie Stuttgart
- Moré J.J., 1977, *The Levenberg-Marquardt Algorithm: Implementation and Theory* In Watson G. (ed.) Numerical Analysis, Vol. 630. Lecture Notes in Mathematics Springer Verlag
- Nayar S.K., 1988, *Sphereo: Determining Depth using Two Specular Spheres and a Single Camera*, In Proceedings of Society of Photo-Optical Engineering (SPIE): Optics, Illumination, and Image Sensing for Machine Vision III, 245-254
- Nayar S.K., 1997a, *Catadioptric Omnidirectional Camera*, In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, 482-488
- Nayar S.K., 1997b, *Omnidirectional Video Camera*, In Proceedings of the DARPA Image Understanding Workshop, 235-241
- Nayar S.K., 1997c, *Omnidirectional Vision*, In Proceedings of the 8th International Symposium of Robotics Research.
- Nayar S.K., Baker S., 1997, *Catadioptric Image Formation*, In Proceedings of the DARPA Image Understanding Workshop, 1431-1437
- Nayar S.K., Bolle R.M., 1996, *Reflectance Based Object Recognition*, International Journal of Computer Vision, 17, 219-240
- Nayar S.K., Gluckman J., Swaminathan R., et al., 1998, *Catadioptric Video Sensor*, In Proceedings of the 4th International Conference on Applications of Computer Vision (WACV), IEEE Computer Society, 236-237

- Nayar S.K., Peri V., 1999, *Folded Catadioptric Cameras*, In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, IEEE Computer Society, 217-223
- Nene S.A., Nayar S.K., 1998, *Stereo with Mirrors*, In Proceedings of International Conference on Computer Vision (ICCV), IEEE Computer Society, 1087-1094
- Oh S.J., Hall E.L., 1987, *Guidance of a Mobile Robot using Omnidirectional Vision Navigation*, In SPIE Proceedings, Vol. 852, Society of Photo-Optical Engineering (SPIE), 288-300
- Okutomi M., Kanade T., 1993, *A Multiple-Baseline Stereo*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 15, 353-363
- Oren M., Nayar S.K., 1994a, *Generalization of Lambert's Reflectance Model*, In Proceedings of SIGGRAPH, 239-246
- Oren M., Nayar S.K., 1994b, *Seeing Beyond Lambert's Law*, In European Conference on Computer Vision (ECCV), 269-280
- Oren M., Nayar S.K., 1996, *A Theory of Specular Surface Geometry*, International Journal of Computer Vision, 24, 105-124
- Otha M., 1994, *Recent Technology of Projection Optics*, In SEMI Technological Symposium., 387-397
- Pajdla T., 2002, *Stereo with Oblique Cameras*, International Journal of Computer Vision, 47, 161-170
- Pérard D., 2000, *Automated visual inspection of specular surfaces with structured-lighting reflection techniques*, Ph.D. thesis, University of Karlsruhe
- Peri V.N., Nayar S.K., 1997, *Generation of Perspective and Panoramic Video from Omnidirectional Video*, In Proceedings of the DARPA Image Understanding Workshop, 243-245
- Piegl L., 1991, *On NURBS: A Survey*, IEEE Transactions on Computer Graphics and Applications, 11, 55-71
- Pope A., Lowe D., 1994, *Vista: A Software Environment for Computer Vision Research*, In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), 768-772
- Ritter R., Hahn R., 1984, *Zur Analyse des Reflexions-Moiré-Effekts*, Forschung im Ingenieurwesen, 50, 87-90
- Roelofs R., 1951, *Distortion, principle point, point of symmetry and calibrated principle point*, Photogrammetria, 7, 49-66
- Rogers P.J., 1999, *Compact Multispectral Optic with Switchable Field of View*, In EUROPTO Conference on Design and Engineering of Optical Systems, Vol. 3737, Society of Photo-Optical Engineering (SPIE), 261-268

- Savarese S., Perona P., 2001, *Local Analysis for 3D Reconstruction of Specular Surfaces*, In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, IEEE Computer Society, 738-745
- Schowalter D., 2002, *Anwenderorientiertes Kalibrierverfahren für eine Stereokamera, Master's thesis*, Fachbereich Maschinenbau und Verfahrenstechnik, Fachhochschule Furtwangen
- Shen J., Paillou P., 1995a, *Trinocular Stereovision by Generalized Hough Transform*, In Society of Photo-Optical Engineering (SPIE), Vol. 2588, 399-407
- Shen J., Paillou P., 1995b, *Trinocular Stereovision by Matching in Parameter Space*, In Society of Photo-Optical Engineering (SPIE), Vol. 2620, 732-739
- Stein G.P., 1997, *Lens Distortion Calibration Using Point Correspondences*, In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), 602-608
- Stewart C.V., 1989, *Trinocular Stereo: Theoretical Advantages and a New Algorithm*, In Sensor Fusion II: Human and Machine Strategies, Vol. 1198, Society of Photo-Optical Engineering (SPIE), 377-391
- Sturm P.F., Maybank S.J., 1999, *On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications*, In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, 432-437
- Swaminathan R., Nayar S.K., 2000, *Non-Metric Calibration of Wide-Angle Lenses*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22, 1172-1178
- Takeya A., Kuroda T., Nishiguchi K., Ichikawa A., 1998, *Omnidirectional Vision System Using Two Mirrors*, In Conference on Novel Optical Systems Design and Optimization II, Vol. 3740, Society of Photo-Optical Engineering (SPIE), 50-60
- Tamura S., Kim E.K., Close R., Sato Y., 1994, *Error Correction in Laser Scanner Three-Dimensional Measurement by Two-Axis Model and Coarse-Fine Parameter Search*, Pattern Recognition, 27, 331-338
- Tingstad J.S., 1991, *Alignment of an Aspheric Mirror Subsystem for an Advanced Infrared Catadioptric System*, In Current Developments in Optical Design and Optical Engineering, Vol. 1527, Society of Photo-Optical Engineering (SPIE), 194-198
- Torrance K., Sparrow E., 1967, *Theory for Off-Specular Reflection From Roughend Surfaces*, Journal of the Optical Society of America, 57, 1105-1114
- Tsai R.Y., 1986, *An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision*, In Proceedings of Conference on Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, 364-374



- Williamson T., Thorpe C., 1998, *Detection of Small Obstacles at Long Range Using Multi-baseline Stereo*, In IEEE International Conference on Intelligent Vehicles, 311-316
- Yamazawa K., Yagi Y., Yachida M., 1993, *Omnidirectional Imaging with Hyperboloidal Projection*, In Proceedings of International Conference on Intelligent Robotics and Systems (IROS), IEEE Computer Society, 1029-1034
- Zhang Z., 1999, *Flexible Camera Calibration By Viewing a Plane From Unknown Orientations*, In Proceedings of International Conference on Computer Vision (ICCV), IEEE Computer Society, 666-673
- Zhuang H., Roth Z.S., 1995, *Modeling Gimbal Axis Misalignments and Mirror Center Offsets in a Single-Beam Laser Tracking Measurement System*, International Journal of Robotics Research, 14, 211-224

## APPENDIX A

---

### **CAD Data Representation of the Engine Hood**

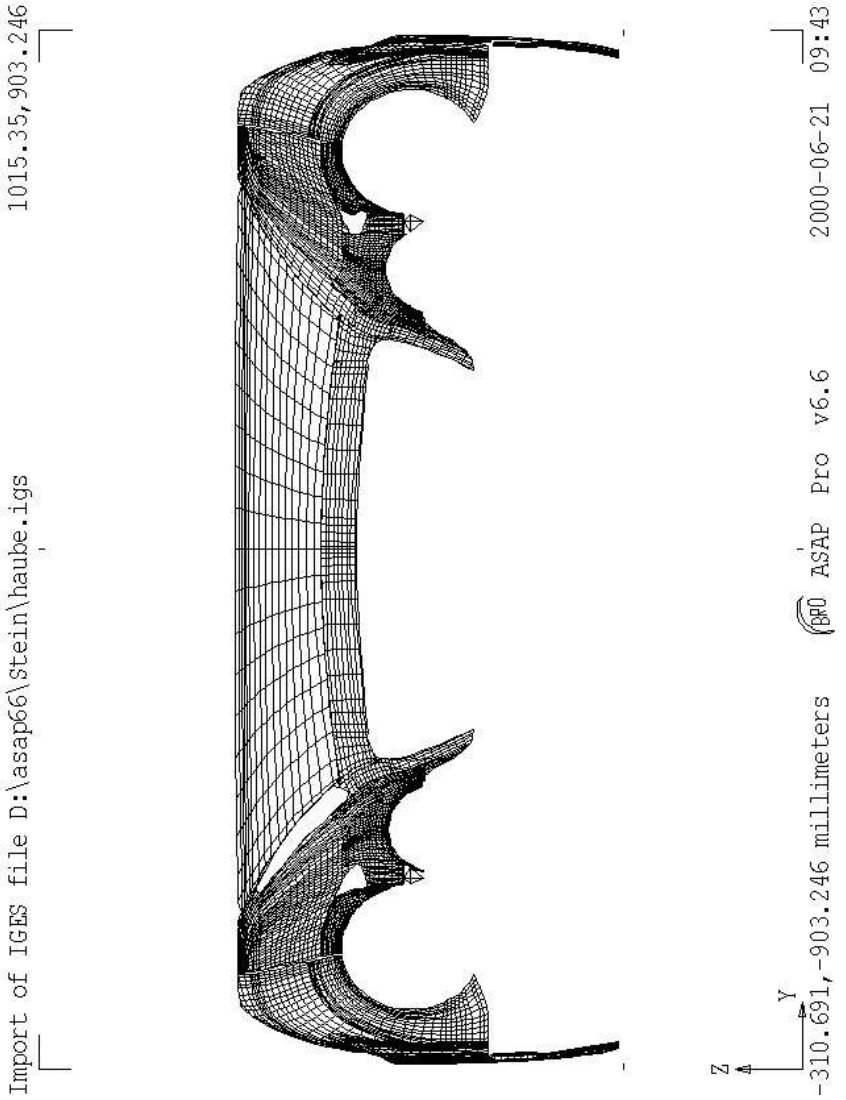


Figure A.1: A CAD Data Representation of the Engine Hood from the Front.

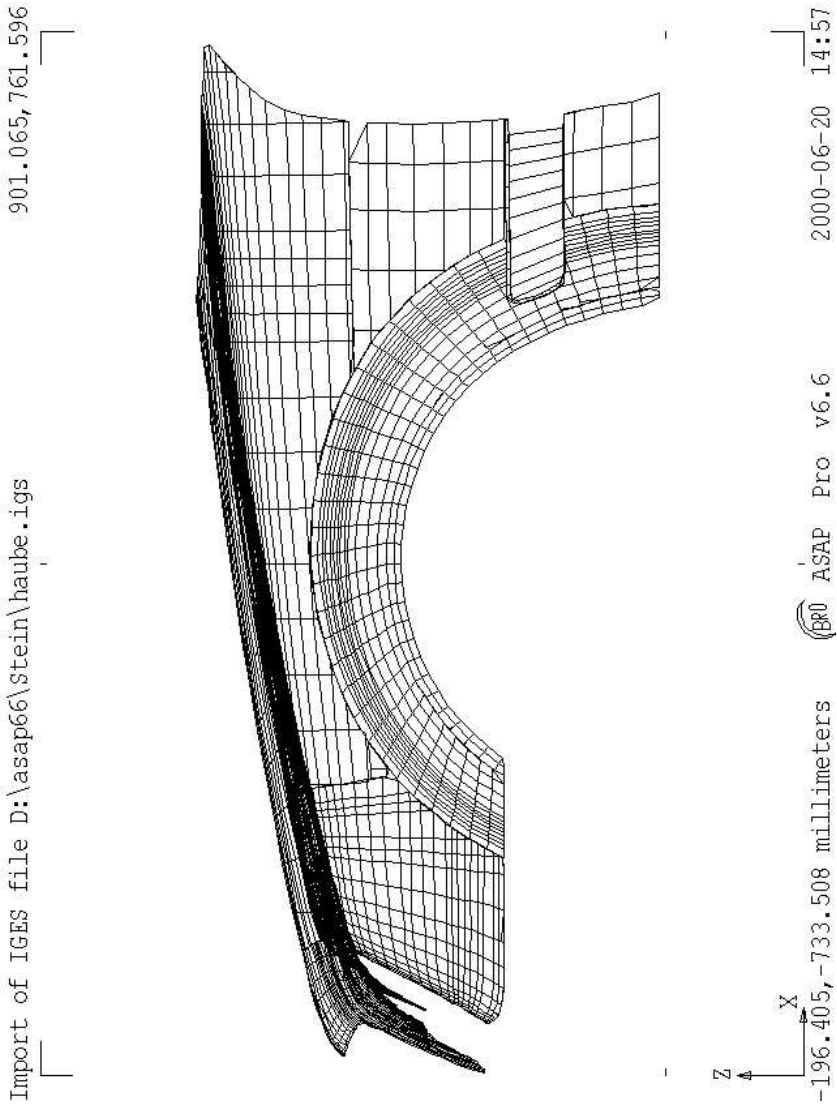


Figure A.2: A CAD Data Representation of the Engine Hood from the Side.

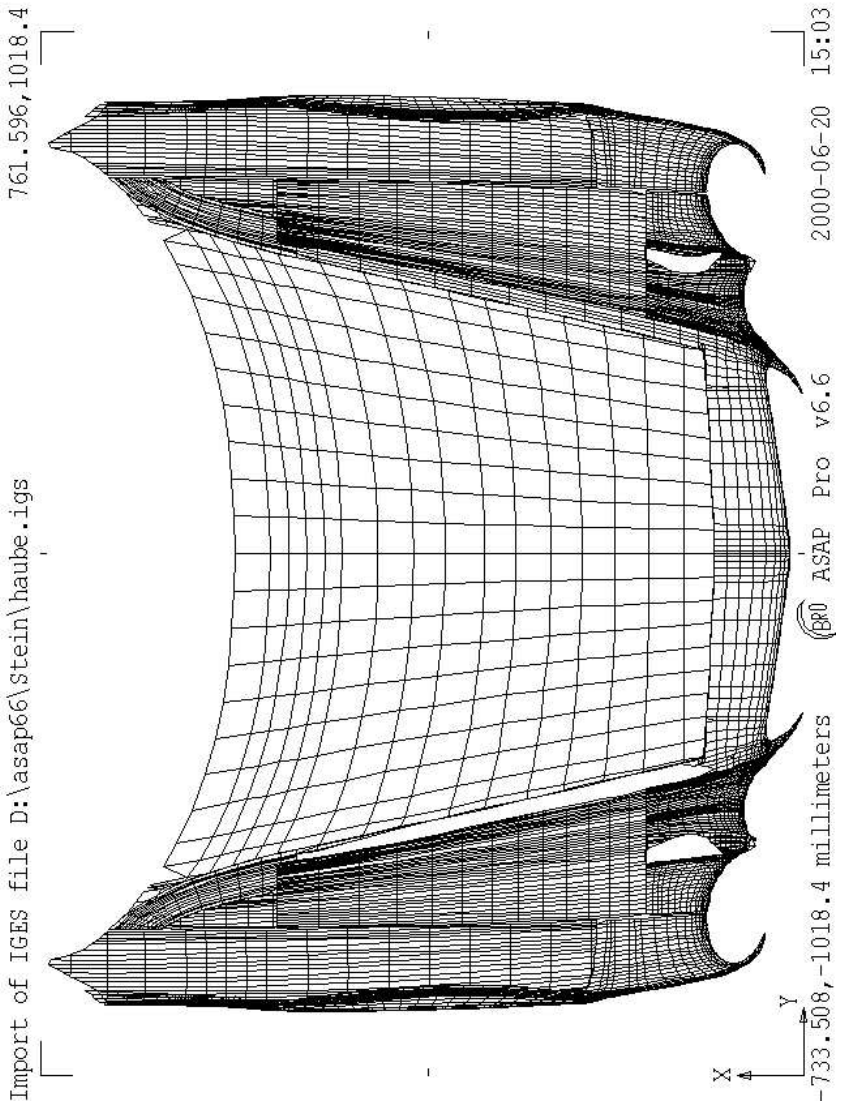


Figure A.3: A CAD Data Representation of the Engine Hood from the Top.

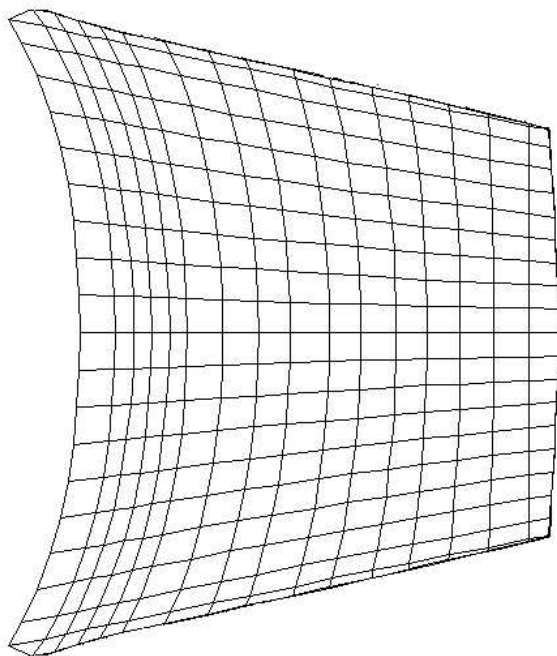


Figure A.4: A CAD Data Representation of the Central Engine Hood: A separate movable part.

## APPENDIX B

---

### **Air Pressure Deformation of the Hood**

To investigate the influence of the air pressure on the hood, a driving speed of 250 km/h is assumed. The hood is represented by CAD data. Blue is a low deformation, red a high one. The maximum deformation of about 2.2 mm occurs in the center of the front part. In the central upper area, where the most interesting objects for the intended application are visible, the deformation is low. In addition, the air pressure deformation effect diminishes with the velocity to the square and the velocity range in the application is up to 50 km/h.

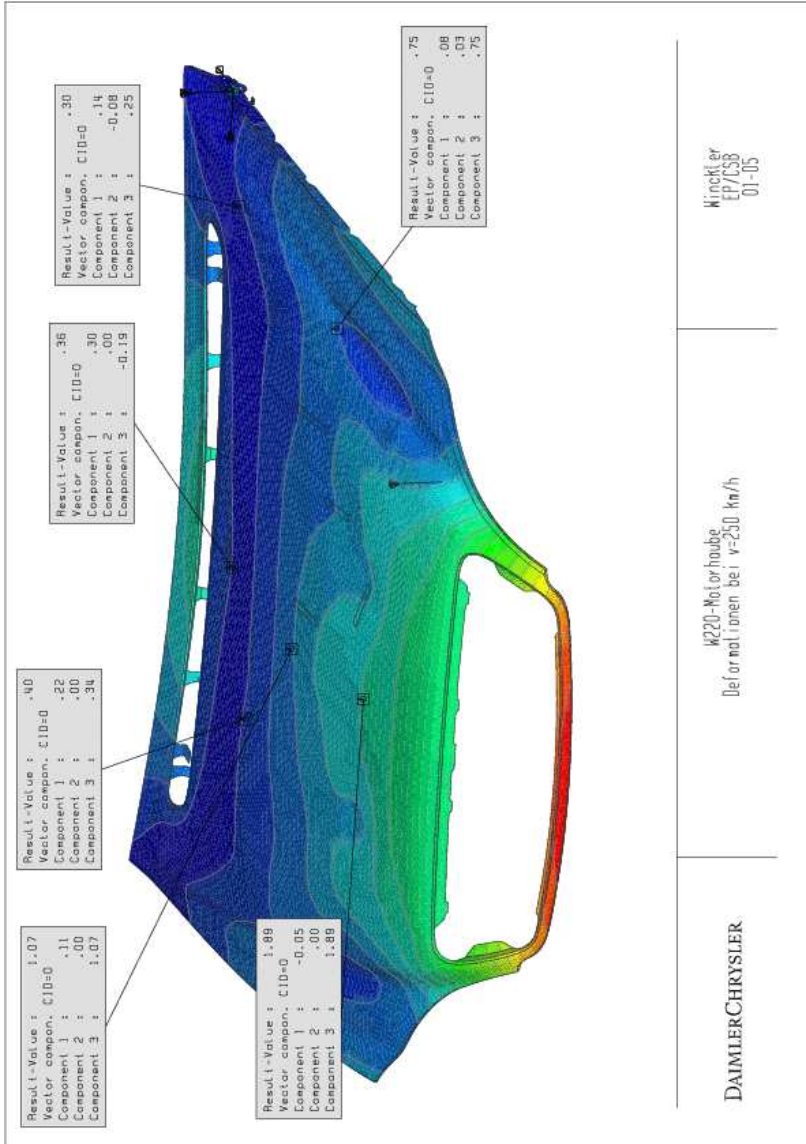


Figure B.1: Air Pressure Deformation of the Hood: Top View. Source: T. Winkler, DaimlerChrysler AG, EP/CSB.



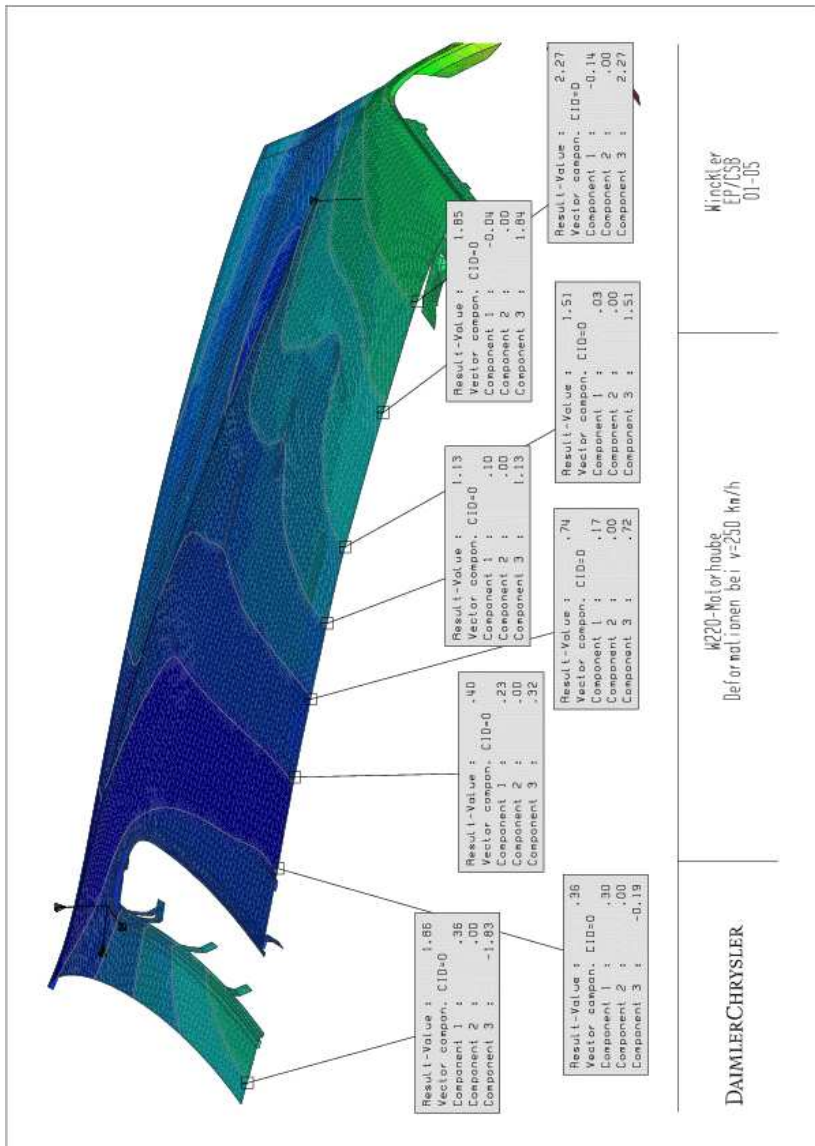


Figure B.2: Air Pressure Deformation of the Hood: Side View. The hood is cut along its axis of symmetry. Source: T. Winkler, DaimlerChrysler AG, EP/CSB.

# Curriculum Vitae

---

**Persönliche Daten** Würz-Wessel, Alexander  
Ziegelstraße 66, 73084 Salach  
geb. am 19. Mai 1971 in Albstadt  
verheiratet, ein Kind

## Ausbildung

08.1977 - 06.1991 Schulausbildung in Albstadt. Abschluß: Abitur  
07.1991 - 06.1992 Grundwehrdienst beim Fernmeldebataillon 10 in Sigmaringen  
10.1992 - 09.1995 Studium der Physik (Diplom) an der Universität Tübingen (Vordiplomprüfung März 1995)  
10.1995 - 09.1996 Studium der Physik (Master) an der University of Massachusetts, Boston, Massachusetts, USA  
10.1996 - 09.1998 Weiterführung des Studiums der Physik (Diplom) an der Universität Tübingen  
09.1998 - 08.1999 Diplomarbeit am Institut für Astronomie und Astrophysik, Abtl. Astronomie, Universität Tübingen, unter Anleitung von Prof. Dr. K. Werner, Dr. N. Kappelmann und Dr. J. Barnstedt zum Thema „Anschluß eines MCP-Detektors an ein Windows NT-System“, Abschluß: Diplom-Physiker  
seit 04.2000 Anfertigung der Dissertation unter Anleitung von Prof. Dr. K. Werner und Dr. F.J. Stein am DaimlerChrysler Forschungsinstitut Esslingen zum Thema „Free-formed Surface Mirrors in Computer Vision Systems“

## Berufliche Tätigkeiten

03.1995 - 03.2000 Wissenschaftliche Hilfskraft in verschiedenen Projekten am Institut für Astronomie und Astrophysik, Universität Tübingen  
seit 04.2000 Mitarbeit in verschiedenen Projekten im Bereich Autonome Fahrsysteme am DaimlerChrysler Forschungsinstitut Esslingen  
seit 03.2001 Sprecher der DaimlerChrysler Doktorandengruppe am Standort Stuttgart

## Akademische Lehrer

Meine akademischen Lehrer waren vor andern die Herren Prof. Dres.:

H. Clement, A. Fäßler, H. Fischer, K.-H. Gauckler, F. Gönnenwein, H. Müther, F. Rex, H. Ruder, R. Staubert, G. Wagner und K. Werner an der Universität Tübingen  
L. Catz, P. Mahon, G. Shane, P. Shah und D. Rao an der University of Massachusetts, Boston, Massachusetts, USA  
S. Carroll am Massachusetts Institute of Technology, Cambridge, Massachusetts, USA