Calibration of Various Configurations of Multiple Cameras for Driver Assistance Systems

Je Hak Ahn

A thesis submitted to
Auckland University of Technology
in fulfillment of the requirements for the degree of
Master of Philosophy (MPhil)

2010

School of Computing and Mathematical Sciences

Primary Supervisor: Dr. Russel Pears

Supervisor: Professor Reinhard Klette

"..They may know the mystery of God, namely, Christ, in whom are hidden all the treasures of wisdom and knowledge." Col 2:2-4

To my family, especially Hye-Won and In-Shik.

Abstract

This thesis is an exploratory and seminal work aimed to select and test available calibration algorithms for efficiency, usability and most importantly, accuracy of multiple camera and lens configurations in the context of DAS (Driving Assistance Systems). The camera and lens configurations considered were basically fixed stereo setups with normal and fish-eye lenses combined with low and high camera sensor resolutions. The stereo camera setup used for calibration and experiments were similar to those used during DAS experiments.

The selected calibration algorithms were four: OpenCV calibration, Bouguet, Mei and Scaramuzza algorithms. The OpenCV calibration was selected and tested for normal lens while Bouguet, Mei and Sacaramuzza algorithms were selected for fish-eye lens.

The methodologies selected and used for testing and comparing calibrations were backprojection error and row misalignment error as well as direct comparison of calibration parameters whenever applicable.

The calibration experiment results showed that OpenCV calibration is a suitable and accurate calibration algorithm for normal lens in the context of DAS. Similarly,Bouguet's fish-eye calibration toolbox seems to be the most appropriate in terms of accuracy and robustness in the context of DAS according these calibrations experiments. Mei's algorithm was second and Scaramuzza was third mostly due to inaccuracy and difficulty to use.

Finally, this research contributed to the utilization of multiple camera calibration in DAS systems as well as to the evaluation and recommendations for best camera configurations for different purposes and environment conditions.

Keywords: Driver assistance, stereo analysis, multiple cameras, calibration, evaluation.

Acknowledgments

I would like to thank my supervisors Professor Reinhard Klette and Dr. Russel Pears, for their comments, suggestions and help writing this thesis.

Especially Professor Reinhard Klette for his support, encouragement and invaluable lessons that made this possible.

Finally, thanks to Jiang Ruyi for his help with experiments and Ralf Haeusler and Ratheesh Kalarot for their comments.

Je Ahn
Auckland
September 27, 2010

Contents

Chapter 1

Introduction

This introductory chapter discusses this exploratory and seminal work in the research area of the calibration of optical systems in the context of vision-driver assistance. Firstly, it provides an overview of Vision-Based Driver Assistance and its importance in modern world followed by an introduction to the .enpeda.. project researching in the area of Vision-Based Driver Assistance. Finally, it presents multiple camera systems utilized in the .enpeda.. project and in driving assistance systems in general.

1.1 Vision-Based Driver Assistance

It has been reported in 2003 that the impact of traffic accidents and related consequent congestion is very significant with over 100,000 deaths and almost 500,000 injuries as a direct result of over 600,000 traffic related accidents in the People's Republic of China alone or direct costs of around 3.4 billion U$S in the United States [32]. There is a projection that deaths and injuries related to road traffic accidents will increase by about 65% in the next 20 years if the current trend continues [22].

Traffic accidents are in general the result of the interaction or combination of these factors: (i) human behavior, (ii) vehicle's capabilities, and (iii) road infrastructure [19]. Firstly, changing human behavior intends to deal with improving driver's behavior such as slowing down or driving safely by keeping enough distance and so on. Secondly, road infrastructure intends to design safer and better roads, road signals and so on. Finally, vehicle's capabilities intend to add or improve the safety and driving capabilities of moving vehicles. Improving vehicle capabilities are the best and easiest candidate for changes and extensive research has been done resulting in many safety features becoming standard car features over the years such as: seat belts, Anti-lock Braking Systems or Electronic Stability Control among others.

The research in the area of *Driving Assistance Systems* (DAS) is very interesting and active in the context of improving vehicle capabilities such as [24], [14] and [28] among many others. DAS can be visualized as a driver assistance system that works intuitively and unobtrusively but also overridable by the human driver [8].

DAS functions could be multiple and varied such as: lane keeping assistant, vision enhancement, adaptive cruise control, forward and intersection collision avoidance among others. Current DAS utilize active systems such as LIDAR (Light Detection and Ranging), radar or passive systems such as optical stereo systems. Active systems are accurate but also have limitations such as being sensitive to poor visibility or problems with hilly roads in the case of LIDAR or being blind to traffic signals or road markings in case of radar.

Thus, computer vision based on optical systems is a good option in term of costs (cost of related equipment such as cameras) and capabilities as for example distinguish road markings and characteristics. Stereo optical systems can determine object distances by triangulation but it is still not robust or discriminative enough to detect different road objects. Optical systems are affected negatively by adverse weather and road conditions but research results have shown that stereo optical systems produces good results and has a lot of potential as the DAS choice of sensors in the future despite its current limitations. There is intense ongoing research in DAS by research groups all over the world such as the *.enpeda..* project in New Zealand which is introduced next.

1.2 The *.enpeda..* Project

The *.enpeda..* project (environment perception and driver) was started in 2007 by Reinhard Klette and focuses on the research of DAS and related algorithms. The *.enpeda..* project group is based at the Tamaki Campus of The University of Auckland and focuses specifically on the area of computer vision which aims to understand or model a 3D environment from multiple image sequences. These image sequences are either recorded or captured live by 'normal' cameras with normal or wide angle optics.

The project also involves multimedia imaging by addressing "...visualization of 3D environments, using recorded stereo sequences or computer graphics for synthesizing 3D scenes or geometry, and human-machine interaction to some extent" [21]. In general, the practical application involves having cameras installed in any vehicle such as a passenger car, wheelchair, forklift among others where the driving is supported or assisted by a computer vision system that analyzes the video sequences from the cameras [21].

The *.enpeda..*'s test vehicle or *ego-vehicle* (vehicle with cameras to capture stereo sequences) named HAKA1 (acronym for High Awareness Kinematic Vehicle number 1), see Fig. 1.1, was facilitated by the partnership with Daimler AG. HAKA1 provides a fully road worthy car as a mobile platform for simulating a passenger car in all driving conditions where such video analysis systems and algorithms can be tested and research conducted. The image data collected with HAKA 1 have been used in several research such as [16] and [15] among others.

HAKA1 can be fitted with an onboard computer for some online graphic processing such as lane detection for example, but more importantly, HAKA1 can be fitted with multiple cameras internally and externally as well as a GPS sensor and motion sensors such as IMU (inertial measurement unit) in addition to car's factory-installed yaw rate and speed sensors. The setup of stereo cameras is particularly interesting and important for DAS applications and it will be explained next.

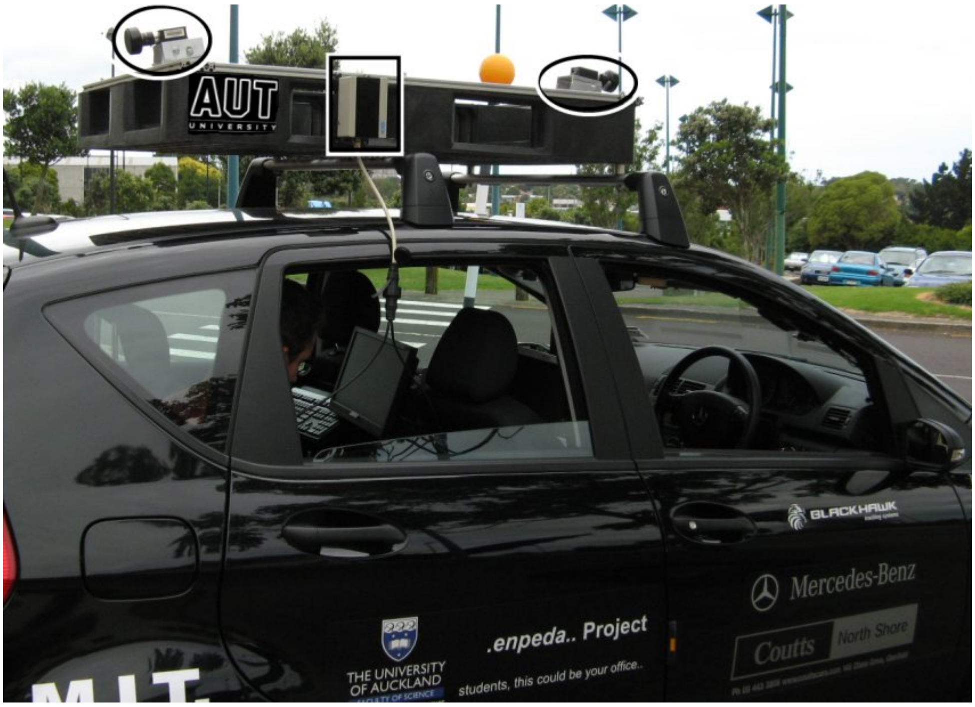

Figure 1.1: HAKA1 mounted with two fish-eye cameras *circled* and LIDAR *rectangled*.

1.3 Multiple Camera Systems

Computer vision utilizes cameras in the context of DAS to recognize and understand the surroundings. Nevertheless, the choice and suitability of a particular configuration depends obviously on the application that is intended such as parking, collision avoidance, lane departure, and so on. There are many options and configurations possible for setting up the cameras such as the number of cameras, the type of lenses or optics, fixed or moving rigs, overlapping field of view among others. It is very important and useful to be able to extract the depth or 3D data and thus those setups that allows depth will be considered here.

The depth or 3D data can be extracted using available *depth cues* . In general, there are several depth cues such as:

1. Binocular disparity (difference in images projected into one and another camera for example).

2. Perspective (convergence of parallel lines).

3. Size (object images get smaller when farther away).

4. Relative motion (object images move slower when farther away).

They are all useful and intuitive for humans but no so easy to implement for computer vision systems except for the *binocular disparity* which is only possible with binocular or *stereo vision*. Binocular disparity is the difference in image localization between both projections due to the different position of cameras or eyes. In simple terms, the left eye sees the same object slightly shifted (more pronouncedly the closer the object) to the left in comparison to the right eye, and vice versa. The distance or depth can be determined by simple triangulation.

A *monocular vision* or single camera is the simplest setup but also useful for lane detection for example. However, for many applications a multicamera setup is more appropriate since 3D information is needed. A *stereo camera* is the minimum multicamera configuration setup that allows to extract depth or 3D data by means of binocular disparity and triangulation. In general, there are also several ways of setting up a stereo system:

1. Single camera with a single lens shifted by translation or rotation (taking one image at one position and the second image shifted).

2. Single camera multiple lens (one camera uses a prism or similar optics to take two images from two separated lenses simultaneously such as the Mars Polar Lander stereo imager).

3. Multiple camera (different camera taking images with their own lenses).

For DAS purposes, the multiple camera is the preferred setup in terms of cost and application requirements. A single camera shifting is not useful since the images can not be synchronized when the camera is moving, and considering the relative low cost of cameras nowadays, it does not justify the complexity of the optics necessary in the single camera with multiple lens setup. Furthermore, the multiple camera setup can be varied such as fixed rigs or active rigs such as *PTU* (pan and tilt unit). PTU has advantages over a fixed rig but also extra cost and more complex to build and calibrate. On the other hand, fixed rig of stereo cameras have proven robust enough and able to provide useful data for DAS applications in general, and they will be the subject of this thesis.

The fixed stereo rig setup can be further varied depending on the choice of optics. Normal lenses are available and used, but fisheye or wide angle lenses have shown to be a good alternative for many applications such as intersection assistance or panoramic view.

The benefits of fisheye lenses are relatively low costs while providing a large field of view (FOV typically around 180°) with a single camera from one point of view at a single moment. Therefore, a stereo fisheye lens camera system can be used efficiently and quite accurately to cover a large field of view for 3D information [1]. Furthermore, experiments with stereo fisheye lenses in DAS vehicles have shown that it has the best calibration stability in comparison to other systems as well as better use of the image area [9].

Regardless of the optics used (normal, fisheye or catadioptrics), the stereo images must be *synchronized* (images taken a the same time) since the cameras could be moving rather fast (cameras are mounted on moving cars in the DAS context) as well as calibrated so the images can be *rectified* and processed easily to facilitate the search for stereo correspondence typically.

Rectification means in simple terms correcting the distortions caused by the optics (radial distortion mostly) and setup (misalignment, rotation among others). Formally, *epipolar rectification* is a geometric transformation of a pair of images resulting in every point in the scene being projected in the same row in both images.

1.4 Methodology

The testing and comparison of the selected algorithms can be done using two simple and straightforward methods: backprojection error and row misalignment error of rectified stereo images. These two methods provides a measure of the accuracy and robustness of the selected four calibration algorithms for normal and fish-eye lens.

Firstly, two different camera setups using the inside mounting for normal lens and the outside mounting for fish-eye lens are to be used for calibrating stereo cameras using the applicable calibration algorithms. Once the calibration is completed, the calibration parameters are to be compared and analyzed using the two methodologies outlined above.

Specifically, the OpenCV experiments setup in the DAS context with two different camera configurations and pixel resolutions using high quality optics (as normal lens) for the pinhole camera model. Additionally, the calibrated focal length parameter can be used to compare against the lens manufacturer's specification values for further confirmation of accuracy.The fish-eye lens calibration methods are tested next using high quality fish-eye optics and high pixel resolution cameras.

The backprojection error in pixel units can be compared directly for the same camera and lens configurations as well as the row misalignment error for which the stereo images are to be undistorted linearly and rectified using the calibration data.

Multiple sets of calibration images will be taken for the same camera and lens configuration to compare and analyze the robustness of the algorithms. The calibration results of the different sets of calibration images can be analyzed for convergence and error from the calibration data.

Finally, the ease of use and suitability of the calibration algorithms, the processes and steps needed to run the different calibration toolboxes while calibrating the experiment calibration images are noted and discussed at the end.

1.5 Organization of this Thesis

Calibration is an important and necessary step for stereo image rectification and the subject of this thesis. Some selected available calibration tools and methods will be compared and analyzed for robustness, accuracy and applicability in the context of DAS applications, and specifically the HAKA1 platform.

Chapter 2 presents a survey underlying the theory of calibration and rectification along with a comparison of recent methods for multiple camera systems. Chapter 3 presents the actual workflow of the calibration methods used for HAKA1. Chapter 4 addresses the methods used to check the accuracy of the calibration. Chapter 5 presents the results of the calibration using different cameras and optics available in HAKA1. Summary and conclusion follow in Conclusions 6.

Chapter 2

Basics for Multiple Camera Systems

This chapter explains the different issues to be dealt with when installing and using multiple cameras on an ego-vehicle, such as HAKA1. Firstly, the issues that arises when installing cameras in general, and camera synchronization issues mounted on fast moving platforms (HAKA1) will be discussed. Finally, it provides an overview of popular calibration methods from the literature for normal and fish eye cameras.

2.1 Camera Installation

The concept of *6D analysis or vision* is of relevance to DAS. 6D vision refers to the analysis of both stereo and motion to bring a consistent scene interpretation as well as extraction of objects from surroundings. Thus, in the context DAS and considering the relevance of 6D, we will need to consider not only the setup of multiple cameras but also the vehicle's movement factor as well.

Cameras are essential components of DAS by capturing the images needed to understand or at least be aware of the vehicle's surroundings while in motion. Thus the proper and stable camera installation is an important factor in 6D analysis as well. Namely, an improper and unstable installation of cameras might negate all the benefits of having high quality optics and camera sensors or even worse, render the images useless because of 'ghosts images' (unwanted and unintended reflections), obstructed views (e.g. rain) for example (see Figure 2.1).

There are available *active camera systems* which are cameras mounted on a motorized device that allows the camera to pan and tilt. Active camera systems have the benefit of allowing to cover a greater area of view than a camera with fixed orientation without sacrificing image resolution. Nevertheless, active systems are far more complicated and outside the scope of the *.enpeda..* project or this thesis so only conventional fixed orientation camera configuration would be considered here. Due to the research nature of the *.enpeda..* project there is an additional challenge in the setup of the cameras by requiring that the camera installation in HAKA1 to be flexible (non-fixed or permanent) but also stable.

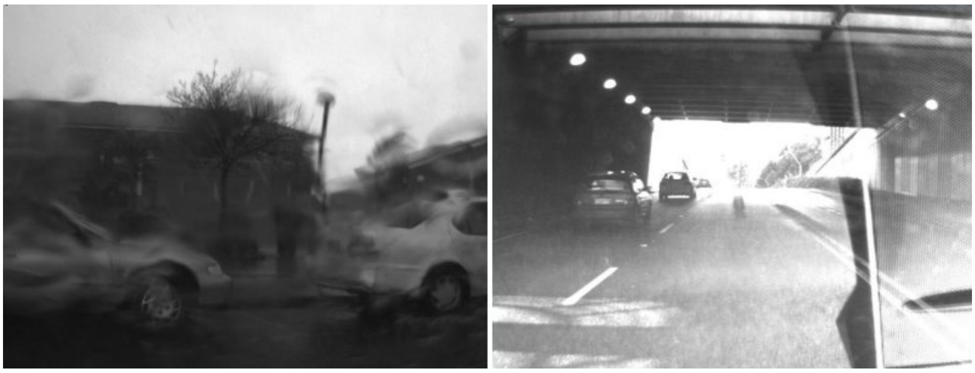

Figure 2.1: *Left:* rain. *Right:* dashboard reflected on the windshield.

The limited research regarding camera installation (see [9] for an example) points out that there are limited locations where stereo cameras can be mounted. Furthermore, the ideal locations of the cameras varies depending on the applications (intersection assistance, pedestrian detection, panoramic view and so forth). Nevertheless, it is obvious that few locations and orientations are possible in practice for installing the stereo camera since the cameras will be used in all weather and road conditions such as sunny, rainy, snowing, paved and unpaved roads and so forth. Additionally, it also has to be used at cruising speeds as well as in intersections which requires a good unobstructed field of view. As an example of limited options for installation is the front view (looking at the road ahead) applications which leaves the area behind the windscreen and inside the wipe clean area as the only viable option.

In short, the camera installation must be robust and stable over time but also practical enough so regular maintenance or mounting and dismounting of cameras or lenses can be done easily as the routine part of the research work. Thus the major factors that need to be considered for a camera setup in HAKA1 are that:

1. It allows a practical installation and maintenance of multiple cameras.

2. It is robust and stable over time and over all road conditions.

3. It provides a solution for different weather conditions or different FOV lenses.

A popular and easy installation solution for stereo cameras is a metal bar (which holds the stereo cameras) mounted to the windshield with suction cups (see Figure 2.2).

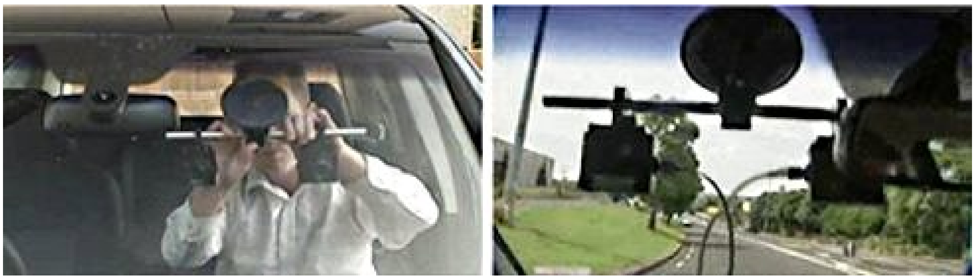

Figure 2.2: Stereo cameras attached with suction pads to windshield.

The benefits of using suction pads are that: (i) It can be installed inside the cabin behind a wipe clean area of the windscreen and (ii)) it is practical and easy to mount and dismount the whole camera setup. However, using the suction pads mounting has the very serious drawback of instability and limitations in the number of cameras. The setup is unstable due to the use of suction pads made of rubber which absorbs vibrations but also sudden roll or tilt movements of the car which are needed to be recorded. Also using suction pads limits the number of cameras that can be mounted simultaneously.

Finally, only interior (inside cabin) camera installation have been considered so far but also outside cabin or external installation are interesting and needed for certain applications such as panoramic view. There is not much research regarding this issue but in general similar requirements and constraints applies. Nevertheless, there is one difference with respect to inside camera installation which is the lack of a wipe clean windscreen to shield and provide a rain-free view. Usually cameras are mounted directly on roof racks which is not the best location for research purposes since it does not provide unobstructed views or allow for a practical and stable mounting of multiple cameras.

2.2 Camera Synchronization

Camera synchronization is an important factor for 6D vision and DAS in general since multiple cameras are used (at least two). Cameras record still images at a given frame rate to capture motion or moving objects. Hence multiple cameras in motion capturing moving images need to be synchronized in order to capture the image them at the same point in time. Synchronization is the equivalent of triggering the opening and closing of the shutter in different cameras simultaneously to ensure the cameras are taking the photo at the same time(or very closely to each other).

As mentioned earlier, synchronization is a very important factor in 6D vision and specifically because of the motion and speeds involved in typical DAS scenarios. And as a result, the displacement of objects is considerable from one shot to the next, so even delays of fractions of seconds in the synchronization can affect greatly. For example, a stereo sequence shot in the motorway where the ego-vehicle is traveling around 80 km/h would mean that even a small time lag between the left camera with respect to the right results in a difference of several meters in traveled distance.

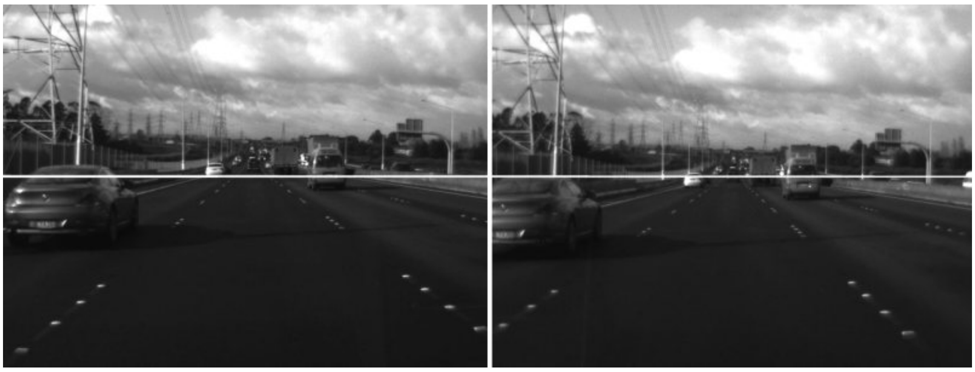

Figure 2.3: Stereo images taken by non-synchronized cameras.

In terms of image rows, the images of objects that should be in the same row in both cameras could be displaced several pixel rows from one camera to the other thus making difficult if not impossible the search of corresponding points along epipolar lines. Note in Figure 2.3 an example of non-synchronized stereo cameras where the white car in the left image is several rows up with respect to the right image.

There are two solutions available to synchronize multiple cameras: (i) Software synchronization (typically works for cameras sharing same Firewire bus but some manufacturers provide for cameras across different firewire buses). (ii) Hardware synchronization or by external trigger (more precise but also more cumbersome). It requires a square pulse generator for example that 'triggers' or signals to the camera to capture an image. It is also more accurate since it users its own clock unlike the software synchronization which depends on the PC operating system.

2.3 Calibration Methods for Pinhole-Type Cameras

As we all know, cameras takes images or 2D representations of real 3D objects projected in the image plane (which is typically the CCD or CMOS image sensors).

Hence a camera model is created to represent and to understand the geometrical and physical properties of a given camera. The most simple camera model is the *pinhole camera model* which uses a perspective projection and the details can be found in the literature such as Hartley and Zisserman's Multiple View Geometry book [12] or Klette, Schlüns and Koschan's Computer Vision book [18].

The pinhole model has however its limitations and assumptions which are not always true but the pinhole model is a good approximation of real normal cameras and is applicable in most cases. As mentioned earlier, each camera model has its own characteristic geometrical and physical properties or parameters such as focal length, principal point and so forth. The process of finding out those parameters in a given camera model is called *calibration*. Calibration has been studied and used in fields such as *photogrammetry* for a long time. Photogrammetry is basically extracting or calculating the 2D or 3D information from photographs (used extensively in aerial photo reconnaissance for example).

There are in general three type or methods of calibration that can be identified:

1. Test-range or *photogrammetric calibration* by correspondence of 3D objects and their 2D image feature points as in a planar *checkerboard* (see Figure 2.4 for examples of planar and 3D *calibration objects*).

2. *Non-metric calibration* by using geometric invariants such as parallel lines, plumb lines, vanishing points among others.

3. *Self calibration* by matching corresponding features from a sequence of images and using only the constraints of the camera model parameters.

Needless to say that no calibration method or technique is perfect and always applicable as all having its advantages and limitations. The issues and limitations generally found in calibration methods are as follows:

1. Requirement for a precise calibration objects with control points whose coordinates are precisely known such as 2D or 3D checkerboards for example.

2. Not extracting all parameters (partial calibration).

3. Noise and error sensitive.

4. Not fully automatic requiring user intervention or complex procedures.

Calibration considered so far has been limited to pinhole cameras but obviously it is also possible to calibrate fish eye cameras as well. The fundamental difference between the pinhole camera and fish eye camera is in the fact that the latter has higher radial distortion. Fish eye lenses in general have significantly more *lens*

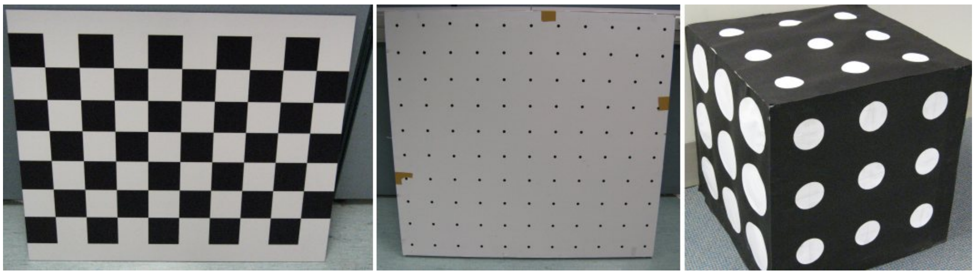

Figure 2.4: *Left:* planar checkerboard. *Center:* planar circular. *Right:* 3D circular.

distortion compared to normal lenses used in pinhole cameras. Distortion refers to any deviation in trajectory of the ray from the ideal model which are results of the lens design by the manufacturer or imperfections during manufacturing or assembly. The most evident distortion is the *radial distortion* such as barrel or pincushion which is a radially symmetric distortion around the center of the image. Also there is *tangential distortion* or asymmetric distortion which is less noticeable in high quality lenses. In general, radial distortion is the most prevalent type of distortion and noticeable even in pinhole cameras in general.

The main difference between the pinhole cameras and fish eye cameras is the tremendous radial distortion characteristic of fish eye lenses and as a result, they also have typically a large FOV. The difference in the magnitude of the distortion is the reason that fish eye lens cameras are so different from pinhole cameras and cannot be approximated with the perspective projection as done fairly well with pinhole cameras. Thus fish eye lenses need a different projection model and consequently a different calibration method or technique than pinhole cameras.

We will consider two examples of photogrammetry calibration using a calibration grid such as the *DLT calibration* method and Zhang's method [31]. Both methods are applicable to pinhole cameras and their main differences are the simplicity of DLT's method (in part due to the fact of being one of the earliest) dealing mostly with linear equations and disregarding the lens distortions (as it is presented here). On the other hand, Zhang's method is more recent and yields more robust and accurate results (including up to the fourth order radial distortion) while simpler and easier in gathering or collecting the calibration data. The DLT method will be described next to give an overview of a calibration method for pinhole cameras.

2.4 DLT's Basic Calibration Method

A basic calibration but well known method by *direct linear transform* or DLT which was developed in 1971 by Aziz and Karara [2] and gives a good overview of the calibration process for the pinhole model. The calibration process starts by taking several pictures of a calibration object such as the checkerboard rig at different poses or attitudes covering the FOV, as much as possible. The next step is to determine the image coordinates of all calibration points either manually or by image analysis.

Let us define the 3D point $\hat{X} = (X, Y, Z)^T$ in world coordinates projected into the image point $\hat{x} = (x, y, -f)^T$ (using the convention of image plane behind the focal point). We obtain the following transformations:

$$\begin{pmatrix} x - c_x \\ y - c_y \\ -f \end{pmatrix} = \begin{pmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{pmatrix} \begin{pmatrix} X + t_x \\ Y + t_y \\ Z + t_z \end{pmatrix} \tag{2.1}$$

where the principal point is = $(c_x, c_y, 0)$ and $x = \frac{f \cdot X_c}{Z_c}$ and $y = \frac{f \cdot Y_c}{Z_c}$.

The image points in camera coordinates X_c and Y_c are obtained from the homography of $\hat{X}$ to $\hat{X_c}$ translated and rotated afterwards as follows:

$$\hat{X_c} = \mathbf{R}(\hat{X} + \mathbf{T}) \qquad \text{with} \qquad \mathbf{T} = (t_x, t_y, t_z)^T = (-X_0, -Y_0, -Z_0)^T \tag{2.2}$$

$\mathbf{R}$ is the matrix representing R_x, R_y, R_z which are the rotations to the X, Y, Z axis respectively.

We obtain two linear equations expressed in terms of $(x - c_x)$ and $(y - c_y)$ from Equation (2.1) as follows:

$$\begin{aligned} x - c_x &= -f \cdot \frac{r_1(X - X_0) + r_2(Y - Y_0) + r_3(Z - Z_0)}{r_7(X - X_0) + r_8(Y - Y_0) + r_9(Z - Z_0)} \\ y - c_y &= -f \cdot \frac{r_4(X - X_0) + r_5(Y - Y_0) + r_6(Z - Z_0)}{r_7(X - X_0) + r_8(Y - Y_0) + r_9(Z - Z_0)} \end{aligned} \tag{2.3}$$

Equation (2.3) can be rewritten as:

$$\begin{pmatrix} X & Y & Z & 1 & 0 & 0 & 0 & 0 & -xX & -xY & -xZ \\ 0 & 0 & 0 & 0 & X & Y & Z & 1 & -yX & -yY & -yZ \end{pmatrix} \begin{pmatrix} L_1 \\ L_2 \\ \vdots \\ L_{11} \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \tag{2.4}$$

To solve the eleven unknown parameters $L_1, L_2, \ldots, L_{11}$, more than six calibration points (X, Y, Z) and their respective image points are needed to solve the Equation (2.4) Since six calibration points give us an overdetermined system of twelve linear equations. In general, about 100 points are needed in practice for stability of the solution.

The transformation parameters $L_1 \dots L_{11}$ which contain the intrinsic and extrinsic parameters can be calculated using the pseudo-inverse matrix of the overdetermined system of equations as follows:

The principal point (c_x, c_y) by:

$$c_x = \frac{L_1L_9 + L_2L_{10} + L_3L_{11}}{{L_9}^2 + {L_{10}}^2 + {L_{11}}^2}$$
$$c_y = \frac{L_5L_9 + L_6L_{10} + L_7L_{11}}{{L_9}^2 + {L_{10}}^2 + {L_{11}}^2}$$

The effective focal length $f = (f_x + f_y)/2$ by:

$${f_x}^2 = -c_x^2 + \frac{L_1^2 + L_2^2 + L_3^2}{({L_9}^2 + {L_{10}}^2 + {L_{11}}^2)^2}$$
$${f_y}^2 = -c_y^2 + \frac{L_5^2 + L_6^2 + L_7^2}{({L_9}^2 + {L_{10}}^2 + {L_{11}}^2)^2}$$

Regarding the extrinsic parameters, the rotation angles in R_x, R_y, R_z are calculated in two steps:

1. The unknown coefficients in matrix **R** are calculated using the equations: $L_1 - L_3, L_5 - L_7, L_9 - L_{11}$.

2. The individual yaw, tilt and roll angles are calculated from the coefficients $r_i, i = 1, \dots, 9$ of the matrix **R** calculated previously.

Finally, the projection center (X_0, Y_0, Z_0) is calculated from Equation (2.3) which gives a system of two equations with three variables.

$$\begin{pmatrix} a_1 & a_2 & a_3 \\ a_4 & a_5 & a_6 \end{pmatrix} \begin{pmatrix} X_0 \\ Y_0 \\ Z_0 \end{pmatrix} = \begin{pmatrix} A \\ B \end{pmatrix} \tag{2.5}$$

The solution for a_i, for $i = 1, \dots, 6$, and A, B is as follows is by taking two calibration points and their respective projected image points:

$$a_1 = (c_x - x)r_7 - fr_1$$
$$a_2 = (c_x - x)r_8 - fr_2$$
$$a_3 = (c_x - x)r_9 - fr_3$$
$$a_4 = (c_y - y)r_7 - fr_4$$

$$\begin{aligned} a_5 &= (c_y - y)r_8 - fr_5 \\ a_6 &= (c_y - y)r_9 - fr_6 \\ A &= Xa_1 + Ya_2 + Za_3 \quad \text{and} \quad B = Xa_4 + Ya_5 + Za_6 \end{aligned} \tag{2.6}$$

Finally, the parameters in Equation (2.1) can be found by solving all previous equations. The DLT method described here does not consider any distortion but a more refined camera model with distortion can be included as well. For the sake of simplicity, the distortion has not been included here though. The next section describes a better and more sophisticated calibration method that includes radial distortion as well.

2.5 Zhang's Planar Calibration

Zhang's method [31] uses at least two images of a planar checkerboard for example in different orientations by moving either the camera or the checkerboard with the advantage the movement could be arbitrary. Again Zhang estimates the initial parameters using a closed-form solution. Next, a non-linear optimization using the maximum likelihood estimation is performed.

The calibration steps can be summarized as follows:

1. Take several images of the checkerboard (in different orientations) and extract the control or feature points from the images.

2. Estimate the intrinsic and extrinsic camera parameters using the closed-form solution.

3. Estimate radial distortion coefficient by solving least-squares.

4. Refine parameters by using optimization tools.

Given the feature or control points as $\widetilde{M} = \begin{bmatrix} X, & Y, & Z, & 1 \end{bmatrix}^T$ and their image points as $\widetilde{\mathbf{m}} = \begin{bmatrix} x, & y, & 1 \end{bmatrix}^T$, the image formation can be expressed in homogeneous normalized coordinates as follows:

$$\widetilde{\mathbf{m}} = \mathbf{A} \begin{bmatrix} \mathbf{R} & \mathbf{T} \end{bmatrix} \widetilde{M} \qquad \text{with} \quad \mathbf{A} = \begin{bmatrix} \alpha & \gamma & c_x \\ 0 & \beta & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2.7}$$

where $\mathbf{A}$ is the camera intrinsic matrix with γ being the skew of image sensors, (c_x, c_y) the principal point coordinates, α and β are the scaling factors in X and Y axes respectively. $\mathbf{R}$ and $\mathbf{T}$ are respectively the rotation and translation parameters of the world coordinate system with respect to the camera coordinate system. In particular, the world coordinate system can be set to $Z = 0$ for all points by setting the model plane on $Z = 0$.

Let us denote r_i as the i^{th} column element of the rotation parameter $\mathbf{R}$ and $\mathbf{T}$ the translation vector and s an arbitrary scale factor, then Equation (2.7) can be expressed as follows:

$$s\widetilde{\mathbf{m}} = \mathbf{A}\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{A}\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{T} \end{bmatrix}\begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = \mathbf{A}\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{T} \end{bmatrix}\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (2.8)$$

Thus, a feature point $\mathbf{M}$ and its image point $\mathbf{m}$ is related by a homography $\mathbf{H}$ which is defined up to a scale factor λ (for simplicity assuming here there is no distortion) as in the following equation:

$$s\widetilde{\mathbf{m}} = \mathbf{H}\widetilde{M} \quad \text{with} \quad \mathbf{H} = \lambda\mathbf{A}\begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{T} \end{bmatrix} \quad (2.9)$$

The homography $\mathbf{H} = \begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix}$ can be computed by a non-linear optimization method such as the Levenberg-Marquardt algorithm which minimizes the Euclidean backprojection error (feature or control points projected into the image plane). Taking advantage that $\mathbf{r}_1$ and $\mathbf{r}_2$ are orthonormal, we obtain the following Equation (2.11) as two basic constraints on the intrinsic parameters using the notation $\mathbf{A}^{-T} = (\mathbf{A}^{-1})^T = (\mathbf{A}^T)^{-1}$:

$$\mathbf{h_1}^T\mathbf{A}^{-T}\mathbf{A}^{-1}\mathbf{h_2} = 0 \quad (2.10)$$

$$\mathbf{h_1}^T\mathbf{A}^{-T}\mathbf{A}^{-1}\mathbf{h_1} = \mathbf{h_2}^T\mathbf{A}^{-T}\mathbf{A}^{-1}\mathbf{h_2} \quad (2.11)$$

In other words, the homography $\mathbf{H}$ is a 3 x 3 matrix with eight-degrees of freedom as a projective transform and since there are 6 extrinsic parameters (three rotation and three translation), we can obtain only two constraints for the intrinsic parameters.

To solve the calibration equations, first the parameters can be estimated by a closed-form solution which is obtained by the symmetric matrix $\mathbf{B} = \mathbf{A}^{-T}\mathbf{A}^{-1}$ or also expressed as a six-dimensional vector $\mathbf{b} = (B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, B_{33})^T$.

We obtain the following equation using $\mathbf{h}_i = (h_{i1}, h_{i2}, h_{i3})^T$ as the i^{th} column vector of $\mathbf{H}$:

$$\mathbf{h}_i^T\mathbf{B}\mathbf{h}_j = \mathbf{v}_{ij}^T\mathbf{b} \quad (2.12)$$

where the six-dimensional $\mathbf{v}_{ij}$ is:

$$\mathbf{v}_{ij} = [h_{i1}h_{j1}, h_{i1}h_{j2}+h_{i2}h_{j1}, h_{i2}h_{j2}, h_{i3}h_{j1}+h_{i1}h_{j3}, h_{i3}h_{j2}+h_{i2}h_{j3}, h_{i3}h_{j3})^T \quad (2.13)$$

The two constraints given in Equation (2.11) map into the following two homogeneous equations:

$$\begin{bmatrix} \mathbf{v}_{12}^T \\ (\mathbf{v}_{11} - \mathbf{v}_{12})^T \end{bmatrix} \mathbf{b} = 0 \tag{2.14}$$

Considering now that there are n images taken, we obtain then n Equation (2.14) which can be written as the following equation where $\mathbf{V}$ is a $2n \times 6$ matrix and the equation gives a solution for $\mathbf{b}$ up to a scale factor when $n \geq 3$):

$$\mathbf{Vb} = 0 \tag{2.15}$$

The solution of Equation (2.15) is known as the eigenvector of $\mathbf{V}^T\mathbf{V}$ associated with the smallest eigenvalue). With the $\mathbf{b}$ estimation it is possible now to calculate the intrinsic parameters from the relation $\mathbf{B} = \lambda\mathbf{A}^{-\mathbf{T}}\mathbf{A}$ as follows:

$$\begin{aligned} \lambda &= B_{33} - [B_{13}^2 + c_y(B_{12}B_{13} - B_{11}B_{23})]/B_{11} \\ \alpha &= \sqrt{\lambda/B_{11}} \\ \beta &= \sqrt{\lambda B_{11}/(B_{11}B_{22} - B_{12}^2)} \\ \gamma &= -B_{12}\alpha^2\beta/\lambda \\ c_x &= \gamma \cdot c_y/\alpha - B_{13}\alpha^2/\lambda \\ c_y &= (B_{12}B_{13} - B_{11}B_{23})/(B_{11}B_{22} - B_{12}^2) \end{aligned} \tag{2.16}$$

and the extrinsic parameters as follows:

$$\begin{aligned} \mathbf{r}_1 &= \lambda\mathbf{A}^{-1}\mathbf{h_1} \\ \mathbf{r}_2 &= \lambda\mathbf{A}^{-1}\mathbf{h_2} \\ \mathbf{r}_3 &= \mathbf{r}_1 \times \mathbf{r}_2 \\ \mathbf{T} &= \lambda\mathbf{A}^{-1}\mathbf{h_3}. \end{aligned} \tag{2.17}$$

The intrinsic and extrinsic parameters obtained previously from Equations (2.16) and (2.17) are the initialization values for a bundle adjustment algorithm (Levenberg-Marquardt for example) of the following function:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \|\mathbf{m}_{ij} - \hat{\mathbf{m}}(\mathbf{A}, \mathbf{R}_i, \mathbf{T_i}, M_j)\|^2 \tag{2.18}$$

where the matrix $\mathbf{R}$ is related to the vector r by the Rodriguez formula and $\hat{\mathbf{m}}(\mathbf{A}, \mathbf{R}_i, \mathbf{T_i}, \mathtt{M}_j)$ is the projection of point M_j in image i as in Equation (2.9).

Finally, the lens distortion has been neglected so far but at this point the radial distortion is taken into account (since it is the predominant distortion while tangent distortion remains neglected). Let (x, y) the coordinates of the ideal (if there were no distortion present) image coordinates and $(\breve{x}, \breve{y})$ the corresponding real image coordinates then the distortion can be expressed as a polynomial:

$$\begin{aligned} \breve{x} &= x + x[k_1(x^2+y^2) + k_2(x^2+y^2)^2] \\ \breve{y} &= y + y[k_1(x^2+y^2) + k_2(x^2+y^2)^2] \end{aligned} \tag{2.19}$$

with the radial distortion coefficients initialized to zero ($k_1 = k_2 = 0$) for initialization and further refined by iterative minimization of the average Euclidean distance of the projected and observed image points using an overdetermined system of linear equations. Alternatively, Zhang proposed to include the distortion parameters into the error in Equation (2.18) to estimate simultaneously due to the slow convergence of the iterative minimization when calculating the distortion coefficients.

2.6 Calibration Method for Fish Eye Cameras

The calibration methods covered so far are limited to the pinhole camera model because it assumes perspective projection. The pinhole camera model for normal lenses models the projection of a 3D scene into an image plane. This modeling is applicable because normal lenses have smaller FOV but it is not appplicable with fish-eye lenses because the resulting image plane is an infinite plane if the perspective projection is used as well.

From the pinhole camera projection equation we have:

$$x = \frac{fX}{Z}, \quad y = \frac{fY}{Z} \qquad \text{with} \quad r = f \tan\phi \tag{2.20}$$

where f = focal length, (x, y) = image point coordinates and (X, Y, Z) = scene point coordinates or r = distance between image point and principal point and ϕ = ray incidence angle (between incoming ray and principal axis). We can see that r goes to ∞ when ϕ gets closer to $\frac{\pi}{2}$. Or it can be interpreted alternatively that because of the tremendous distortion of fish-eye lenses, the approximation by perspective projection no longer holds as in the case of the pinhole camera model.

Lens manufacturers design lenses according to the application with different optical characteristics as FOV, *projection model*, *distortion model* and forth. The projection model refers to the ideal projection while distortion model describes the actual deviation from the (ideal) projection model.

The most common projection models are as follows:

$$r = f \tan\phi \tag{2.21}$$

$$r = f \sin \phi \tag{2.22}$$

$$r = f\phi \tag{2.23}$$

$$r = 2f \tan(\phi/2) \tag{2.24}$$

$$r = 2f \sin(\phi/2) \tag{2.25}$$

where (2.21) is perspective, (2.22) is orthogonal, (2.23) is equidistance, (2.24) is stereographic, and (2.25) is equisolid perspectives respectively. Fish-eye lenses usually follow equidistance, stereographic or equisolid projection models but not exactly so in general fish-eye lenses can be approximated as well by a polynomial equation:

$$r(\phi) = k_1\phi + k_2\phi^3 + k_3\phi^5 + \dots \tag{2.26}$$

where $k_1, k_2, k_3, \dots$ are parameters to be determined by calibration. The traditional polynomials such as the Zhang's method (pinhole camera model) fourth order polynomial is insufficient due to the tremendous distortion in fish-eye lenses. Instead, higher order polynomial is needed for fish-eye lenses, up to ninth order in the case of Equation (2.26).

Using higher order polynomial for modeling the distortion in fish-eye lenses results in more unknown coefficients to be determined by calibration. Namely, modeling the distortion with Equation (2.26) leads to five radial coefficients $k_1, \dots, k_5$ to be determined by calibration instead of just two k_1, k_2 as in Zhang's method. Nevertheless, there is a better alternative to polynomial model which is the division polynomial model since it requires fewer terms or unknown coefficients to be determined (one or two coefficients are good enough for most cases). The division model

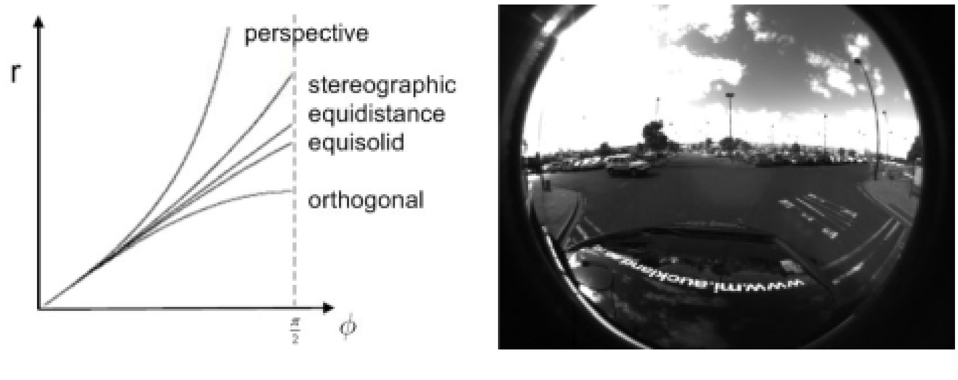

Figure 2.5: *Left:* projections for fish-eye. *Right:* fish-eye image.

proposed by Fitzgibbon [7] has the form:

$$\mathbf{x_u} = \frac{\mathbf{x_d}}{1 + k_1 r^2 + k_2 r^4 + \ldots + k_p r^{(2p)}} \tag{2.27}$$

where $r^2 = (x_d^2 + y_d^2)$ is the pixel radius, $\mathbf{x_u} = (x_u, y_u)$ and $\mathbf{x_d} = (x_d, y_d)$ are the images points undistorted and distorted respectively and $2p$ is the polynomial order in the general form.

In general there are two approaches for fish-eye lenses calibration which are: (i)using the pinhole camera and correcting fish-eye lens distortion (with higher order polynomials for example) afterwards and (ii) by modeling the fish-eye projection directly. The first method (i) is limited to relatively small FOV (less than π since when ϕ approaches $\pi/2$ the image point goes to ∞). Therefore, the second approach (ii) of modeling directly the fish-eye lens projection is more general. Kannala and Brandt [17] proposed a generic camera model for fish-eye lenses that utilizes the polynomial model (using division model would yield to fewer parameters but the calibration method remains basically the same).

This generic model calibration procedure consists of four steps:

1. Initialization of internal parameters.

2. Computation of homography (back-projection).

3. Initialization of external parameters.

4. Minimization of projection error.

The initialization of internal parameters $k_1, k_2, k_3, k_4, u_0, v_0$ can be done by finding the k_1, k_2 parameters in the model $r(\phi) = k_1\phi + k_2\phi^3$ that fits best the lens focal length and FOV data, k_3, k_4 by estimating from $k_3 = x_{max}/r_{max}$ and $k_4 = y_{max}/r_{max}$ when the image is a full frame for example, and finally the c_x and c_y center of distortion as approximately the image center.

The computation of homography is done by backprojecting the image points $m_j^i = (u^i, v^i)^T$ into the points $\tilde{x}_j^i$ located on the unit sphere centered in the camera coordinates as follows:

$$s\tilde{x}_j^i = \mathbf{H_j} x_p^i \tag{2.28}$$

where $\mathbf{H_j}$ is the planar homography and $x_p^i = (X^i, Y^i, 1)^T$ the calibration's control points coordinates for the view j. The homography $\mathbf{H_j}$ can be calculated by backprojecting the control points as mentioned earlier and estimating initially $\mathbf{H_j}$ from the correspondence $\tilde{x}_j^i \leftrightarrow x_p^i$ and defining $\hat{x}_j^i = \mathbf{H_j} x_p^i / \|\mathbf{H_j} x_p^i\|$ as the image of x_p^i by $\mathbf{H_j}$. Finally, refining $\mathbf{H_j}$ by minimization of $\sum_{j=1}^{N} \sin^2 \alpha_j^i$ (where α_j^i is the angle between $\tilde{x}_j^i$ and $\hat{x}_j^i$).

The initialization of external parameters from the homographies $\mathbf{H_j} = \left[\mathbf{r}_j^1, \mathbf{r}_j^2, \mathbf{r}_j^3 \mathbf{T}_j\right]$ up to a scale and:

$$\begin{aligned}\mathbf{r}_j^1 &= \lambda_j h_j^1 \\ \mathbf{r}_j^2 &= \lambda_j h_j^2 \\ \mathbf{r}_j^3 &= r_j^1 \times r_j^2 \\ \mathbf{T}_j &= \lambda_j h_j^3\end{aligned}$$

where $\lambda_j = (\mathbf{H_j}^{3,3}) / \|h_j^i\|$ and using the singular value decomposition (SVD) to calculate the closest orthogonal matrices (Fröbenius norm as metric) as the initial values of $\left[\mathbf{r}_j^1, \mathbf{r}_j^2, \mathbf{r}_j^3\right]$.

Finally, the camera parameters are refined by using Levenberg-Marquardt to minimize of the projection error or sum of squared distances d (between the measured and calculated control points) as follows:

$$\sum_{i=1}^{M} \sum_{j=1}^{N} d(m_j^i, \hat{m}_j^i)^2 \tag{2.29}$$

2.7 Rectification

As mentioned earlier, the concept of image rectification is important in 6D vision since it simplifies the finding of stereo correspondence in images. The rectification is well known and researched topic, as for example [11]. When using stereo cameras for example, the ideal case is to have them in *standard stereo geometry* to get better results. Standard or coplanar stereo geometry refers to the ideal setup of two cameras (pinhole cameras for example) that have:

1. Same focal length,
2. Collinear image rows,
3. Coplanar image plane,
4. Same image plane's size,
5. Parallel optical axes.

Standard stereo geometry gives better results because the stereo correspondence in rectified images is reduced to a one-dimensional search due the *epipolar constraint*. Epipolar constraint refers to the fact that image rows or *epipolar lines* are collinear or more specifically that the row y in the left image is collinear with row y' in the right image for example.

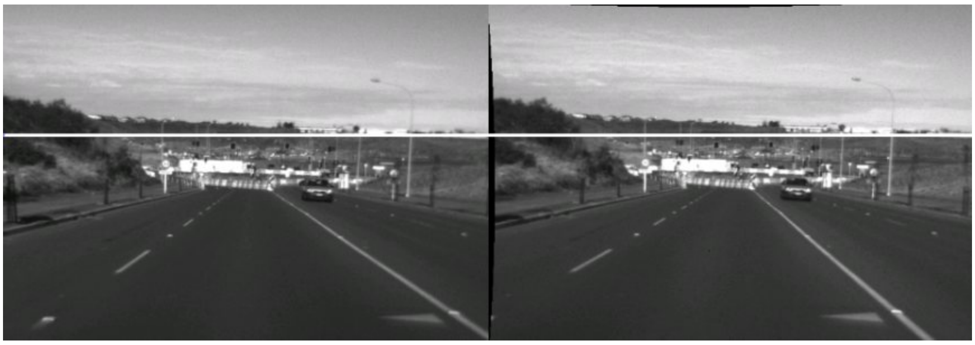

Figure 2.6: Rectified stereo image with an epipolar line (drawn as white line).

In standard stereo geometry a scene or a 3D point $\hat{X} = (X, Y, Z)$ is projected into the image $\hat{x}$ where b is the *baseline* (a line that intersects the optical centers of both cameras) as follows:

$$\hat{x}_l = \frac{f}{Z}(X, Y) \qquad \text{and} \qquad \hat{x}_r = \frac{f}{Z}((X - b), Y) \tag{2.30}$$

More formally, the *epipolar geometry* relates to the geometric relationship between two perspective views of the same 3D scene and created by the intersection of the image planes and the planes that have the baseline as axis. Thus, epipolar line is the intersection between an epipolar plane and the image plane with the epipolar lines intersecting in the epipole. An *epipolar plane* is any of the planes that intersect the baseline.

Before rectification can be done, the stereo cameras are assumed to be calibrated by any of the calibration methods presented in Section 2.3 to determine the camera internal parameters and relative positions.

Rectification in simple terms is finding two new *virtual cameras* $\widetilde{\mathbf{H}}$ and $\widetilde{\mathbf{H}'}$ from $\mathbf{H}$ and $\mathbf{H}'$, where $\mathbf{H} = \mathbf{A}[\mathbf{RT}]$ and $\mathbf{H}$ and $\mathbf{H}'$ are the given homographies of the real stereo camera (see Figure 2.7). Virtual cameras are ideal cameras with parallel optical axes, no distortion, and independent of the real cameras used so the resulting rectified images are independent as well. Rectification can be visualized as the 3D retroprojection or mapping of the real images into virtual stereo cameras. Alternatively, rectification is a process that moves the real cameras over their optical centers to make them coplanar resulting in epipoles at infinite and parallel epipolar lines. Also, bringing the baseline parallel to the horizontal axis ensures that the epipolar lines are horizontal as well.

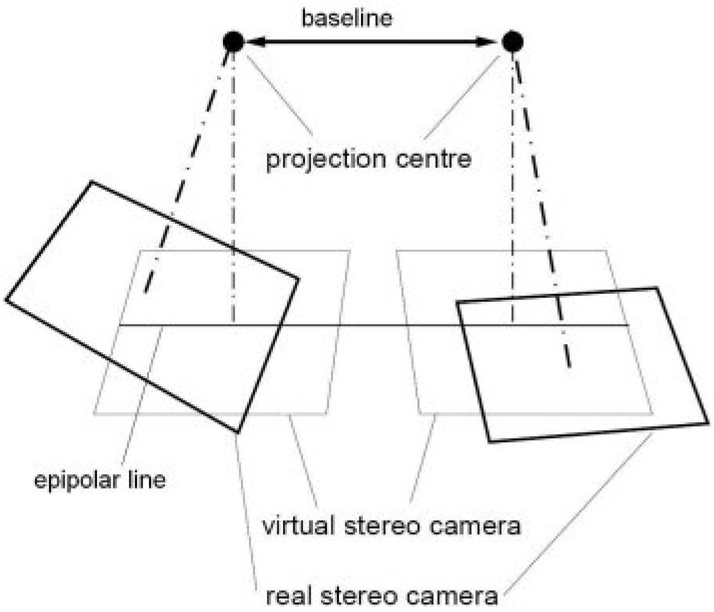

Figure 2.7: Rectification.

Let us examine next a common stereo rectification method used in Bouguet's calibration toolbox [5]. In this toolbox, the rectification aims to both minimize the retroprojection changes and maximize the common view area. Let us define the rotation and translation matrices between stereo images as $(\mathbf{R}, \mathbf{T})$. The changes in the retroprojection is minimized by splitting the matrix $\mathbf{R}$ in half between the left and right camera, which produces the rotation matrices r_l, r_r for the left and right camera respectively. The rotation r_l, r_r makes the cameras coplanar since their principal rays became parallel but still not in standard stereo geometry.

Let us define $\mathbf{R_{rect}} = [\mathbf{e_1}, \mathbf{e_2}, \mathbf{e_3}]^T$ the rotation matrix that will take the left camera into the virtual camera position but also align the epipolar lines. The epipole vector e_1 can be calculated by taking the principal point (c_x, c_y) as the origin of the left image with the direction of the epipole along the translation vector $\mathbf{T}$ while e_2 is chosen to be orthogonal to the principal ray, and finally e_3 is orthogonal to e_1 and e_2 as follows:

$$\mathbf{e_1} = \frac{\mathbf{T}}{\|\mathbf{T}\|}$$

$$\mathbf{e_2} = \frac{[-T_y, T_x, 0]^T}{\sqrt{T_x^2 + T_y^2}}$$
$$\mathbf{e_3} = \mathbf{e_1} \times \mathbf{e_2}$$

The epipolar row alignment is done by calculating $\mathbf{R_l}, \mathbf{R_r}$ from $\mathbf{R_{rect}}$ as follows:

$$\mathbf{R_l} = \mathbf{R_{rect}} \cdot r_l$$
$$\mathbf{R_r} = \mathbf{R_{rect}} \cdot r_r$$

Finally, the rectified left and right camera matrices $\mathbf{A_{rect-l}}, \mathbf{A_{rect-r}}$ can be calculated as follows:

$$\mathbf{H_l} = \mathbf{A_{rect-l}}\mathbf{\hat{H}_l} = \begin{bmatrix} f_{xl} & \gamma_l & c_{xl} \\ 0 & f_{yl} & c_{yl} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$
$$\mathbf{H_r} = \mathbf{A_{rect-r}}\mathbf{\hat{H}_r} = \begin{bmatrix} f_{xr} & \gamma_l & c_{xr} \\ 0 & f_{yr} & c_{yr} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & T_x \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

where $\mathbf{H_l}, \mathbf{H_r}$ are the homography of left and right cameras respectively, and similarly γ_l, γ_r the skew factors, f_l, f_r the focal lengths, and c_l, c_r the principal points.

Chapter 3

Advanced Camera Configuration Calibration Methods

This chapter informs about the standard routines to be carried out when setting up a multi-camera system (e.g., in a test vehicle such as HAKA1) for stereo image analysis. The steps to be performed include the installation of the cameras, their geometric calibration, the derived rectification procedures, and some testing of image recording for the different options of multi-camera set-ups.

3.1 Camera Installation

In chapter 2 we presented the three requirements for a successful multiple camera installation setup:

1. It allows a practical installation of multiple cameras.
2. It is robust and stable over time and all road conditions.
3. It provides a solution for different weather conditions or different FOV lenses.

As pointed out earlier in Section 2.1, a suction pad is not a good solution for multiple cameras since it is unstable among other issues. Installing multiple cameras (more than two cameras) would require an optical bench to provide with an unobstructed view for the cameras but also without blocking the driver's view as well. Mounting more than two cameras on suction pad is an impractical solution as we can see in Figure 2.2. Alternative camera mounting locations have been used such as the car's front grill to overcome the visibility obstruction issue [4] .

We learned from experience while working on the *enpeda..* project that despite the challenges, the multiple camera configuration is needed as such since it is necessary to have multiple cameras installed simultaneously even though not all cameras might be used or recording at the same time. Also because we wish very often to keep some or all cameras 'fixed' in a particular configuration or installation in order

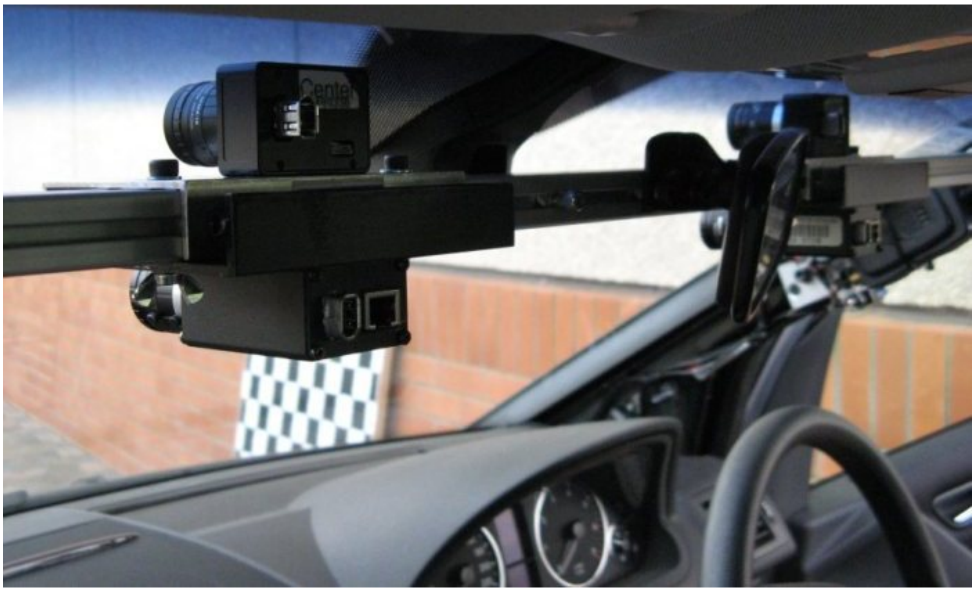

Figure 3.1: Calibration setup for Basler and Firefly MV stereo cameras with 9 mm lenses.

to reuse the cameras over time for comparison or just to keep them in the same configuration and installation for different weather conditions or multiple experiments.

At the same time, the camera installation also must not be permanent but practical and accessible in order to service the cameras from time to time such as adjusting their focus or changing lenses, or even for dismounting or mounting the cameras. As it is also noted in Section 2.1, there are certain applications such as panoramic view that requires necessarily an external installation because of the visual obstructions inside the cabin caused by the 'pillars' or unwanted reflections of the glass windows. Obviously, the external camera installation must satisfy the same requirements (internal cameras installation) of a stable and practical platform for the installation of multiple cameras.

Additionally, it would be also greatly desirable that external installation offered the same protection and rain-free view offered by the windshield but in practice it is still a challenging issue. The reason the rain remains a challenge is because the wiper provides a rain-free view but at the cost of a view intermittently obstructed by the wiper's travel. The wiper's visual obstruction has not been solved satisfactorily yet and it is one of the areas for future work, since rain affects vision and DAS should

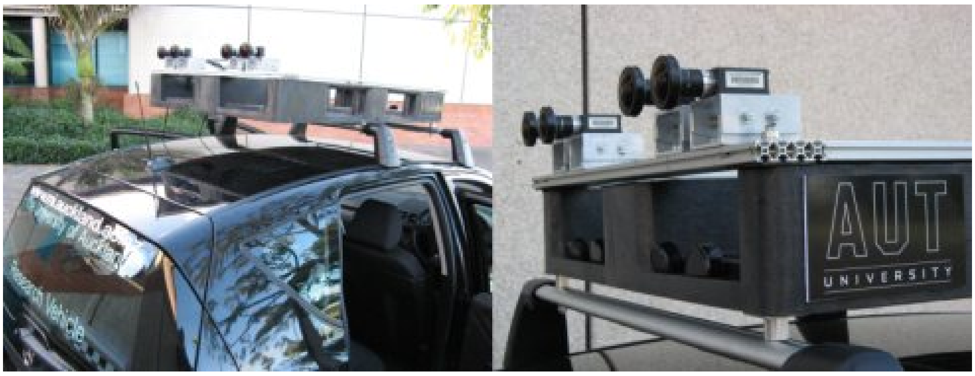

Figure 3.2: Calibration setup for Basler and Firefly MV stereo cameras with fish eye lens on roof mounting.

handle rainy conditions as well.

In summary, a solution that addresses the issues above have been implemented in HAKA1 with two setups manufactured with prefabricated aluminum bars with groves. This aluminum bar is called ′channel′ in the industry jargon.

1. The first setup inside the cabin for all weather condition is an horizontal mounting bar made of a 20mm by 40mm channel mounted between the ′A′ pillars (columns that support the roof and windshield). Figure 3.1 shows the final camera mounting bar.

2. The second setup for external camera mounting is a rectangular frame made of four 20mm by 80mm channels forming a rectangular frame resting horizontally on top of the roof racks (Figure 3.2.)

The design and location of the interior mounting bar and roof frame meets the three requirements outlined at the beginning. However, the design was not straightforward but evolved as unanticipated challenges became apparent during the experiments. The challenges are: (i)unwanted reflections inside cabin, (ii) wobbling or cantilevering movements in the roof mounting frame. The reflection issue (i) was underestimated at the beginning and addressed by designing a mounting bracket that allows the cameras as closest as possible to the windshield since doing that minimizes the dashboard reflections. The challenge was that the windshield is curved unlike the mounting bar which is straight. So the solution for mounting the cameras closest to the windshield regardless of its location (middle or end of windshield for example)was to design a mounting bracket with multiple fittings that allows the camera to be closer to the windshield as possible (Figure A.3. A close view of the bracket design and how is mounted on the mounting bar).

Nevertheless, the mounting bracket solution did not eliminate the unwanted reflections satisfactorily (Figure 2.1). The unwanted reflections affected quite significantly while evaluating algorithms and the scene when recording with the cameras installed inside the cabin. Thus a final solution for eliminating or reducing the reflections was needed urgently and finally the current solution was implemented. The current solution is covering the underside of the camera up to the windshield by stretching a piece of black, uniform texture fabric held with small suction cups (Figure 3.3). The fabric solution works better than paper (another common solution) since fabric can be folded or removed easily to access the lenses.

The second challenge (ii) while designing the external camera mounting is the vibration or cantilevering of the long section of the frame. Cantilevering is the flexing movement when the long and thin channels are supported on thin support points

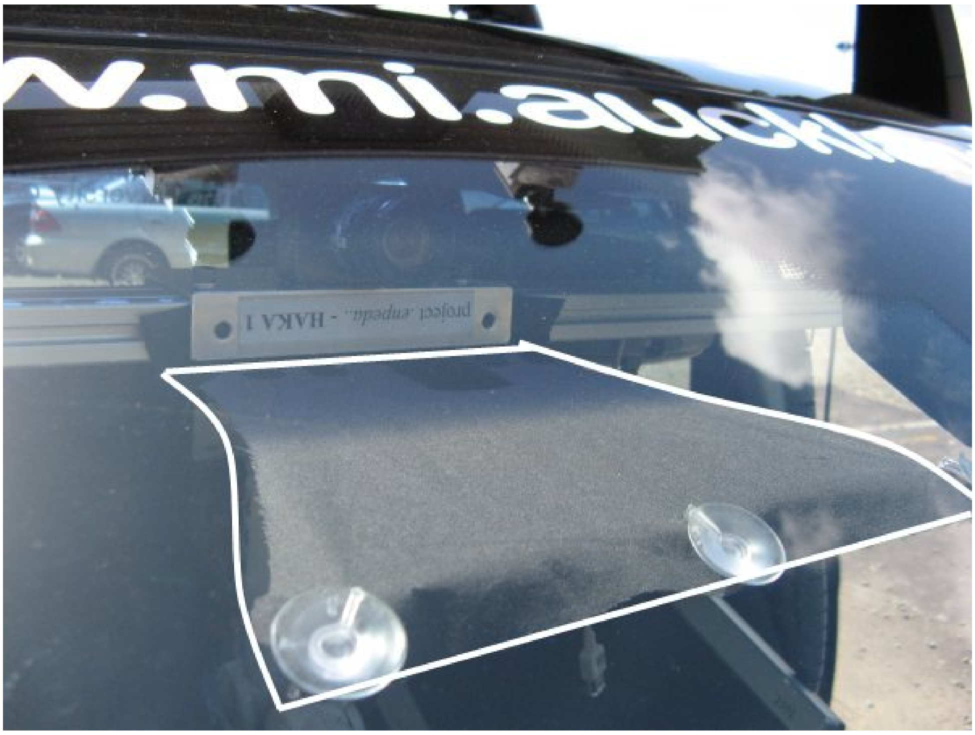

Figure 3.3: Black fabric for anti-reflection (marked with white borders).

such as the car roof racks. Even small weights such as the cameras and the camera brackets fitted at the end of frame amplified this flexing movements or vibrations. The solution was to mount the channel frame on top of another rigid but lightweight support frame such as the black plastic pallet visible in Figure 3.2 (roof frame made of channels for mounting cameras externally). The support frame also provided rigidity to torsion along the opposite corners as well as cantilevering.

Finally, special brackets have been for the externally mounted cameras to allow for some limited rotation (tilt and yaw) to orient the cameras front and down to the road (the area of interest). Furthermore, the external frame allows for all around view and mounting of cameras and currently two sets of stereo cameras with fish-eye lenses can be mounted facing forward and backward for full stereo view of HAKA1′s front and rear (see Figures A.4 and A.5).

3.2 Camera Synchronization

Camera synchronization is a very important for DAS since the multiple cameras are mounted on vehicles that can travel at high speed on motorways for example (as explained in Section 2.1). The synchronization problem emerged early on when the rectified stereo images taken while driving showed that object′s image displaced several pixel rows.

The *synchronization* of multiple cameras is normally checked by taking a shot or series of shots of a CRT screen that displays a special calibrated test pattern (see Figure 3.4). The test pattern is drawn by a software readily and freely available online (we use one called *synctest* created by Peter Wimmer [30]). Knowing the refresh rate of the CRT monitor and some parameters entered by the user, it is easy to calculate the time delay between multiple cameras. The synchronization issue depends on the context and its particular application. Few fractions of a seconds might be good enough in general but unacceptable for DAS applications which require ranges of less than 1/100 s to be considered ′synchronized′ for DAS purposes.

There are two solutions available to the synchronization issue as mentioned in Section 2.2:(i) *software synchronization*, (ii) *hardware triggering*. The software synchronization solution (i) was implemented first in the *enpeda..* project. One software solution was available thanks to one of the *enpeda..* project partners. They made available their camera or image recording application which synchronizes multiple cameras within milliseconds range. But there are limitations with this specific software application:

1. Limited to standard video formats (DCAM Format 7 excluded).
2. All cameras must be connected to the same port.

3. The overall bandwidth (resolution and frame rate) limited due to multiple cameras sharing the same port. For example, for a Firewire A (400 Mbps) port with stereo cameras connected, only about half of that is available for each camera and so on (more precisely, only around 160 Mbps is available for data because of the 20% overhead in the Firewire A protocol).

Another software solution is provided by manufacturer PointGrey worked for

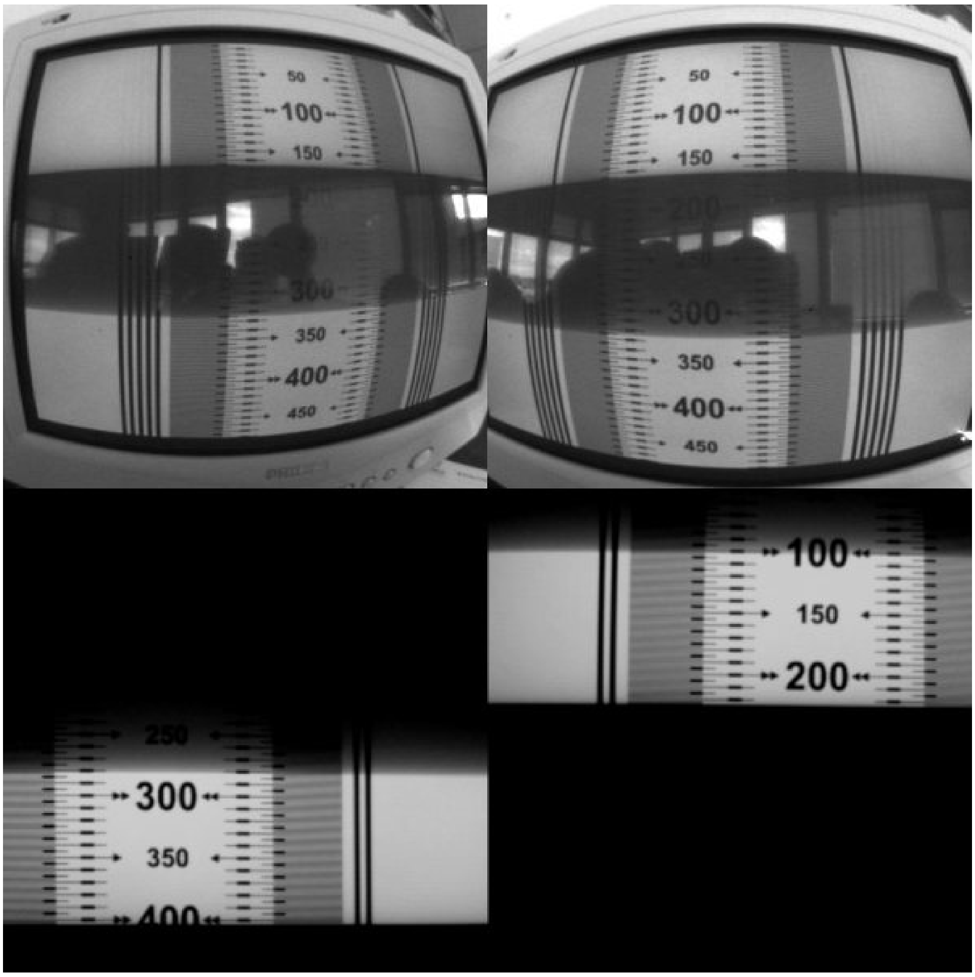

Figure 3.4: *Top:* stereo Basler AF602 cameras with fish-eye lenses synchronized with DT9817-H. *Bottom:* sample of non-synchronized stereo images.

its line of cameras such as the Firefly MV for all video formats and across different ports but limited to Windows OS only. On the other hand, Basler does not provide any software synchronization solution for its line of cameras.

However, we needed a universal and accurate solution to synchronize all the cameras (PointGrey and Basler) in a multiple camera configuration connected across different Firewire ports (ideally one camera connected to one Firewire port to maximize the resolution and frame rate). The solution that meets these requirements is by hardware triggering (ii) with the an A/D signal generator utilized as external trigger. The external trigger device used is the *DT9817-H* module manufactured by Datatranslation (a multi function A/D signal acquisition and generator) and controlled by an application created using Datatranslation's SDK. In particular, the Basler AF602 cameras required an external trigger for synchronization since they must be connected to different ports to use full resolution available at highest frame rate possible. The test results shows that the synchronization is in the order of milliseconds when using DT9817-H as the external trigger in four Basler cameras.

The external trigger DT9817-H module controls the frame rate and the synchronization by sending square pulses at the frequency set in the application. Additionally, DT9817-H allows to control the shutter time by varying the length or duration of the pulse that controls the frame rate. But the results of experiments have shown that it is better to keep the square pulse length or duration constant (about half of the length or duration of the total length from pulse to pulse) and set the shutter time by software instead. The reason is that the dynamic range in illumination from a very sunny afternoon to late afternoon is to great. Thus a settings that works fine in low light conditions such as late afternoon could be too bright to control in sunny conditions causing the cameras to flicker and lose images.

3.3 Overview of Pinhole Camera Calibration Methods

The calibration of pinhole cameras (a static setup and in the context of DAS) is practical and quite robust and accurate using planar checkerboard as the large number of available methods and recent research shows, with two very well known methods shown earlier in Section 2.3. One of the most popular calibration method is the OpenCV calibration not only because of its widespread usage but also practical by using planar checkerboard and more importantly, shown to be robust and accurate. OpenCV made its calibration toolbox readily available since it comes included with the *OpenCV* library (a popular open source computer vision library developed by Intel).

The OpenCV calibration method is based on the work of Bouguet's calibration

[5] which has been also implemented in Matlab. Thus, the Bouguet's calibration (either in the OpenCV or Matlab implementation) is the widely used for the calibration of pinhole model cameras in many settings such as laboratory, autonomous robots, DAS among others.

Finally, Bouguet extended this calibration toolboxes later on to calibrate fish-eye cameras as well (which is included as one of the fish-eye calibration methods in Section 3.5). Bouguet's calibration in the OpenCV implementation form will be covered next not only because of its popularity but also since it is the basis for one of the selected fish-eye calibration methods.

3.4 OpenCV Calibration

The OpenCV calibration implementation was greatly influenced by Zhang's calibration method (shown earlier in Section 2.5) and in lesser form by Heikkila's four step calibration method [13]. Bouguet's camera model (which is utilized in the OpenCV calibration) is similar to Heikkila's camera model. On the other hand, Bouguet's method consists of two steps similar to Zhang's method: (i) closed form solution of the calibration parameters, (ii) non-linear optimizations. The first step (i) is very similar to Zhang's initial estimation of the planar homography except here it exploits the orthogonality of vanishing points.

The second step calculates the parameters by minimizing the Euclidean back-

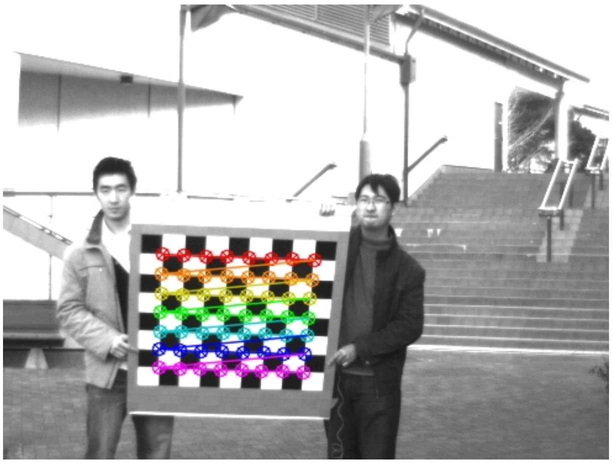

Figure 3.5: Checkerboard corner detection with OpenCV calibration toolbox.

projection error also similar to Zhang's method but using a different optimization (example of corner detection Figure 3.5).

OpenCV calibration solves four intrinsic parameters f_x, f_y, c_x, c_y and five distortion parameters composed of three radial k_1, k_2, k_3 and two tangential p_1, p_2 (similar to Heikkila's internal camera model). The calibration parameters are calculated first disregarding any distortion. Let us define the homography $\mathbf{H}$ as column vectors for each view where r_1, r_2 are the two rotation matrix columns, $\mathbf{T}$ the translation vector, s a scale factor, and $\mathbf{A}$ the camera intrinsic matrix.

$$\mathbf{H} = \left[\begin{array}{ccc} \mathbf{h_1} & \mathbf{h_2} & \mathbf{h_3} \end{array} \right] = s\mathbf{A} \left[\begin{array}{ccc} r_1 & r_2 & \mathbf{T} \end{array} \right] \tag{3.1}$$

We obtain from Equation (3.1) the following equations for $r_1, r_2, \mathbf{T}$ (where $\lambda = \frac{1}{s}$):

$$r_1 = \lambda \mathbf{A}^{-1}\mathbf{h_1}, \quad r_2 = \lambda \mathbf{A}^{-1}\mathbf{h_2}, \quad \mathbf{T} = \lambda \mathbf{A}^{-1}\mathbf{h_3} \tag{3.2}$$

Since the rotation vectors r_1, r_2 are orthonormal, then $r_1^T r_2 = 0$ and also $\|r_1\| = \|r_2\|$ or $r_1^T r_1 = r_2^T r_2$. Thus, from Equation (3.1) and by exploiting the orthonormality of r_1, r_2, the first constraint is derived as:

$$\mathbf{h_1}^T \mathbf{A}^{-T}\mathbf{A}^{-1}\mathbf{h_2} = 0 \tag{3.3}$$

the second constraint can be derived by replacing r_1 and r_2 as follows:

$$\mathbf{h_1}^T \mathbf{A}^{-T}\mathbf{A}^{-1}\mathbf{h_1} = \mathbf{h_2}^T \mathbf{A}^{-T}\mathbf{A}^{-1}\mathbf{h_2} \tag{3.4}$$

To simplify notation, let us set:

$$\mathbf{B} = \mathbf{A^{-T}A^{-1}} = \begin{bmatrix} B_{11} & B_{12} & B_{13} \\ B_{12} & B_{22} & B_{23} \\ B_{13} & B_{23} & B_{33} \end{bmatrix} \tag{3.5}$$

where the general closed-from solution of $\mathbf{B}$ is:

$$\mathbf{B} = \begin{bmatrix} \frac{1}{f_x^2} & 0 & -c_y/f_x^2 \\ 0 & \frac{1}{f_x^2} & -c_y/f_y^2 \\ \frac{-c_x}{f_x^2} & \frac{-c_y}{f_y^2} & (\frac{c_x^2}{f_x^2} + \frac{c_y^2}{f_y^2} + 1) \end{bmatrix} \tag{3.6}$$

By replacing with $\mathbf{B}$ in Equation (3.4) and also noting that $\mathbf{B}$ is symmetric, we obtain that:

$$\mathbf{h_i^T B h_j} = \mathbf{v_{ij}^T b} = \begin{bmatrix} h_{i1}h_{j1} \\ h_{i1}h_{j2} + h_{i2}h_{j1} \\ h_{i2}h_{j2} \\ h_{i3}h_{j1} + h_{i1}h_{j3} \\ h_{i3}h_{j2} + h_{i2}h_{j3} \\ h_{i3}h_{j3} \end{bmatrix}^T \begin{bmatrix} B_{11} \\ B_{12} \\ B_{13} \\ B_{13} \\ B_{23} \\ B_{33} \end{bmatrix}^T \tag{3.7}$$

the two constraints can be rewritten as

$$\begin{bmatrix} \mathbf{v_{12}^T} \\ (\mathbf{v_{11}} - \mathbf{v_{12}})^\mathbf{T} \end{bmatrix} \mathbf{b} = 0 \tag{3.8}$$

Therefore, we can stack K equations from K checkerboard images as $\mathbf{Vb} = 0$, where $\mathbf{V}$ is a $2K \times 6$ matrix, and there exists a solution for $\mathbf{b}$ when $K \geq 2$.

The intrinsic parameters are calculated from the closed-form solution Equation (3.6) directly as follows:

$$\begin{aligned} \lambda &= B_{33} - (B_{13}^2 + c_y(B_{12}B_{13} - B_{11}B_{23}))/B_{11} \\ f_x &= \sqrt{\lambda/B_{11}} \\ f_y &= \sqrt{\lambda B_{11}/(B_{11}B_{22} - B_{12}^2)} \\ c_x &= -B_{13}\alpha^2/\lambda \\ c_y &= (B_{12}B_{13} - B_{11}B_{23})/(B_{11}B_{22} - B_{12}^2) \end{aligned} \tag{3.9}$$

The extrinsic parameters are calculated from Equation (3.2) (where $\lambda = 1/\|\mathbf{A^{-1}h_1}\|$) as follows:

$$\begin{aligned} r_1 &= \lambda \mathbf{A^{-1}h_1} \\ r_2 &= \lambda \mathbf{A^{-1}h_2} \\ r_3 &= r_1 \times r_2 \\ \mathbf{T} &= \lambda \mathbf{A^{-1}h_3} \end{aligned} \tag{3.10}$$

The second part of calibration is a non-linear optimization by minimization of the total retroprojection error, using Levenberg-Marquardt algorithm for example, over the calibration parameters. The objective function to be minimized is $J(\theta)$; see Equation (3.11), where $\mathbf{R}$ is the covariance matrix of the observation error, $\mathbf{y}(\theta)$ is the vector of the estimated error (geometric distance) and θ is a vector containing the nine intrinsic parameters (focal length, principal point and distortion coefficients) and six extrinsic parameters.

$$J(\theta) = \mathbf{y}^T(\theta)\mathbf{R}^{-1}\mathbf{y}(\theta) \tag{3.11}$$

The OpenCV distortion model was originally introduced by Brown in 1966 and is the same distortion model used by Heikkila later on. Unlike Zhang's method that considers only radial distortion with two coefficients, this distortion model includes both radial and tangential distortions as follows:

$$x_r = x_d(1 + k_1r^2 + k_2r^4 + k_3r^6) \tag{3.12}$$

$$y_r = y_d(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \tag{3.13}$$

$$x_p = x_r + (2p_1 y + p_2(r^2 + 2x^2)) \tag{3.14}$$

$$y_p = y_r + (2p_2 x + p_1(r^2 + 2y^2)) \tag{3.15}$$

The tangential distortion [Equations (3.14)and (3.15)] deals with decentering and lens defects while the radial distortions [Equation (3.12) and (3.13)] is a polynomial function of 6-th order even though for most cases the 4-th order suffices ($k_3 = 0$).

3.5 Calibration Methods for Fish-Eye Cameras

In Figure 3.6 we can notice immediately the disastrous results of trying to calibrate fish-eye lens camera images using pinhole camera calibration methods. Calibration methods for fish-eye must be able to handle the enormous radial distortions at least in order to be useful. In that respect, one approach to calibrate fish-eye lens cameras is to reuse the pinhole camera model while compensating for the distortions. Bouguet's fish-eye calibration toolbox [5] uses this approach and an early popular (still popular) calibration method for fish eye lenses. This calibration method is based on his previous work on calibration for pinhole camera model but available only in the Matlab version and not implemented in the OpenCV library. Bouguet's

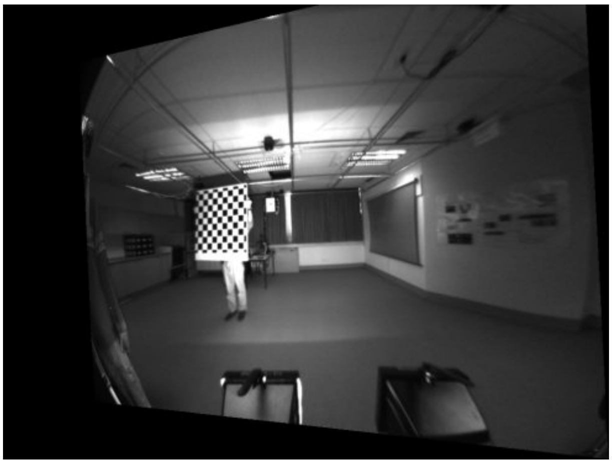

Figure 3.6: Calibration results of fish-eye lens image undistorted using OpenCV for pinhole cameras.

fish eye calibration is not radically different from the pinhole calibration method but gives surprisingly good calibration results and became popular as well.

The other approach for fish-eye calibration is using a projection model different than pinhole camera. The calibration methods developed by Scaramuzza [23] and Mei [20] are two examples of calibration using a different projection model. Interestingly, these calibration methods by Scaramuzza and Mei are intended for omnidirectional but they can handle both catadioptric and wide lenses despite the physical and optical differences between fish-eye lens cameras and omnidirectional cameras. These calibration methods make use of a generic projection model in order to accept a wide variety of mirror shapes for catadioptric systems as well as fish-eye lenses. Furthermore, Scaramuzza and Mei have based or borrowed many functions in their calibration toolbox from Bouguet's calibration toolbox as a statement of the popularity of Bouguet's toolbox.

The calibration methods by Scaramuzza and Mei will be explained later in the chapter while Bouguet's fish eye calibration will be presented next.

3.6 Bouguet's Planar Calibration

The Bouguet's fish-eye calibration toolbox [5] is a minor reworking of his pinhole camera calibration toolbox using the same calibration steps and algorithms except for the obvious change needed in distortion modeling.This calibration method also consist of two steps:(i) a closed-form solution of the calibration parameters and (ii) non-linear optimizations. The calibration algorithm and procedure remain unchanged (except for some minor differences) with respect to the pinhole camera calibration toolbox. The differences are in the distortion modeling as well as few a extra steps to compensate for the distortion.

Regarding the distortion modeling, in the pinhole camera model both radial and tangential distortions with five coefficients ($\kappa_1, \kappa_2, \kappa_3, p_1, p_2$) are considered in the OpenCV calibration method. In particular, the radial distortion in the pinhole calibration toolbox is a polynomial of 6^{th} order but typically 4^{th} order suffices.

For the fish-eye lens calibration on the other hand, the tangential distortion disregarded completely and the radial distortion is also a polynomial function gain but up to the 8^{th} order instead. The resulting distortion function with 4 coefficient has the form:

$$1 + \kappa_1\rho^2 + \kappa_2\rho^4 + \kappa_3\rho^6 + \kappa_4\rho^8 \quad \text{where} \quad \rho = \sqrt{x^2 + y^2} \tag{3.16}$$

Obviously, just replacing the radial distortion function will not do it but more changes are needed which are minor nonetheless. The changes are the extra steps introduced to compensate for the enormous distortions typical of fish-eye lenses as well as different parameter initialization process as thereof. Firstly, the initialization for parameters is more primitive and simpler than the OpenCV calibration implementation where DLT's equations and the vanishing point properties were exploited. For the fish-eye calibration toolbox, the principal point is estimated with the image center and the focal length using the approximation $f = \rho/\phi$ instead.

Nevertheless, these estimations produces bad approximations so in practice the calibration works best using the initial approximations entered by the user. The calibration toolbox can read the principal point estimated by the user manually by finding visually the principal point from a raw sample image and the focal length from the lens manufacturer's spec sheet.

Finally, the fish-eye distortion is compensated before the checkerboard corner detection step and later before the calibration process itself using the initial parameters estimation. As a result of the introduction the fish-eye distortion compensation step, the calibration algorithms and the calibration process in general remains largely unchanged since both calibration methods work in normalized coordinates. Experiments have shown the critical importance of entering good parameters estimations for the initialization as rough estimates causes the checkerboard corner detections to fail miserably (Figure 3.5).

3.7 Scaramuzza's Planar Calibration

A generic model for calibration of omnidirectional (catadioptric and wide angle lenses) was proposed by Scaramuzza and Siegwart in [23]. This generic model applies to camera with *central projection* (with a single effective viewpoint or single focal point). But fish-eye lenses which are generally *non-central projection* (with multiple focal points depending on incidence angle) can be approximated as well with this generic model (Figure 3.7).

This generic model assumes radial or axial symmetry of the optics and approximates the projection model with a parametric function rather than a specific projection function. The method is generic since this model uses a parametric function as projection function regardless of the type of mirror (parabolic, hyperbolic among others) or lens in the optics. The calibration method also uses a 2D calibration object with known geometry such as planar checkerboard and shown at different arbitrary poses or positions and orientations.

Firstly, this generic camera model has two planes: (i) camera image plane (u',v') (ii) sensor image plane (u", v"). The image plane(i) is really the camera sensor plane and the so called 'sensor plane' (ii) is actually an imaginary plane parallel to the image plane(i) and located at the origin of the coordinate system in the

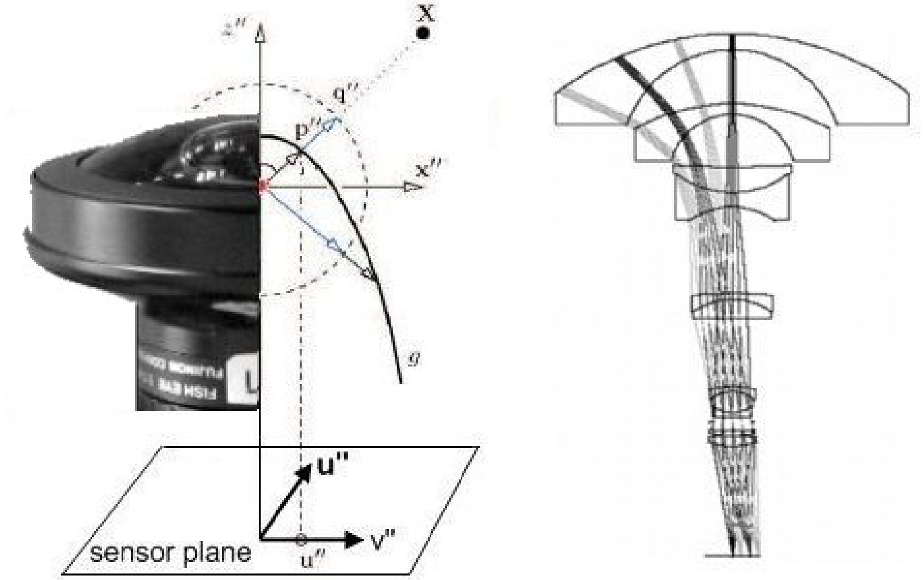

Figure 3.7: *Left:* fish-eye lens mapping of scene point X into the sensor point u″ [23]. *Right:* fish-eye lens as a non-central system [23].

case of catadioptric. Let X be scene point and $\mathbf{u}'' = [u'', v'']^T$ its projection on the sensor plane(ii), and $\mathbf{u}' = [u', v']^T$ its projection onto the camera plane(i). In general, both projection into planes (i) and (ii) are related by an homography $\mathbf{A}$ so that $\mathbf{u}'' = \mathbf{A}\mathbf{u}' + \mathbf{T}$. The imaging function $\mathbf{g}(u'', v'')$ that relates to a point $\mathbf{u}''$ in the sensor plane and the scene point $\mathbf{X}$ as follows:

$$\lambda p = \lambda \mathbf{g}(u'') = \lambda \mathbf{g}(\mathbf{A}\mathbf{u}' + \mathbf{T}) = \mathbf{P}\mathbf{X} \tag{3.17}$$

where λ is a constant, $\mathbf{P}$ the perspective projection matrix but actually $\mathbf{g}(u'', v'') = (u'', v'', f(u'', v''))^T$ where f normally depends on the particular shape of the mirror in catadioptric case. Here f is generalized here with a parametric function of the form $f(u'', v'') = a_0 + a_1\rho'' + a_N\rho''^N$ where $\rho'' = \sqrt{u''^2 + v''^2}$ (Figure 3.8).

The calibration will determine the coefficients a_i and the matrices $\mathbf{A}$ and $\mathbf{t}$ that satisfy Equation (3.17). The calibration is done in two stages where $\mathbf{A}, \mathbf{T}$ are estimated first and the coefficients $a_1, \ldots, a_N$ afterwards.

Hence, firstly the parameters $\mathbf{A}, \mathbf{T}$ of the homography are estimated iteratively assuming initially that the camera and sensor plane coincide or that $\mathbf{A} = I$ and $\mathbf{T} = 0$. $\mathbf{A}$ is refined by a non linear method such as Levenberg-Marquardt algorithm

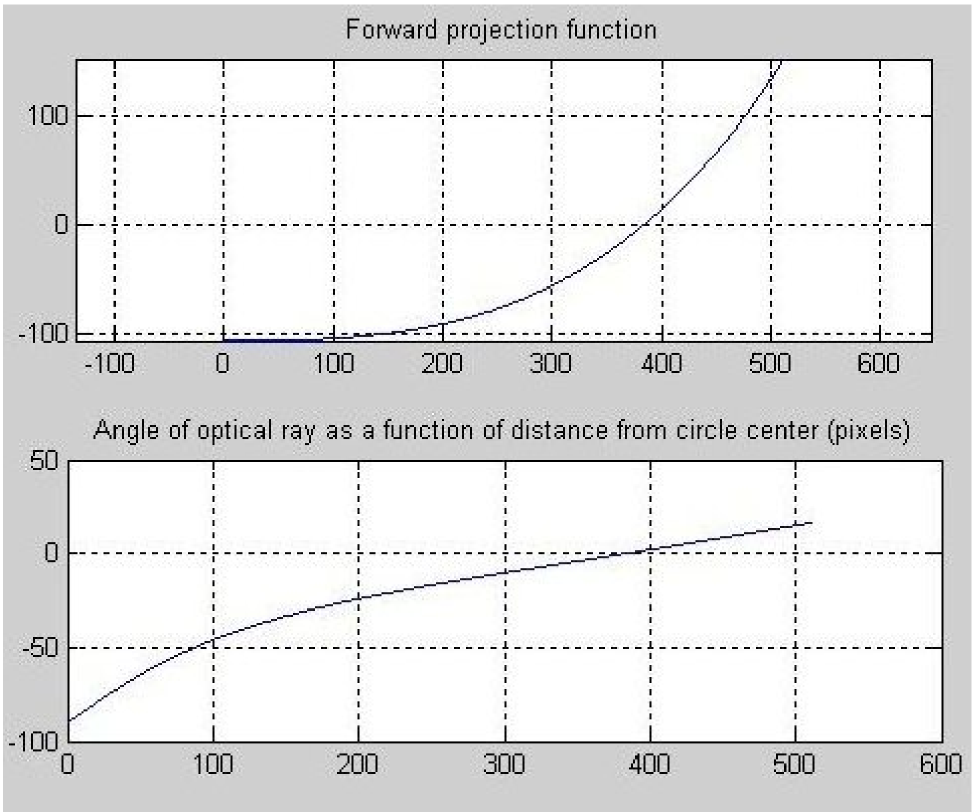

Figure 3.8: *Top:* parametric function f. *Bottom:* optical ray angle θ of the corresponding 3D with respect to the horizon.

and $\mathbf{t}$ refined by a iterative search algorithm.

Thus Equation (3.17) can be rewritten as (where (u', v') is pixel coordinate of image point $\mathbf{u}'$):

$$\lambda p'' = \lambda \begin{bmatrix} u' \\ v' \\ a_0 + a_2 {\rho'}^2 + \ldots + a_N {\rho''}^N \end{bmatrix} = \mathbf{P} \cdot \mathbf{X} \tag{3.18}$$

Let I^i be the observed image point and $\mathbf{M_j^i} = [X_j^i, Y_j^i, Z_j^i]$ its coordinates and $\mathbf{m_j^i} = [u_j^i, v_j^i]$ its corresponding image point coordinates in the image plane (the index i indicates the observed checkerboard and index j the j-th point on the i-th checkerboard. Since the calibration object is a planar checkerboard, $Z_j^i = 0$ and

Equation (3.18) can be simplified to (where $[\mathbf{r^i_1}, \mathbf{r^i_2}, \mathbf{r^i_3}] = \mathbf{R^i}$ is the rotation matrix and $\mathbf{T}$ the translation matrix):

$$\lambda^i_j \cdot p^i_j = \lambda^i_j \begin{bmatrix} u^i_j \\ v^i_j \\ a_0 + a_2 {\rho^i_j}^2 + \ldots + a_N {\rho^i_j}^N \end{bmatrix} = \mathbf{P^i} \cdot \mathbf{X^i_j} \tag{3.19}$$

$$\mathbf{P^i} \cdot \mathbf{X^i_j} = \begin{bmatrix} \mathbf{r}^i_1 & \mathbf{r}^i_2 & \mathbf{r}^i_3 & \mathbf{T}^i \end{bmatrix} \begin{bmatrix} X^i_j \\ Y^i_j \\ 0 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} \mathbf{r}^i_1 & \mathbf{r}^i_2 & \mathbf{T}^i \end{bmatrix} \begin{bmatrix} X^i_j \\ Y^i_j \\ 1 \end{bmatrix} \tag{3.20}$$

The Equation (3.20) can be further simplified by removing constant λ^i_j (depth scale) by multiplying both sides of the equation with $\mathbf{p^i_j}$ as follows:

$$\lambda^i_j p^i_j \times p^i_j = p^i_j \times \begin{bmatrix} \mathbf{r}^i_1 & \mathbf{r}^i_2 & \mathbf{T}^i \end{bmatrix} \begin{bmatrix} X^i_j \\ Y^i_j \\ 1 \end{bmatrix} = 0$$

$$\begin{bmatrix} u^i_j \\ v^i_j \\ a_0 + a_2 {\rho^i_j}^2 + \ldots + a_N {\rho^i_j}^N \end{bmatrix} \times \begin{bmatrix} \mathbf{r}^i_1 & \mathbf{r}^i_2 & \mathbf{T}^i \end{bmatrix} \begin{bmatrix} X^i_j \\ Y^i_j \\ 1 \end{bmatrix} = 0 \tag{3.21}$$

The extrinsic parameters will be also determined for each checkerboard pose but let us focus in particular a pose i first. From Equation (3.21), each point $\mathbf{p_j}$ on the checkerboard contributes to three homogeneous equations (without the index i for readability and with $g(\rho_j) = a_0 + a_2 {\rho^i_j}^2 + \ldots + a_N {\rho^i_j}^N$):

$$v_j(r_{31}X_j + r_{32}Y_j + t_3) - g(\rho_j)(r_{21}X_j + r_{22}Y_j + t_2) = 0 \tag{3.22}$$

$$g(\rho_j)(r_{11}X_j + r_{12}Y_j + t_1) - u_j(r_{31}X_j + r_{32}Y_j + t_3) = 0 \tag{3.23}$$

$$u_j(r_{21}X_j + r_{22}Y_j + t_2) - v_j(\rho_j)(r_{11}X_j + r_{12}Y_j + t_1) = 0 \tag{3.24}$$

The variables X_j, Y_j, Z_j as well as u_j, v_j are known and Equation (3.24) is linear with respect to $r_{11}, r_{12}, r_{21}, r_{22}, t_1, t_2$. By stacking all unknowns in this Equation (3.24) into a vector, we obtain a system of linear equations for L points in the checkerboard:

$$\mathbf{M} \cdot \mathbf{H} = 0 \quad \text{where} \quad \mathbf{H} = [r_{11}, r_{12}, r_{21}, r_{22}, t_1, t_2]^T \tag{3.25}$$

$$\text{and} \quad \mathbf{M} = \begin{bmatrix} -v_1 X_1 & -v1Y_1 & u_1 X_1 & u_1 Y_1 & -v_1 & u_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ -v_L X_L & -vLY_L & u_L X_L & u_L Y_L & -v_L & u_L \end{bmatrix} \tag{3.26}$$

By using SVD, a linear estimation of H is obtained by minimization of least-squares or min $\|\mathbf{MH}\|^2$ with $\|\mathbf{H}\|^2 = 1$. Since the vectors $\mathbf{r_1}, \mathbf{r_2}$ are orthonormal, the Equation (3.25) is unique up to a scale factor and also the unknowns r_{31}, r_{32} be calculated. So far, the calibration has calculated the rotation parameters $r_{11}, r_{12}, r_{21}, r_{22}$ and translation parameters t_1, t_2 as well as r_{31}, r_{32} for each pose i. The missing translation t_3 will be calculated next along with the intrinsic parameters.

We can find now the intrinsic parameters $a_1, \ldots, a_N$ by replacing the estimated values in Equation (3.22),(3.23) with the calculated values obtained previously and calculating the t_3^i for each pose. Again we obtain the following system of linear equations by stacking up all unknown variables in Equations (3.22),(3.23) into a vector and including all K poses of the checkerboard:

$$\begin{bmatrix} A_1 & A_1\rho_1^2 & \ldots & A_1\rho_1^N & -v_1 & 0 & \ldots & 0 \\ C_1 & C_K\rho_1^2 & \ldots & C_1\rho_1^N & -u_1 & 0 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots & \vdots & \ldots & \vdots \\ A_K & A_K\rho_K^2 & \ldots & A_K\rho_K^N & 0 & 0 & \ldots & -v_K \\ C_K & C_K\rho_1^2 & \ldots & C_K\rho_K^N & 0 & 0 & \ldots & -u_K \end{bmatrix} \begin{bmatrix} a_0 \\ a_2 \\ \vdots \\ a_N \\ t_3^1 \\ t_3^2 \\ \vdots \\ t_3^K \end{bmatrix} = \begin{bmatrix} B_1 \\ D_1 \\ \vdots \\ B_K \\ D_K \end{bmatrix} \tag{3.27}$$

$$\begin{aligned} \text{where} \qquad A_i &= r_{21}^i X^i + r_{22}^i Y^i + t_2^i \\ B_i &= v^i (r_{31}^i X^i + r_{32}^i Y^i) \\ C_i &= r_{11}^i X^i + r_{12}^i Y^i + t_1^i \\ D_i &= u^i (r_{31}^i X^i + r_{32}^i X^i). \end{aligned}$$

The linear least-squares solution of Equation (3.27) is obtained through the pseudo-inverse matrix method as well as the parameters $a_1, \ldots, a_N$. The best fitting polynomial f is calculated by iteratively increasing from $N = 2$ until it stops when the retroprojection error is less than a given error e.

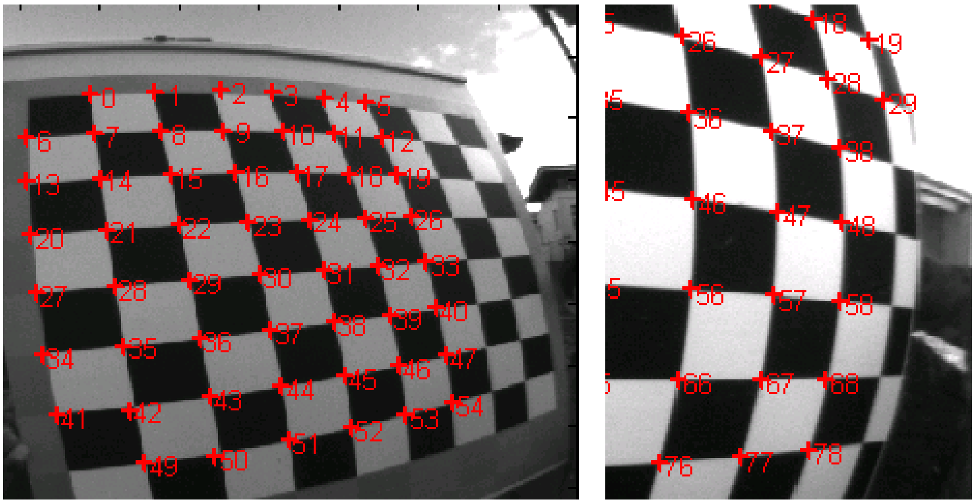

Figure 3.9: Automatic corner detection. *Left:* not detecting correctly. *Right:* missing corners.

The parameters calculated previously are to be further refined by a combination of: (i) linear and (ii) non-linear (maximum likelihood criterion) minimization. The linear refinement (i) is described below:

1. Recalculate all extrinsic parameters $r_{11}, r_{12}, r_{21}, r_{22}, r_{31}, r_{32}, t_1, t_2, t_3$ by using the estimated intrinsic parameters $a_0, a_2, \ldots, a_n$ and solving using SVD (up to a scale factor) the resulting system of linear homogeneous equations from Equations (3.24),(3.22),(3.23). The scale factor is determined later on due the orthogonality of $\mathbf{r_1}, \mathbf{r_2}$.

2. Refine in turn the intrinsic parameters using the recalculated extrinsic parameters earlier by solving the resulting linear system of Equations (3.23),(3.22) using the pseudo-inverse matrix methods.

Nevertheless, the previous linear solution minimizes an algebraic distance that is not physically meaningful. Thus, the parameters are further refined as mentioned earlier by a non linear minimization of maximum likelihood criterion as follows:

$$E = \sum_{i=1}^{K} \sum_{j=1}^{L} \left\| \mathbf{u}_j^i - \hat{\mathbf{u}}(\mathbf{R}^i, \mathbf{T}_i, \mathbf{A}, \mathbf{O}_c, a_0, a_2, \ldots, a_N, \mathbf{X}_j^i) \right\|^2 \qquad (3.28)$$

where there are K images of the checkerboard with L control or corner points in the checkerboard and $\mathbf{u}(\mathbf{R}^i, \mathbf{T}_i, \mathbf{A}, \mathbf{O}_c, a_0, a_2, \ldots, a_N, \mathbf{X}_j^i)$ is the retroprojection of the scene point $\mathbf{X_j^i}$ on the i-th pattern Equation (3.20) and $\mathbf{R^i}$ and $\mathbf{T^i}$ the rotation and translation matrices.

The final calibration parameters are calculated by minimizing the retroprojection error Equation (3.28) using Levenberg-Marquardt algorithm. Nevertheless, the main hindrance with using this toolbox is the automatic corner dectection feature not working correctly most of the time as shown in Figure 3.9.

3.8 Mei's Calibration Method

The third calibration method introduced here, by Mei and Rives [20] is also for single view point omnidirectional cameras using checkerboards and is an improvement on checkerboard and corner detection over the Scaramuzza's calibration method.

The camera projection model will be introduced first, followed by an explanation of initialization of parameters and the calibration steps at the end.

This calibration method uses an *unified projection model* as camera projection model. The unified projection model based on a previous work by Geyer and Barreto [10],[3] and is used as an exact model with small corrections. The unified projection model proposed by Mei and Rives (see Figure 3.10), simplifies the calibration by reducing the number of parameters as well overcoming the difficulty of estimating the initial values as in the polynomial approximation model by Scaramuzza presented in Section 3.7. The unified model uses a reference system different than the used in [10],[3] (see Figure 3.11) and the values of parameters ξ(mirror parameter) and η depends on the actual mirror equation used, in particular for fish-eye lenses the parameter $\xi = 1$.

The image scene points in the mirror are projected into the sphere as follows: $(\chi)_{F_m} \rightarrow (\chi_s)_{F_m} = \frac{\chi}{\|X\|} = (X_s, Y_s, Z_s)$ and the projected image points are then referenced with respect to $\mathbf{C}_p = (0, 0, \xi), (\chi_s)_{F_m} \rightarrow (\chi_s)_{F_p} = (X_s, Y_s, Z_s + \xi)$. This point is projected further into the normalized plane $\mathbf{m} = (\frac{X_s}{Z_s+\xi}, \frac{Y_s}{Z_s+\xi}, 1) = \hat{h}^{-1}(\chi_s)$ where the normalized image point is finally projected using the projection matrix $\mathbf{K}$ as follows:

$$\mathbf{p} = \mathbf{K}\mathbf{m} = \begin{bmatrix} f_1\eta & f_1\eta\alpha & c_x \\ 0 & f_2\eta & c_y \\ 0 & 0 & 1 \end{bmatrix} \mathbf{m} = k(\mathbf{m}) \tag{3.29}$$

with focal length = $[f_1, f_2]^T$ or f, principal point = (c_x, c_y), and skew γ.

This the final projection of this unified projection model uses a *generalized camera* (represented by $\mathbf{K}$) which considers the camera sensor and mirror as one device. The importance of using the generalized camera for calibration is that the estimation of (f, η) is not independent.

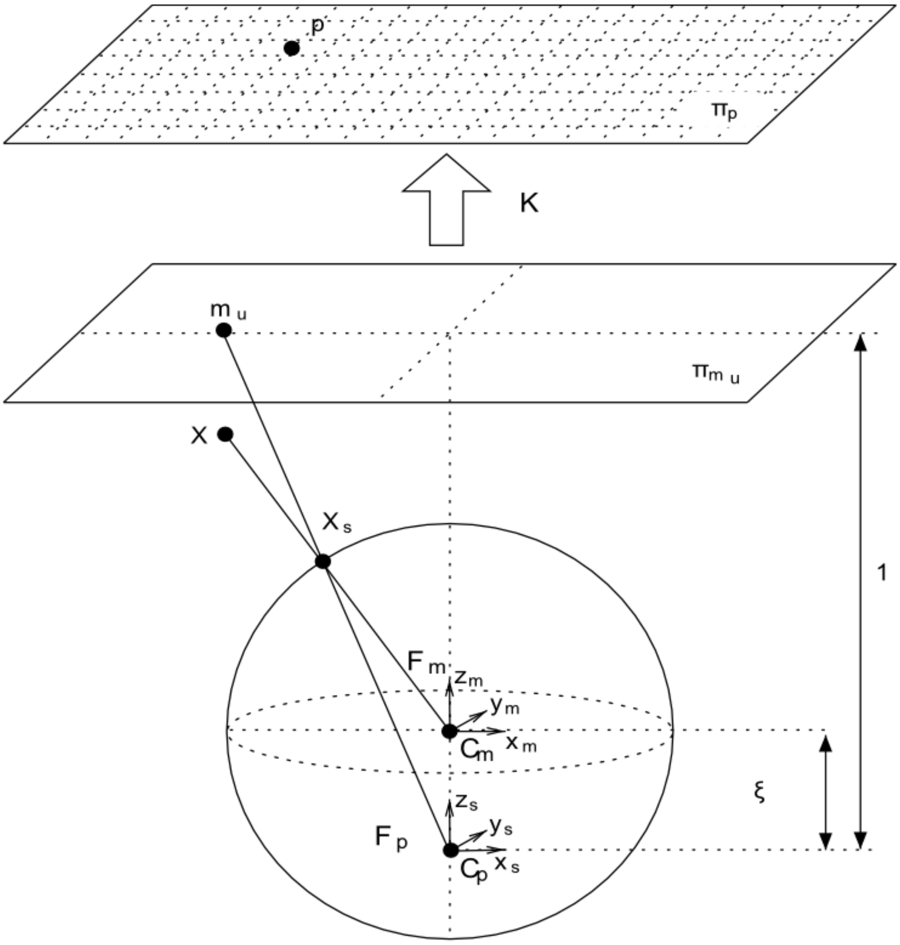

Figure 3.10: Unified projection model [20].

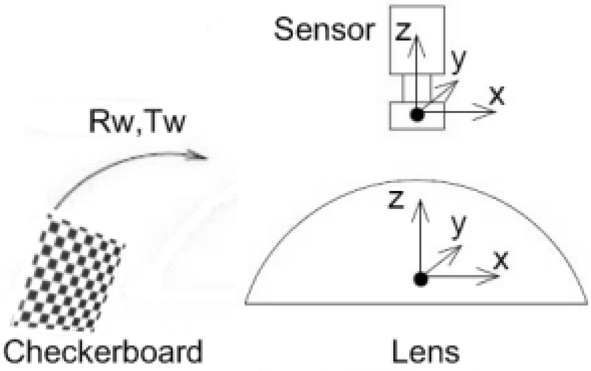

Figure 3.11: Coordinate reference system.

The calculation of χ_s for a given image point p is as follows:

$$\hat{h}^{-1}(\mathbf{m}) = \begin{bmatrix} \tau x \\ \tau y \\ \tau - \xi \end{bmatrix} \quad \text{with} \quad \tau = \frac{\xi + \sqrt{1 + (1-\xi^2)(x^2+y^2}}{x^2+y^2+1}) \tag{3.30}$$

The projection model considers both radial $f_d(\rho)$ and tangential $f_t(x, y)$ distortion (where $\rho = \sqrt{x^2+y^2}$), similarly to the OpenCV calibration's model as follows:

$$\begin{aligned} f_d(\rho) &= 1 + k_1\rho^2 + k_2\rho^4 + k_5\rho^6 \\ f_t(x,y) &= \begin{bmatrix} 2k_3xy + k_4(\rho^2 + 2x^2) \\ 2k_4xy + k_3(\rho^2 + 2y^2) \end{bmatrix} \end{aligned} \tag{3.31}$$

The following notation will be used for the distortion function D and distortion parameters $V^3 = \begin{bmatrix} k_1 & k_2 & k_3 & k_4 & k_5 \end{bmatrix}$ and the generalized camera projection as $P(X, V^4)$, with $\mathbf{V} = \begin{bmatrix} V^1 & V^2 & V^3 & V^4 \end{bmatrix}$ where $V^4 = \begin{bmatrix} \alpha & f_1 & f_2 & c_x & c_y \end{bmatrix}$, and $V^2 = \|\xi\|$, and $V^1 = \begin{bmatrix} q_{w1} & q_{w2} & q_{w3} & q_{w4} & t_{w1} & t_{w2} & t_{w3} \end{bmatrix}$ the quaternions to parametrise the rotation Eq. 3.37 and W the corresponding transformation.

Finally, let G be the composition of all projection functions as follows:

$$G = P \circ D \circ H \circ W \tag{3.32}$$

Thus, by using a non-linear minimization algorithm such as Levenberg-Marquardt, the last step of the calibration will minimize the following function:

$$F(x) = \frac{1}{2}\sum_{i=1}^{m}(G(\mathbf{V}, g_i) - e_i)^2 \tag{3.33}$$

where m is the number of checkerboard images with g_i points each, and e_i the values of projected images.

The validity of the unified projection model for fish-eye lenses in particular is shown next. The perspective projection of an image point $\mathbf{m}_u = (x, y, 1) = (\frac{X}{Z}, \frac{Y}{Z}, 1)$ can be expressed in the unified projection model with $\xi = 1$ (where the parameter ξ is not very meaningful here since a fish-eye lens does not have a mirror) as follows:

$$\mathbf{m}_d = (\frac{X}{Z + \|\chi\|}, \frac{Y}{Z + \|\chi\|}, 1) \tag{3.34}$$

Finally, from Equation 3.34, the polynomial division model equations for fish-eye lens cameras are obtained by algebraic manipulation as follows:

$$\rho_u = \frac{2\rho_d}{1 - \rho_d^2} \qquad \text{with} \qquad \rho = \sqrt{m_x^2 + m_y^2} \tag{3.35}$$

The initialization step of parameters for the calibration method will be presented next. The parameters ξ, f, c_x, c_y are estimated as follows: Firstly, the principal point is initialized by approximating it with the image center or alternatively using the circular image border if full frame image is available. Secondly, ξ is initialized as $\xi = 1$ and furthermore the authors found that inaccuracies in ξ, f not affecting significantly the corner extraction step.

Finally, the generalized focal length f estimated linearly from at least three image points selected from a non-radial line image as follows from Equation 3.30:

$$\hat{h}^{-1}(\mathbf{m}) \sim \begin{bmatrix} x \\ y \\ f(x,y) \end{bmatrix} \quad \text{where} \quad f(x,y) = \frac{1}{2}(1 - (x^2 + y^2)) \tag{3.36}$$

Let $\mathbf{p} = (u, v)$ be a image plane point and $\mathbf{p}_c = (u_c, v_c)$ the corresponding point, then by the relation $f : \mathbf{p}_c = f\mathbf{m}$ (projection on the normalized plane) we obtain now:

$$\hat{h}^{-1}(\mathbf{m}) \sim \begin{bmatrix} x \\ y \\ g(u_c, v_c) \end{bmatrix} \quad \text{where} \quad g(\mathbf{m}) = \frac{1}{2f}(f^2 - (u_c^2 + v_c^2)) \tag{3.37}$$

We obtain for points belonging to a line image l defined by the normal $\mathbf{N}$ the following relation:

$$\mathbf{N} = \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix}, \quad \hat{h}^{-1}(\mathbf{m})^T\mathbf{N} = 0 \Leftrightarrow \begin{cases} 0 = n_x u_c + n_y v_c + \frac{1}{2}(a - b(u_c^2 + v_c^2)) \\ a = f n_z \\ b = n_z\, f \end{cases} \tag{3.38}$$

In general, for n points $\mathbf{p_1}, \mathbf{p_2}, \ldots, \mathbf{p_n}$ belonging to the same line l, we obtain the least square solution by (SVD) $\mathbf{P} = \mathbf{USV}^T$ from the last column of $\mathbf{V}$ associated to the minimal singular value from this system:

$$\mathbf{P}_{n\times 4} C_{4\times 1} = 0 \tag{3.39}$$

where $\mathbf{P} = \begin{bmatrix} u_{c1} & v_{c1} & 1/2 & -(u_{c1}^2 + v_{c1}^2)/2 \\ \vdots & \vdots & \vdots & \vdots \\ u_{cn} & v_{cn} & 1/2 & -(u_{cn}^2 + v_{cn}^2)/2 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \end{bmatrix}$

The value of focal length f is calculated from the selection of at least three image points located on a non-radial line of focal as follows: (i) let us define $t = c_1^2 + c_2^2 + c_3 c_4$ first and $d = \sqrt{1/t}, n_x = c_1 d$ with $n_y = c_2 d$. (ii) for non-radial lines ($n_z = \sqrt{1 - n_x^2 - n_y^2}$) we obtain then from $\mathbf{C}$:

$$f = \frac{c_3 d}{n_z} \tag{3.40}$$

Finally, the calibration procedure consists of three steps: (i) initialization of parameters, (ii) finding the homography, and finally (iii) non-linear refinement of parameters. During the first step (i) initialization of parameters, at least three non-radial point in the image are selected to estimate focal length f with Equation 3.40. The homography calculation step (ii) is performed by retroprojecting the checkerboard images as usual. Lastly, the final step (iii) involves the non-linear minimization of the cost function Equation 3.33 (an euclidean distance between the checkerboard projection and the images) which is used to refine the parameters as usual again.

Chapter 4

Evaluation Techniques

This chapter informs about two ways of evaluating accuracies of calibration in the context of DAS, and these techniques will then be applied in Chapter 5. The back projection error is simply defined by mapping recorded calibration patterns back into 3D space and identifying the places where they would "hit" the shown calibration object. The row misalignment error is evaluating the accuracy of having matching image lines in the calculated rectified images.

4.1 Specifics of a Driver Assistance Context

Camera calibration involves many intrinsic and extrinsic parameters as seen in earlier chapters. Furthermore, the accuracy of camera calibration depends on several factors such as: How the system is setup, baseline in stereo systems, quality of optics, resolution of the camera sensor, and so forth. In general, calibration results are generally sensitive to the resolution of camera sensors such that better calibration results can be obtained with higher resolution cameras. However, it is also possible

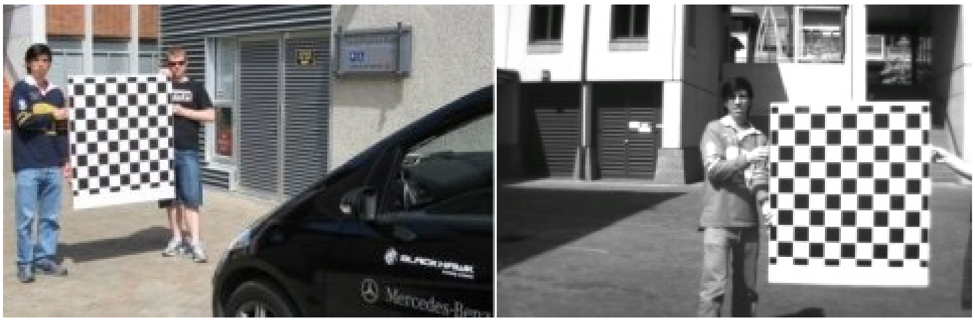

Figure 4.1: Calibrating pinhole-type cameras in HAKA1. *Left:* as seen by an observer. *Right:* as seen by a gray-level camera in HAKA1.

to get more accurate results with a good calibration of low resolution cameras than with a bad calibration of high resolution cameras. Thus, a good calibration method can be said as the method that produces or calculates the best or optimal overall solution for all parameters.

Calibration in the DAS context is not only different from calibrating in a laboratory by being outdoors, but also influenced by other factors. Thus, a brief introduction of those differences and factors affecting calibration in DAS will be introduced next before presenting the calibration evaluation methods used.

In theory, it should be possible to calibrate a camera setup in a laboratory but in reality that is not always possible or practical since the cameras must be firmly and securely fitted in the car outside a lab.

Thus, calibration in the DAS context requires that calibration is performed on the ego-vehicle in a dynamic and ever changing outdoor space rather than in a controlled laboratory setup. Thus, the analysis and comparison of different calibration methods must be performed in similar conditions and setup as the cameras are normally operated. Performing calibration outdoors in a different scale imposes different requirements and challenges. For example, the checkerboard used in *.enpeda..* project is quite large (at least $90cm \times 100cm$) with large squares made of lambertian materials to deal with outside bright lightning conditions. Also, the typical baseline of stereo cameras for DAS is around 40cm and the distance of objects are in the order of several meters (Figure 4.1 shows a typical calibration session of normal lens cameras mounted in HAKA1).

Furthermore, calibration of fish-eye lens cameras in a DAS context poses even more challenges. The same pinhole-type camera calibration methods can not be

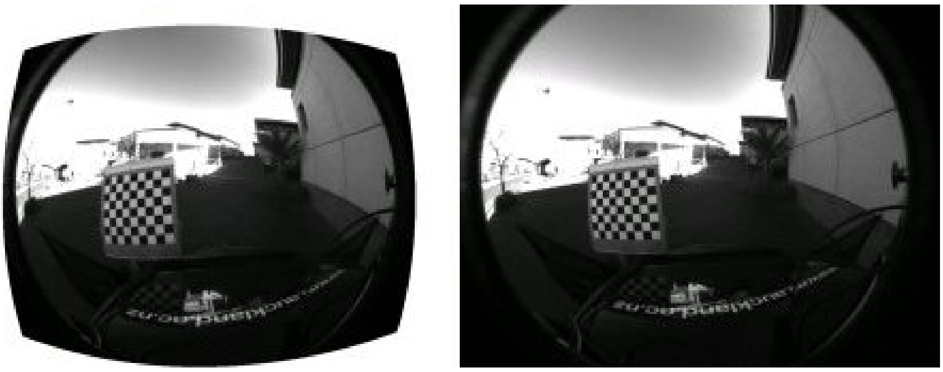

Figure 4.2: Checkerboard image too small. *Left:* calibration result. *Right:* raw image.

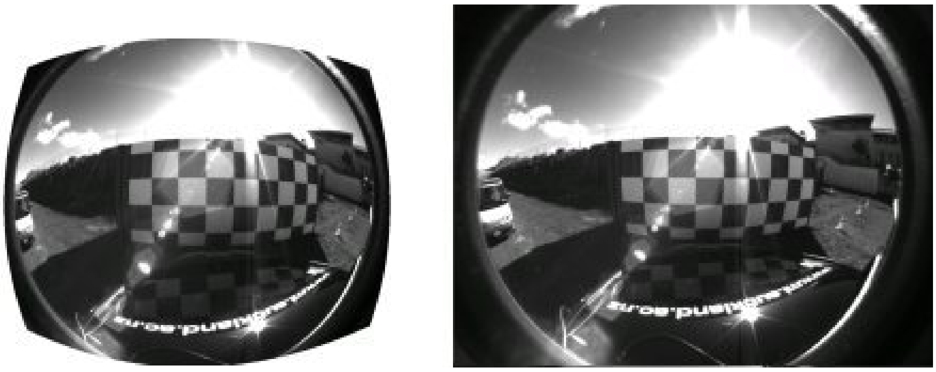

Figure 4.3: Checkerboard image too far away. *Left:* calibration result. *Right:* raw image.

directly applied here, as attempts to use the same large (90 cm by 90 cm) checkerboard in a similar fashion was unsuccessful (see Figure 4.2) as well as attempts to use a much larger checkerboard (approximately $2.4m \times 6m$) made no improvements (see Figure 4.3).

The previous definition of a good calibration method can be further refined as a method that also provides accurate results as the camera system and in the context that is being used. So firstly, the obvious and meaningful metric of measuring the back projection error will be considered first. Secondly, the calibration errors in row misalignments (applicable for stereo images) will be finally considered, since the calibration results in the DAS context are used very often for stereo vision analysis.

In summary, the selected evaluation methods are back projection error and row misalignment error. The row misalignment error can be estimated using common static stereo analysis such as disparity match of epipolar lines with Dynamic Programming as well as comparing the calibrated row component of the calibrated principal point. The back projection error will be presented next.

4.2 Backprojection Error

The process of mapping 2D object images into the 3D space, by inverting the perspective transform, is generally referred to as *backprojection*. For example, see [18] for the backprojection of surface patches. The simplest case of backprojection is given by mapping a point p in an image "along the ray starting at this point" into the 3D space, where the direction of that ray is defined by the central projection of the camera (i.e., the line connecting the center of projection with the given point p. A point p can not identify a corresponding point P in 3D space but it requires an intersection of the calculated ray with a plane in 3D space instead. Backprojection is similar to the corner detection process but reversed as it calculates the the back-projected P from the detected corner p in the image sensor. Thus, the formulas to calculate the backrprojection depends on the projection model used.

The calculated intrinsic and extrinsic parameters (being the result of the considered calibration) define a central projection; we also take a set of specified, say N *calibration points* (e.g., selected corners of squares of the used calibration checkerboard) for defining a set of projected points in the image; central projection and those N points define now a set of N rays pointing into the 3D space. Now we intersect those N rays with the calibrated 3D pose of the shown calibration checkerboard (i.e., with a plane). The resulting N points in 3D (i.e., on that plane) should now ideally coincide with the corresponding calibration points. To visualize this situation, we may map the texture of the calibration checkerboard on the rectangle used as calibrated pose of this board, and also indicate the intersecting N rays by N points on that textured rectangle.

Altogether, this maps a set of N points into an image of the calibration board. Calibration points appear in this image in the form of some “noisy” pixel patterns. It remains to identify distances between one of those N points and, say, the centroid of its corresponding calibration point.

Let us define N the total number of calibration points $(\hat{X}_i, \hat{Y}_i)$ representing the real (distorted) image world coordinates of the checkerboard, and (X_i, Y_i) the image points calculated by the calibrated camera model. Thus, mean error, maximum error, and standard deviation are defined as follows:

$$
\begin{aligned}
E_{mean} &= \frac{1}{N}\sum_{i=1}^{N} D_i && \text{(mean error)}\\
E_{max} &= \max_{1\leq i\leq N} D_i && \text{(maximum error)}\\
\sigma &= \left(\frac{1}{N}\sum_{i=1}^{N}(D_i - E_{mean})^2\right)^{\frac{1}{2}}
\end{aligned}
$$

where $D_i = \sqrt{(\hat{X}_i - X_i)^2 + (\hat{Y}_i - Y_i)^2}$.

A good calibration produces back-projection error is in the subpixel range. But that is not always the case as we can see in Figure 4.4 and 4.5 an example of bad fish-eye calibration using Bouguet’s method (back-projection error of each checkerboard corner plotted).

As a result, back-projection error provides an intuitive and meaningful evaluation tool while comparing the results of different calibration methods for the same camera setups. However, the results of back-projection error can not be compared across different cameras since the back-projection error is sensitive to the camera sensor's resolution, FOV and object to camera distances. The deal with these issues of portability across different cameras and camera setups, the normalized calibration error can be used as a general evaluation method.

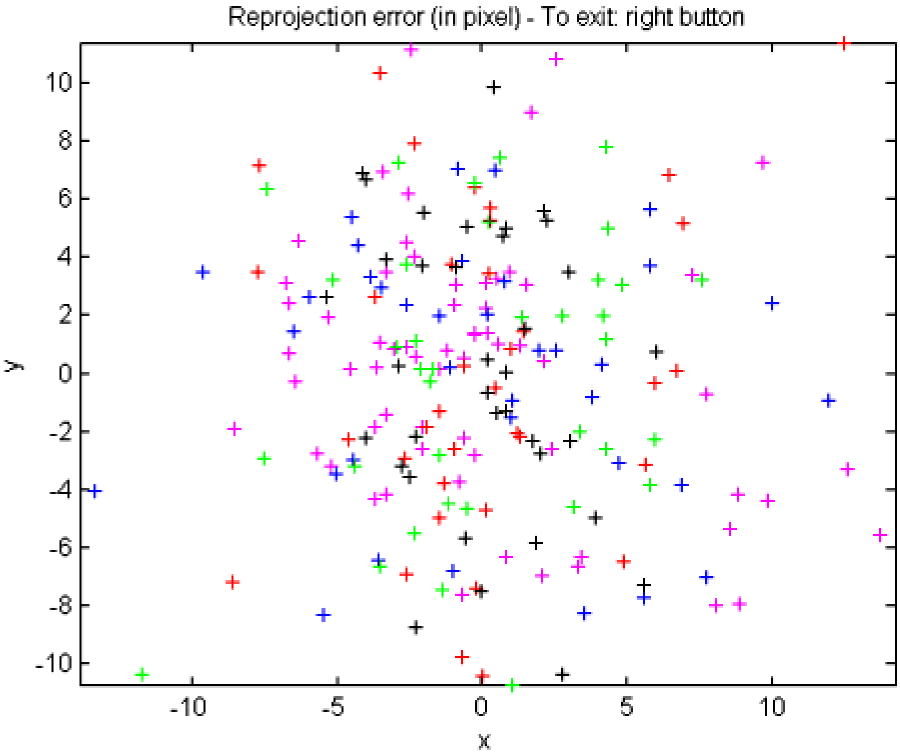

Figure 4.4: Back-projection errors of a bad fisheye calibration using Bouguet's method. Crosses (+) indicates the pixel error in x-axis and y-axis for each calibration points.

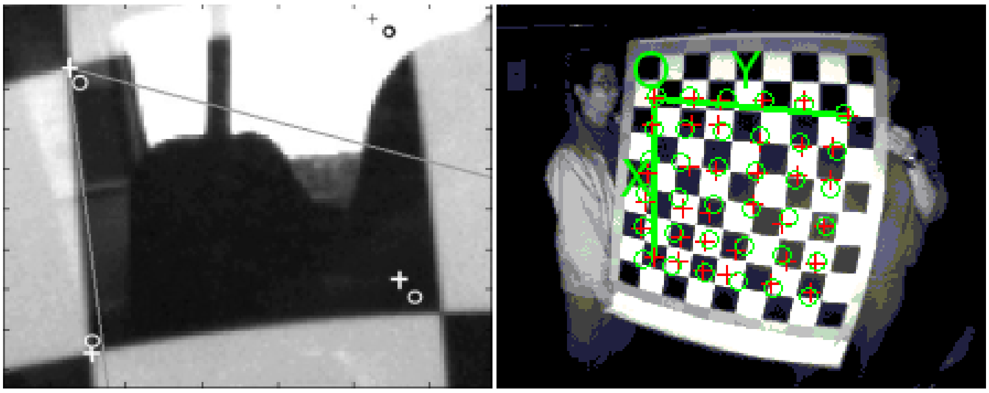

Figure 4.5: *Left:* closeup of bad backprojected corners (circles) and incorrectly detected corners (crosses). *Right:* Back-projected corners of one of the calibration images shown in Fig. 4.4 as red crosses.

4.3 Row Misalignment Error

The calibration data is needed to rectify the raw images, as we have seen previously, which in turn simplifies the stereo correspondence task for example. *Stereo correspondence* refers to the matching of 3D points in stereo images (from two different camera views). Namely, a scene point P visible in both cameras is projected onto two *corresponding image points* p_l and p_r (left and right respectively). A vector can be defined from p_l as *disparity* as follows:

$$\Delta = (p_l - p_r)^T = (x_l - x_r, y_l - y_r)^T \tag{4.1}$$

For a stereo cameras set in the standard stereo geometry, the disparity becomes just a *scalar disparity* instead as follows:

$$\Delta_s = \sqrt{(x_l - x_r)^2 + (y_l - y_r)^2} \tag{4.2}$$

Finally, a *disparity map* $\Delta(x, y)$ can be created from calculating the scalar disparities of image points corresponding to the point P with 3D coordinates $P = (X, Y, Z)$ calculated as follows (where f, b are focal length and baseline respectively):

$$\begin{aligned} x_r &= \frac{f(X-b)}{Z} \\ x_l &= \frac{fX}{Z} \quad \text{with} \quad y = \frac{fY}{Z} \end{aligned}$$

$$\text{we obtain} \quad X = \frac{b \cdot x_l}{x_l - x_r}, Y = \frac{b \cdot y}{x_l - x_r}, Z = \frac{b \cdot f}{x_l - x_r}$$

In general, the correspondence analysis in stereo images can done by analyzing the intensity values using a *similarity measure* of a window of size or block of neighboring pixels along the epipolar lines and as a result it is called *intensity-based correspondence analysis*.

One such similarity measure is the *mean square error (MSE)*, which compares the intensity values of block of pixels in the stereo images, can be defined as follows:

$$\text{MSE(x,y,d)} = \frac{1}{m \cdot n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} \left(E_l(x+i, y+j) - E_r(x-d+i, y+j)\right)^2 \tag{4.3}$$

with $d > 0$ and multiple of m, an offset of $x_l - x_r$ of the column positions in the left and right images for blocks of size $m \times n$. The scalar disparity in the standard stereo geometry $\Delta_{ssg}(x, y)$ can be defined as the difference in column positions for the pair of blocks with minimum MSE value. Finally, the correspondence analysis along epipolar lines can be done by the matching process of the MSE similarity measure for example using Dynamic Programming (DP). The proposed method of evaluating the different calibration method is by comparing the results of matching epipolar lines using DP or Dynamic Programming (a detail explanation of matching of epipolar lines using Dynamic Programming can be found in [18].

For the rectified stereo images of accurately calibrated cameras should give good results when running this proposed evaluation method (see Figure 4.6). The rationale is that more accurate calibration methods should give better disparity matching results because of more accurately rectified images. The proposed evaluation tool of using disparity matching with DP will provide an indirect measure of the row misalignment errors by means of the visual inspection of the resulting disparity match. Figure 4.6 shows an example comparing the results of the disparity matches between raw images and rectified ones.

Once a stereo camera system has been calibrated and resulting images rectified, depth maps (which are useful in DAS) can be generated as well using stereo correspondence and simple triangulation of disparities. There are many factors and errors that affect the evaluation process such as: mismatch and correlation errors but we are interested only in the calibration or pointing error p .

Finally, when the same calibration images are used the row misalignment error can also be easily and quickly estimated by comparing the row component $y - axis$ of the calibrated principal point. Since row errors will affect all rows including the principal point.

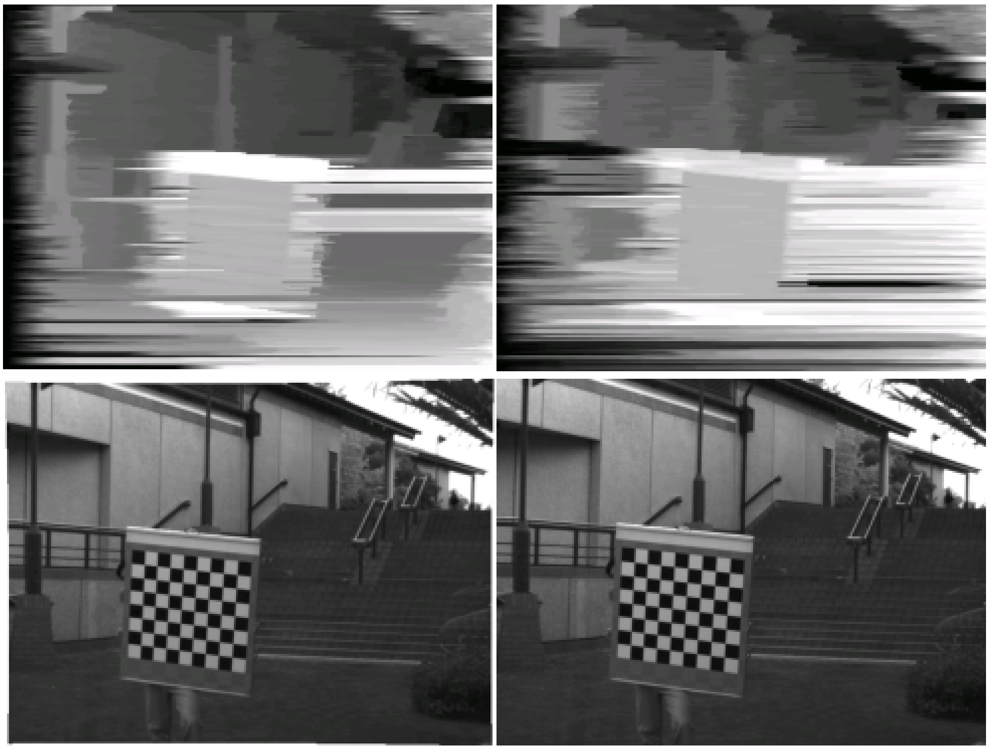

Figure 4.6: Row misalignment error of stereo images by DP (Basler 9mm lens). *Top left:* left camera raw image. *Top right:* same image rectified. *Bottom left:* disparity match of raw images. *Bottom right:* disparity match of the same rectified images.

Chapter 5

Experiments

This chapter informs about the experimental results of different camera configurations using the calibration methods described in Chapter 3 and using the evaluation techniques described in Chapter 4.

5.1 Introduction

This section presents the results of the experiments using different camera and lens configurations to test and compare all four selected calibration methods introduced in Chapter 3.

The details about the experiment setups and components used are as follows, the experiments were performed with the stereo cameras setup in HAKA 1 as they would be normally done during normal operations and the details about the camera and lens configurations, calibration grid and images are detailed next.

Two camera configurations have been used consisting of two camera models at different image resolutions: (i) PointGrey Firefly MV with 1/3″ sensor cameras at 640 × 480 resolution, and (ii) Basler AF622 with 2/3″ sensor cameras at 1280 × 960 resolution at least.

The selected lenses used in the experiments were: (i) Fujinon HF9HA-1B 9mm focal length tv lens for pinhole camera experiments and (ii) Fujinon FE185C057HA-1 fish eye lens for fisheye experiments. Furthermore, the pinhole camera experiments were done with cameras mounted inside while the fisheye experiments were done with the cameras setup on the roof mounting similar to the setup in Figures 3.2, 3.1.

The calibration grid used was the planar calibration grid or checkerboard of 90 cm × 100 cm in size with at least 9 × 8 squares (Figure 4.1). The calibratiom images used were similar to those typically used for the calibration in the DAS context and in these experiments consisted of approximately 20 stereo images which can be used as four smaller sets of five to 20 images as well. These four sets of calibration images have been used to test the performance and accuracy of each calibration method with varying number of calibration images.

The following sections are organized as follows: the results of the evaluation techniques explained in Chapter 4 are presented by presenting the experimental data of the calibration methods in tables, graphs the backprojection error σ as calibration pixel error figures and figures of disparity matches of selected calibration methods as well.

The experiments using OpenCV calibration method will be presented next followed by Bouguet, Mei and Scaramuzza's methods using the fisheye lens afterwards.

5.2 OpenCV Calibration's Experiments

OpenCV calibration method used two sets of 25 calibration images taken with Firefly MV cameras (lower resolution) first and followed by the second set taken with Basler cameras (high resolution images). The experiments for testing the pinhole camera were slightly different than the rest since only one pinhole calibration method was considered. Five sets of calibration images instead of four and a third smaller set of calibration images were used (Figures A.8, A.9, A.10).

The results of the calibration (Tables 5.1, 5.2, 5.3) shows the average backprojection errors $error_x$,$error_y$(pixels) for the x-axis and y-axis respectively and the calibration parameters (intrinsic parameters f_x, f_y, c_x, c_y in pixels and distortion parameters (tangential p_1, p_2 and radial $\kappa_1, \kappa_2, \kappa_3$ where $\kappa_3 = 0$). The calibration results using Firefly MV with 9mm lens are presented in Table 5.1 and graphs of backprojection error for the left and right cameras are presented in Figure 5.1. The row misalignment error can be estimated visually by comparing the disparity matches

Table 5.1: Calibration data for left Firefly MV camera with 9mm lens at 640 × 480.

Images	5	10	15	20	25
$error_x$	0.064	0.061	0.059	0.056	0.054
$error_y$	0.081	0.075	0.072	0.069	0.067
f_x	1532.5 ± 10.2	1533.5 ± 7.0	1533.0 ± 6.4	1532.6 ± 5.8	1533.0 ± 5.4
f_y	1534.0 ± 10.4	1535.0 ± 7.0	1534.4 ± 6.4	1534.0 ± 5.8	1534.3 ± 5.3
c_x	358.7 ± 12.0	360.2 ± 8.7	364.3 ± 7.2	361.8 ± 6.4	361.4 ± 6.1
c_y	270.7 ± 9.8	271.7 ± 7.3	269.7 ± 6.3	270.6 ± 5.6	271.3 ± 5.3
$p_1(10^{-2})$	0.2 ± 0.2	0.2 ± 0.1	0.1 ± 0.1	-0.1 ± 0.1	0.1 ± 0.1
$p_2(10^{-2})$	0.3 ± 0.2	0.2 ± 0.1	0.3 ± 0.1	0.2 ± 0.1	0.2 ± 0.1
$\kappa_1(10^{-2})$	-13.2 ± 7.1	-8.6 ± 5.7	-10.1 ± 4.4	-10.5 ± 3.4	-10.8 ± 3.2
$\kappa_2(10^{-1})$	-27.0 ± 27.0	-42.6 ± 22.9	-43.8 ± 18.5	-43.9 ± 13.3	-43.2 ± 12.3

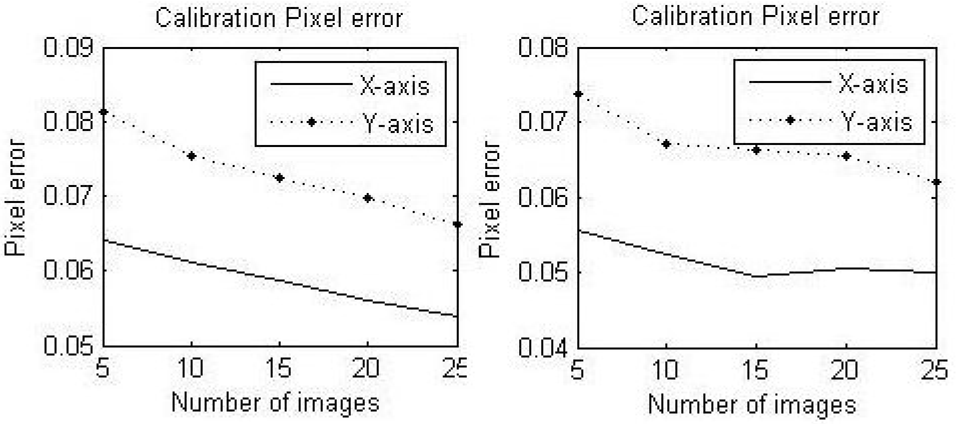

Figure 5.1: Firefly MV camera with 9mm lens backprojection error. *Left:* left camera. *Right:* right camera.

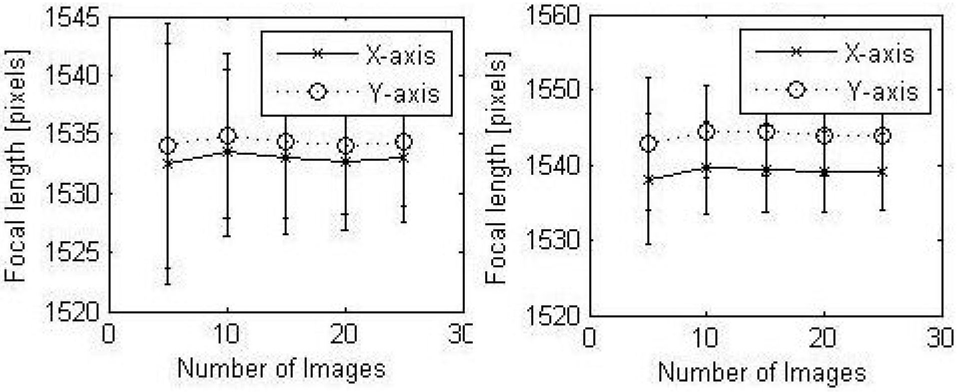

Figure 5.2: Firefly MV camera w/9mm lens focal length values and error. *Left:* left camera. *Right:* right camera.

of the raw and rectified images in Figure 5.3 and finally the graph of focal length as function of calibration set number is shown in Figure 5.2. The calibration results with Basler cameras fitted with 9mm lens using the calibration images (Figures A.9) are presented in Table 5.2 and graphs of backprojection error for the left and right cameras are presented in Figure 5.5. Finally, Figure 5.6 shows the disparity match between the raw and rectified images and the graph of focal length as function of

Table 5.2: Calibration data for left Basler camera with 9mm lens at 1280 × 960.

Images	5	10	15	20	25
$error_x$	0.064	0.057	0.055	0.060	0.078
$error_y$	0.074	0.072	0.075	0.086	0.100
f_x	1371.1 ± 18.3	1380.0 ± 10.7	1381.1 ± 10.0	1378.4 ± 3.1	1377.8 ± 2.8
f_y	1371.6 ± 19.5	1380.2 ± 10.7	1381.0 ± 10.1	1378.4 ± 3.1	1380.1 ± 2.8
c_x	645.1 ± 16.0	645.0 ± 11.3	646.9 ± 8.2	644.6 ± 4.8	640.5 ± 3.2
c_y	473.8 ± 12.5	472.4 ± 9.6	473.9 ± 6.6	466.1 ± 4.0	465.7 ± 3.0
$p_1(10^{-2})$	-0.1 ± 0.2	-0.1 ± 0.1	-0.1 ± 0.1	-0.1 ± 0.04	0.10 ± 0.03
$p_2(10^{-2})$	0.2 ± 0.3	0.1 ± 0.2	0.1 ± 0.1	0.2 ± 0.1	-0.10 ± 0.03
$\kappa_1(10^{-2})$	-2.9 ± 15.6	-16.5 ± 4.1	-20.6 ± 1.8	-19.9 ± 1.3	-23.3 ± 0.5
$\kappa_2(10^{-2})$	-1143 ± 1115	-90 ± 69	-8 ± 12	-12 ± 10	17 ± 3

Table 5.3: Calibration data with set of five and 10 images.

Images	5	10
$error_x$	0.101	0.095
$error_y$	0.117	0.126
f_x	1373.6 ± 5.4	1376.9 ± 3.6
f_y	1376.6 ± 5.4	1379.9 ± 3.5
c_x	646.5 ± 5.8	641.9 ± 4.4
c_y	462.6 ± 5.0	465.8 ± 4.0
$p_1(10^{-2})$	0.16 ± 0.05	0.14 ± 0.04
$p_2(10^{-2})$	-0.21 ± 0.11	-0.06 ± 0.05
$\kappa_1(10^{-2})$	-21.8 ± 1.1	-23.3 ± 0.6
$\kappa_2(10^{-2})$	-13.8 ± 4.0	16.6 ± 3.2

calibration set number is shown in Figure 5.4. In addition, another two extra set of five and 10 calibration images (Figure A.10) selected from the last 10 images from the original 25 calibration set) was used in the calibration experiments as well. Interestingly, these calibration sets yielded a bigger pixel error but the calibration results are better or comparable to the results of the previous 20 image set calibration as shown in Table 5.3. These results are due to the fact that in these calibration sets, the checkerboard were closer to the camera and larger as a result compared to the other calibration sets. This phenomenon resulted in larger backprojection error as the number of calibration images increased to 15, 20 and 25 as it can be seen in the Figure 5.5. However, the larger checkerboard images do not affect negatively the accuracy of the calibration but they actually improve it. As a result, the set of

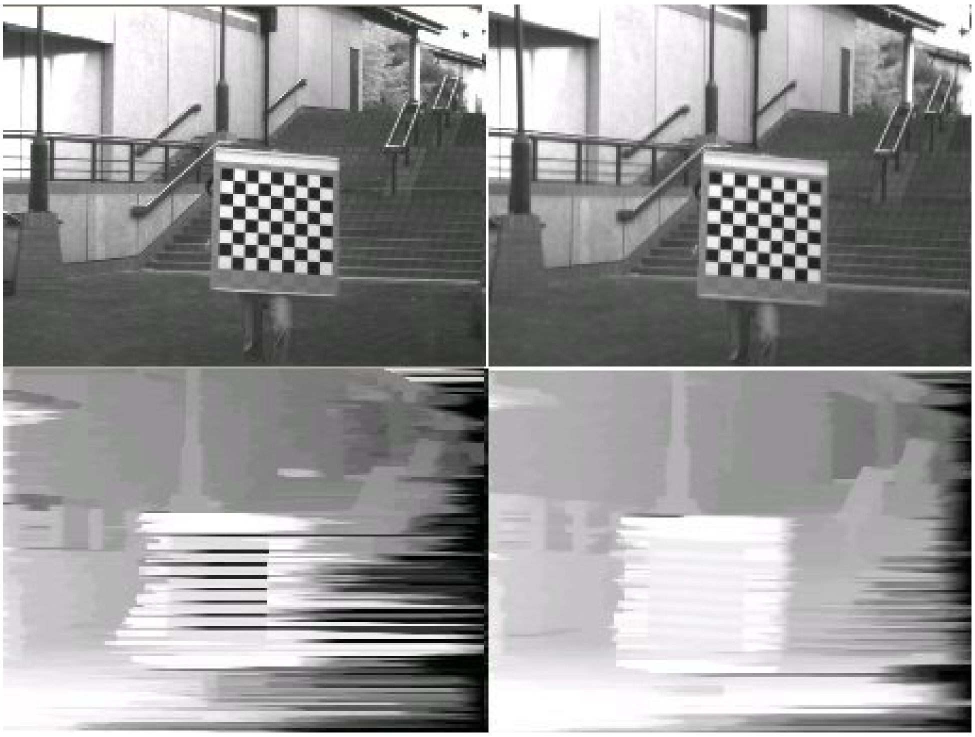

Figure 5.3: Row misalignment error of stereo images by DP (Firefly MV 9mm lens). *Top left:* left camera raw image. *Top right:* same image rectified. *Bottom left:* disparity match of raw images. *Bottom right:* disparity match of the same rectified images.

just five or 10 calibration with closer checkerboard images yielded very good results for such a small set of calibration images and confirmed by the calibration data in Table 5.3. Another indicator of good calibration is the focal length which the (Figures: 5.2,5.4) shows that the backprojection error is within a subpixel range and also good indicator of accuracy not only visual inspection of the disparity matching in rectified images but also comparing with the calibrated focal length against the value specified by the lens manufacturer which can be calculated knowing the camera sensor pixel size. For example, the pixel size of Firefly MV is $6\mu m \times 6\mu m$ thus $1532.93pixels \times 6\mu m \approx 9.19mm$ which is similar to the reported $9mm$ focal length of the Fujinon HF9HA-1B lens.

Thus, the OpenCV calibration method is found to be an accurate and suitable pinhole calibration method as the good convergence of the calibrated focal length

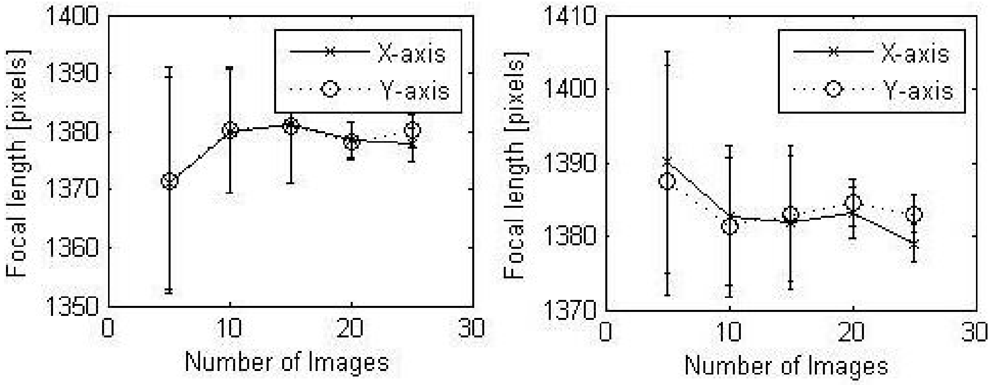

Figure 5.4: Basler at 1280 × 960 with 9mm lens focal length values and error. *Left:* left camera. *Right:* right camera.

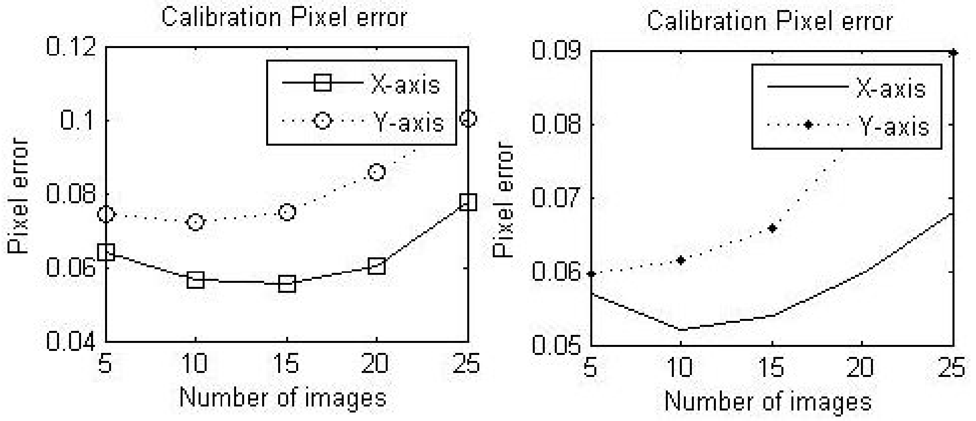

Figure 5.5: Basler at 1280 × 960 resolution with 9mm lens backprojection error. *Left:* left camera. *Right:* right camera.

values shows in Figures 5.2 and 5.4. The focal length was selected as another reference for the quality of calibrations because it is a parameter that can be compared with across the same model of lenses and it is independent of pixel resolutions.

5.3 Bouguet's Calibration Experiments

Bouguet's calibration method for fisheye used calibration images taken with Basler cameras at 1280 × 1024 resolution and another taken with Firefly MV cameras at 640 × 480 (Figures A.11, A.12).

Figure 5.8 shows the disparity match of two linearly rectified images. Table 5.5 shows the result of the calibration with four sets of calibration images from five to 20 images (Figures A.12). The table displays average backprojection errors $error_x$, $error_y$ (pixels) for the x-axis and y-axis respectively and the calibration parameters (intrinsic parameters f_x, f_y, c_x, c_y in pixels and radial distortion parameters $\kappa_1, \kappa_2, \kappa_3, \kappa_4$). The calibration results using Firefly MV with fisheye lens are presented in Table 5.4 and graphs of backprojection error in Figure 5.9.

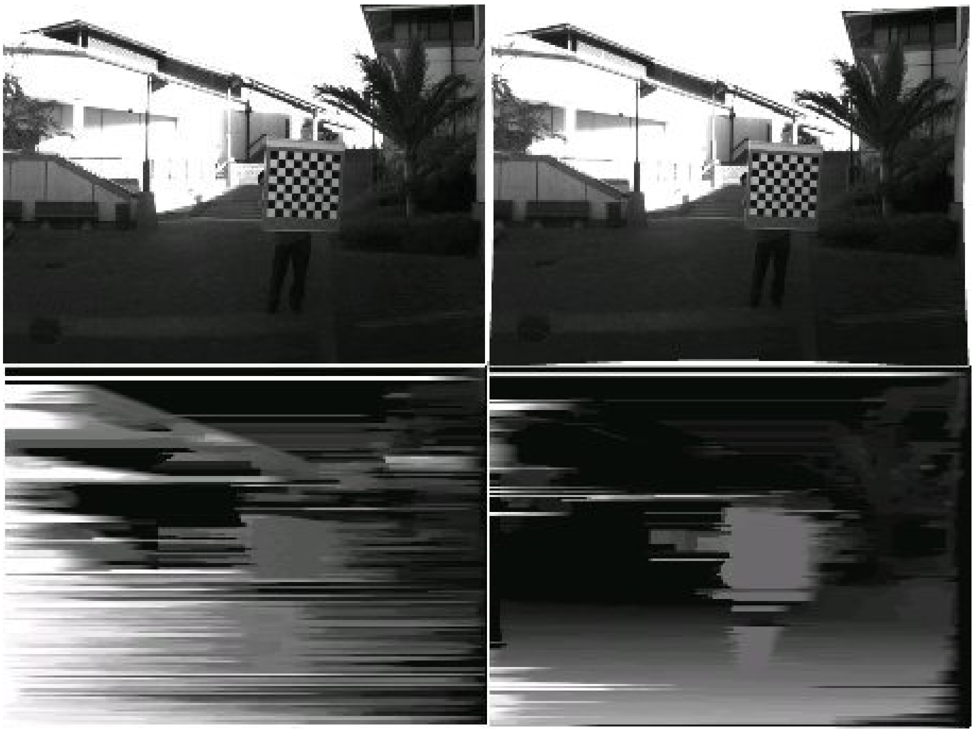

Figure 5.6: Row misalignment error of stereo images by DP (Basler 9mm lens). *Top left:* left camera raw image. *Top right:* same image rectified. *Bottom left:* disparity match of raw images. *Bottom right:* disparity match of the same rectified images.

Table 5.4: Calibration data for left Firefly MV camera with fisheye lens at 640 × 480.

Images	5	10	15	20
$error_x$	0.064	0.069	0.067	0.069
$error_y$	0.070	0.077	0.078	0.080
f_x	297.0 ± 0.7	296.7 ± 0.6	296.6 ± 0.5	296.8 ± 0.4
f_y	296.4 ± 0.7	296.1 ± 0.6	296.1 ± 0.5	296.3 ± 0.4
c_x	341.4 ± 0.5	341.6 ± 0.3	341.6 ± 0.3	341.6 ± 0.2
c_y	240.0 ± 0.4	240.3 ± 0.3	240.3 ± 0.2	240.3 ± 0.2
κ_1	0.027 ± 0.015	0.024 ± 0.011	0.022 ± 0.007	0.021 ± 0.007
κ_2	-0.085 ± 0.079	-0.067 ± 0.063	-0.056 ± 0.031	-0.05 ± 0.030
κ_3	0.137 ± 0.166	0.107 ± 0.134	0.081 ± 0.055	0.067 ± 0.052
κ_4	-0.080 ± 0.120	-0.062 ± 0.098	-0.042 ± 0.037	-0.033 ± 0.031

The calibration results using Basler with fisheye lens are presented in Table 5.5 and graphs of backprojection error in Figure 5.10. We can also compare the calibrated focal length with the manufacturer's specification as follows: the pixel size of Basler AF622 is $6.7\mu m \times 6.7\mu m$ thus $296.55 pixels \times 6.7\mu m \approx 1.98mm$ which is similar to the reported $1.8mm$ focal length of Fujinon FE185C057HA-1.

In summary, this method shows to be precise with a subpixel backprojection error, accurate as the disparity match shows and robust as it can be seen by cal-

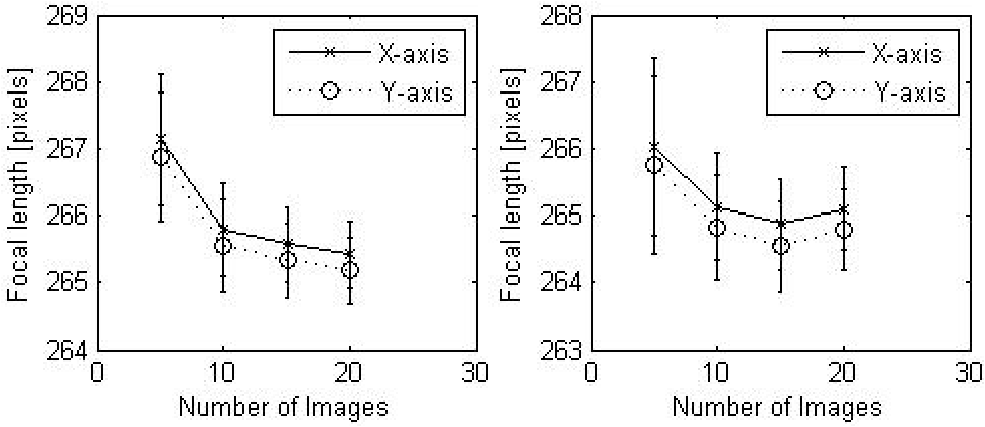

Figure 5.7: Bouguet's calibration data with Basler at 1280 × 1024 with fisheye lens focal length values and error. *Left:* left camera. *Right:* right camera.

ibrated focal length values are converging nicely (Figure 5.7) but more importantly the graphs comparing the calibrated principal point by different methods (Figures 5.11, 5.12) which show the values calculated by Bouguet's method is consistent and confirmed by the results of the other methods.

However, the result of the disparity match (Figure 5.8) is poorer compared to the results of OpenCV (Figures 5.6, 5.3) since in the case of the fisheye lens experiments a small central area of the image have been selected and magnified during the rectification process. Additionally, the Bouguet's calibration method requires a manual initialization of the focal lengths and principal point in order to successfully detect the corners which is cumbersome and not very practical. An initial calibration using the manually entered aforementioned parameters are good initialization parameters

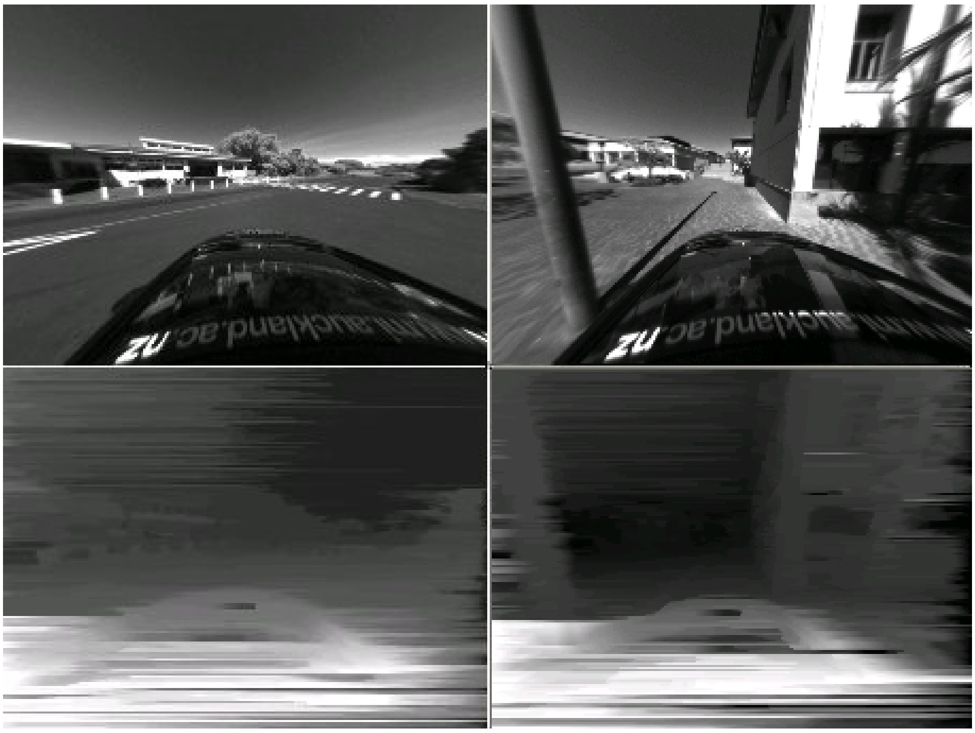

Figure 5.8: Row misalignment error of stereo images by DP (Basler w/fisheye lens). *Top left:* left camera raw image. *Top right:* another raw image. *Bottom left:* disparity match of top left image rectified. *Bottom right:* disparity match of top right image rectified.

Table 5.5: Calibration data for left Basler camera with fisheye lens at 1280 × 1024.

Images	5	10	15	20
$error_x$	0.068	0.070	0.073	0.070
$error_y$	0.070	0.076	0.078	0.075
f_x	267.1 ± 1.0	265.8 ± 0.7	265.6 ± 0.6	265.4 ± 0.5
f_y	266.9 ± 1.0	265.5 ± 0.7	265.3 ± 0.6	265.2 ± 0.5
c_x	632.1 ± 0.2	632.2 ± 0.2	632.3 ± 0.1	632.3 ± 0.1
c_y	488.2 ± 0.2	488.3 ± 0.1	488.1 ± 0.1	488.1 ± 0.1
κ_1	0.001 ± 0.007	0.009 ± 0.005	0.013 ± 0.005	0.014 ± 0.004
κ_2	0.021 ± 0.012	0.007 ± 0.016	-0.005 ± 0.014	-0.008 ± 0.012
κ_3	-0.028 ± 0.023	-0.014 ± 0.019	0.001 ± 0.017	0.005 ± 0.015
κ_4	0.010 ± 0.003	0.005 ± 0.008	-0.001 ± 0.007	-0.002 ± 0.006

for the real calibration which produces accurate calibration data as shown in the experiments. In overall, Bouguet's calibration method is found to be suitable, precise and accurate enough to be used in the DAS context according to these experiments using these sets of calibration images under the described conditions. The need for manual initialization of the some parameters is not a problem since the same lens is used over and over again and just the camera setup needing a re-calibration.

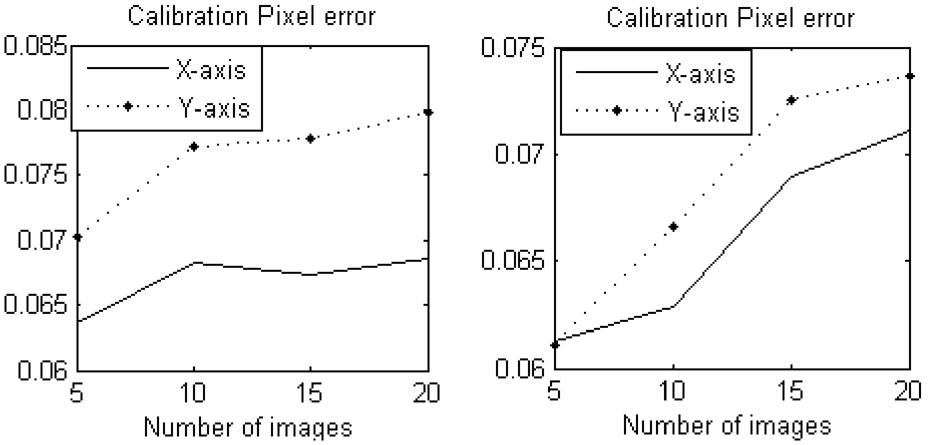

Figure 5.9: Firefly MV at 640 × 480 resolution with fisheye lens backprojection error. *Left:* left camera. *Right:* right camera.

5.4 Mei's Calibration Experiments

Mei's calibration method for fisheye used the same set of calibration images as the previous Bouguet's calibration method (Figures A.11, A.12). Tables 5.6, 5.7 below show the result of the calibrations with average backprojection errors $error_x$,$error_y$ (pixels) for the x-axis and y-axis respectively and the calibration parameters (intrinsic parameters f_x, f_y, c_x, c_y in pixels and radial distortion parameters κ_1, κ_2 as well as tangential distortion parameters κ_3, κ_4).

The calibration results using Firefly MV with fisheye lens are presented in Table 5.6 and graphs of backprojection error in Figure 5.13.

The calibration results using Basler with fisheye lens are presented in Table 5.7 and graphs of backprojection error for the left and right cameras (Figure 5.14). It is interesting to note that the backprojection errors are also in the similar range to those obtained by Bouguet's method. Nevertheless, the difference in the generalized focal length f_x, f_y and principal point are greater than those by Bouguet's method but consistent with them indicating they are accurate but not as precise (Figures 5.11, 5.12). Mei's methods uses the generalized focal length according to the unified projection model which has not physical meaning in the case of fisheye lenses.

Thus, the f_x, f_y values can not be compared to the calibrated focal length by Bouguet's method. Another difference with Bouguet's method implementation is

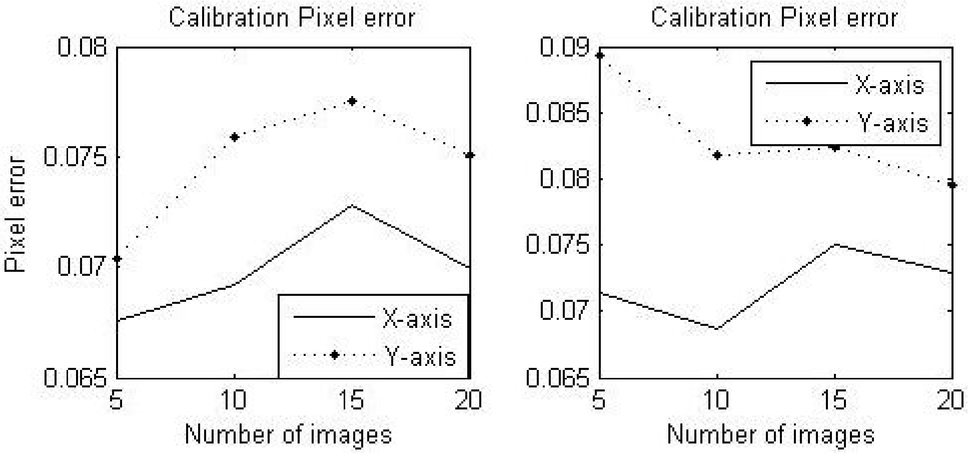

Figure 5.10: Basler at 1280 × 1024 resolution with fisheye lens backprojection error. *Left:* left camera. *Right:* right camera.

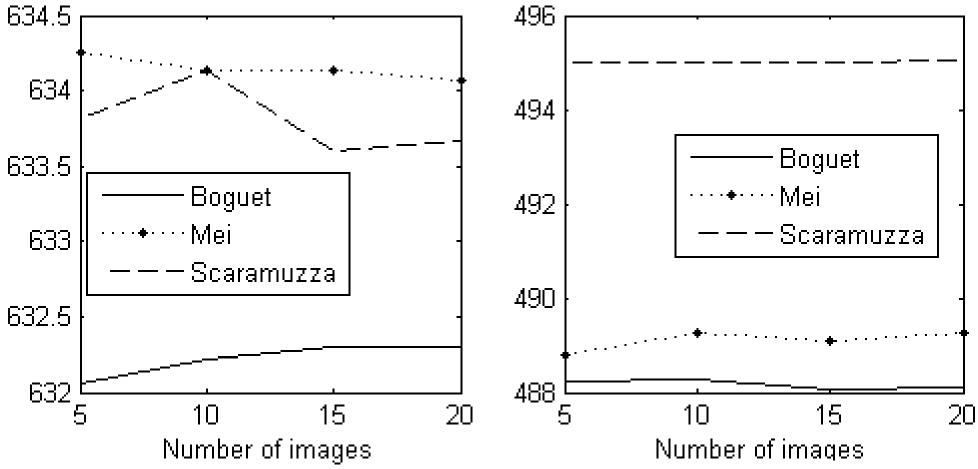

Figure 5.11: Comparison of principal points (left Basler at 1280 × 1024 res.) as function of calibration images. *Left:* x-axis values. *Right:* y-axis value.

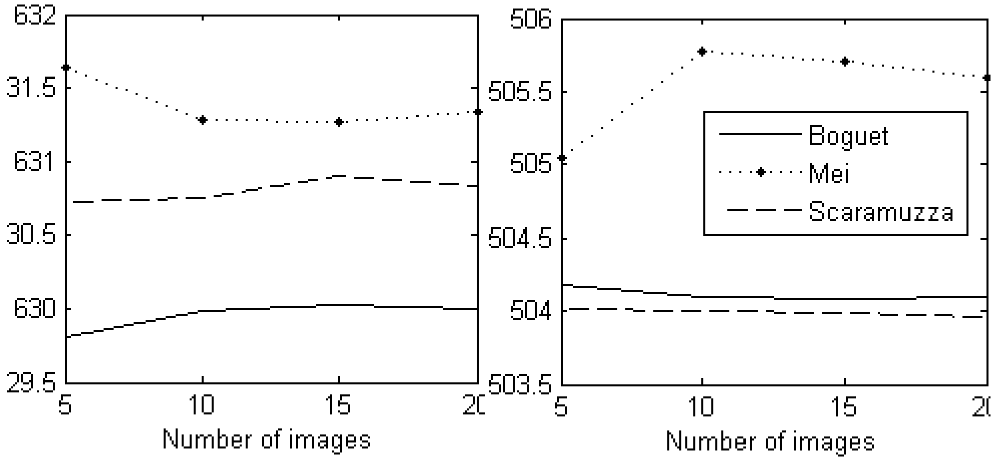

Figure 5.12: Comparison of principal points (right Basler at 1280 × 1024 res.) as function of calibration images. *Left:* x-axis values. *Right:* y-axis value.

the semi-automatic initialization of intrinsic parameters which performed quite well and easily. However, the fact that the calibration results are not as precise as Bouguet's method makes it less preferable than Bouguet's method as well as the need for many

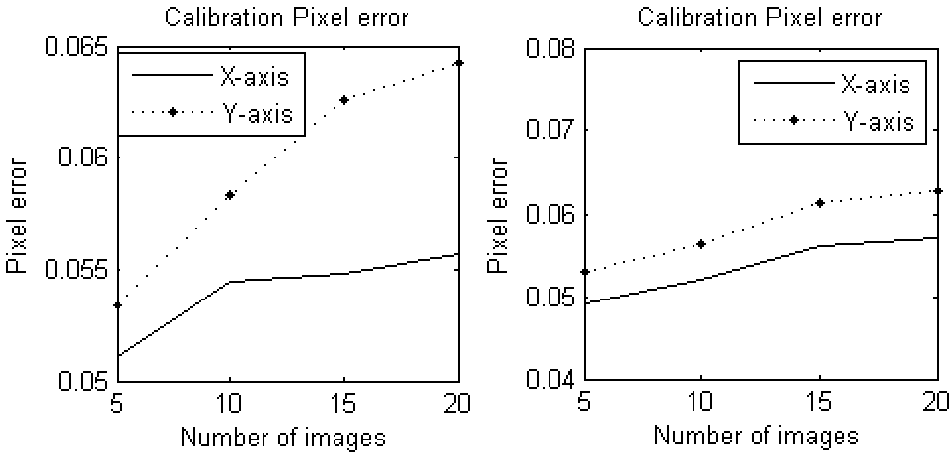

Figure 5.13: Firefly MV camera with 9mm with fisheye lens backprojection error. *Left:* left camera. *Right:* right camera.

more calibration images as we can observe the slower convergence by comparing the values of f_x, f_y calibrated with five or ten images against those calibrated with 20 images. In overall, Mei's calibration method is the second best behind Bouguet's method according to these experiments using these sets of calibration images under the described conditions.

Table 5.6: Calibration data for left Firefly camera with fisheye lens at 640 × 480.

Images	5	10	15	20
$error_x$	0.051	0.054	0.055	0.056
$error_y$	0.053	0.058	0.063	0.064
f_x	417.5 ± 46.3	427.5 ± 45.4	547.4 ± 82.7	527.4 ± 47.1
f_y	416.7 ± 46.2	426.8 ± 45.3	546.4 ± 82.6	526.4 ± 47.0
c_x	342.6 ± 0.5	342.7 ± 0.3	342.7 ± 0.3	342.6 ± 0.2
c_y	241.4 ± 0.4	241.6 ± 0.3	241.7 ± 0.3	241.7 ± 0.2
κ_1	-0.341 ± 0.007	-0.343 ± 0.005	-0.313 ± 0.040	-0.321 ± 0.018
κ_2	0.136 ± 0.012	0.138 ± 0.010	0.108 ± 0.036	0.113 ± 0.016
κ_3	-0.0003 ± 0.0003	-0.0004 ± 0.0002	-0.0003 ± 0.0003	-0.0003 ± 0.0002
κ_4	-0.0005 ± 0.0005	-0.0002 ± 0.0003	-0.0005 ± 0.0004	-0.0005 ± 0.0003

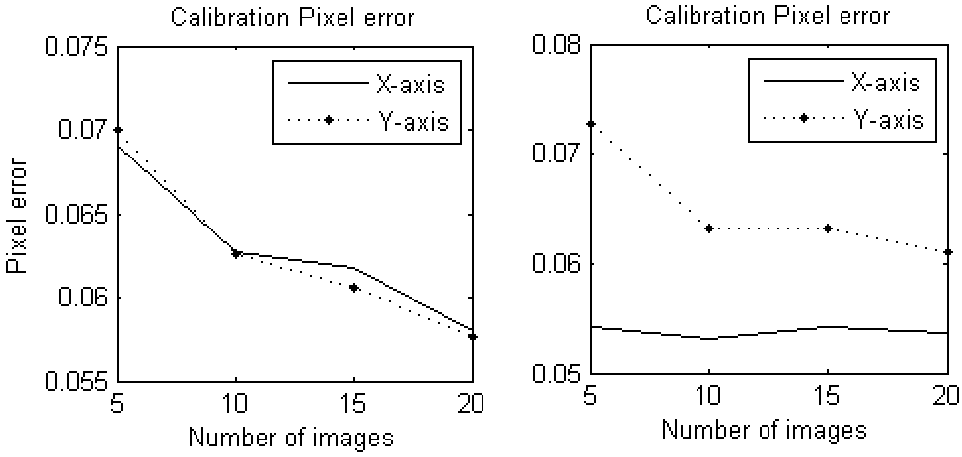

Figure 5.14: Basler at 1280 × 1024 resolution with fisheye lens backprojection error. *Left:* left camera. *Right:* right camera.

5.5 Scaramuzza's Calibration Experiments

Scaramuzza's calibration method for fisheye used the same set of calibration images as the previous Bouguet's and Mei's calibrations methods (Figures A.11, A.12).

The tables 5.9, 5.8 show the results of the fisheye calibrations similarly to previous experiments with the single backprojection error as $error$ (pixels) as the average of $error_x$,$error_y$ and the calibration parameters (principal point coordinates c_x, c_y in

Table 5.7: Calibration data for left Basler camera with fisheye lens at 1280 × 1024.

Images	5	10	15	20
$error_x$	0.069	0.063	0.062	0.058
$error_y$	0.070	0.063	0.061	0.058
f_x	496.4 ± 56.8	501.0 ± 39.6	500.6 ± 33.0	506.8 ± 30.3
f_y	496.0 ± 56.7	500.5 ± 39.6	500.1 ± 33.0	506.3 ± 30.3
c_x	634.3 ± 0.5	634.1 ± 0.3	634.1 ± 0.2	634.1 ± 0.2
c_y	488.8 ± 0.6	489.3 ± 0.4	489.1 ± 0.3	489.3 ± 0.3
κ_1	-0.294 ± 0.031	-0.302 ± 0.021	-0.301 ± 0.017	-0.302 ± 0.016
κ_2	0.097 ± 0.017	0.101 ± 0.012	0.101 ± 0.010	0.102 ± 0.010
κ_3	0.0005 ± 0.0009	-0.0003 ± 0.0007	-0.0001 ± 0.0005	-0.0004 ± 0.0004
κ_4	-0.0014 ± 0.0008	-0.0013 ± 0.0004	-0.0013 ± 0.0003	-0.0012 ± 0.0002

Table 5.8: Calibration data for left Firefly camera with fisheye lens at 640 × 480.

Images	5	10	15	20
$error$	0.0979	0.1042	0.1060	0.1078
c_y	241.7	241.7	241.7	241.7
c_x	344.2	344.1	344.1	344.1
a_1	100.0	100.0	100.0	100.0
a_2	-2.96	-2.96	-2.96	-2.96
a_3	0	0	0	0
a_4	9.62×10^{-5}	1.30×10^{-5}	1.02×10^{-5}	1.02×10^{-5}
a_5	1.10×10^{-8}	4.51×10^{-8}	4.34×10^{-8}	8.14×10^{-8}
a_6	1.49×10^{-11}	8.30×10^{-14}	2.18×10^{-11}	2.90×10^{-11}

pixels and the distortion parameters $a_1, a_2, a_3, a_4, a_5, a_6$. The focal length is a function of ρ thus not included in the tables.

The calibration results are presented in Table 5.8, 5.9 and graph of backprojection error in Figure 5.15. The backprojection error is noticeable greater especially in the left Basler camera experiments (Figure 5.15) which translates into less accurate calibration.

Because the same calibration images were used it is possible to estimate the row mismatch by comparing the row component of the calibrated principal point c_y as mentioned in Section 4.3 since c_y value is representative of any inaccuracies in row

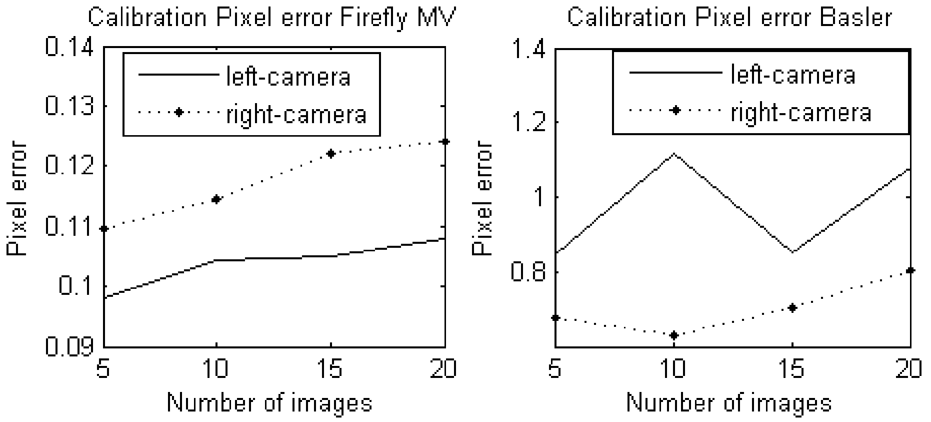

Figure 5.15: Backprojection error. *Left:* Firefly MV at 640 × 480. *Right:* Basler at 1280 × 1024 resolution.

Table 5.9: Calibration data for left Basler camera with fisheye lens at 1280 × 1024.

Images	5	10	15	20
$error$	0.844	1.117	0.851	1.082
c_y	495.0	495.0	495.0	495.0
c_x	633.8	634.1	633.6	633.7
a_1	100.0	100.0	100.0	100.0
a_2	-2.72	-2.70	-2.71	-2.73
a_3	0	0	0	0
a_4	1.56×10^{-5}	1.48×10^{-5}	1.39×10^{-5}	1.30×10^{-5}
a_5	-2.43×10^{-8}	-1.93×10^{-8}	-1.06×10^{-8}	-8.00×10^{-9}
a_6	5.61×10^{-11}	4.79×10^{-11}	2.65×10^{-11}	2.51×10^{-11}

mismatches.

As Table 5.10 shows, the difference in c_y values between Bouguet and Mei's methods are approximately $1.17 pixels$ and $1.5 pixels$ for the left and right cameras respectively. However, a comparison between Bouguet and Scaramuzza's methods yields a difference in c_y of approximately $6.94 pixels$ and $0.14 pixels$ for the left and right cameras respectively (see Figure 5.16). We can conclude that Bouguet and Mei's values are quite accurate while the Scaramuzza's left camera calibration value is not. This row mismatching inaccuracy is confirmed again by the poorer results seen in Figure 5.18 when compared to the similar disparity match in using Bouguet's calibration data (Figure 5.8).

The reasons for the inaccuracies and greater errors are most likely due to the automatic corner detection feature which it does not perform well as noted earlier (Figure 3.9). In fact, it was almost not practical that for the lower resolution images all the corners had to be detected manually which makes it very impractical and time consuming task. Additionally, the tables 5.8, 5.9 shows no improvement or

Table 5.10: Comparison of backprojection error and row errors (fisheye calibration of Basler cameras at 1280 × 1024 resolution with 20 calibration images.

Camera	Parameter	Bouguet	Mei	Scaramuzza
Left	$error$	0.073	0.058	1.082
	c_y	488.1 ± 0.1	489.3 ± 0.3	495.0
	c_x	632.3 ± 0.1	634.1 ± 0.2	633.7
Right	$error$	0.076	0.057	0.80
	c_y	504.1 ± 0.1	505.6 ± 0.2	504.0
	c_x	632.0 ± 0.2	631.3 ± 0.2	630.8

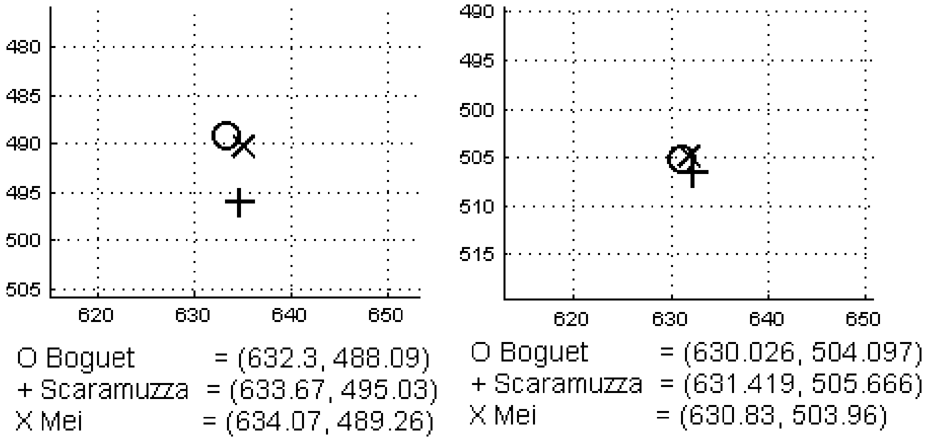

Figure 5.16: Comparison of principal points calibrated by different methods (Basler at 1280 × 1024). *Left:* left camera. *Right:* right camera.

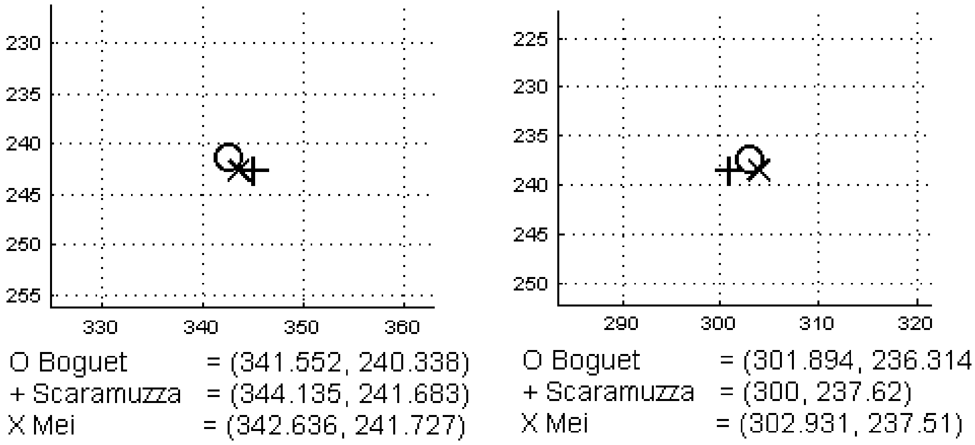

Figure 5.17: Comparison of principal points calibrated by different methods (Firefly MV at 640 × 480). *Left:* left camera. *Right:* right camera.

very slow convergence of the intrinsic parameters when increasing the number of calibration images unlike the other methods where there are improvements.

In summary, the results of comparing these three fisheye calibration methods using the same Basler cameras setup and set of calibration images shows in Table 5.10

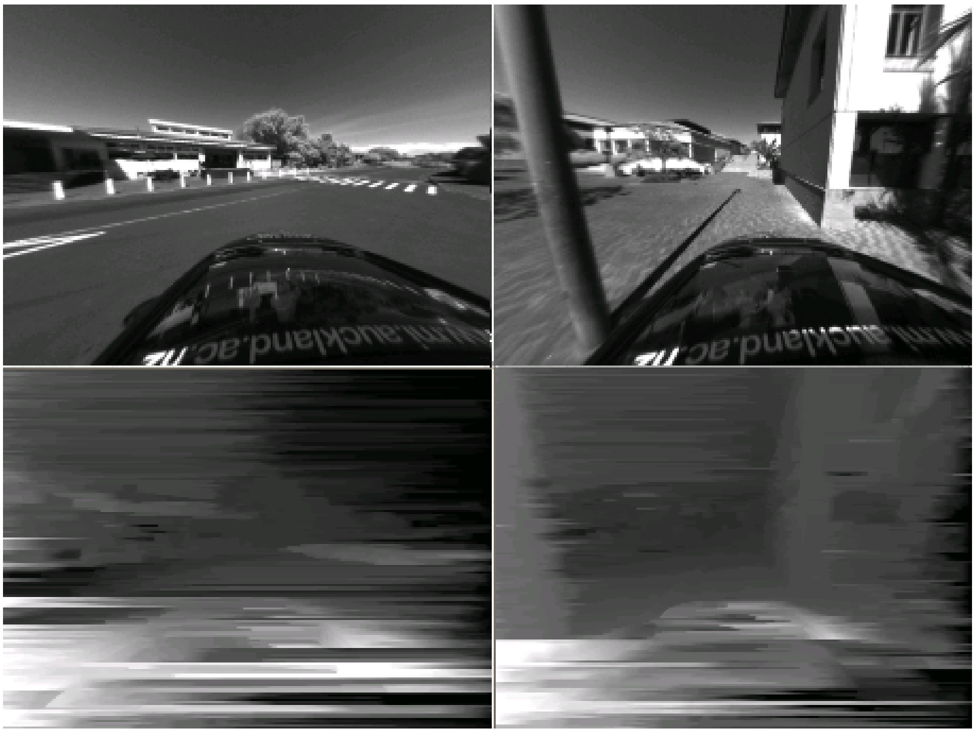

Figure 5.18: Row misalignment error of stereo images by DP (Basler w/fisheye lens). *Top left:* left camera raw image. *Top right:* another raw image. *Bottom left:* disparity match of top left image rectified. *Bottom right:* disparity match of top right image rectified.

the backprojection errors as well as the calibrated principal points results shown in Figure 5.16 confirms the results that the Bouguet's method performed better (closely followed by Mei) in terms of accuracy as the error in the principal points are smaller even though the backprojection error was slightly larger than Mei's values.

Finally, Scaramuzza's calibration method performed the worst in accuracy, robustness and suitability according to these experiments using these sets of calibration images under the described conditions.

Chapter 6

Conclusions

This chapter summarizes the results of the experiments to test and compare the selected calibration methods for various types of lenses in different configurations in DAS context. Finally, possible areas for future work and improvements are identified.

6.1 Results

The OpenCV experiments setup in the DAS context with two different camera configurations and resolutions confirmed its accuracy and robustness with even only ten calibration images.

The focal length as a function of used calibration images (Figures: 5.2 and 5.4) shows that the backprojection error is within a subpixel range and a good indicator of accuracy which is also confirmed by the visual comparison of disparity matching of raw images and rectified images.

Thus, the OpenCV calibration method is confirmed to be an accurate and suitable pinhole calibration method based on the backprojection error results, disparity matches and the good convergence of the calibrated focal length values shows in Figures 5.2 and 5.4. The focal length was selected as another reference for the quality of calibrations because it is a parameter that can be compared with across the same model of lenses and it is independent of pixel resolutions. Therefore it is a good indicator of accuracy for other pinhole camera calibrations.

The experiments with fisheye lens calibration methods were tested next with Bouguet's fisheye calibration method presented. This method proved to be accurate as well as precise within a subpixel backprojection error. Additionally, the calibrated focal length values are converging well when increasing the number of calibration images (Figure 5.7). Finally, the comparison of the principal points with other calibration methods as well the disparity matches of linearly rectified images show that the calibration data to be precise enough. As stated earlier, the need for a manual initialization is not a problem since the camera setups are reused often in the

DAS context and the re-calibration is needed just as a measure of normal maintenance due to vibrations and movements of normal driving. In overall, Bouguet's calibration method and implementation seems most suitable, and accurate enough according to these experiments using the given sets of calibration images. Mei's calibration method came second mostly because of its of accuracy and finally Scaramuzza's calibration method proved almost impractical to use but more importantly the calibration results were not improving or converging very slowly as well as great differences noted in the principal point of one of the cameras. The focal length is not comparable in the fisheye lenses since it depends on the particular camera and projection model used. However, the principal point can be used as reference since the same calibration images were used in all fisheye calibration methods thus they should return the same or very similar principal point values.

In conclusion, OpenCV calibration method is an accurate and suitable calibration method for vision-based DAS even using small number of calibration images of a checkerboard in both low and high resolution images. As a result, the experiments shows that the optimal number of calibration images for OpenCV is around ten as a good balance between accuracy and calibration errors for the smallest set of calibration images. On the other hand, for fisheye lenses configurations, Bouguet's method seems accurate and robust enough as well as suitable performing well with also a small set of calibration images. Mei's method came second in accuracy and requiring a large number of calibration images but easy to use because the semi-automatic intrinsic parameters initialization performs quite well. Scaramuzza's method did not require any manual input nor parameters initialization but the automatic corner detection which made it almost impractical but most importantly failed in robustness, accuracy and precision.

6.2 Future Work

Three areas can be identified for further improvements and extensions. Firstly, using different checkerboards that might improve the performance of Scaramuzza's calibration method since an accurate corner detection provides accurate data for calibration. It seems plausible that the poor performance of automatic corner detection affected significantly on the accuracy of Scaramuzza's calibration method. Secondly, another evaluation technique independent of image resolution such as the normalized calibration error used to comparing the calibration methods. Finally, a robust and accurate automatic or self-calibration method for all lenses (pinhole and fisheye lenses) would be not only convenient but necessary as normal production cars incorporates stereo cameras for DAS.

Appendix A

Appendix

Additional figures and graphs with closeups of the inside camera mounting bar and external roof mounting, set of calibration images, experiments' focal length and extrinsic parameters plotted among others.

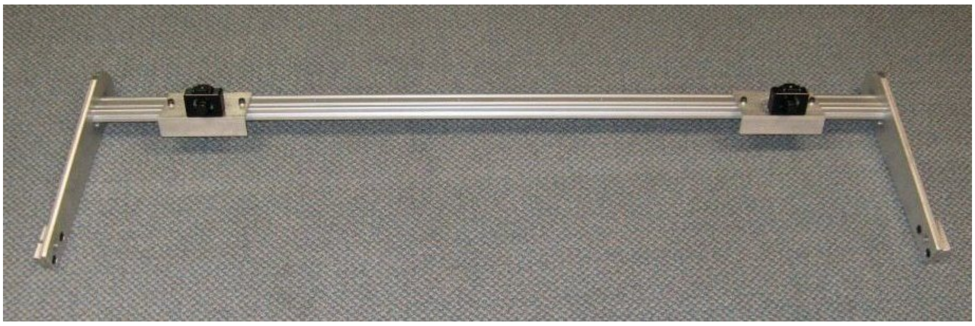

Figure A.1: Inside cabin cameras mounting bar (ready to be fitted in HAKA1).

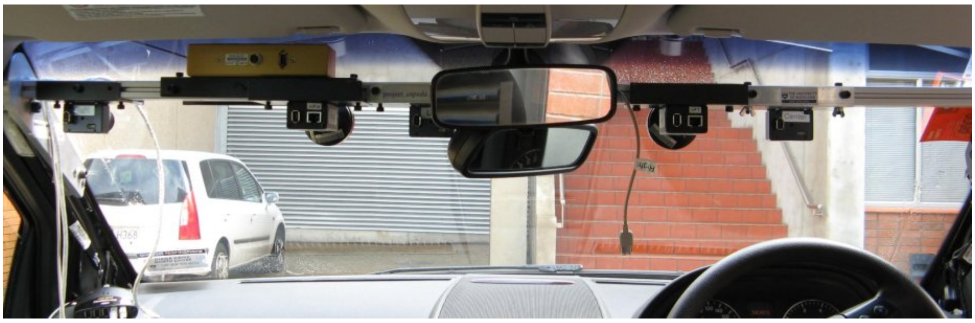

Figure A.2: Inside cabin camera's mounting bar (six cameras fitted including stereo Bumblebee camera).

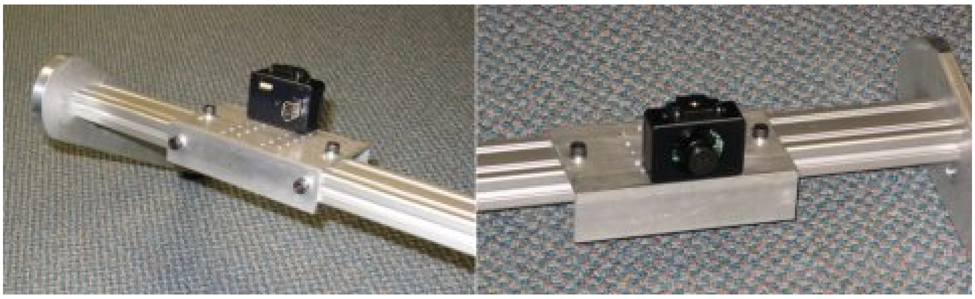

Figure A.3: Close view of a bracket and the mounting bar (Firefly MV camera with 9mm lens).

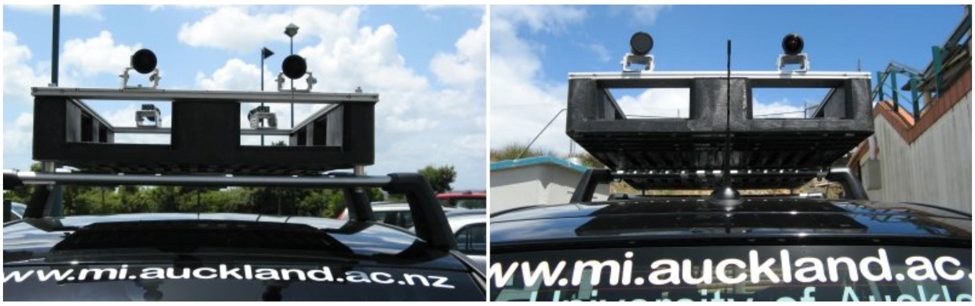

Figure A.4: View of externally mounted fish eye cameras. *Left:* front view of HAKA1. *Right:* rear view of HAKA1.

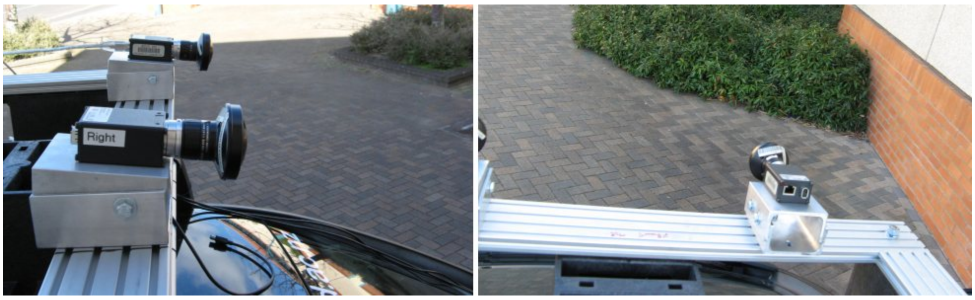

Figure A.5: Close view of the brackets for externally mounted fish eye cameras.

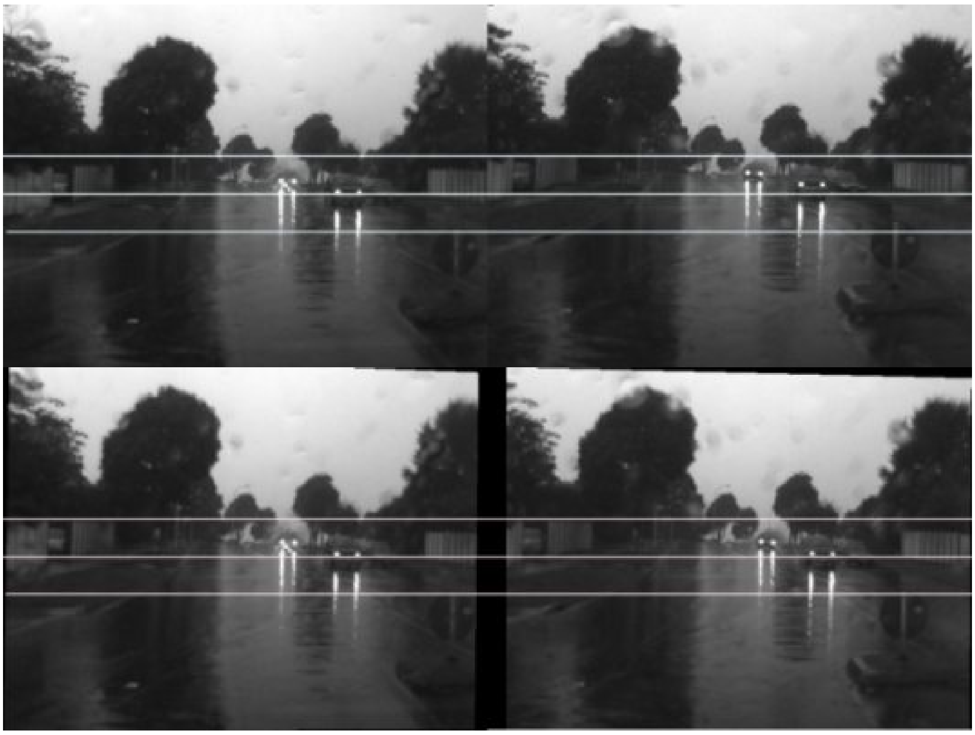

Figure A.6: Pinhole camera stereo rectification (9mm lens). *Top:* stereo image before rectification. *Bottom:* rectified stereo images.

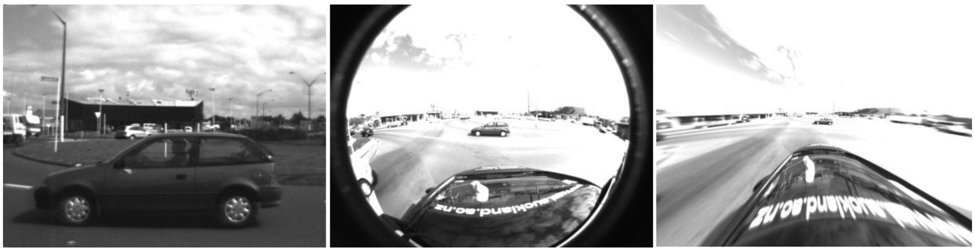

Figure A.7: Comparison of pinhole and fish eye lens for the same scene. *Left:* normal 9mm lens pinhole camera. *Center:* raw image taken with fisheye 1.8mm lens. *Right:* same image rectified.

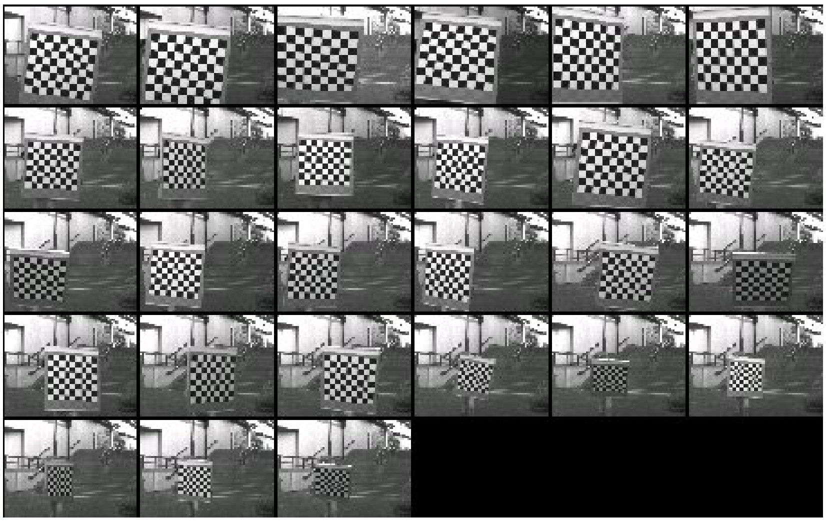

Figure A.8: Calibration images of the left Firefly MV camera with 9mm lens.

Figure A.9: Calibration images of the left Basler camera with 9mm lens at 1280 × 960.

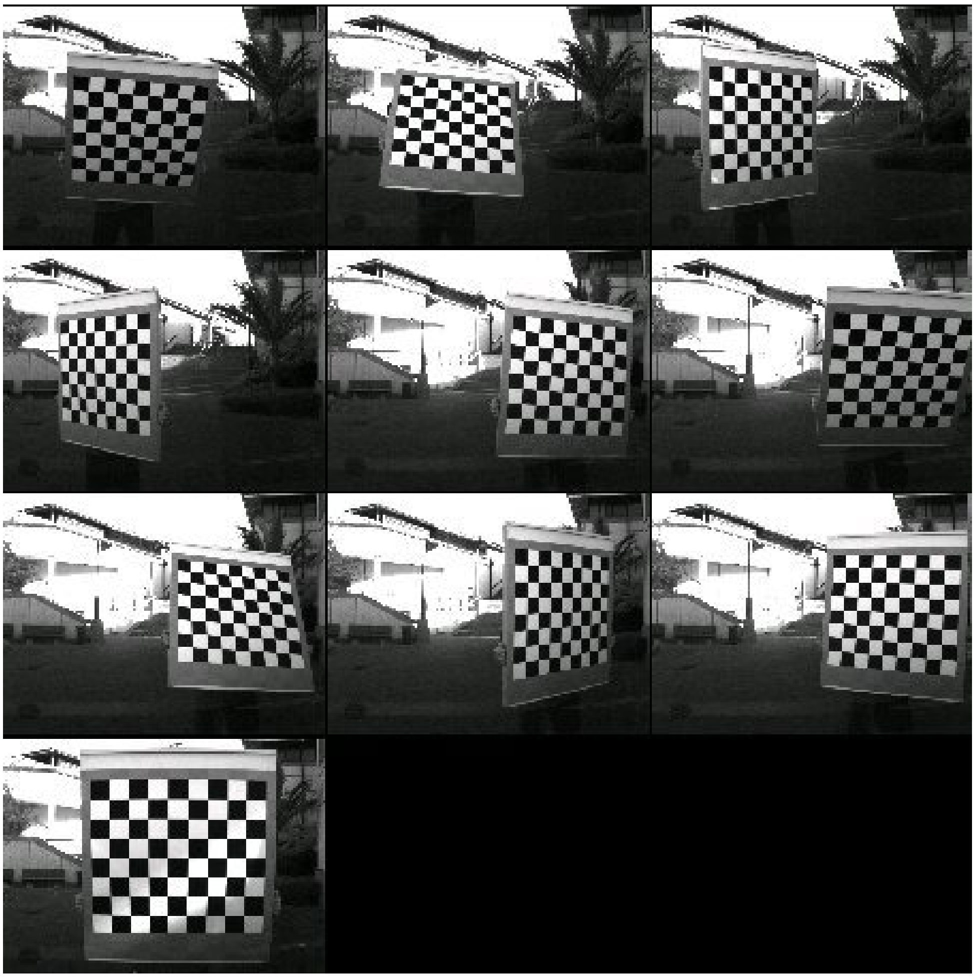

Figure A.10: Set of 10 calibration images (with closer checkerboard images) taken with Basler camera at 1280 × 960

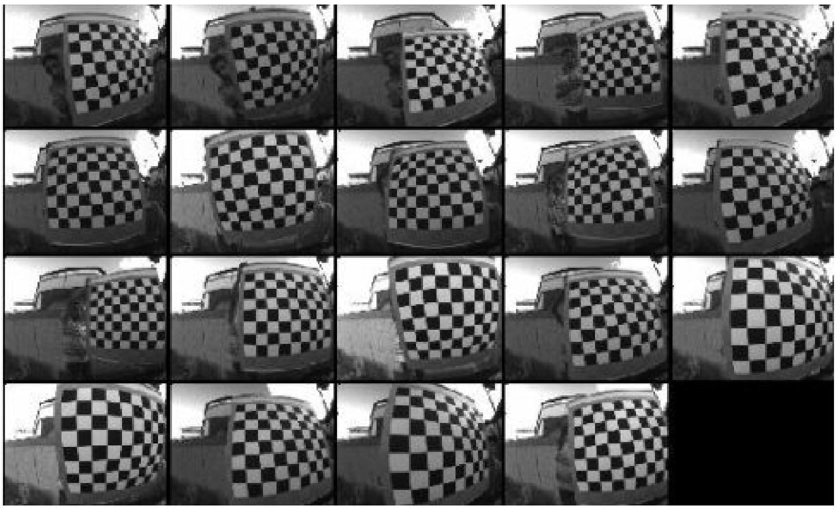

Figure A.11: Calibration images of the Firefly MV camera with fisheye lens.

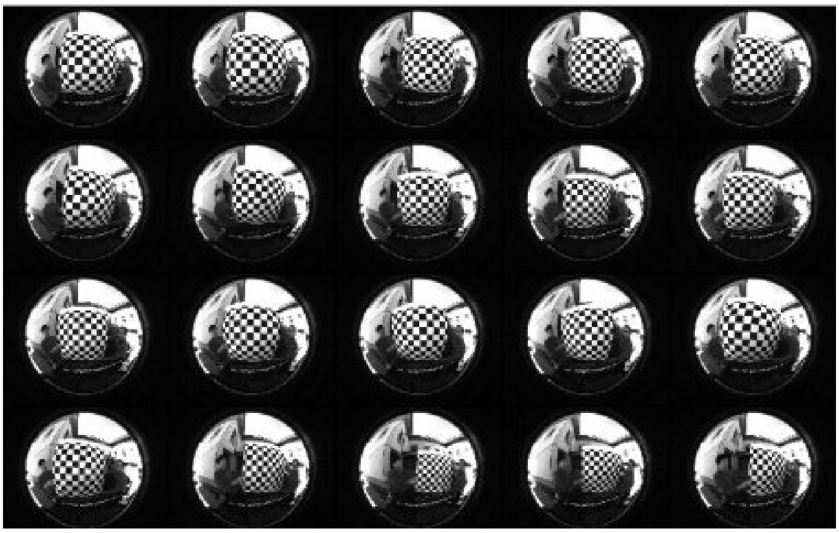

Figure A.12: Calibration images of the left Basler camera with fisheye lens at 1280 × 1024.

Bibliography

[1] Abraham, S., Förstner, W.: Fish-eye-stereo calibration and epipolar rectification. *ISPRS Journal Photogrammetry Remote Sensing*, **59**:278–288, 2005.

[2] Abdel-Aziz, Y. I., Karara, H. M.: Direct linear transformation into object space coordinates in close-range photogrammetry. In Proc. *Symposium Close-Range Photogrammetry*, pages 1–18, 1971.

[3] Barreto, P., Araujo, H.: Issues on the geometry of central catadioptric image formation. In Proc. *IEEE Conf. Computer Vision Pattern Recognition*, volume 1, pages 422-427, 2001.

[4] Bertozzi, M., Broggi, A., Fascioli, A., Graf, T., Meinecke, M.: Pedestrian detection for driver assistance using multiresolution infrared vision. *IEEE Trans. on Vehicular Technology*, **5**:1666–1678, 2004.

[5] Bouguet, J.: Camera calibration toolbox for Matlab. Retrieved on February 14, 2009 from `www.vision.caltech.edu/bouguetj/calib_doc/`

[6] Dang, T., Hoffmann, C.: Tracking camera parameters of an active stereo rig. In Proc. *DAGM*, LNCS 4174, pages 627–636, 2006.

[7] Fitzgibbon, A.: Simultaneous linear estimation of multipleview geometry and lens distortion. In Proc. *IEEE Conf. Computer Vision Pattern Recognition*, volume 1, pages 125–132, 2001.

[8] Fletcher, L., Loy, G., Barnes, N., Zelinsky, A.: Correlating driver gaze with the road scene for driver assistance systems. *Robotics Autonomous Systems*, **52**:71–84, 2005.

[9] Gehrig, S. K.: Large-field-of-view stereo for automotive applications. In *OmniVis 2005*, Beijing, 2005.

[10] Geyer, C., Daniilidis, K.: A unifying theory for central panoramic systems and practical implications. In Proc. *European Conf. Computer Vision*, pages 445–461, 2000.

[11] Hartley, R.: Theory and practice of projective rectification. *Int. Journal of Computer Vision*, **35**: 115–127, 1999.

[12] Hartley, R., Zisserman, A.: *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, 2000.

[13] Heikkila, J., Silven,O.: A four-step camera calibration procedure with implicit image correction. In Proc. *IEEE Conf. Computer Vision Pattern Recognition*, pages 1106–1112, 1997.

[14] Huang, P., Chen, C., Hsaio, P., Fu, L.: On-board vision system for lane recognition and front-vehicle detection to enhance driver's awareness. In Proc. *IEEE Int. Conf. on Robotics and Automation*, volume 3, pages 2456–2461, 2004.

[15] Jiang, R., Klette, R., Wang, S., Vaudrey, T.: New lane model and distance transform for lane detection and tracking. In Proc. *CAIP*, LNCS 5702, pages 1044–1052, 2009.

[16] Morales, S., Woo, Y., Klette, R., Vaudrey, T.: A study on stereo and motion data accuracy for a moving platform. In Proc. *PSIVT*, LNCS 5744, pages 292–300, 2009.

[17] Kannala, J., Brandt, S.: A generic camera calibration method for fish-eye lenses. In Proc. *Int. Conf. Pattern Recognition*, volume 1, pages 10–13, 2004.

[18] Klette, R., Schlns, K., Koschan, A.: *Computer Vision: Three-Dimensional Data from Images*. Springer, Singapore, 1998.

[19] Lu, M., Wevers, K., Van Der Heijden, R.: Technical feasibility of advanced driver assistance systems (ADAS) for road traffic safety. *Transportation Planning Technology*, **28**:167–187, 2005.

[20] Mei, C., Rives, P.: Single view point omnidirectional camera calibration from planar grids. In Proc. *IEEE Conf. Robotics Automation*, pages 3945–3950, 2007.

[21] Multimedia Imaging Portal, *.enpeda..* project description. Retrieved February 15, 2009 from `www.mi.auckland.ac.nz/enpeda`.

[22] Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A., Jarawan, E., Mathers, C.: *World Report on Road Traffic Injury Prevention*, World Health Organization, Geneva, 2004.

[23] Scaramuzza, D., Siegwart, R.: A practical toolbox for calibrating omnidirectional cameras. In *Vision Systems Applications* (G. Obinata and A. Dutta, editors), pages 297–310, I-Tech Education and Publishing, Vienna, 2007.

[24] Stein, G., Rushinek, E., Hayun, G., Shashua, A.: A computer vision on a chip: a case study from the automotive domain. In Proc. *Computer Vision and Pattern Recognition Workshop*, page 130, 2005.

[25] Sun, W., Cooperstock, J. R.: Requirements for camera calibration: must accuracy come with a high price? In Proc. *IEEE Workshop Applications Computer Vision*, volume 1, pages 356–361, 2005.

[26] Thirthala, S., Pollefeys, M.: The radial trifocal tensor: A tool for calibrating the radial distortion of wide-angle cameras. In Proc. *IEEE Conf. Computer Vision Pattern Recognition*, volume 1, pages 321–328, 2005.

[27] Tsai, R.: A versatile camera calibration technique for high-accuracy 3d machine vision metrology. using-of-the- shelf tv cameras and lenses. *IEEE J. Robotics Automation*, **3**:323–344, 1987.

[28] Wang, C., Huang, S., Fu, L.: Driver assistance for lane detection and vehicle recognition with night vision. In Proc. *IEEE Int. Conf. on Robotics and Automation*, pages 3530–3535, 2005.

[29] Weng, J., Cohen, P., Hernoiu, M.: Camera calibration with distortion models and accuracy evaluation. *IEEE Trans. Pattern Analysis Machine Intelligence*, **14**: 965–980, 1992.

[30] Wimmer, P.: Camera Sync Tester User Guide Retrieved on September 8, 2010 from `http://www.3dtv.at/knowhow/Synctest_en.aspx`

[31] Zhang, Z.: Flexible camera calibration by viewing a plane from unknown orientations. In Proc. *Int. Conf. Computer Vision*, pages 666–673, 1999.

[32] Zheng, N. N., Tang, S., Cheng, H., Li, Q., Lai, G., Wang, F. Y.: Toward intelligent driver-assistance and safety warning systems. *IEEE Intelligent Systems*, **19**(2):8–11, 2004.

Index