Multi-Camera Calibration for Accurate Geometric Measurements in Industrial Environments

R. Usamentiaga[a,*], D. F. Garcia[a]

[a]*Department of Computer Engineering, University of Oviedo, Campus de Viesques, 33204 Gijón, Spain*

Abstract

Accurate camera calibration is a challenging task required for 3D reconstruction sensors. In structured light sensors based on laser line projection, it is necessary to determine the position and orientation of the camera relative to the laser plane. The standard approach based on laser plane fitting requires a difficult setup, particularly in multi-camera configurations. This work proposes a calibration method based on a calibration plate with protruding cylinders. The projection of the laser line on the calibration plate produces a set of partial contours of ellipses, resulting from the projection of the cross section of the cylinders. These contours are used to calculate the position and orientation of the camera accurately. The proposed procedure is divided in two steps: a coarse calibration based on a robust estimator, which provides a rough approximation, and a fine calibration that iteratively optimizes the solution. The result is an accurate and robust procedure that can be applied to multiple cameras simultaneously on devices with limited computational power. Extended tests are used to validate the procedure with test pieces. Results indicate the calibration can be performed in 30 ms, producing a calibration error of only 0.027 mm.

Keywords: Calibration target, Camera projection, Camera calibration, Geometric measurement

1. Introduction

3D reconstruction is one of the most challenging fields in computer vision [1]. The goal is to extract accurate geometric information about the objects in the scene using images acquired from one or more cameras. There is a wide variety of techniques to achieve this goal, which can be broadly classified as passive or active [2]. Among passive techniques, stereo vision is one of the most popular and widely studied techniques. In stereo vision, 3D reconstruction is achieved by finding matching pixels from images acquired from different points of view. The 2D positions of these pixels are used to triangulate the corresponding 3D position in the scene. This approach presents robustness problems, especially in the presence of surfaces with uniform textures where finding matching features is difficult. Active techniques use an external light source that projects a structured light pattern onto the scene, greatly improving the performance of the 3D reconstruction. The projected shape of the light pattern on the objects of the scene is used to reconstruct 3D surfaces [3]. The projected light also increases the number of matching features, providing a dense 3D surface.

Active 3D reconstruction sensors that project a single laser line onto the objects are considered one of the most reliable 3D reconstruction sensors [4, 5]. The laser line is projected across the objects in the scene, producing a deformed profile according to the shape of the objects it is illuminating. Optical triangulation is then used to estimate the corresponding 3D coordinates of the illuminated pixels in the profile. In order to reconstruct the whole 3D surface of the object, a relative movement is required between the object and the structured light sensor, which consists of a camera and a laser line projector. This 3D reconstruction technique is very accurate and robust, as it does not depend on changing illumination conditions [6]. Therefore, it has been widely used in many different industrial applications, such as weld inspection [7], flatness measurement [8] or metrological analysis [9].

Calibration is a crucial aspect of structured light sensors. The extraction of metric information from the images requires an accurate estimation of the parameters that control the projection of light onto the camera sensor. The calibration procedure is generally performed in two steps: the calibration of the internal camera parameters, and the calibration of the position and orientation of the camera. The calibration of the internal camera parameters, including distortions, is a well-known procedure widely studied in the field [10]. The most common approach is the observation of a flat calibration plate with different orientations. The observations of control markers on the plate are used to estimate the coefficients of the projection model. The second step required for 3D reconstruction using laser line projectors is the determination of the position and orientation of the camera with respect to the laser light plane, commonly known as the pose. One of the most accurate methods is based on laser plane fitting. This procedure requires the observation of the laser line projected on the same calibration plate used for the internal calibration. At least two images acquired from the cali-

*Corresponding author

Email address: rusamentiaga@uniovi.es (R. Usamentiaga)
URL: http://www.atc.uniovi.es/rusamentiaga (R. Usamentiaga)

bration plate at different positions or orientations are necessary. The coordinates of the projected laser lines in the images are extracted and used to fit a plane. The result of this procedure is the equation of the laser plane in world coordinates, which indicate the position and orientation of the laser plane relative to the camera. The detailed procedure can be found in [11]. Recent applications of this same procedure can be found in [12, 13].

The plane fitting method is very accurate. However, it requires moving the calibration plate in front of the camera. Also, at least two images of the laser line projected on the calibration plate are required. Moreover, the resulting calibration is only valid while the relative position between the camera and the laser projector remains constant. Applying this procedure in industrial environments has a significant number of drawbacks. In industrial environments, machines and sensors are affected by vibrations. Eventually, this vibration loosen the screws, changing the relative position of the camera and the projector. These displacements are very small, tenths of a millimeter, but they significantly decrease the performance of the 3D reconstruction sensor. Therefore, in these conditions, calibration must be performed periodically to ensure maximum performance. The calibration using the plane fitting is time consuming, and hence not suitable for periodic repetition. Moreover, in industrial environments it is not easy to place the calibration target in the field of view of the camera, where heavy machinery is installed. Also, the calibration plate includes printed calibrated markers in black and white. Using this type of calibration plate in industrial environments would require cleaning it every time it is used to prevent dirt and dust from covering the markers. Another important drawback of the plane fitting method is that it is designed for a single camera and laser projector. Industrial inspection often requires more than one camera to cover the whole surface of the inspected product. In order to inspect long products, such as blooms or beams, three, four, or even five cameras can be required depending on the shape of the product. Therefore, the required time for the complete calibration of the system would be greatly increased using the plane fitting method.

This paper presents a robust procedure to calibrate and measure long products, suitable for multi-camera configurations in industrial environments. It is assumed that the internal camera calibration is performed previously using a standard calibration target. The proposed procedure deals with the estimation of the extrinsic calibration, that is, the estimation of the position and orientation of the cameras with respect to the corresponding laser light plane of each camera. In order to perform the calibration, a calibration plate containing protruding cylinders is used. The information about the geometry of the calibration plate and the projected laser lines is used to calculate the required extrinsic calibration. The proposed procedure is designed to work with as many cameras as required. Moreover, the calibration of all the cameras is carried out at the same time, which reduces the time required to complete the calibration of the system, and facilitates periodic calibration with minimum downtime. The proposed calibration procedure is designed for robustness because in industrial environments dust, oil or water frequently provoke noise in the signals. Moreover, as accuracy is a funda-
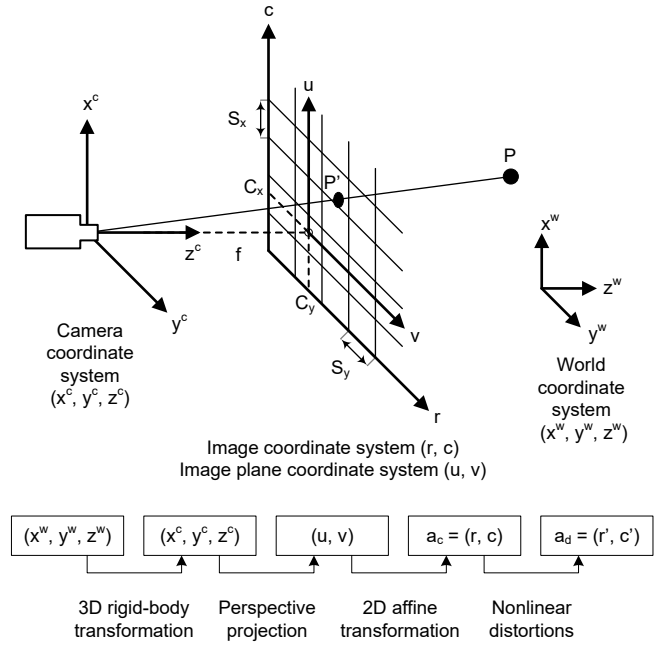

Figure 1: Camera projection model

mental aspect of 3D reconstruction sensors, extreme accuracy is also a requirement of the proposed method. Furthermore, the proposed procedure is designed to be applied with minimal computational requirements, which makes the application suitable for time-constrained systems.

The remainder of this paper is organized as follows. Section 2 introduces the fundamentals of camera projection geometry; Section 3 presents the proposed approach; Section 4 discusses the results obtained; and finally, Section 5 reports conclusions.

2. Camera projection geometry

The projection of a point in the scene onto the image can be described with a series of geometric transformations [14], as can be seen in Figure 1. In general, it can be broken down into four transformations when considering lens distortion.

The first transformation is a 3D rigid-body transformation from world coordinates in the scene to camera coordinates, where the camera is at the origin and the z-axis is perpendicular to the image plane. This transformation involves three rotations (α, β, γ) and three translations (t_x, t_y, t_z). These six parameters, usually referred to as extrinsic camera parameters, depend on the position of the camera in the scene and will change if the camera is affected by any type of movement or vibration. This transformation is represented using a 4×4 matrix usually denoted Rt. This transformation can be expressed as (1), where $P^w = (x^w, y^w, z^w)^T$ and $P^c = (x^c, y^c, z^c)^T$ represent the coordinates of points in world and camera coordinates, and r_{ij} are obtained from the rotation angles. The rotations and translations express the orientation and position of the world coordinate sys-

tem with respect to the camera.

$$\begin{pmatrix} x^c \\ y^c \\ z^c \\ 1 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x^w \\ y^w \\ z^w \\ 1 \end{pmatrix} \tag{1}$$

The second transformation is a perspective projection from camera coordinates to the image plane based on the pinhole camera model. This transformation can be expressed as (2), where $(u, v)^T$ is the corresponding point in the image plane coordinate system and f is the focal length.

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x^c \\ y^c \\ z^c \\ 1 \end{pmatrix} \tag{2}$$

The third transformation is a 2D affine transformation from image plane coordinates to image coordinates or pixel coordinates. This transformation can be expressed as (3), where $(c, r)^T$ are the column and row of the pixel, S_x and S_y are scale coefficients depending on the size of the pixels, and C_x and C_y are translation coefficients that compensate for displacements of the central pixel.

$$\begin{pmatrix} c \\ r \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{S_x} & 0 & C_x \\ 0 & \frac{1}{S_y} & C_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \tag{3}$$

The second and third transformations are usually combined in a general form as (4).

$$\begin{pmatrix} c \\ r \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{f}{S_x} & 0 & C_x & 0 \\ 0 & \frac{f}{S_y} & C_y & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x^c \\ y^c \\ z^c \\ 1 \end{pmatrix} \tag{4}$$

The final transformation is nonlinear to compensate for optical distortions. This transformation is sometimes neglected when high accuracy is not strictly required [15]. This transformation is usually expressed as (5), where $a_c = (c, r)^T$ represents the correct coordinates, $a_d = (c', r')^T$ the distorted coordinates, $\mathcal{F}$ is a nonlinear distortion function, and δ represents a set of coefficients that describe distortion.

$$a_c = a_d + \mathcal{F}(a_d, \delta) \tag{5}$$

Different methods have been proposed to model optical distortions [16, 17]. In general, the models consider three types of distortion: radial distortion, tangential distortion, and prism distortion. The number of coefficients required depend on the complexity of the model. In the polynomial model 5 coefficients are used: 3 for radial distortion and 2 for tangential distortion. More complex models that consider all distortion types require 14 coefficients [18].

The third and final transformations are sometimes considered in a different order, as can be seen in [10] and [18]. Distortions can also be considered from correct to distorted coordinates [17]. These different approaches change the coefficients but not the resulting transformation.

The parameters that control the transformation from camera coordinates to pixels are referred to as intrinsic camera parameters, as they only depend on the camera internals and not on the position or orientation of the camera.

The transformation from world coordinates to pixel coordinates is called forward projection. The inverse transformation, the backward projection, is an ill-defined problem because pixels correspond to projection lines in world coordinates, as many different points in world coordinates are projected onto the same pixel. 3D reconstruction techniques solve this problem by using stereo vision, where the intersection of the projection lines of the same pixel in two cameras are used to calculate the corresponding position in world coordinates. A different approach is to consider a single measurement plane in the scene. In this case, the projection line of the pixel is intersected with this plane to calculate the corresponding position in world coordinates.

2.1. Camera calibration

Camera calibration is the procedure used to estimate the parameters that control the projection of points in the scene onto pixels in the image according to the mathematical model that describes the transformations. There are two types of parameters: intrinsic and extrinsic. The intrinsic parameters model the internal projection of light onto the image sensor through the lens. The extrinsic parameters model the position and orientation of the camera with respect to a user-defined coordinate system. Metric information can only be extracted from images when the two types of parameters are accurately estimated. Because the two types of parameters model different aspects of the projection, two different camera calibration procedures are generally applied: intrinsic camera calibration [19] and extrinsic camera calibration [20].

The most common calibration methods are based on a two-phase technique. They first apply a liner approximation in order to obtain an initial guess of the projection parameters. Then, an iterative algorithm is used to optimize the estimation. This nonlinear procedure requires a set of matched points in world coordinates and pixel coordinates which are generally obtained from a calibration object with known geometric information. The most commonly used calibration object is a flat calibration plate printed with a calibration target that contains squares or circles. The calibration plate is observed by the camera from different orientations to obtain multiple correspondences [21]. The same calibration target can be used to calibrate the intrinsic and extrinsic parameters. In order to perform the extrinsic calibration using this calibration target, it must be exactly aligned with the user-defined coordinate system of the scene. In the case of the extrinsic calibration for laser line projection sensors, the extrinsic calibration is usually performed using the plane fitting method.

3. Proposed method

The proposed method is based on a calibration object. Laser lines are projected across the object and the deformed laser profiles are extracted from the images. The calibration object has
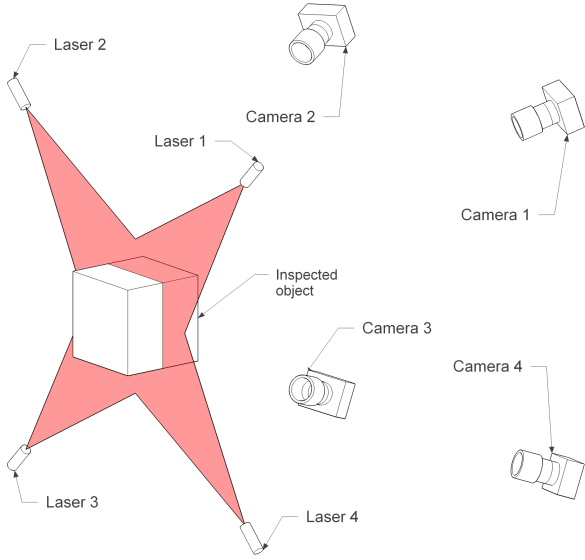
Figure 2: Architecture of the system with four cameras and four corresponding laser projectors

been accurately designed and measured. Thus, the geometric information about the shape of the object is known. Consequently, it can be used to calibrate the pose of the cameras.

Figure 2 shows an illustration of the architecture of the system, including the laser projectors and the cameras. In this example, four sensors are considered, each consisting of a camera and a laser projector. The four laser lines are projected on the same plane. Adjacent lasers and cameras are configured to use a different light wavelength. Thus, they do not interfere with each other. Four cameras is the most common approach to cover the whole shape of the object, although more cameras could be necessary in some cases depending on the shape of the inspected product. The proposed calibration procedure is designed to work with as many sensors as required.

The geometry of the calibration target can be seen in Figure 3. It consists of 13 protruding cylinders on a plate of approximately 200 mm×200 mm made of aluminum. The aluminum calibration target is lightweight and durable. Moreover, it has very good corrosion resistance in most environments. In order to avoid reflections, a dark anti-reflective coating is applied to the surface. The projection of the laser lines on the calibration target is observed simultaneously by all the cameras. Therefore, it is possible to calibrate all the cameras at the same time using a common reference system. After fabrication, the calibration plate is measured using a coordinate measuring machine, which is a device to accurately measure the physical geometry of an object. In this work, the Renishaw Renscan is used, with a reported accuracy of 1 μm.

The grid lines used in Figure 3b are represented with a spacing of 10 mm. The same spacing is used throughout the text, even when the scale of the figure is different.

The laser plane projected across the calibration target produces a profile consisting of 13 circles. However, these circles are projected as ellipses in the images. Figure 4 shows the cal-

ibration target observed by each camera. As can be seen, there are occlusions. Thus, only a partial view of the ellipses is visible. The visible part of the ellipses depend on the position of the cylinders and the camera.

The objective of the calibration is to find the optimal 3D rigid-body transformation that can be used to transform the observation of the calibration target in Figure 4 with the geometry in Figure 3b. The proposed method is applied in two steps. First an initial estimation is calculated using a coarse calibration. Then, an iterative algorithm is used to optimize the calibration. Figure 5 shows a summary of the proposed procedure.

The proposed procedure is designed to work in industrial environments where adverse conditions are the norm. The consequences of these conditions include variable luminance, reflections, or uneven laser light. Thus, the calibration procedure is designed to be robust.

3.1. Coarse calibration

The coarse calibration is applied to calculate a good initial calibration in order to guarantee convergence in the fine calibration procedure applied next.

Industrial applications based on laser triangulation are affected by reflections in the images. These reflections are usually interpreted as part of the laser line, producing noise and outliers. Moreover, when laser projectors work for extended periods of time, the intensity of the laser light decreases. This reduced performance occurs unevenly, affecting some parts of the projected laser line more than others. Consequently, the laser stripe extraction procedure fails to obtain the coordinates of the projected laser line in parts of the image [22]. The proposed procedure is designed to deal with these issues in order to produce a robust calibration, even when some of these adverse conditions affect the quality of the image.

The first step required for calibration is the extraction of the laser line, i.e., the determination of the coordinates of the laser line in the images. In this work a precise subpixel method for the detection of curvilinear structures is used [23]. This method provides the coordinates of the laser line accurately [24]. The method is also applied to other applications with excellent results, such as road detection in aerial images or vein detection in medical imaging. Figure 6 shows the extracted laser profile from the images in Figure 4.

The extracted coordinates of the laser line are grouped using a clustering method based on the distance, producing a set of curvilinear segments. In this way, the projection of the laser line on each cylinder is distinguished. A filtering procedure is applied to the calculated segments based on their lengths. Short segments are removed, as they are considered reflections or noise. The remaining segments are fitted to an ellipse [25]. Other more accurate methods have been proposed for ellipse fitting [26], however extreme accuracy is not relevant in this step, as the fine calibration is used to improve these coarse results. These ellipses represent the projection of the cross section of the cylinders, i.e., circles, into the image plane. The results of this process for the images acquired by each camera can be seen in Figure 7. A filtering procedure is also applied to the
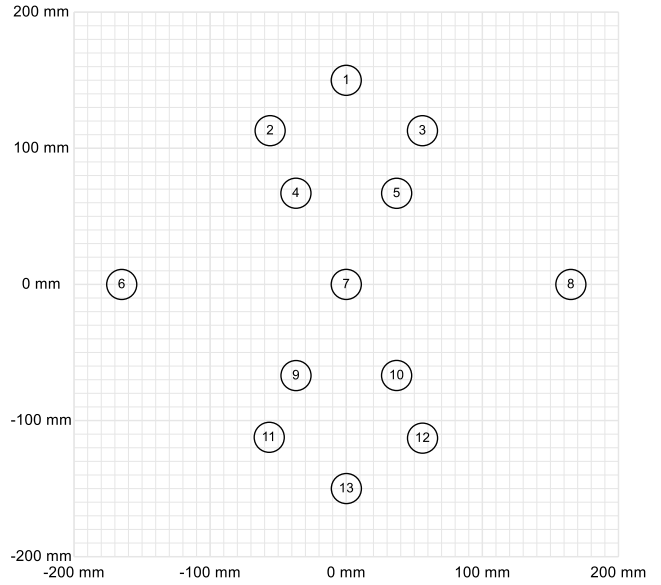
Figure 3: Calibration target used for calibration. (a) Image of the calibration target. A coating in areas where the laser is projected is shown in red. (b) 2D Plan of the cylinders on the calibration plate including position and size.

Figure 4: Observation of the calibration target. (a) Camera 1. (b) Camera 2. (c) Camera 3. (d) Camera 4.

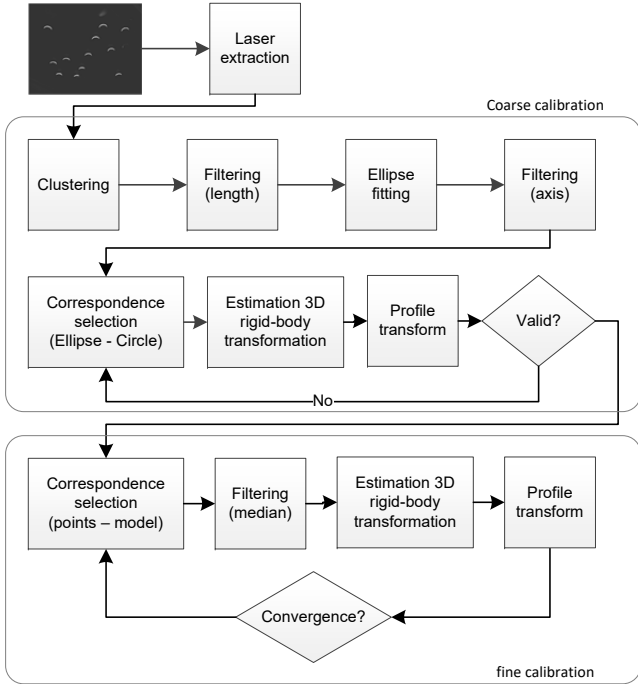Figure 5: Architecture of the system with four cameras and four corresponding laser projectors

calculated ellipses based on the lengths of the minor and major axes. Ellipses in which the length of the minor or major axis is outside a specific allowed range are removed because they are considered noise.

The centers of the calculated ellipses, marked in Figure 7, represent a good approximation for the centers of the cylinders of the calibration target. However, in this work these positions are assumed to be corrupted by noise, either because there are reflections in the image or because some cylinders are missing. Moreover, depending on the orientation and position of the camera with respect to the calibration target, the number of occluded cylinders changes. Thus, no restrictions are imposed on the calculated centers for the proposed procedure to work. Consequently, no prior assumptions can be made about the correspondence between the calculated centers of the ellipses and the centers of the circles in the calibration target.

The extrinsic calibration problem can be solved using four correspondences between images and world coordinates. Therefore, it is necessary to find four valid correspondences between the center of the calculated ellipses and the centers of the circles in the calibration target. Without prior information, and assuming data can be corrupted by noise, the proposed method looks for inliers to estimate the extrinsic parameters using a iterative approach inspired by RANSAC [27].

A possible solution to the problem can be obtained by creating a random correspondence between four ellipses and four circles in the calibration target. Figure 8 shows two examples for camera 1: one invalid and one valid. Using the extrinsic calibration obtained from these two correspondence sets produces the results shown in Figure 9. This figure shows the transformation of the extracted profile in Figure 7 for camera 1 from image

coordinates to world coordinates. As can be seen, the estimated correspondences are valid only when the obtained extrinsic calibration can transform the profiles according to the circles in the calibration target. An invalid correspondence produces a completely incorrect result.

Whether a solution is valid or not can be assessed by calculating the distance from the centers of circles in the calibration target to the centers of the ellipses after they have been transformed to world coordinates. When more than half of the centers of the circles are within a threshold distance to the transformed centers of the ellipses, the extrinsic calibration is considered to be valid, i.e., an inlier. This approach ensures robustness, as outliers are not considered for the validation test. For example, the number of circles that are within a threshold distance to the transformed center of the ellipses (inliers) in Figure 9a is 0, which indicates that the solution in Figure 8a is not valid. On the other hand, the number of inliers in Figure 9b is 12, which indicates that the solution in Figure 8b is valid. An ambiguity problem arises to distinguish the left and right cameras. The solution using this symmetrical target is to include a constraint depending on the camera that guarantees that the resulting transformation corresponds to the correct side. A similar problem occurs with the images obtained from the top and bottom cameras, which are symmetrical around the y axis. Thus, the estimated calibration for one camera can be confused with the other. The solution to this ambiguity is to indicate the expected rotation around the z axis for each camera. If a valid solution is found with an unexpected rotation around the z axis, it is discarded, as it is considered invalid for that camera.

The number of possible combinations of four ellipses and their corresponding four circles depends on the number of detected ellipses, as the number of circles does not change. Considering N ellipses are detected, the number of all possible combinations of ellipses taken four at a time without repetition can be mathematically expressed as $C(N,4)$, which is also equal to the binomial coefficient. For each of these combinations, all possible correspondences to the circles must be calculated. In this case, order does not matter because the correspondences of two ellipses with two circles in different orders are different solutions. Therefore, considering M circles in the calibration target, the partial permutations of circles taken four at a time are calculated, i.e., $P(M,4)$. Consequently, the total number of possible solutions is $C(N,4) \times P(M,4)$.

Among all the possible combinations, there are multiple valid solutions. The number can be easily estimated when no outliers are present, as it is the same as the number of combinations of ellipses. Therefore, the probability of finding a valid solution after testing only one case can be expressed as (6). This probability increases as more cases are tested. If the experiment is repeated r times, the probability of finding a valid solution can be calculated using (7). Solving (7) for r results in (8), which indicates the number of tests required to obtain a valid solution with probability P_r.

$$P = \frac{C(N,4)}{C(N,4) \times P(M,4)} \qquad (6)$$

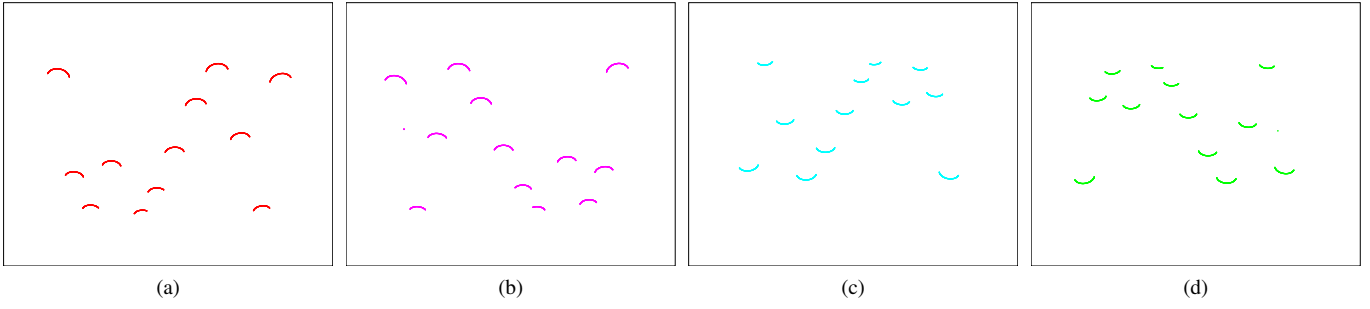$$P_r = 1 - (1 - P)^r \qquad (7)$$

Figure 6: Extracted laser profile. (a) Camera 1. (b) Camera 2. (c) Camera 3. (d) Camera 4.
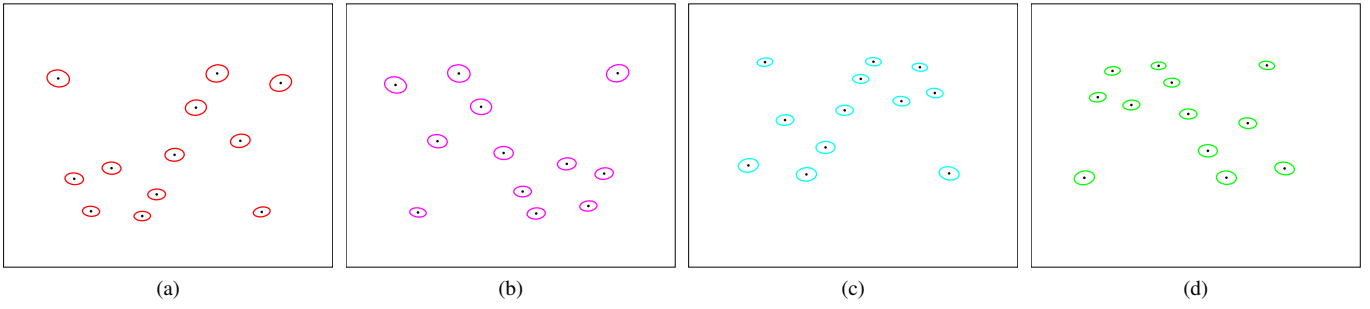

Figure 7: Ellipse fitting. (a) Camera 1. (b) Camera 2. (c) Camera 3. (d) Camera 4.
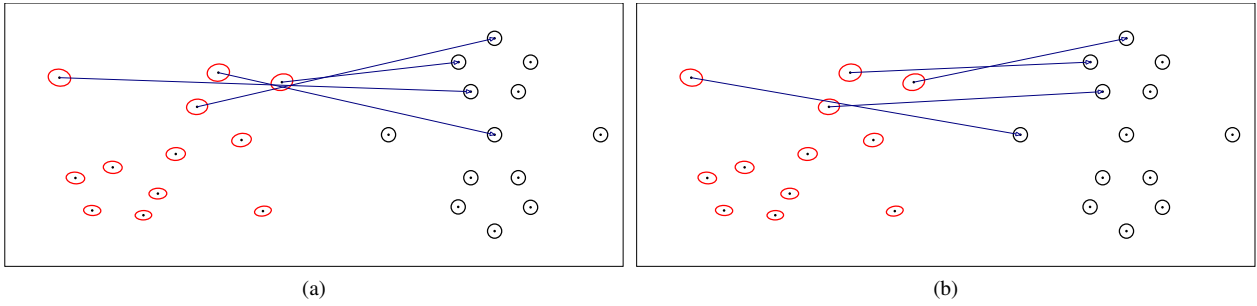

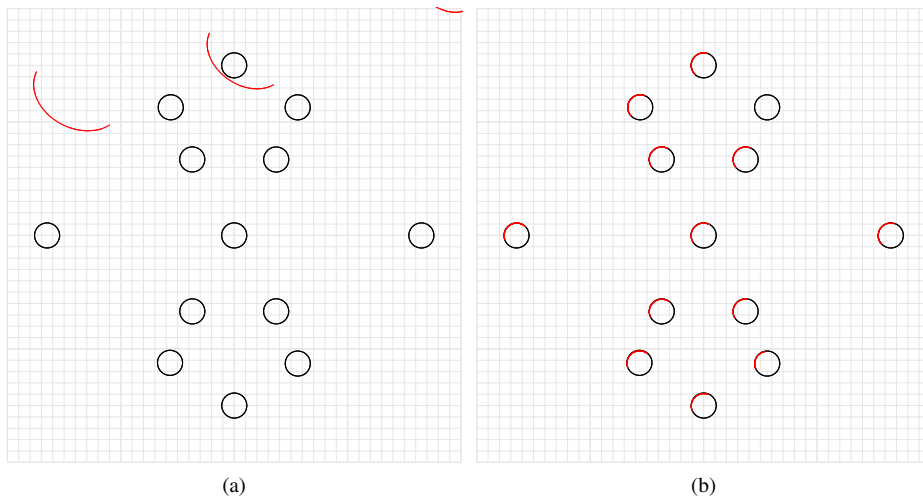Figure 8: Correspondence between ellipses and circles. (a) Invalid correspondence. (b) Valid correspondence.

Figure 9: Result of the calibration after random correspondence estimation. (a) Invalid correspondence. (b) Valid correspondence.

$$r = \frac{log(1 - P_r)}{log(1 - P)} \qquad (8)$$

In the example considered in Figure 8, 12 ellipses are detected. Therefore, taking into account that the calibration target contains 13 cylinders, the total number of possible combinations is $C(12, 4) \times P(13, 4) = 8\,494\,200$. Among these combinations, only 495 are valid. The probability of finding a valid solution after testing 10 000 combinations is 0.44. In order to find a valid solution with 99% confidence the number of repetitions is 79 022.

In the worst case scenario, the number of tests required to find a valid solution is very high. Thus, in this work a simple yet effective heuristic is proposed to increase the probability of finding a valid solution in fewer tests. The proposed procedure consists in sorting all the possible combinations in order to ensure solutions with a high probability of valid solution being tested first. The proposed heuristic is based on the design of the calibration target. As can be seen in Figure 3b, the cylinder on the left (number 6) and the cylinder on the right (number 8) are located much further away than the rest. When the calibration target is observed by the cameras, these two cylinders are clearly distinguishable in the images. In order to detect these ellipses in the images, the distances from each center of each ellipse to all the others are calculated. The two most distant centers from the closest neighbors are the left and right cylinders in the calibration target, or vice versa. The results can be seen in Figure 10. This heuristic may be incorrect when noise is present in the images, but in most cases this is a very important clue that can be used to increase the probability of finding a valid solution in fewer tests.

The application of this heuristic to the example considered in Figure 8 does not reduce the search space, but the probability of finding a valid solution increases sharply. Including the heuristic, the probability of finding a valid solution after testing 100 combinations is 0.36. In order to find a valid solution with 99% confidence the number of repetitions is 1 012, around 79 times faster. Therefore, the proposed procedure can be applied in real-time.

The results of the proposed procedure for the coarse calibration can be seen in Figure 11. This figure shows the transformation of the laser profile observed by each camera, shown in Figure 6, to world coordinates. As can be seen, the ellipses observed in the images are correctly transformed into segments of circles in the calibration target. Therefore, the objective of this procedure is achieved successfully, producing a good initial guess for the calibration. This step is required for the fine calibration applied next.

3.2. Fine calibration

The coarse calibration produces a robust estimation about the position and orientation of the camera with respect to the calibration target. However, the result is not accurate. The main reason is that ellipse fitting is very sensitive to noise, especially when the whole contour of the ellipse is not visible. The extracted profile from the images only contains part of the ellipse contour due to occlusion, which decreases the accuracy of the

fitting procedure. Moreover, the center of the projected ellipse does not correspond to the center of the circle in the calibration target, it is just an approximation.

The proposed procedure for the fine calibration is based on the Iterative Closest Point (ICP) algorithm [28], which is used for registration between cloud points. This algorithm presents an iterative optimization procedure that finds the best transformation between sets of points. The adaptation of this algorithm to the proposed fine calibration procedure is applied in four consecutive steps repeated until convergence: transformation of profile to world coordinates using the current extrinsic calibration, estimation of correspondences, filtering of correspondences, and calculation of improved extrinsic calibration.

The coarse extrinsic calibration is used to transform the extracted profile of the laser line to world coordinates. Figure 12 shows a close view of the results of the coarse calibration for the left, right, top and bottom circles of the calibration target (numbers 6, 8, 1 and 13). As can be seen, some segments of the transformed profile are misaligned with respect to the calibration target, especially the bottom circle represented in Figure 12d. On average, the distance from the transformed profile to the calibration circles is approximately 0.3 mm.

Figure 13a shows the top part of the bottom circle represented in Figure 12d after the transformation to world coordinates using the coarse calibration. Next, correspondences between the transformed points and the circles in the calibration target are estimated. To speed up this process an R-tree is used to determine which circle corresponds to each point in the profile [29]. Then, the closest point in the corresponding circle to the each point is calculated. This approach is very efficient, as only the circles close to the points are used to calculate the closest point operation, greatly improving the execution speed. The result of the correspondence estimation can be seen in Figure 13a, with arrows representing the correspondences between points in the profile and points in the circle. The average distance between the transformed profile and the circle at this stage is 0.659 mm, as indicated in the figure.

The next step is the filtering of correspondences. Some points in the profile can be produced by noise data. Considering these points for the calibration would lead to inaccurate results. The proposed solution is to remove these points based on the median distance between potential correspondences. Assuming the distribution of the distances is Gaussian, the median absolute deviation is a robust procedure to detect and remove these outliers [30].

The calculation of the extrinsic calibration based on the estimated correspondences is performed next. This calculation is performed based on the filtered correspondences. The result of the correspondence estimation is a set of 3D correspondences between points in the transformed profile and points in the calibration target. Each point in the transformed profile corresponds to a point in the profile extracted from the image. These points, in image coordinates, and the corresponding points in the calibration target, in world coordinates, are used to calculate the extrinsic calibration. The result of this new extrinsic calibration can be used to transform the points in the profile to world coordinates again. The result can be seen in Figure 13b.
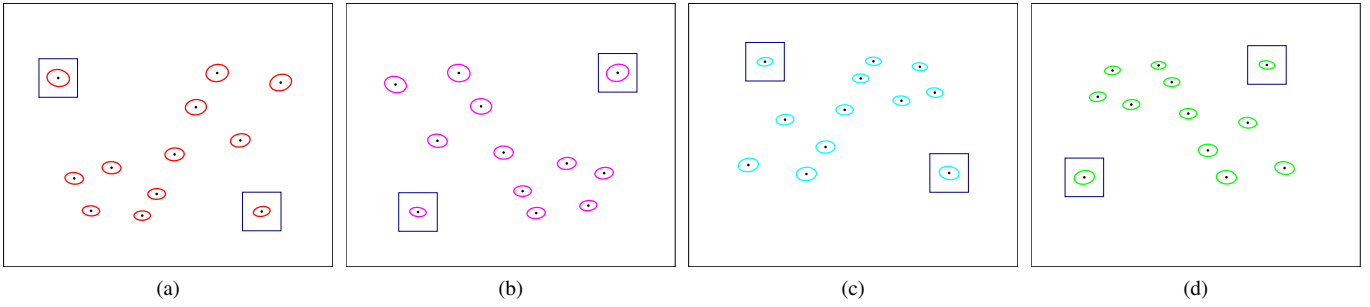
Figure 10: Left and right cylinders detected in the images. (a) Camera 1. (b) Camera 2. (c) Camera 3. (d) Camera 4.
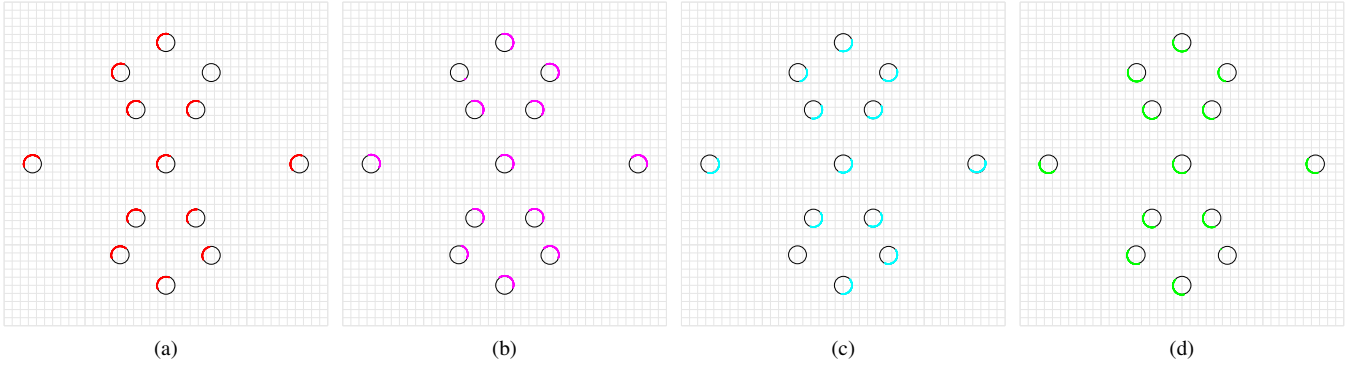

Figure 11: Transformed profile to world coordinates using the coarse calibration. (a) Camera 1. (b) Camera 2. (c) Camera 3. (d) Camera 4.
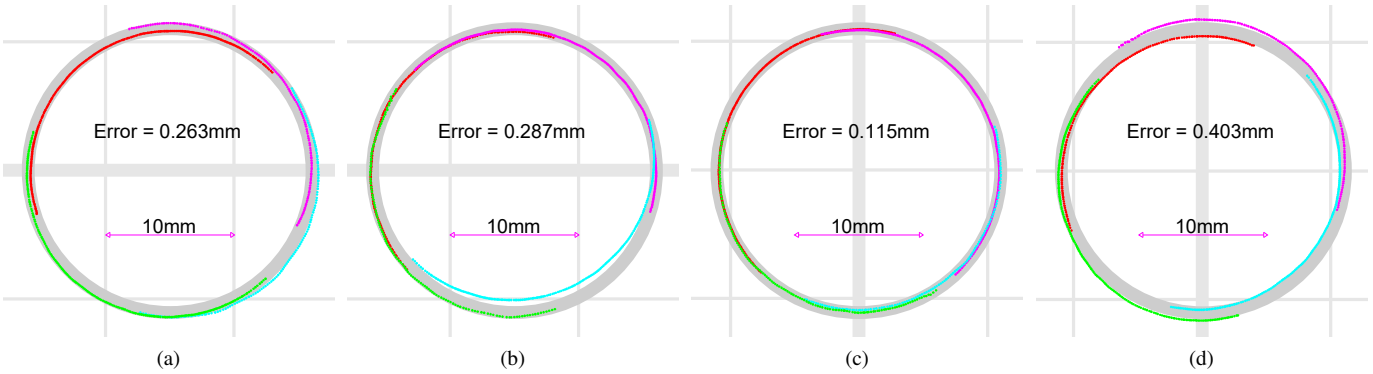

Figure 12: Close view of the transformed profile to world coordinates using the coarse calibration. (a) Left circle (6). (b) Right circle (8). (c) Top circle (2). (d) Bottom circle (13).
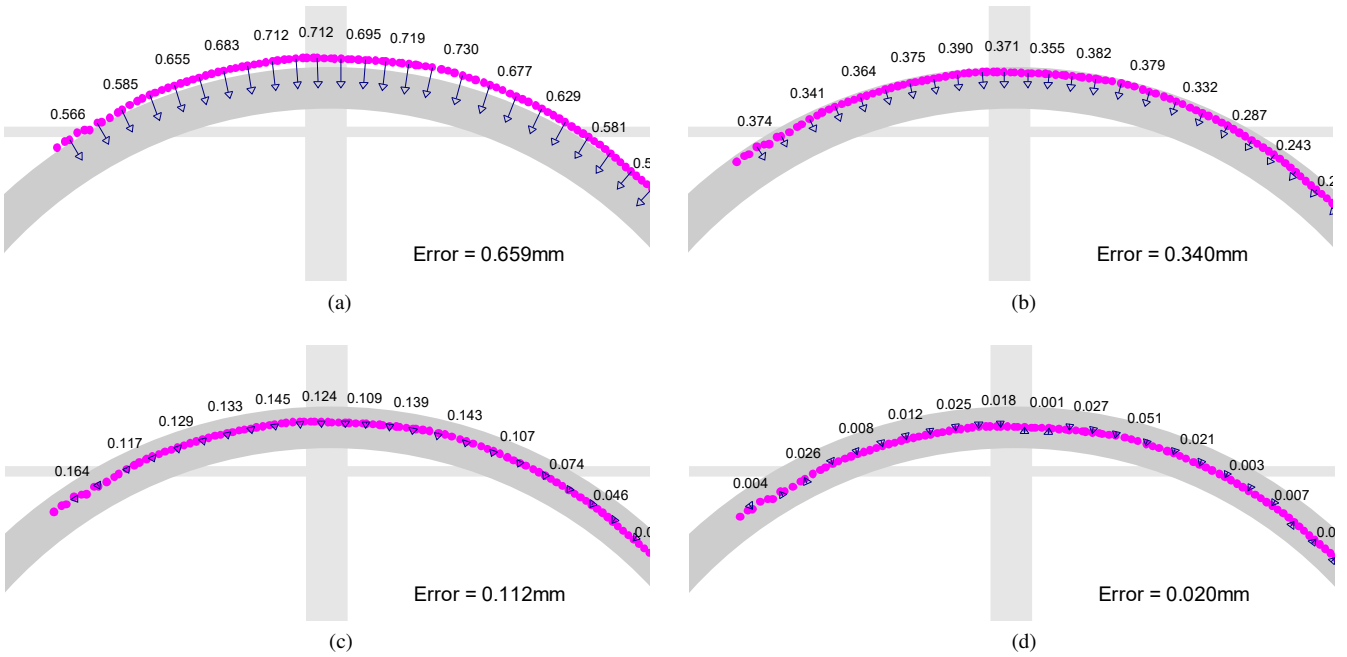
Figure 13: Steps of the fine calibration. (a) Initial transformation. (b) After iteration 1. (c) After iteration 3. (d) After iteration 20.

No knowledge about the exact correspondence between points is available. Thus, the result is not completely accurate. However, as can be seen, the points pull the profile closer to the calibration circle, reducing the distance to 0.340 mm.

The repetition of these four steps improves the extrinsic calibration in each iteration. Figures 13c and 13d show the results after iterations number 3 and 20, where the reduction in the distance between the transformed profile and the calibration circle are clearly observed, producing a final average distance of 0.0020 mm.

The convergence criterion used to determine when the iterative procedure can be stopped is based on the movement of the transformed profile from one iteration to the next. When the transformed profile does not move significantly with respect to the previous iteration, the procedure is stopped, as the calibration is not improving. The movement can be measured very effectively by subtracting the transformation matrices between consecutive iterations.

The results of the fine calibration can be observed in Figure 14. This figure can be compared with Figure 12. As can be seen, the alignment between the transformed profile and the calibration circles has greatly improved. The average distance between the transformed profiles and the circles is reduced approximately by a factor of 10, resulting in an average distance error of 0.03 mm.

In industrial environments, sensors and machines are greatly affected by vibrations. These movements change the position of the cameras with respect to the inspected product, and make frequent re-calibration necessary. The proposed fine calibration procedure requires an initial approximation. The first time the system is calibrated, the coarse calibration provides the required approximation. However, successive calibrations can use the last successful calibration as their initial approximation, and hence the coarse calibration is not required. The proposed procedure will converge even the when the initial approximation is inaccurate. Only under extreme circumstances, such a change in the position of the camera, is it necessary to repeat the coarse calibration.

4. Results and discussion

4.1. Coarse calibration

The performance of the calibration can be measured by calculating the average distance from the laser profile in world coordinates to the circles in the calibration target. The results can be seen in Figure 15. The average distance for all cameras is 0.24 mm. Although this is not a good calibration, it can be considered a good approximation.

Figure 16 shows the number of combinations that had to be tested before a valid solution was found for each camera. The tests were executed on an Intel Core i7 4770 running at 3.4 GHz, which can test 4000 combinations per second approximately. This execution speed indicates that testing all possible combinations can take around 40 minutes. However, as the probability analysis indicated, it is only necessary to test a fraction of the total number of combinations to find a valid solution. The time required to find a valid solution for each camera is represented in Figure 17, which corroborates the previous statement. In all cases the time required for the coarse calibration to find a valid solution was below 4.5 seconds. These results were obtained without considering any heuristic. However, when the proposed heuristic based on the left and right cylinders was applied, the number of tests and time required to find a valid solution decreased sharply. As can be seen in
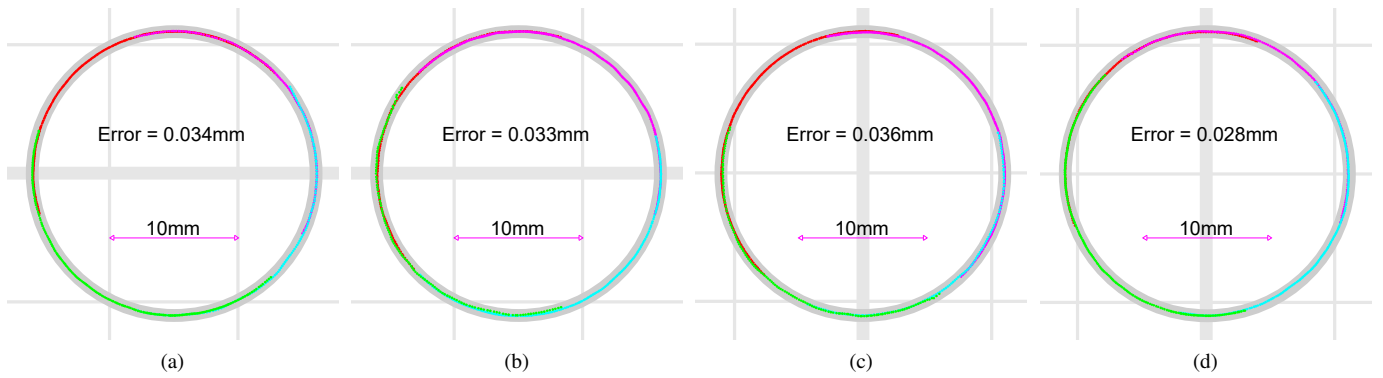
Figure 14: Close view of the transformed profile to world coordinates using the fine calibration. (a) Left circle (6). (b) Right circle (8). (c) Top circle (2). (d) Bottom circle (13).
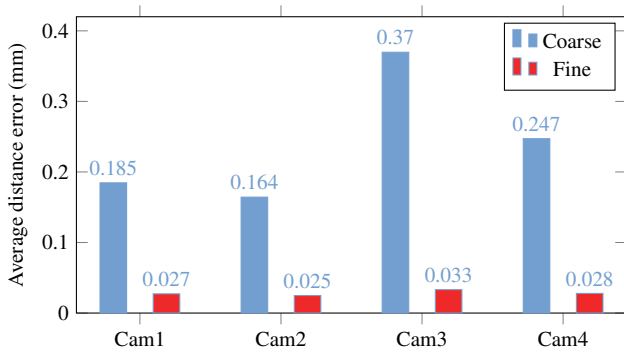

Figure 15: Average distance from the profile in world coordinates to the calibration target for each camera after calibration.
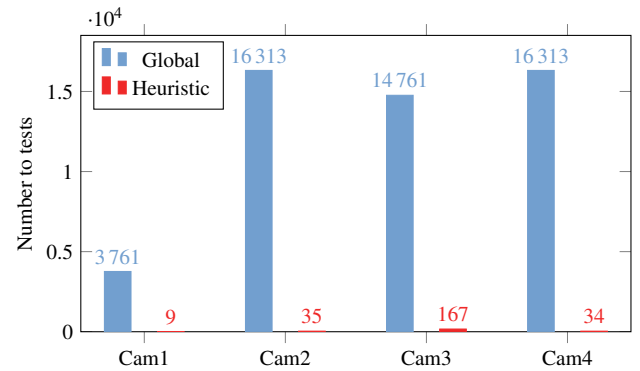

Figure 16: Number of tests required to find a valid solution in the coarse calibration.

the figures, in the worst case only 167 tests were required, producing the result in 38 milliseconds. The proposed heuristic is around 100 times faster. Using the proposed heuristic, in order to find a valid solution with 99% confidence, the time required is 250 ms considering the obtained execution speed.

In industrial environments adverse conditions can provoke noise in the images. Therefore, the proposed procedure for the coarse calibration is designed to deal with noise in the observed profile of the calibration target. In order to assess the robustness of the proposed coarse calibration, an observed profile of the calibration target is corrupted with synthetic noise, as can be seen in Figure 18a. A clustering procedure is applied to the points in the profile. Short segments are removed and the remaining segments are fitted to an ellipse. The result can be seen in Figure 18b. The noise provokes outliers in the detected ellipses. Some of these ellipses are removed because the length of the axes is outside the considered valid range. Nevertheless, some incorrect ellipses are still considered for the next step, as can be seen in Figure 18c. The result of the coarse calibration can be seen in Figure 18d, where the coordinates of the profile are transformed to world coordinates. This result demonstrates that the proposed procedure is not affected by noise, producing a correct result even under heavy noise conditions. Moreover, finding a valid solution only took 14 863 tests, performed in 3.7 seconds.
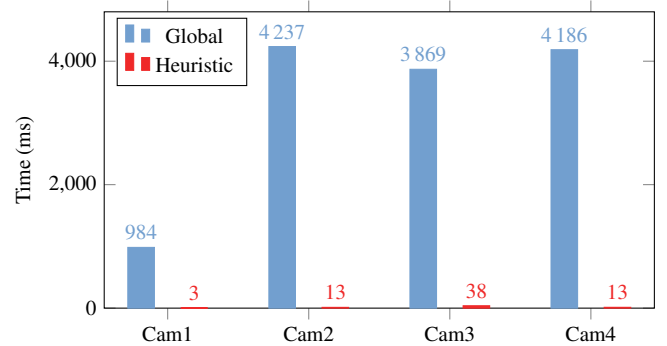

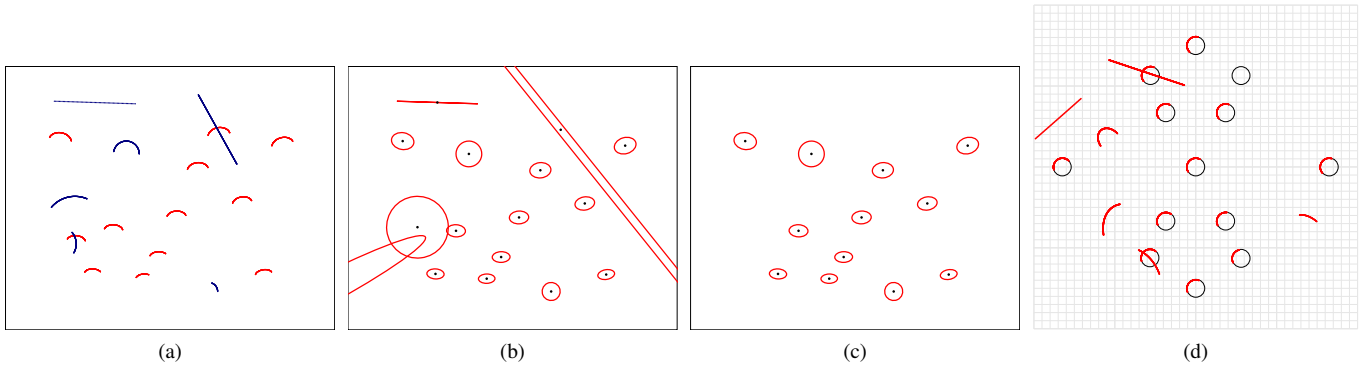Figure 17: Time required to find a valid solution in the coarse calibration.

Figure 18: Coarse calibration with noise. (a) Noise added to the profile. (b) Ellipse fitting. (c) Ellipse Filtering. (d) Transformation to world coordinates.

4.2. Fine calibration

Figure 15 shows the average distance from the laser profile in world coordinates to the circles in the calibration target after the fine calibration. The calibration error is reduced by a factor of 10 when compared with the results of the coarse calibration. On average the calibration error for all cameras is only 0.027 mm, much better than the 0.249 mm obtained after the coarse calibration.

Figure 19 shows the time required for the fine calibration to converge. In all cases, convergence is obtained in less than 80 ms. Moreover, these measurements are obtained using all the points in the profiles. Reducing the number of points per profile can reduce the required time to converge without affecting the final obtained accuracy. Figure 20 shows the results of a calibration experiment reducing the number of points in the profiles obtained by each camera. The value on the *x* axis represents the degree of subsampling applied. For example, a value of 10 indicates that only one point in every 10 is used for the fine calibration, that is, only 10% of the points. The results indicate that using only 10% of the points in the profile produces the same results in terms of accuracy. However, using this number of points reduces the required time for calibration to only 12 ms. Therefore, using this level of subsampling, the complete calibration procedure, applying the coarse calibration first and the fine calibration after that, only requires 30 ms. This low computational requirement makes the proposed calibration procedure suitable for real-time applications or resource-constrained embedded systems.

Figure 21 shows the results of the fine calibration when the profile is affected by noise. This figure shows a close view of the result of the fine calibration when using the profile corrupted with noise. The noise added is the same used for Figure 18. As can be seen, the proposed method to filter correspondences correctly detects noise (represented with a different color in the figure) based on the distance. These points are not used in the calibration, and hence the resulting calibration is still very accurate. This indicates that the proposed fine calibration is not affected by adverse conditions that can provoke noise in the signal.
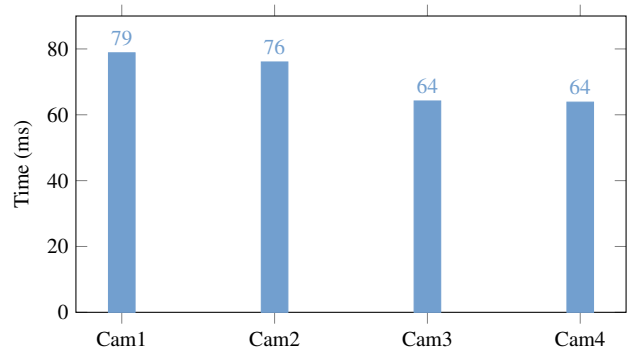

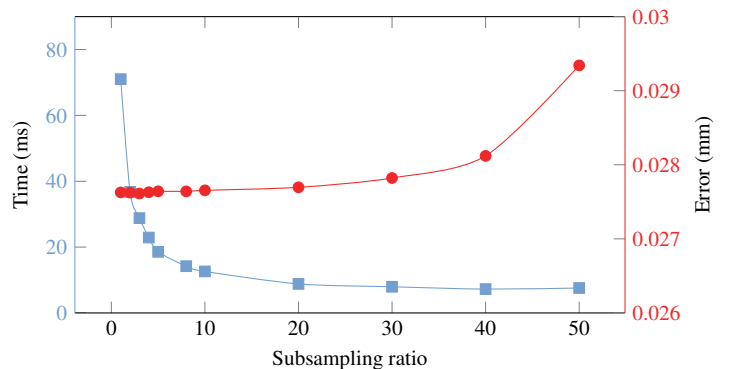Figure 19: Time required for the fine calibration to converge.

Figure 20: Influence of the subsampling strategy on the fine calibration accuracy and time required to converge.
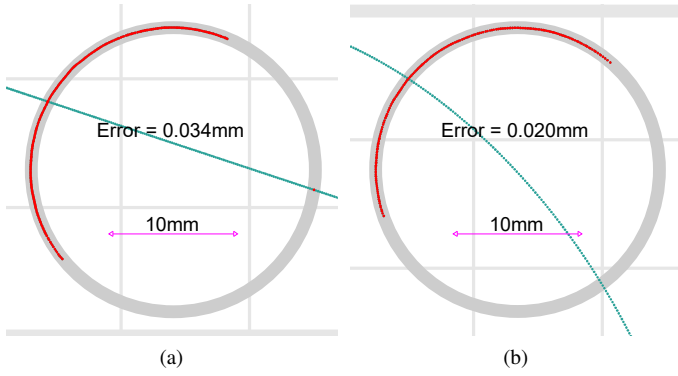
12

Figure 21: Close view of the transformed profile to world coordinates using the fine calibration with added noise. (a) Circle number 2. (b) Circle number 11.
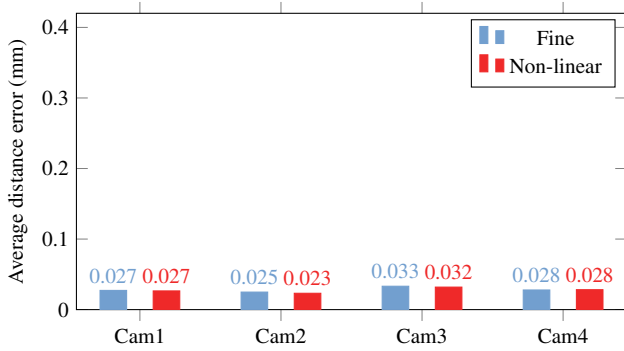

Figure 22: Average distance from the profile in world coordinates to the calibration target for each camera after the calibration using the non-linear solver.

4.3. Comparison with non-linear solver

The proposed fine calibration iteratively improves the estimation of the position and orientation of the camera based on the correspondences between the transformed points and the closest points in the calibration target. A different approach to solve the problem is to consider the estimation of the optimal position and orientation as an optimization problem. The objective is to minimize the distance from the transformed profile in world coordinates to the calibration target. The coefficients of the model are the extrinsic camera parameters (three translations and three rotations). This problem can be described as a Non-linear Least Squares problem. Therefore, it can be solved using one of the many tools available. In this work the Ceres solver is used [31]. This open source library is used for modeling and solving large, complicated optimization problems. It has been used in production in many different applications, where it has been extensively tested.

The considered solver requires an initial estimation. Thus, the initial coarse calibration is used. Consequently, the results can be directly compared with the fine calibration, as they start the optimization procedure from the same data. The results can be seen in Figure 22. The difference with the fine calibration is negligible. On average the difference is approximately $1\,\mu m$.

The improvement of the calibration accuracy using the non-linear solver is negligible. Moreover, the non-linear solver takes much longer to find the optimal solution than the proposed fine
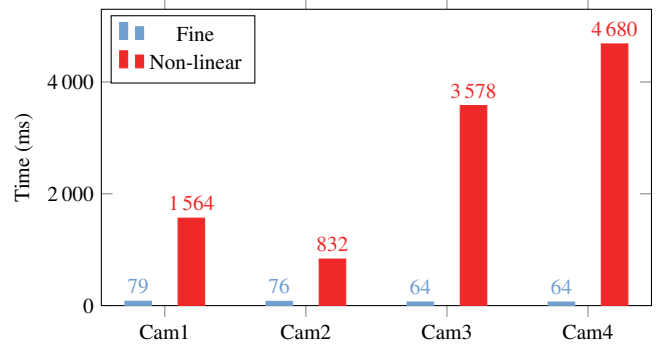

Figure 23: Time required for the calibration using the non-linear solver to converge.

Table 1: Errors in the accuracy test using the calibration plate. All values are given in millimeters

| | Position | | |
	-10 mm	+10 mm	Center
Camera 1	0.031	0.028	0.027
Camera 2	0.030	0.035	0.024
Camera 3	0.042	0.036	0.032
Camera 4	0.028	0.036	0.028
Average	0.033	0.034	0.027

calibration procedure. Figure 23 shows a comparison of the time required by the two considered methods. The non-linear solver requires, in the worst case, 70 longer to achieve the same solution. Therefore, the proposed fine calibration procedure performs better when considering the computational cost required to converge.

Different configurations of the Ceres were tested with similar results in the time required to converge and the accuracy of the resulting calibration, including the dense Cholesky solver or the dense QR solver.

4.4. Accuracy tests with the calibration target

In order to test the accuracy of the proposed calibration procedure, the first experiment uses the same calibration plate used for calibration. In the experiment, the calibration plate is moved to the left (-10 mm from the center), to the right (+10 mm from the center), and back to the center. In each of these three positions, images are acquired and the laser profiles are extracted and converted to world coordinates. Then, a registration procedure is applied in order to align the resulting combined profile in world coordinates from all cameras with the model of the calibration plate[29]. Figure 24 shows the calibration plate illuminated by the laser during the experiment.

The results of the experiment can be seen in Table 1. The table shows the average distance from the profiles acquired with each camera to the model of the calibration plate after alignment. On average, the distance error is 0.031 mm with a standard deviation of 0.005 mm for the four cameras. These results are very similar to those obtained in the calibration residuals.
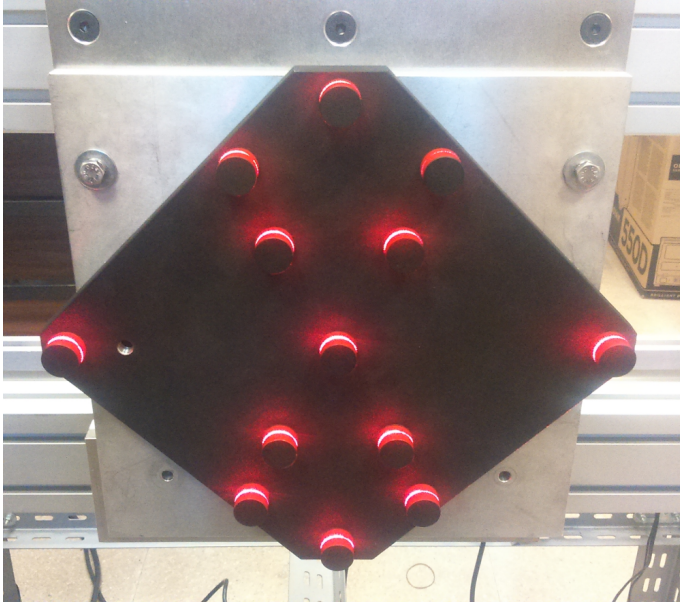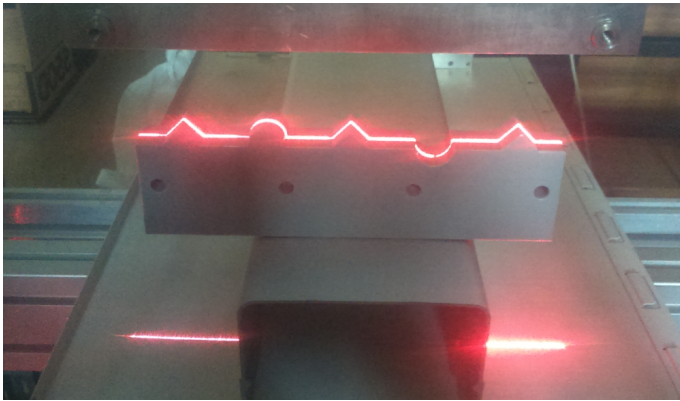
13

Figure 24: Calibration plate in the accuracy test

Figure 25: Test piece in the accuracy test

4.5. Accuracy tests with plastic piece

Extended accuracy tests were performed using a test piece. This test piece consists of a combination of straight lines and curves, designed to test the accuracy with combined geometry. Figure 25 shows the test piece illuminated by the laser during the experiment. The measurement reference of the test piece is performed in a similar manner as the calibration plate. A coordinate measuring machine with a reported accuracy of $1\,\mu m$ (Renishaw Renscan) is used to accurately measure the geometry of the test piece and create the model. Figure 25 shows the test piece illuminated by the laser during the experiment.

In this experiment, only the two top cameras are used (Camera 1 and 2). The images with the laser projected on the test piece are processed, extracting and converting the laser profiles to world coordinates. Then, the profiles from each of the two cameras are combined into one, as they are in the same reference system. Finally, an alignment procedure is applied to the combined profile. The result can be seen in Figure 26. Some small parts of the laser line are missing due to occlusions.
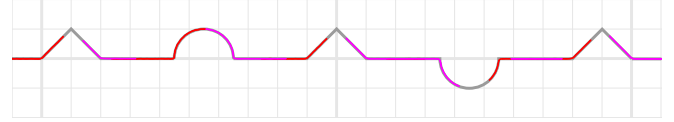

Figure 26: Test piece model and reconstruction

Table 2: Errors in the accuracy test using the test piece. All values are given in millimeters

	Position		
	-45 mm	+45 mm	Center
Camera 1	0.026	0.030	0.033
Camera 2	0.040	0.045	0.043
Average	0.033	0.037	0.038

In order to calculate the accuracy of the reconstruction, the distance from each point in the profile to the model of the test piece is calculated. The experiment is repeated three times, moving the test piece vertically: top, center and bottom. In the top position the test piece is moved 45 mm from the center upwards (+45 mm); in the bottom position the test piece is moved 45 mm from the center downwards (-45 mm). The center position corresponds to the position of the central circle of the calibration plate (circle 7).

The results of the accuracy experiment with the test piece can be seen in Table 2. On average, the distance error is 0.036 mm with a standard deviation of 0.0026 mm for the two cameras. These results are slightly worse than those obtained in the calibration residuals. One of the reasons is that the test piece is made of plastic, where the Lambertian reflectance is worse than in the calibration plate.

4.6. Accuracy tests with rails

The final test is performed with samples of rails. Three samples of rails are considered: A2, A3 and A4. In the three cases, rails are type 54E1 (UIC 54), a rail model defined in the UNE EN13674-1 standard. Figure 27 shows an image of the rails used for this experiment.

The shape of the head of the rails is measured using two different sensors: Calipri [32] and Miniprof [33]. Calipri is a non-contact measurement device based on laser projection that can be used to measure any profile or body contour. Miniprof is a

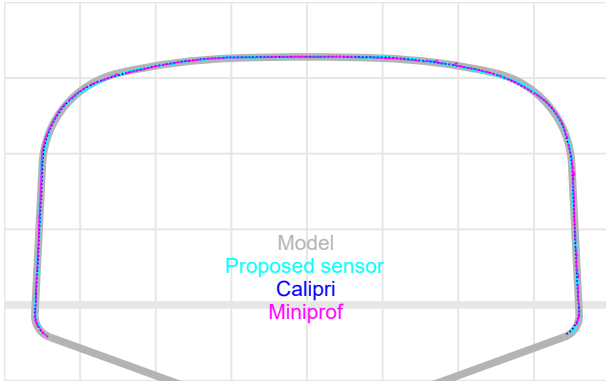Figure 27: Rails used for testing

14

Figure 28: Rail model and reconstruction

Table 3: Distances in the accuracy test using the rail samples. All values are given in millimeters

| | Rail sample | | |
	A2	A3	A4
Proposed sensor	0.085	0.083	0.074
Calipri	0.073	0.072	0.079
Miniprof	0.103	0.099	0.074

contact sensor for supervising railway material and performing cross-sectional rail profile measurements. The two sensors are used to estimate the shape of the head of the three considered rail samples. These measurements are compared with the profile obtained by the proposed system. Only the two top cameras are used (Camera 1 and 2).

Figure 28 shows the model of one of the rail samples and the measurements of the shape performed by the three sensors: the proposed system, Calipri and Miniprof. Visually, there are no apparent differences until the figure is enlarged.

Using a similar procedure to previous experiments, the distances from the profiles to the rail model are calculated. In this case, the rail model does not represent a true reference because rails are rarely fabricated exactly as the model indicates, there is always a geometric tolerance. Thus, the most interesting results in this experiment are the comparison between the measurements performed by the three sensors. The results can be seen in Table 3. The results between the proposed sensor and Calipri are very similar, within 0.01 mm distance. The results of the Miniprof sensor are in some cases more distant compared with both Calipri and the proposed sensor. However, the Miniprof sensing device must be pressed against the object manually, which can provoke outliers if not placed correctly. This issue can be seen in some points in Figure 28, in the top-right area. In general, it can be concluded that the measurements provided by the sensor using the proposed calibration procedure are consistent with the two other calibrated devices.

5. Conclusions

Calibration is the key to providing accurate 3D reconstruction in structured light sensors based on laser triangulation. In this work, a robust calibration procedure is proposed for multi-camera configurations, particularly designed for industrial environments where adverse conditions can affect the quality of images and signals. The proposed procedure is based on a calibration plate with protruding cylinders where the laser lines are projected. Cameras acquire images of the projected lasers and extract profiles that contain a set of partially visible contours of ellipses, resulting from the projection of the cross-section of the cylinders. The profiles are used to accurately obtain the position and orientation of the camera relative to the laser plane. The proposed calibration procedure is applied in two steps: a coarse calibration based on a robust estimator that provides a rough approximation, followed by a fine calibration that iteratively improves the calibration until convergence. Experimental results show excellent performance:

- On average the calibration error for all cameras is only 0.027 mm. This accuracy demonstrates that the proposed procedure can be applied in the most demanding applications.

- Measurement tests corroborate the performance, obtaining an accuracy similar to calibration when reconstructing the surface of test pieces.

- It is demonstrated that heavy noise in the profiles does not affect the calibration results. Both the coarse and the fine calibration include steps to detect and filter outliers. Thus, the calibration is noise insensitive, which is a major requirement for industrial applications.

- The proposed calibration procedure can be applied very quickly, it requires only 30 ms. Therefore, it can be applied in real-time applications or resource-constrained embedded systems.

The proposed calibration procedure is extremely accurate and robust. It can be applied to multiple cameras at the same time on devices with limited computational power. Therefore, it allows for periodic calibration in industrial environments to achieve maximum performance with minimal maintenance.

Acknowledgements

References

References

[1] R. I. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, 2nd Edition, Cambridge University Press, ISBN: 0521540518, 2004.

[2] R. Szeliski, Computer vision: algorithms and applications, Springer Science & Business Media, 2010.

[3] J. Salvi, S. Fernandez, T. Pribanic, X. Llado, A state of the art in structured light patterns for surface profilometry, Pattern recognition 43 (8) (2010) 2666–2680.

[4] J. Zhang, A. Djordjevich, Study on laser stripe sensor, Sensors and Actuators A: Physical 72 (3) (1999) 224–228.

[5] G. Xu, Z. Hao, X. Li, J. Su, H. Liu, L. Sun, An optimization solution of a laser plane in vision measurement with the distance object between global origin and calibration points, Scientific reports 5 (2015) 11928.

[6] G. Zhang, Z. Wei, A novel calibration approach to structured light 3d vision inspection, Optics & Laser Technology 34 (5) (2002) 373–380.

[7] Y. Li, Y. F. Li, Q. L. Wang, D. Xu, M. Tan, Measurement and defect detection of the weld bead based on online vision inspection, IEEE Transactions on Instrumentation and Measurement 59 (7) (2010) 1841–1849.

[8] R. Usamentiaga, J. Molleda, D. F. Garcia, F. G. Bulnes, J. Entrialgo, C. M. S. Alvarez, Flatness measurement using two laser stripes to remove the effects of vibrations, IEEE Transactions on Industry Applications 51 (5) (2015) 4297–4304.

[9] M. Gavrikov, R. Sinetsky, D. Knyazev, Technology and software of system of the metrological analysis of cellular cylindrical surfaces of large-size products, in: Industrial Engineering, Applications and Manufacturing (ICIEAM), International Conference on, IEEE, 2016, pp. 1–4.

[10] J. Salvi, X. Armangué, J. Batlle, A comparative review of camera calibrating methods with accuracy evaluation, Pattern recognition 35 (7) (2002) 1617–1635.

[11] R. Usamentiaga, J. Molleda, D. F. Garcia, Structured-light sensor using two laser stripes for 3d reconstruction without vibrations, Sensors 14 (11) (2014) 20041–20063.

[12] Q. Sun, Y. Hou, Q. Tan, G. Li, A flexible calibration method using the planar target with a square pattern for line structured light vision system, PloS one 9 (9) (2014) e106911.

[13] S. Chi, Z. Xie, W. Chen, A laser line auto-scanning system for underwater 3d reconstruction, Sensors 16 (9) (2016) 1534.

[14] O. Faugeras, Three-dimensional computer vision: a geometric viewpoint, MIT press, 1993.

[15] R. Usamentiaga, Easy rectification for infrared images, Infrared Physics & Technology 76 (2016) 328–337.

[16] R. Tsai, A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses, IEEE Journal on Robotics and Automation 3 (4) (1987) 323–344.

[17] J. Weng, P. Cohen, M. Herniou, et al., Camera calibration with distortion models and accuracy evaluation, IEEE Transactions on pattern analysis and machine intelligence 14 (10) (1992) 965–980.

[18] D. Douxchamps, K. Chihara, High-accuracy and robust localization of large control markers for geometric camera calibration, IEEE transactions on pattern analysis and machine intelligence 31 (2) (2009) 376–383.

[19] M. A. Penna, Camera calibration: A quick and easy way to determine the scale factor, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (12) (1991) 1240–1245.

[20] C.-C. Wang, Extrinsic calibration of a vision sensor mounted on a robot, ieee Transactions on Robotics and Automation 8 (2) (1992) 161–175.

[21] Z. Zhang, A flexible new technique for camera calibration, Pattern Analysis and Machine Intelligence, IEEE Transactions on 22 (11) (2000) 1330–1334.

[22] R. Usamentiaga, J. Molleda, D. F. García, Fast and robust laser stripe extraction for 3d reconstruction in industrial environments, Machine Vision and Applications 23 (1) (2012) 179–196.

[23] C. Steger, An unbiased detector of curvilinear structures, IEEE Transactions on pattern analysis and machine intelligence 20 (2) (1998) 113–125.

[24] Y. Li, J. Zhou, F. Huang, L. Liu, Sub-pixel extraction of laser stripe center using an improved gray-gravity method, Sensors 17 (4) (2017) 814.

[25] A. Fitzgibbon, M. Pilu, R. B. Fisher, Direct least square fitting of ellipses, IEEE Transactions on pattern analysis and machine intelligence 21 (5) (1999) 476–480.

[26] K. Kanatani, Y. Sugaya, Y. Kanazawa, Ellipse fitting for computer vision: implementation and applications, Synthesis Lectures on Computer Vision 6 (1) (2016) 1–141.

[27] M. A. Fischler, R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the ACM 24 (6) (1981) 381–395.

[28] P. J. Besl, N. D. McKay, A method for registration of 3-d shapes, IEEE Trans. Pattern Anal. Mach. Intell. 14 (2) (1992) 239–256. doi:10.1109/34.121791.

[29] R. Usamentiaga, D. F. García, J. Molleda, Efficient registration of 2d points to cad models for real-time applications, Journal of Real-Time Image Processing (2015) 1–19.

[30] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel, Robust statistics: the approach based on influence functions, Vol. 114, John Wiley & Sons, 2011.

[31] S. Agarwal, K. Mierle, Others, Ceres solver, http://ceres-solver.org.

[32] Calipri sensor, http://www.nextsense.at/en/calipri/index.php, accessed: 2017-07-04.

[33] Miniprof sensor, https://www.greenwood.dk/miniprof.php, accessed: 2017-07-04.