Optical Engineering

OpticalEngineering.SPIEDigitalLibrary.org

Calibration routine for a telecentric stereo vision system considering affine mirror ambiguity

Rüdiger Beermann
Lorenz Quentin
Markus Kästner
Eduard Reithmeier

SPIE.

Rüdiger Beermann, Lorenz Quentin, Markus Kästner, Eduard Reithmeier, "Calibration routine for a telecentric stereo vision system considering affine mirror ambiguity," *Opt. Eng.* **59**(5), 054104 (2020), doi: 10.1117/1.OE.59.5.054104

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Calibration routine for a telecentric stereo vision system considering affine mirror ambiguity

Rüdiger Beermann,* Lorenz Quentin, Markus Kästner, and Eduard Reithmeier
Leibniz Universität Hannover, Institut für Mess- und Regelungstechnik,
Fakultät Maschinenbau, Hannover, Germany

Abstract. A robust calibration approach for a telecentric stereo camera system for three-dimensional (3-D) surface measurements is presented, considering the effect of affine mirror ambiguity. By optimizing the parameters of a rigid body transformation between two marker planes and transforming the two-dimensional (2-D) data into one coordinate frame, a 3-D calibration object is obtained, avoiding high manufacturing costs. Based on the recent contributions in the literature, the calibration routine consists of an initial parameter estimation by affine reconstruction to provide good start values for a subsequent nonlinear stereo refinement based on a Levenberg–Marquardt optimization. To this end, the coordinates of the calibration target are reconstructed in 3-D using the Tomasi–Kanade factorization algorithm for affine cameras with Euclidean upgrade. The reconstructed result is not properly scaled and not unique due to affine ambiguity. In order to correct the erroneous scaling, the similarity transformation between one of the 2-D calibration plane points and the corresponding 3-D points is estimated. The resulting scaling factor is used to rescale the 3-D point data, which then allows in combination with the 2-D calibration plane data for a determination of the start values for the subsequent nonlinear stereo refinement. As the rigid body transformation between the 2-D calibration planes is also obtained, a possible affine mirror ambiguity in the affine reconstruction result can be robustly corrected. The calibration routine is validated by an experimental calibration and various plausibility tests. Due to the usage of a calibration object with metric information, the determined camera projection matrices allow for a triangulation of correctly scaled metric 3-D points without the need for an individual camera magnification determination. *© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI.* [DOI: 10.1117/1.OE.59.5.054104]

Keywords: structured light; fringe projection; telecentric lens; affine camera; stereo camera pair; calibration; affine mirror ambiguity; factorization algorithm.

Paper 191817 received Dec. 30, 2019; accepted for publication May 8, 2020; published online May 26, 2020.

1 Introduction

Fringe projection profilometry is a state-of-the-art method in order to characterize the geometry information of three-dimensional (3-D) objects, as it allows a noncontact, fast, and areal data acquisition in the micrometer range.[1–3] If a measurement setup with a small field-of-view (FOV) is required, telecentric lenses can be employed either in stereo vision (with[4,5] or without additional projector[6,7]) or in single camera–projector configurations (with entocentric[8–10] or telecentric projector[11,12]) or telecentric Scheimpflug approaches.[13,14]

The calibration of a telecentric structured light sensor is not as straightforward as in the entocentric case, as a telecentric camera cannot be modeled by the pinhole camera but requires the introduction of the so-called affine camera model instead. As a telecentric lens ideally only maps parallel light onto the camera sensor, the projection center lies at infinity (cf. Ref. 15, p. 166, 173). A distance change along the optical axis of the camera will not result in a dimensional change of the mapped object.

*Address all correspondence to Rüdiger Beermann, E-mail: ruediger.beermann@imr.uni-hannover.de

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

The need for accurate calibration strategies for affine structured light sensors and cameras resulted in a variety of publications in this field. Therefore, in order to motivate this paper and to correctly categorize the derived approach, a short overview on existing calibration strategies is given. The overview is similar to the one provided by Chen et al.,[6] but extended by recent developments and adapted or shortened when considered reasonable. For example, phase-height-based methods such as given in Ref. 16 are not covered, as they are not considered relevant for the derived calibration strategy reported in this paper. Also, calibration techniques based on 3-D objects with exactly measured feature locations (e.g., cubes with markers) are not covered, as the manufacturing of such objects is extremely expensive, and therefore not considered to be practical. Specially adapted calibration techniques for telecentric sensors in Scheimpflug arrangement, as found in Refs. 13 and 14, are not covered as well, as they do not apply to the used hardware setup.

1.1 *Planar-Object-Based Methods*

In this category, strategies are summarized, which use two-dimensional (2-D) calibration planes to calibrate affine cameras.

Lanman et al.[17] presented an approach to reconstruct 3-D surface data based on the motion of an object's depth discontinuities when viewed under orthographic projection. To this end, the authors introduce a model-based calibration approach for a telecentric camera using a planar checkerboard, modified with a pole of known height in order to recover the ambiguity in sign, when estimating the extrinsic parameters for a specific calibration pattern pose. The camera calibration uses a factorization approach inspired by Zhang[18] in order to provide start values for the camera intrinsics and extrinsics. The parameters are further refined in a Levenberg–Marquardt optimization. The authors do not consider lens distortion.

Chen and Liao et al.[6,19] presented a two-step calibration approach for a telecentric stereo camera pair, which comprises a factorization method to determine the initial camera parameters similar to the approach found in Ref. 17. The parameters are refined in a nonlinear optimization routine. The sign ambiguity problem when recovering the rotation matrix is solved with help of a micropositioning stage used to capture two calibration plane poses under known translational displacement. Moreover, the approach considers radial distortion. The authors suggest the acquisition of as many target poses as possible in order to avoid degeneracy and in consequence an "ill calibration" (Ref. 6, p. 88).

Li et al.[11,20] proposed a calibration method for a single camera based on an analytical camera description in order to model the distortion of a telecentric lens correctly (namely radial, decentering, and thin prism distortions) and developed it into an approach to calibrate a structured light sensor with telecentric camera and projector. It is not fully clear how the authors solve the problem of sign ambiguity, when recovering the extrinsics. In their literature review, Li and Zhang[9] state that "it is difficult for such a method to achieve high accuracy for extrinsic parameters calibration [. . .]."

Yao and Liu[21] introduced an approach where again an additional stage is used to solve for the extrinsic sign ambiguity. After a camera start value determination based on a distortion-free camera model, two nonlinear optimization steps are executed. In the first step, the calibration plane coordinates are optimized to allow the usage of cheap print patterns. Second, all camera parameters are refined, including radial and tangential lens distortion, and also the distortion center. The approach provides a greater flexibility, as the distortion center is not necessarily fixed to the middle of the sensor. Nevertheless, a comparison between calibration results based on a printed and a precisely manufactured pattern shows great difference in the estimated distortion parameters. The authors argument that the distortion is generally small for telecentric lenses. Therefore, small differences in the optimization procedure result in great parameter differences. Another reason could be the missing re-estimation of the calibration plane coordinates in the second nonlinear optimization step. The distortion-free camera model is considered ground truth when estimating the calibration points.

Hu et al.[22] presented an approach for a single camera calibration based on the results by Yao et al., but provided a method to gain an initial estimation for the distortion center to avoid local minima. The distortion center and the parameters are further refined in a subsequent nonlinear

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

full-parameter optimization. The authors consider both radial and tangential distortion coefficients. Their approach is developed into a full calibration and reconstruction routine for a microscopic stereo vision system.[5]

Li and Zhang[9] introduced a calibration routine for a hardware setup comprising an entocentric projector and a telecentric camera and used the absolute coordinate frame of the projector as a reference for the telecentric camera. In the first step, the projector is calibrated with the standard camera pinhole model. The necessary correspondences are provided by the uncalibrated telecentric camera, capturing multiple calibration plane poses with and without vertical and horizontal phasemap, respectively (cf. concept of image capturing projector in Ref. 23). The feature correspondences used for the projector calibration are then projected back into 3-D (in the projector's coordinate frame) to calibrate the affine camera. This approach is very stable but requires an entocentric projector, which might not be available in a sensor setup.

1.2 *Affine Autocalibration*

This category comprises so-called autocalibration approaches for affine cameras. As most autocalibration approaches require structure-from-motion results as input, exemplary developments in this field are covered as well.

According to Hartley et al., "auto-calibration is the process of determining internal camera parameters directly from multiple uncalibrated images" (cf. Ref. 15, p. 458), without using specially designed calibration devices with known metric distances, or scene properties such as vanishing points. The derivation of the camera intrinsics might be directly connected to the reconstruction of 3-D scene points, upgrading a nonunique projective or affine reconstruction to a Euclidean reconstruction by applying special constraints. Such a constraint could be the assumption of fixed camera intrinsics for all images.

The basic theory for autocalibration of a perspective projection camera is formulated by Faugeras et al.[24] Well-known classical structure-from-motion approaches under orthography are suggested for the two-view scenario by Koenderink and van Doorn,[25] and for at least three views by Tomasi and Kanade, namely the factorization algorithm.[26] The camera is moved around an object and captures images from different positions under orthographic projection. Detected feature correspondences in the sequential images are used to recover the scene's shape and the camera motion in affine space. Appropriate boundary conditions allow for the reconstruction of Euclidean structure up to scale.

The affine 3-D reconstruction result is used as input in the generalized affine autocalibration approach by Quan.[27] The authors introduced metric constraints for the affine camera, comprising orthographic, weak perspective, and paraperspective camera model.

An important precondition for the applicability of the Tomasi–Kanade factorization algorithm is the visibility of the used point correspondences in all views. Using data subsets, Tomasi and Kanade enable the factorization approach to handle missing data points. The subset-based reconstructed 3-D coordinates are projected onto the calculated camera positions in order to obtain a complete measurement matrix. This method nevertheless requires feature points that are visible in all views (the data subsets). It allows patching of missing matrix entries, rather than providing an approach for sparse data sets.

Brandt derived a more flexible structure-from-motion approach, as "no single feature point needs to be visible in all views" (cf. Ref. 28, p. 619). The approach comprises two iterative affine reconstruction schemes, and a noniterative, linear method, using four noncoplanar reference points visible in all views. Brandt and Palander[29] furthermore presented a statistical method to recover the camera parameters directly from provided point correspondences without the necessity of an affine reconstruction. As solution, a posterior probability distribution for the parameters is obtained.

Guilbert et al. proposed an approach for sparse data sets using an affine closure constraint, which allows "to formulate the camera coefficients linearly in the entries of the affine fundamental matrices" (cf. Ref. 30, p. 317), using all available information of the epipolar geometry. The authors claim that the algorithm is more robust against outliers compared to factorization algorithms. Moreover, they present an autocalibration method and directly compare it to Quan's

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

method. The so-called contraction mapping scheme shows a 100% success rate in reaching the global minimum and a lower execution time.

Horaud et al.[31] described a method to recover the Euclidean 3-D information of a scene when capturing scene data with an uncalibrated affine camera mounted to a robot's end effector. The authors use controlled robot motions, in order to remove affine mirror ambiguity and guarantee a unique affine reconstruction solution. The camera intrinsics are obtained by performing an QR-decomposition according to Quan.[27]

An approach of motion recovery from weak-perspective images is presented by Shimshoni et al.[32] The authors reformulate the motion recovery problem to a search for triangles on a sphere, offering a geometric interpretation of the problem.

Further information on the concepts of affine autocalibration in general can be found in Ref. 33, p. 163 et seq.

1.3 *Hybrid Method*

Liu et al.[12] combined the Tomasi–Kanade factorization algorithm with a 3-D calibration target in order to retrieve the parameters of a fringe projection system with telecentric camera and projector. The authors use a 3-D calibration target with randomly distributed markers. The target consists of two 2-D planes, forming a rooftop structure. As the marker positions on the planes are not required to be known beforehand, the target manufacturing requirements are low.

The suggested approach is basically a two-step routine: the 3-D calibration target is captured by the camera in different orientations, with and without two sets of gray code patterns, generated by the projector. The approach of the so-called image capturing projector by Zhang et al.[23] allows now to solve the correspondence problem between camera, projector, and circular dots on the target. First, the dots' image coordinates are extracted for camera and projector. Then, using the Tomasi–Kanade algorithm and an appropriate upgrade scheme from affine to Euclidean space, an initial guess for the calibration targets shape (3-D coordinates of the circular dots) and the corresponding projection matrices are obtained. As the point cloud data can only be reconstructed up to scale, the camera's effective magnification has to be provided in order to reconstruct metric 3-D data of the circular dots. As no metric distances are defined on the 3-D calibration target, the authors suggest the additional usage of a simple 2-D target in a plane-based calibration routine, such as given in Ref. 21. In the second step, the initial guesses are used as start parameters in a nonlinear bundle adjustment scheme to minimize the total projection error. Next to the target poses, also the projector-camera rig parameters and the 3-D coordinates of the calibration target are refined.

1.4 *Contributions in this Paper*

The approach by Liu et al. is an alternative to the routines discussed in Sec. 1.1, avoiding among others planarity-based degeneracy problems [e.g., as reported by Chen et al. in Ref. 6 (p. 88) or in general by Collins et al. in Ref. 34]. The approach does not rely on the usage of a plane with linear stage or a pole but on a 3-D rooftop calibration target. The Tomasi–Kanade algorithm provides a good estimation of the camera rotations (even with a relatively low number of captured object poses), which allows for a robust convergence of the subsequent nonlinear refinement.

Nevertheless, in order to obtain a fully calibrated measurement system, the magnification factor has to be determined separately in an individual step, which is cumbersome. Also, the authors do not address the problem of the so-called mirror ambiguity, which is still present when reconstructing affine point data with the Tomasi–Kanade algorithm [cf. Ref. 35 (p. 415), Ref. 36 (p. 7–8), and Ref. 31 (p. 1576)]. As the reconstructed 3-D data might be mirrored, the start values for nonlinear optimization are also estimated based on a mirrored point cloud, resulting in mirror-based camera locations (for further clarification see Sec. 3.2.5). Although the subsequent nonlinear optimization might still converge, triangulated geometry results might be mirrored, as the camera – projector – arrangement is potentially inverted.

The mirror ambiguity is especially in a stereo camera setup problematic. Two individual affine reconstruction schemes for both cameras can result in start values, that are both based

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

on a mirrored and nonmirrored point cloud. A combination of the camera start values in a single stereo optimization directly affects its robustness. The optimizer might converge toward a local minimum or not converge at all.

Therefore, we propose an adapted calibration procedure for a structured light sensor comprising a telecentric stereo camera pair and an entocentric projector as feature generator. The projector is not meant to be used for the calibration of the affine cameras to allow for a direct calibration. Hence, the suggested routine is also valid for a simple stereo camera setup without projector. As the triangulation is conducted between the two cameras, the hardware setup is equivalent to the setup presented by Liu et al. (two telecentric lenses are used for triangulation).

Our routine is also based on the Tomasi–Kanade factorization algorithm to determine the start values. The application of a more recent affine reconstruction and autocalibration scheme might be interesting in the scope of this paper, but the additional effort for the algorithm implementation will prove not to be necessary, as the proposed calibration scheme works just fine. The feature visibility restriction will not prove to be an obstacle in the suggested approach, as the number of detectable features in all views is large enough by introducing an appropriate calibration target.

The contributions of this paper can be summarized to the following points:

- Our calibration approach uses a 3-D calibration target combining two 2-D planes with defined dot patterns. The designed approach allows for a complete calibration of the presented telecentric stereo camera system without the need for an additional magnification factor determination.
- Although a 3-D target is used, the target fabrication is only slightly more expensive than in the 2-D case. This is due to the fact that the rigid body transformation between two 2-D planes is optimized together with the sensor parameters. Only the planes have to be manufactured with high precision. Prior information on the plane orientation in relation to each other is not necessary. The calibration routine yields a metric 3-D calibration object.
- We introduce an Aruco marker-based detection strategy as introduced by Garrido-Jurado et al.[37] in order to distinctly differentiate between the two plane marker patterns of the 3-D calibration object.
- The estimated rigid body transformation between the two 2-D planes is also used to test the reconstructed 3-D points for affine mirror ambiguity. If the points are mirrored, a simple matrix operation is suggested to correct the erroneous start values.
- We directly include a distortion model into the calibration routine.
- In order to facilitate the acquisition process of calibration images, only one stereo image of the same target pose is required. This pose determines the measurement coordinate frame. The motivation for this procedure is similar to the one given by Chen et al.[6] It is not easy to capture a large number of target orientations, which are on the one hand fully representative for a specific camera and allow for a robust determination of intrinsics, and on the other hand are simultaneously viewable by both cameras. An extreme target pose, which might be helpful for a robust calibration of camera one, is potentially not perfectly observable by camera two.

2 Affine Camera Model

The mathematical model of the affine camera is defined as found in Ref. 6:

$$\underbrace{\begin{pmatrix} {}_{C}\mathbf{u} \\ 1 \end{pmatrix}}_{{}_{C}\mathbf{u}_h} = \underbrace{\begin{bmatrix} \frac{m}{s_x} & -\frac{m\cot(\rho)}{s_x} & c_x \\ 0 & \frac{m}{s_y \sin(\rho)} & c_y \\ 0 & 0 & 1 \end{bmatrix}}_{\mathbf{K}} \underbrace{\begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{{}^{C}\tilde{\mathbf{T}}_{O}} \underbrace{\begin{pmatrix} {}_{O}\mathbf{X} \\ 1 \end{pmatrix}}_{{}_{O}\mathbf{X}_h.}, \tag{1}$$

The model defines a mapping of an arbitrary homogeneous 3-D object point ${}_{O}\mathbf{X}_h$ onto the camera sensor. The point is transformed by a truncated rigid body matrix ${}^{C}\tilde{\mathbf{T}}_{O}$ into the 2-D coordinate frame $\{C\}$ of the camera. The multiplication with the affine camera matrix $\mathbf{K}$ maps

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

the resulting homogeneous 2-D point ${}_{C}\mathbf{X}_{h}$ onto the sensor in location ${}_{c}\mathbf{u}$ (in px) in the coordinate frame $\{c\}$.

The pixel sizes in the x- and y-directions are parametrized by s_x and s_y, respectively (in metric length per pixel, e.g., $\frac{\mathrm{mm}}{\mathrm{px}}$), the magnification is defined by m (no unit). Skew is considered in terms of skew angle ρ. The origin of the image coordinate system is fixed to the middle of the camera sensor to define a center for a telecentric lens distortion model according to $c_x = w/2$ and $c_y = h/2$, with sensor width w and height h.

The affine projection can also be formulated in a compact, inhomogeneous form according to

$$\underbrace{\begin{pmatrix} {}_{c}u \\ {}_{c}v \end{pmatrix}}_{{}_{c}\mathbf{u}} = \underbrace{\begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \end{bmatrix}}_{{}^{c}\mathbf{M}_{O}} \underbrace{\begin{pmatrix} {}_{O}X \\ {}_{O}Y \\ {}_{O}Z \end{pmatrix}}_{{}_{O}\mathbf{X}} + \underbrace{\begin{bmatrix} p_{14} \\ p_{24} \end{bmatrix}}_{{}_{c}\mathbf{p},}, \tag{2}$$

with ${}^{c}\mathbf{M}_{O}$ and ${}_{c}\mathbf{p}$ holding the entries of the matrix multiplication result $\mathbf{K}^{C}\tilde{\mathbf{T}}_{O}$ as given by

$$\mathbf{K}^{C}\tilde{\mathbf{T}}_{O} = \begin{bmatrix} & {}^{c}\mathbf{M}_{O} & & {}_{c}\mathbf{p} \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{3}$$

A distortion model is introduced considering radial and tangential distortion based on the approach by Brown et al. (cf. Refs. 38–40) and is defined as

$${}_{C}X_{d} = (1 + k_1 \cdot R^2 + k_2 \cdot R^4)_{C}X + 2p_1 \cdot {}_{C}X \cdot {}_{C}Y + p_2(R^2 + 2 \cdot {}_{C}X^2), \tag{4}$$

$${}_{C}Y_{d} = (1 + k_1 \cdot R^2 + k_2 \cdot R^4)_{C}X + 2p_2 \cdot {}_{C}X \cdot {}_{C}Y + p_1(R^2 + 2 \cdot {}_{C}Y^2). \tag{5}$$

${}_{C}(X_d, Y_d)$ parametrizes a distorted and ${}_{C}(X, Y)$ an undistorted point in the affine camera coordinate frame $\{C\}$. R defines the radial distance to the distortion center with $R = \sqrt{{}_{C}X^2 + {}_{C}Y^2}$. The coefficients are combined in distortion vector $\mathbf{k}_C = (k_1, k_2, p_1, p_2)^T$.

For perspective cameras, the distortion model is applied upon so-called normalized image points (ideal image plane), in order to avoid numerical instability, when estimating the parameters. As this ideal image plane does not exist for affine cameras, the distortion is added in coordinate frame $\{C\}$. Although this leads to values of larger magnitude compared to the normalized image coordinates for perspective cameras [especially due to the R^4-term in Eqs. (4) and (5)], the distortion vector $\mathbf{k}_C$ could be optimized robustly.

3 Calibration Routine

In the first step, the initial parameter values for the affine camera matrices, the truncated rigid body transformation, and the transformation from the first to the second 2-D calibration plane are estimated. To this end, according to the approach introduced by Liu et al.,[12] the Tomasi–Kanade factorization algorithm[26] is used in order to reconstruct the 3-D data of the calibration target coordinates. In contrast to the approach by Liu et al., two equidistant marker grids with defined distances are used, instead of randomly distributed markers. The additionally provided distance information is exploited to determine the cameras' magnification values to obtain camera projection matrices that allow for metric 3-D measurements. Moreover, the presented routine allows to correct mirrored start values, by distinctly solving the affine mirror ambiguity. The start values are determined for each camera independently, meaning that the complete procedure according to Sec. 3.2 has to be executed twice.

In the second step, the initial parameter values for both cameras are refined together via nonlinear stereo optimization, in which also the distortion parameters are estimated.

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

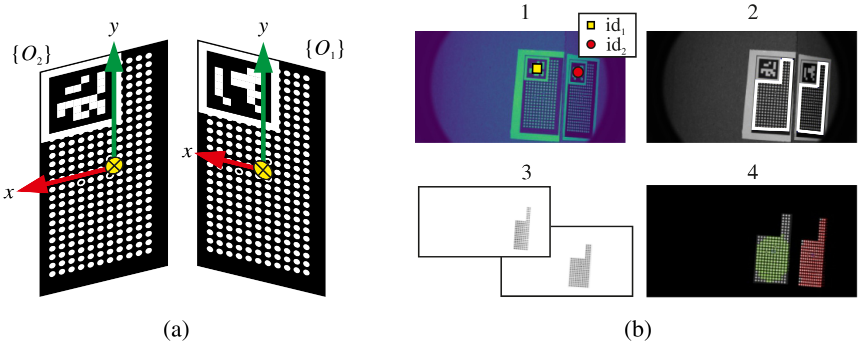


Fig. 1 (a) Layout of calibration target with two individual coordinate systems $\{O_1\}$ and $\{O_2\}$. (b) Detection procedure. Based on the detected Aruco markers [(id1) and (id2) dots, (b, 1)], the regions of interest (ROI) for each plane are determined (b, 2). The ROIs allow for a planewise masking (b, 3) and dot marker detection [green and red, respectively, (b, 4)].

3.1 *Calibration Target and Marker Detection*

The layout of the 3-D calibration target is shown in Fig. 1(a). The rooftop structure was introduced by Liu et al., but the random dot distribution is substituted by two defined planar dot patterns with individual coordinate frames $\{O_1\}$ and $\{O_2\}$. It is necessary to differentiate between the two patterns. To this end, Aruco markers[37] are printed in the left upper corner of each plane. The markers allow for a distinct and robust marker detection [Fig. 1(b, 1)], which permits the masking of everything except for the associated plane data [Fig. 1(b, 2–3)]. After approximate plane detection, the circle markers are identified by a detection algorithm, and the image-plane-correspondences are obtained [Fig. 1(b, 4)].

It is important to notice that at this point, the correspondences of both planes are given in the two individual coordinate frames $\{O_1\}$ and $\{O_2\}$. There is no information on the rigid body transformation which allows for a marker point formulation in a single coordinate frame. The z coordinate for all detected features—independently of the chosen plane—is zero. The necessary transformation will be estimated in the subsequent calibration routine. The advantage is that single planes with individual marker coordinate frames are easier to manufacture than a single 3-D calibration target.

3.2 *Start Value Determination*

3.2.1 *Tomasi–Kanade algorithm*

The factorization algorithm by Tomasi and Kanade[26] is used to reconstruct 3-D coordinates in affine space based on at least four point correspondences over i affine camera images. There is no need for a calibrated camera, or known distances between the corresponding points in the different camera views. The obtained 3-D data is reconstructed up to scale.

The approach was originally introduced in order to obtain shape information from affine image streams but can also be applied if not the camera, but the object itself is moved relatively to the camera. The camera projection matrices ${}^{c}\mathbf{M}_{T_1,i}$ (that project a point from the 3-D frame $\{T_1\}$ onto the 2-D frame of the camera sensor), the translational part ${}_{c}\mathbf{p}_i$, and the 3-D points ${}_{T_1}\mathbf{X}_j$ can be obtained by minimizing cost function e_c:

$$e_c = \sum_{i=1}^{m}\sum_{j=1}^{n} \|{}_{c}\mathbf{u}_{ij} - {}_{c}\hat{\mathbf{u}}_{ij}\|^2 = \sum_{i=1}^{m}\sum_{j=1}^{n} \|{}_{c}\mathbf{u}_{ij} - ({}^{c}\mathbf{M}_{T_1,i}\,{}_{T_1}\mathbf{X}_j + {}_{c}\mathbf{p}_i)\|^2, \tag{6}$$

w.r.t. ${}^{c}\mathbf{M}_{T_1,i}$, ${}_{c}\mathbf{p}_i$, and ${}_{T_1}\mathbf{X}_j$. $\|{}_{c}\mathbf{u}_{ij} - {}_{c}\hat{\mathbf{u}}_{ij}\|$ is the geometric error with ${}_{c}\hat{\mathbf{u}}_{ij}$ as point projection based on the optimized model parameters. i is the number of recorded object poses and j is the number of point correspondences. To reduce the number of parameters, the pixel data are centered by the centroid ${}_{c}\boldsymbol{\omega}_i = {}_{c}(\omega_x, \omega_y)_i^T = {}_{c}(\frac{1}{n}\sum_{j=1}^{n} {}_{c}u_j, \frac{1}{n}\sum_{j=1}^{n} {}_{c}v_j)_i^T$ of the corresponding

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

image points according to ${}_c\mathbf{u}_{\text{centr},i} = {}_c\mathbf{u}_i - {}_c\boldsymbol{\omega}_i$, which yields w.r.t. the new centered data ${}_c\mathbf{p}_i = \mathbf{0}$ and therefore

$$e_c = \sum_{i=1}^{m}\sum_{j=1}^{n} \|{}_c\mathbf{u}_{\text{centr},ij} - {}^c\mathbf{M}_{T_1,i\,T_1}\mathbf{X}_j\|^2. \tag{7}$$

As the point correspondences are corrupted by noise, a solution for ${}^c\mathbf{M}_{T_1,i}$ and ${}_{T_1}\mathbf{X}_j$ can only be approximated. By introducing a measurement matrix $\mathbf{W}$, Eq. (7) is reformulated with the Frobenius norm as

$$e_c = \|\mathbf{W} - \hat{\mathbf{M}}\hat{\mathbf{X}}_1\|_F^2, \tag{8}$$

with

$$\mathbf{W} := \begin{bmatrix} {}_cu_{11} & \cdots & {}_cu_{1n} \\ \vdots & \ddots & \vdots \\ {}_cu_{m1} & \cdots & {}_cu_{mn} \\ \vdots & \ddots & \vdots \\ {}_cv_{11} & \cdots & {}_cv_{1n} \\ \vdots & \ddots & \vdots \\ {}_cv_{m1} & \cdots & {}_cv_{mn} \end{bmatrix}_{(2m)\times n}, \quad \hat{\mathbf{M}} := \begin{bmatrix} {}^c\mathbf{m}_{T_1,11} \\ \vdots \\ {}^c\mathbf{m}_{T_1,m1} \\ \vdots \\ {}^c\mathbf{m}_{T_1,12} \\ \vdots \\ {}^c\mathbf{m}_{T_1,m2} \end{bmatrix}_{(2m)\times 3} \quad \text{and,} \quad \hat{\mathbf{X}}_1 := [{}_{T_1}\mathbf{X}_1 \quad \cdots \quad {}_{T_1}\mathbf{X}_n]_{3\times n}.$$

Measurement matrix $\mathbf{W}$ holds the centered pixel information ${}_c\mathbf{u}_{\text{centr},ij}$. The motion matrix $\hat{\mathbf{M}}$ holds m projection matrices ${}^c\mathbf{M}_{T_1,i} = ({}^c\mathbf{m}_{T_1,i1}, {}^c\mathbf{m}_{T_1,i2})^T$, whereas first rows ${}^c\mathbf{m}_{T_1,i1}$ and second rows ${}^c\mathbf{m}_{T_1,i2}$ are sorted according to the definition of $\hat{\mathbf{M}}$. The shape matrix $\hat{\mathbf{X}}_1$ holds n reconstructed 3-D points. Index 1 indicates the first version of the shape matrix, prior to further transformations.

$\hat{\mathbf{M}}$ and $\hat{\mathbf{X}}_1$ can be obtained by a singular value decomposition (SVD) of $\mathbf{W}$ [refer to Ref. 26 (p. 141) and Ref. 15 (p. 438) for more detailed information on the decomposition]. Until now, the 3-D data are only reconstructed in affine space.

Due to affine ambiguity, motion and shape matrix are not reconstructed uniquely. An arbitrary matrix $\mathbf{Q}$ can be introduced into $\hat{\mathbf{W}} = \hat{\mathbf{M}}\,\hat{\mathbf{X}}_1 = \hat{\mathbf{M}}\,\mathbf{Q}\,\mathbf{Q}^{-1}\hat{\mathbf{X}}_1$, without changing the resulting measurement matrix estimation $\hat{\mathbf{W}}$.

The reconstructed affine 3-D data $\hat{\mathbf{X}}_1$ can be upgraded to Euclidean space, if appropriate metric constraints are imposed upon the motion matrix. To this end, different approaches have been presented, depending on the type of affine camera model.[27] Tomasi and Kanade hypothesized a simple orthographic projection, with a fixed scaling factor of one for each camera view and no additional skew factor. Although the introduced camera model according to Eq. (1) considers skew and a data scaling larger than one (e.g., as expressed by $\frac{m}{s_x}$), the approach by Tomasi–Kanade is suitable. In the parameter refinement step, nonzero skew is allowed, as well as arbitrary magnification values. The constraints of the orthographic model yield matrix $\mathbf{Q}$, which is used to transform the 3-D points $\hat{\mathbf{X}}_1$ from affine to Euclidean space according to

$$\hat{\mathbf{X}}_2 = [{}_{T_2}\mathbf{X}_1 \quad \cdots \quad {}_{T_2}\mathbf{X}_n] = \mathbf{Q}^{-1}\hat{\mathbf{X}}_1. \tag{9}$$

The transformation by matrix $\mathbf{Q}$ requires the definition of a new coordinate frame $\{T_2\}$. The transformed 3-D points $\hat{\mathbf{X}}_1$ now only differ from the absolute metric points by a scaling factor (except for potential skew and assuming the same scaling in x and y directions), as so far no ground truth information with known metric positions was used to recover the exact object scaling.

The transformed motion matrix $\hat{\mathbf{R}} = \hat{\mathbf{M}}\mathbf{Q}$ holds the data on the truncated rotation matrices for each camera view. The truncated rotation matrix for the i'th camera view ${}^c\tilde{\mathbf{R}}_{T_2,i}$ can be obtained from $\hat{\mathbf{R}}$ by resorting the row entries according to

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

$$ {}^{c}\tilde{\mathbf{R}}_{T_2,i} = \begin{bmatrix} {}^{c}\mathbf{r}_{T_2,i1} \\ {}^{c}\mathbf{r}_{T_2,i2} \end{bmatrix}, \quad \text{with } i = 1, \dots, m. \tag{10} $$

The metric constraints for the orthographic model are stated in Ref. 26. Additional information on Euclidean upgrading for affine cameras can be found in Refs. 27, 33 (p. 167), and 41.

3.2.2 *Scaling factor and telecentric magnification*

In order to obtain the metric calibration marker coordinates in 3-D, the data scaling has to be determined. This is achieved using ground truth information in terms of the 2-D marker distance on the planes. The relationship between the 3-D points in $\{T_2\}$ and the 2-D points in $\{O_1\}$ of the first plane can be formulated by an affine transformation matrix ${}^{T_2}\mathbf{A}_{O_1}$ according to

$$ {}_{T_2}\mathbf{X}_{k,h} = {}^{T_2}\mathbf{A}_{O_1\,O_1}\mathbf{X}_{l_1,h} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}_{O_1} \mathbf{X}_{l_1,h}, \quad \text{with } k = l_1 = 1, \dots, n_1. \tag{11} $$

The point data are defined in homogeneous coordinates. Index k only addresses points that correspond to the first plane, n_1 is the total number of detected points on the first plane.

The 12 parameters of the affine matrix are estimated using the method of least squares (e.g., as given in Ref. 42), and the known data sets ${}_{T_2}\mathbf{X}_{k,h}$ and ${}_{O_1}\mathbf{X}_{l_1,h}$. The z coordinate of ${}_{O_1}\mathbf{X}_{l_1,h}$ is zero (degenerate input), the least squares optimization will not provide a solution for the parameters a_{13}, a_{23}, and a_{33}. This is not a problem, as not all parameters need to be known in order to determine the scaling factor s. It can be directly obtained from vector $(a_{11}, a_{21}, a_{31})^T$ by calculating its Euclidean length. It is also possible to obtain s from vector $(a_{12}, a_{22}, a_{32})^T$, as the scaling in x and y directions is approximately equal (square pixel, zero skew assumption with $\rho = 90$ deg). This is due to the data input. Basically, a similarity transformation (rigid body transformation and scaling) with seven parameters is enough to parametrize the transformation between ${}_{T_2}\mathbf{X}_{k,h}$ and ${}_{O_1}\mathbf{X}_{l_1,h}$. Therefore, the average of both s-values is used.

Once s is determined, a scaling matrix can be defined according to $\mathbf{S} = s\mathbf{I}$ with $\mathbf{I}$ as identity matrix. The metric 3-D points of the calibration target are now obtained as

$$ \hat{\mathbf{X}}_3 = \mathbf{S}^{-1}\hat{\mathbf{X}}_2. \tag{12} $$

Some remarks on the estimation of scaling factor s:

- As the points ${}_{T_2}\mathbf{X}_{k,h}$ are more or less exactly defined on a plane, it is possible to transform them into a 2-D coordinate system with $z = 0$. This allows to estimate a full 2-D affine transformation (no degeneracy) and to derive s.
- It is also possible to use the point data of the second calibration plane to obtain the scaling factor.
- The scaling matrix $\mathbf{S}$ is not applied upon the motion matrix $\hat{\mathbf{M}}$. The requirement of $\hat{\mathbf{W}} = \hat{\mathbf{M}}\mathbf{S}\mathbf{S}^{-1}\hat{\mathbf{X}}_2$ is met by introducing the truncated rigid body matrices $\tilde{\mathbf{T}}_i$ for each pose and the camera matrix $\mathbf{K}$ into the equation (cf. Sec. 3.2.4).

3.2.3 *Estimation of rigid body transformation between calibration planes*

In order to provide a start value for the rigid body transformation ${}^{O_1}\mathbf{T}_{O_2}$ (cf. Fig. 2), the transformations ${}^{T_2}\mathbf{T}_{O_1}$ and ${}^{T_2}\mathbf{T}_{O_2}$ between the plane data and the reconstructed 3-D calibration points have to be estimated. The relationship between the points is given as

$$ {}_{T_2}\mathbf{X}_{k,h} = {}^{T_2}\mathbf{T}_{O_1\,O_1}\mathbf{X}_{l_1,h}, \quad \text{with } k = l_1 = 1, \dots, n_1, \tag{13} $$

$$ {}_{T_2}\mathbf{X}_{k,h} = {}^{T_2}\mathbf{T}_{O_2\,O_2}\mathbf{X}_{l_2,h}, \quad \text{with } \begin{cases} k = n_1 + 1, \dots, n \\ l_2 = 1, \dots, n_2 \end{cases}. \tag{14} $$

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

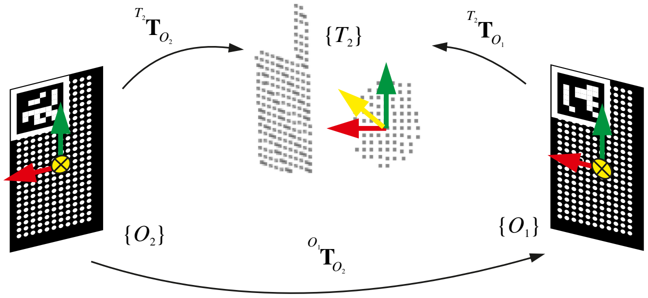

${}^{T_2}\mathbf{T}_{O_2}$
$\{T_2\}$
${}^{T_2}\mathbf{T}_{O_1}$
$\{O_2\}$
$\{O_1\}$
${}^{O_1}\mathbf{T}_{O_2}$

Fig. 2 Rigid body transformations between the reconstructed 3-D data of the calibration target given in $\{T_2\}$ and the coordinate frames of the calibration planes $\{O_1\}$ and $\{O_2\}$.

${}_{T_2}\mathbf{X}_{k,h}$ is considered to be scaled according to Eq. (12)—resulting in a metric point cloud—without introducing an additional index indicating scaling. In accordance with the previous section, the total number of calibration points is $n = n_1 + n_2$. The number of points on the first plane is n_1 and on the second plane n_2.

The rigid body transformations ${}^{T_2}\mathbf{T}_{O_1}$ and ${}^{T_2}\mathbf{T}_{O_2}$ are obtained by an SVD (e.g., as given in Ref. 43), since ${}_{T_2}\mathbf{X}_{k,h}$ and the corresponding calibration plane points ${}_{O_1}\mathbf{X}_{l_1,h}$ and ${}_{O_2}\mathbf{X}_{l_2,h}$ are known.

The desired transformation is then determined according to

$$ {}^{O_1}\mathbf{T}_{O_2} = ({}^{T_2}\mathbf{T}_{O_1})^{-1}\,{}^{T_2}\mathbf{T}_{O_2} = {}^{O_1}\mathbf{T}_{T_2}\,{}^{T_2}\mathbf{T}_{O_2}. \tag{15} $$

3.2.4 *Determination of initial camera matrix and truncated rigid body transformations*

The scaling factor s according to Sec. 3.2.2 can directly be entered into the camera matrix, if the skew factor is supposed to be close to zero ($s \approx \frac{m}{s_x} \approx \frac{m}{s_y}$). As aforementioned, the origin of the image coordinate system is fixed to the middle of the camera sensor. The initial camera matrix is therefore

$$ \mathbf{K} = \begin{bmatrix} s & 0 & w/2 \\ 0 & s & h/2 \\ 0 & 0 & 1 \end{bmatrix}. \tag{16} $$

The (2×3)-truncated rotation matrices ${}^{C}\tilde{\mathbf{R}}_{T_2,i}$ need to be extended to (3×4)-truncated transformation matrices ${}^{C}\tilde{\mathbf{T}}_{T_2,i}$, as a formulation according to Eq. (1) is required. (As now a scaled projection is hypothesized with scaling factor s due to the introduction of the camera matrix, the small index c is changed to a capital C for the extrinsics (e.g., ${}^{c}\tilde{\mathbf{R}}_{T_2,i}$ to ${}^{C}\tilde{\mathbf{R}}_{T_2,i}$) in order to differentiate between the unscaled points in $\{C\}$ and the scaled points on the sensor in $\{c\}$.)

The original sensor data of the i'th camera view were shifted by its centroid ${}_{c}\boldsymbol{\omega}_i \underset{c}{=} (\omega_x, \omega_y)_i^T$. This shift has to be considered when ${}^{C}\tilde{\mathbf{T}}_{T_2,i}$ is computed. Furthermore, the image coordinate system is meant to be fixed to the sensor middle—the necessary shift by $w/2$ and $h/2$ has to be considered as well. The start values for the truncated rigid body matrices can therefore be determined according to

$$ {}^{C}\tilde{\mathbf{T}}_{T_2,i} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \multicolumn{3}{c}{{}^{C}\tilde{\mathbf{R}}_{T_2,i}} & \begin{matrix} \frac{{}_{c}\omega_{x,i} - w/2}{s} \\ \frac{{}_{c}\omega_{y,i} - h/2}{s} \end{matrix} \\ 0 \quad 0 \quad 0 & 1 \end{bmatrix}. \tag{17} $$

As the cameras are meant to be calibrated in coordinate frame $\{O_1\}$, the truncated matrices have to transformed according to

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

$$ {}^{C}\tilde{\mathbf{T}}_{O_1,i} = {}^{C}\tilde{\mathbf{T}}_{T_2,i}\,{}^{T_2}\mathbf{T}_{O_1}. \tag{18} $$

${}^{T_2}\mathbf{T}_{O_1}$ is known from the previous section.

3.2.5 *Affine mirror ambiguity*

Due to the so-called mirror ambiguity of the affine projection, the reconstructed 3-D points obtained by the Tomasi–Kanade factorization algorithm are potentially not accurate but might be mirrored.[35,36] For further clarification Fig. 3(a) is given (inspired by Ozden et al.[44]): a mirror reflection of a 3-D calibration object (here defined by the points $A'B'C'$) w.r.t. a plane, which is in parallel to the image sensor (mirror plane), will have the same affine projection result in camera 1 as the original object (ABC). (In Fig. 3, the sensor plane for camera 1 and the mirror plane are equal.) Therefore, based on multiple views of the calibration object, two different 3-D reconstructions are valid: the mirrored and the original and nonmirrored point cloud.

In consequence, the truncated rigid body transformations for the different camera poses might have been estimated based on a mirrored 3-D point cloud. Both camera poses according to Fig. 3(a) (cam 2′ and 2) result in the exact same image coordinates, when projecting the points ABC or $A'B'C'$ onto the sensor. This can be shown with help of the inhomogeneous affine projection formulation according to Eq. (2). For the sake of simplicity, the camera matrix $\mathbf{K}$ is set to the identity matrix ($\frac{m}{s_x} = \frac{m}{s_y} = 1$, $c_x = c_y = 0$, $\rho = 90$ deg), and the translational shift is supposed to be zero ($t_x = t_y = 0$), yielding a simple orthographic projection according to

$$ \begin{pmatrix} {}_{C}u \\ {}_{C}v \end{pmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \end{bmatrix} \begin{pmatrix} {}_{O}X \\ {}_{O}Y \\ {}_{O}Z \end{pmatrix}. \tag{19} $$

If Eq. (19) is expanded by a (3×3) mirror matrix $\boldsymbol{Q}_{\mathrm{mir}}$ (point reflection about xy-plane) and its inverse, nothing is changed (as $\mathbf{Q}_{\mathrm{mir}}\mathbf{Q}_{\mathrm{mir}}^{-1} = \mathbf{I}$), yielding

$$ \begin{aligned} \begin{pmatrix} {}_{C}u \\ {}_{C}v \end{pmatrix} &= \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{pmatrix} {}_{O}X \\ {}_{O}Y \\ {}_{O}Z \end{pmatrix} \\ &= \begin{bmatrix} r_{11} & r_{12} & -r_{13} \\ r_{21} & r_{22} & -r_{23} \end{bmatrix} \begin{pmatrix} {}_{O}X \\ {}_{O}Y \\ -{}_{O}Z \end{pmatrix}. \end{aligned} \tag{20} $$

In consequence, object point ${}_{O}\mathbf{X}$ is mirrored, and the r_{13} and r_{23} components of the truncated matrix are changed in sign [cf. Ref. 36 (p. 7–8)]. Still, ${}_{O}\mathbf{X}$ is imaged onto the same sensor coordinates, as (exemplary given for ${}_{C}u$)

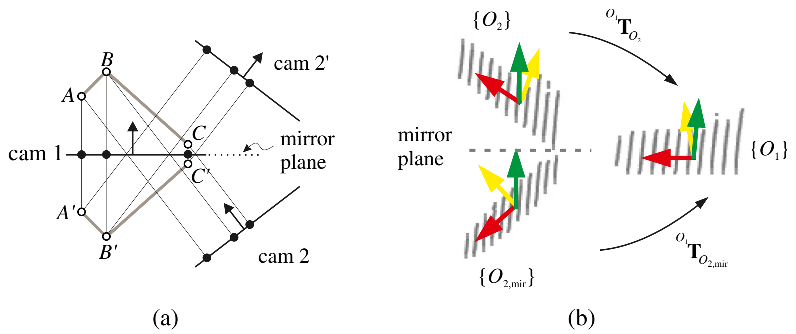


Fig. 3 Mirror ambiguity of affine projection. (a) Principle outline (based on Ref. 44). The optical axes are indicated by black arrows. (b) Transformations between mirrored and original point clouds for the calibration target.

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

$$ {}_{c}u = r_{11} \cdot {}_{O}X + r_{12} \cdot {}_{O}Y + r_{13} \cdot {}_{O}Z = r_{11} \cdot {}_{O}X + r_{12} \cdot {}_{O}Y + (-r_{13}) \cdot (-{}_{O}Z). \tag{21} $$

Therefore, two mathematically equal solutions exist (global minima; in the scope of this paper, the term global minimum stands for a solution with realistic camera intrinsics, but which potentially differs from the physically correct pose estimate due to mirror ambiguity. It is used in distinction to a local minimum, which corresponds to a solution with physically unrealistic intrinsic estimates.), when camera poses (in terms of truncated rigid body matrices ${}^{C}\tilde{\mathbf{T}}_{O_1,i}$) and the shape of the calibration target (in terms of ${}^{O_1}\mathbf{T}_{O_2}$) are estimated—one corresponds to the mirrored, the other to the nonmirrored solution.

A yaw–pitch–roll decomposition of ${}^{O_1}\mathbf{T}_{O_2}$ with rotation angles α, β, and γ can help to identify whether a mirrored scenario is present or not. In case of a mirrored scenario, the transformation is based on the mirrored coordinate system $\{O_{2,\mathrm{mir}}\}$ and not on the nonmirrored system $\{O_2\}$ [cf. Fig. 3(b)], resulting in a different yaw–pitch–roll decomposition: α and γ differ in sign.

In summary, in case of an erroneous, mirror-based start value determination, an elementwise sign correction is mandatory for ${}^{O_1}\mathbf{T}_{O_2}$ and ${}^{C}\tilde{\mathbf{T}}_{O_1,i}$, with help of corrective matrix $\mathbf{T}_{\mathrm{mir}}$

$$ \mathbf{T}_{\mathrm{mir}} = \begin{bmatrix} 1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 \\ -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{bmatrix}. \tag{22} $$

The elementwise sign correction is realized by the Hadamard product (symbol $\circ$) according to

$$ {}^{C}\tilde{\mathbf{T}}_{O_1,i} = {}^{C}\tilde{\mathbf{T}}_{O_1,\mathrm{mir},i} \circ \mathbf{T}_{\mathrm{mir},\,[\cancel{3,\mathrm{row}}]}, \tag{23} $$

$$ {}^{O_1}\mathbf{T}_{O_2} = {}^{O_1}\mathbf{T}_{O_2,\mathrm{mir}} \circ \mathbf{T}_{\mathrm{mir}}. \tag{24} $$

Additional information on the necessary matrix correction is given by Shimshoni et al.[32]

3.3 *Nonlinear Parameter Refinement*

Once the start parameters for both cameras are determined, a nonlinear refinement is executed based on a Levenberg–Marquardt optimization by minimizing

$$ e_{\mathrm{stereo}} = \sum_{i=1}^{m_{c_1}} \left[\sum_{j=1}^{n_{c_1}} \| {}_{c_1}\mathbf{u}_{ij} - {}_{c_1}\hat{\mathbf{u}}_{ij} \|^2 \right] + \sum_{i=1}^{m_{c_2}} \left[\sum_{j=1}^{n_{c_2}} \| {}_{c_2}\mathbf{u}_{ij} - {}_{c_2}\hat{\mathbf{u}}_{ij} \|^2 \right], \tag{25} $$

with

$$ {}_{c_1}\hat{\mathbf{u}}_{ij} = f_1[\mathbf{K}_1, \mathbf{k}_1, {}^{C_1}\tilde{\mathbf{T}}_{O_1,i}, \mathbf{X}_{O_1,j}({}^{O_1}\mathbf{T}_{O_2})], $$
$$ {}_{c_2}\hat{\mathbf{u}}_{ij} = f_2[\mathbf{K}_2, \mathbf{k}_2, {}^{C_2}\tilde{\mathbf{T}}_{O_1,i}, \mathbf{X}_{O_1,j}({}^{O_1}\mathbf{T}_{O_2})]. $$

To differentiate between the two stereo cameras, indexes c (and C) are extended to c_1 and c_2 (C_1 and C_2), respectively, whereas the other parameters are distinguished by indices 1 or 2 (e.g., $\mathbf{k}_1$ as the first camera's distortion coefficients). As the number of correspondences and of captured poses per camera might differ, camera-specific numbers are defined by n_{c_1} or n_{c_2} (correspondences) and m_{c_1} or m_{c_2} (poses), respectively. e_{stereo} is the sum of the squared geometric errors between the matched feature points ${}_{c_1}\mathbf{u}_{ij}$ (or ${}_{c_2}\mathbf{u}_{ij}$) and the corresponding projected points ${}_{c_1}\hat{\mathbf{u}}_{ij}$ (or ${}_{c_2}\hat{\mathbf{u}}_{ij}$) (based on the estimated model). The mean absolute projection error $e_{\mathrm{abs,mean}}$ is given in pixel and is defined in the camera sensor coordinate frames $\{c_1\}$ and $\{c_2\}$, respectively, and defined as (here given for the first camera)

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

$$e_{\text{abs,mean}} = \frac{\sum_{i=1}^{m_{c_1}} \sum_{j=1}^{n_{c_1}} \sqrt{\left({}_{c_1}u_{ij} - {}_{c_1}\hat{u}_{ij}\right)^2 + \left({}_{c_1}v_{ij} - {}_{c_1}\hat{v}_{ij}\right)^2}}{m_{c_1} \cdot n_{c_1}}. \tag{26}$$

The camera matrices $\mathbf{K}_1$ and $\mathbf{K}_2$ (three parameters per camera), the distortion vectors $\mathbf{k}_1$ and $\mathbf{k}_2$ (four parameters per camera), the truncated rigid body transformations ${}^{C_1}\tilde{\mathbf{T}}_{O_1,i}$ and ${}^{C_2}\tilde{\mathbf{T}}_{O_1,i}$ (five parameters per view and camera, the Rodrigues' formula is used to express the rotation), and the rigid body transformation ${}^{O_1}\mathbf{T}_{O_2}$ (six parameters, coupling the errors of camera one and two) are optimized simultaneously, resulting in a total number of $2 \cdot 3 + 2 \cdot 4 + 5 \cdot (m_{c_1} + m_{c_2}) + 6 = 20 + 10 \cdot m$ parameter, if $m = m_{c_1} = m_{c_2}$.

It should be noted, that a large difference between the camera pose number and/or marker number can result in an unequal weighting of the cameras' relevance in the optimization. Therefore, it is required that $m_{c_1} \approx m_{c_2}$ and $n_{c_1} \approx n_{c_2}$. Otherwise an appropriate error weighting approach should be introduced.

4 Experiment

In this section, an exemplary calibration result is presented. To this end, the hardware setup is introduced, along with the calibration target. The calibration result is analyzed with help of plausibility tests, comparing the estimated camera intrinsics and setup extrinsics to data sheet values and experimental boundary conditions.

Finally, the marker locations of the calibration target are triangulated based on the sensor calibration result.

4.1 *Hardware Setup: Sensor and Calibration Target*

The structured light sensor is shown in Fig. 4(a), comprising two monochromatic cameras (Allied Vision Manta G-895B POE) with telecentric lenses (Opto Engineering TCDP23C4MC096 with modified aperture) and a projector with entocentric lens (Wintech Pro4500 based on Texas Instrument's DLP LightCrafter 4500). The projector is only used as feature generator, not used in the calibration routine and is therefore not addressed in this section.

The telecentric lenses allow for the application of two cameras per lens, offering different magnification values. In the present scenario, the magnification $m = 0.093$ is used, theoretically offering an FOV of ∼152.54 mm by 80.72 mm, when used with a 1 in CMOS sensor with a resolution of 4112 pixel by 2176 pixel and a pixel size of 3.45 μm. The hardware configuration results in a pixel size on object side of ∼37 μm. The sensor is not completely illuminated, as the lens offers a smaller aperture. The lenses' DOF is ∼50 mm, the telecentric range is smaller (about 20 mm), and the working distance is 278.6 mm according to the data sheet. The triangulation angle is manually adjusted to ∼45 deg.

The calibration target is shown in Fig. 4(b). The target's basis is formed by a stiff cardboard structure, forming a roof. Two simple planar plastic tiles with circle pattern are fixed on the

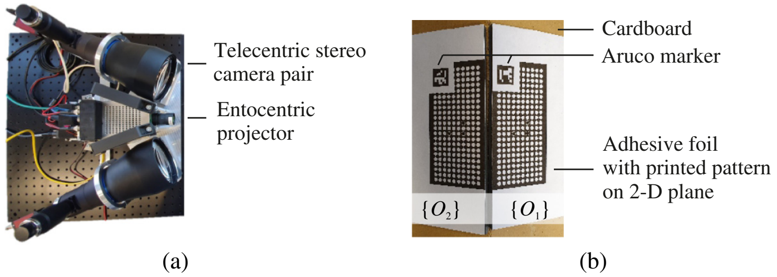

Telecentric stereo camera pair
Entocentric projector
(a)
Cardboard
Aruco marker
Adhesive foil with printed pattern on 2-D plane
$\{O_2\}$
$\{O_1\}$
(b)

Fig. 4 (a) Structured light sensor with telecentric stereo camera pair and entocentric projector as feature generator. (b) Experimental calibration target.

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

rooftop sides with double-faced adhesive tape. The target patterns are printed onto an adhesive foil on a standard ink-jet printer and are adhered to the tiles. The dot marker pitch is 3 mm and the diameter is 2.25 mm.

4.2 *Calibration Results*

The calibration target is captured in different poses (at least three poses per camera). It is not mandatory that both cameras acquire all images based on the exact same target poses as long as at least one image pair of the same pose exists. This image pair is necessary as it will be used to define the measurement coordinate system based on $\{O_1\}$. In the present scenario, $m_{c_1} = 11$ poses are captured for the first and $m_{c_2} = 13$ for the second camera. The marker number for camera one is $n_{c_1} = 282$ per pose, and for camera two $n_{c_2} = 281$ per pose. In consequence, an unequal error balancing due to a large difference in point correspondences can be excluded, but nevertheless should be checked by comparing the individual mean absolute projection error per camera. The first target pose is equal and captured by both cameras, being basis for the measurement coordinate system. The start values for the nonlinear refinement are determined for each camera independently.

4.2.1 *Scenario one: no start value correction*

In the first scenario, the necessity of a potential start value correction is not monitored. Hereby, the effect of erroneous start values on the nonlinear refinement is meant to be illustrated. The corresponding calibration result is given in Fig. 5. The start values are listed in the left column, the refinement result in the right column. For the sake of readability and brevity, only exemplary parameters are given.

${}^{O_1}\mathbf{T}_{O_2}$ is estimated independently for both cameras in the start value determination and should be ideally equal, as the target geometry is not changed in between the image acquisition for both cameras. A comparison of ${}^{O_1}\mathbf{T}_{O_2,1}$ and ${}^{O_1}\mathbf{T}_{O_2,2}$ shows a difference in sign [cf. to red (dot underline) and blue (wave underline) boxed values in Fig. 5]. It follows that ${}^{O_1}\mathbf{T}_{O_2,1} \approx {}^{O_1}\mathbf{T}_{O_2,2} \circ \mathbf{T}_{\mathrm{mir}}$, indicating that a mirrored point cloud either for the first or second camera was used to estimate the start values. (The approximately equal sign is used here, as a simple sign correction does only ideally result in the same matrices. Even in case of nonmirrored conditions, the different experimental data sets for both cameras result in slightly different matrix entries.) In the present scenario, the first camera's point cloud is mirrored, which can be concluded from a yaw–pitch–roll decomposition (cf. Sec. 3.2.5). The nonlinear refinement based on Eq. (25) requires the choice of a single ${}^{O_1}\mathbf{T}_{O_2}$—either ${}^{O_1}\mathbf{T}_{O_2,1}$ or ${}^{O_1}\mathbf{T}_{O_2,2}$. This leads to large deviations when starting the optimization, as either the

Parameter	Start parameters	Refined parameters
Camera matrix $\mathbf{K}_1$	$\begin{bmatrix} 27.063 & 0 & 2056 \\ 0 & 27.063 & 1088 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 27.062 & -0.0802 & 2056 \\ 0 & 27.061 & 1088 \\ 0 & 0 & 1 \end{bmatrix}$
Truncated rigid body transformation ${}^{C_1}\tilde{\mathbf{T}}_{O_1}$ for first target pose	$\begin{bmatrix} -0.7017 & 4.079 \times 10^{-2} & 0.7113 & 35.908 \\ -3.826 \times 10^{-5} & -0.9985 & 5.720 \times 10^{-2} & 4.917 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -0.7019 & 3.795 \times 10^{-2} & 0.711 & 35.915 \\ 1.672 \times 10^{-3} & -0.9985 & 5.492 \times 10^{-2} & 4.905 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Truncated rigid body transformation ${}^{C_2}\tilde{\mathbf{T}}_{O_1}$ for first target pose	$\begin{bmatrix} -1.0001 & 8.240 \times 10^{-3} & \boxed{\underline{1.723 \times 10^{-2}}} & 9.283 \\ -9.063 \times 10^{-3} & -0.9987 & \boxed{\underline{-5.340 \times 10^{-2}}} & 4.409 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -0.9998 & 6.3380 \times 10^{-3} & \boxed{\underline{-1.854 \times 10^{-2}}} & 9.287 \\ -7.268 \times 10^{-3} & -0.9987 & \boxed{\underline{5.055 \times 10^{-2}}} & 4.395 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Rigid body transformation ${}^{O_1}\mathbf{T}_{O_2,1}$	$\begin{bmatrix} 0.7339 & 5.567 \times 10^{-3} & \boxed{0.679 \times 10^{-3}} & 35.001 \\ -1.570 \times 10^{-3} & 0.99998 & \boxed{-6.499 \times 10^{-3}} & -7.792 \times 10^{-3} \\ \boxed{-0.679} & \boxed{3.703 \times 10^{-3}} & 0.734 & \boxed{-10.800} \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.7338 & 5.463 \times 10^{-3} & \boxed{0.679 \times 10^{-3}} & 35.002 \\ -1.004 \times 10^{-3} & 0.99998 & \boxed{-6.956 \times 10^{-3}} & 1.494 \times 10^{-2} \\ \boxed{-0.679} & \boxed{4.422 \times 10^{-3}} & 0.734 & \boxed{-10.811} \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Rigid body transformation ${}^{O_1}\mathbf{T}_{O_2,2}$	$\begin{bmatrix} 0.7341 & 5.864 \times 10^{-3} & \boxed{-0.679 \times 10^{-3}} & 34.978 \\ -1.826 \times 10^{-3} & 0.99998 & \boxed{6.663 \times 10^{-3}} & -7.218 \times 10^{-2} \\ \boxed{0.679} & \boxed{-3.651 \times 10^{-3}} & 0.7341 & \boxed{10.804} \\ 0 & 0 & 0 & 1 \end{bmatrix}$	see refined ${}^{O_1}\mathbf{T}_{O_2,1}$
errors $e_{abs,mean,1}$/pixel and $e_{abs,mean,2}$/pixel	–	0.274, 0.287

Fig. 5 Calibration result for exemplary parameters for scenario one. The start values for the first camera are estimated based on a mirrored point cloud and not corrected. ${}^{O_1}\mathbf{T}_{O_2,1}$ is used as start value for the stereo optimization.

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

${}^{C_1}\tilde{\mathbf{T}}_{O_1,i}$ or ${}^{C_2}\tilde{\mathbf{T}}_{O_1,i}$ matrices do not fit to the chosen calibration target shape in terms of ${}^{O_1}\mathbf{T}_{O_2}$. If ${}^{O_1}\mathbf{T}_{O_2,1}$ is selected (mirrored point cloud), the refinement results according to Fig. 5 are obtained. In this case, the nonlinear refinement starts with a mean projection error of 36.04 pixel.

A direct comparison of all start and refined values shows no great difference for the chosen parameters but reveals a change in sign for the truncated rigid body transformation ${}^{C_2}\tilde{\mathbf{T}}_{O_1}$ for the first target pose [cf. to black (solid underline) boxed values in Fig. 5]. This change in sign happens also for all other target poses and matrices ${}^{C_2}\tilde{\mathbf{T}}_{O_1,i}$, whereas the erroneously chosen start value ${}^{O_1}\mathbf{T}_{O_2,1}$ does only slightly change in the optimization procedure and the critical signs do not change at all [cf. to red (dot underline) and green (dash underline) boxed values in Fig. 5]. The only way to reduce the projection error (if ${}^{O_1}\mathbf{T}_{O_2,1}$ is not changing) is therefore the adaption of the second camera's locations in relation to the target, now in conformance with the mirrored point cloud. This is why the truncated matrices ${}^{C_2}\tilde{\mathbf{T}}_{O_1,i}$ change in sign [cf. Sec. 3.2.5, Eq. (20)].

Still, the resulting projection errors $e_{\text{abs,mean},1}$ and $e_{\text{abs,mean},2}$ are with about 0.28 pixel low (corresponds to $\sim 10\ \mu$m on object side). Also, the similar error results for both cameras indicate a balanced weighting of the cameras' relevance in the nonlinear optimization. An analysis of the error histogram (not shown here) indicates a good model fitting and allows the conclusion that the optimizer converged into the desired minimum. This assumption is further supported as also the estimated lens magnification is consistent with the data sheet value of 0.093. The calibrated magnification can be obtained with help of the scaling factor s and the pixel size s_x, resulting in $m = s \cdot s_x = 27.063\ \frac{\text{px}}{\text{mm}} \cdot 0.00345\ \frac{\text{mm}}{\text{px}} = 0.09336$.

In this scenario, the minimum with erroneous signs is estimated and the mirror ambiguity problem not resolved (according to the results in Fig. 5). (According to Sec. 3.2.5, there are two mathematically equivalent minimum: one for a mirrored and one for the correct, nonmirrored sensor arrangement, just differing in signs.) The r_{13} and r_{23} components of ${}^{C_2}\tilde{\mathbf{T}}_{O_1}$ change during parameter refinement (from mirrored to nonmirrored), resulting in a erroneous camera pose estimation [as outlined in Fig. 3(a) (cam 2′, instead 2)]. This is due to the choice of the mirrored target point cloud as start value in terms of ${}^{O_1}\mathbf{T}_{O_2,1}$ and the absence of a sign change during refinement (cf. to ${}^{O_1}\mathbf{T}_{O_2,1}$ in Fig. 5: r_{13}, r_{23}, r_{31}, r_{32}, t_z do not change in sign). In consequence, a mirrored sensor arrangement is estimated, and all following measurements will result in mirrored point clouds.

If ${}^{O_1}\mathbf{T}_{O_2,2}$ is chosen as start value (describing a nonmirrored calibration point cloud), the initial mean projection error is higher (111.70 pixel). The optimizer is not converging toward a global minimum, and the routine is aborted with a mean absolute projection error of about 13 pixel. This was not to be expected, as also in this case a sign adaption (in this case, for the matrices ${}^{C_i}\tilde{\mathbf{T}}_{O_1,i}$) would result in a global minimum; this time even for the accurate sensor arrangement. The result indicates a basic problem when ignoring affine mirror ambiguity. The necessary sign adaption is not always successful.

4.2.2 *Scenario two: start value correction*

In the second scenario, $\mathbf{T}_{\text{mir}}$ is used to correct the start values of the first camera. The corresponding calibration result is given in Fig. 6 and extended by the distortion parameters for the first camera. The result is obtained when using the corrected ${}^{O_1}\mathbf{T}_{O_2,1}$ as start value, resulting in an initial mean projection error of 0.75 pixel, around 35 pixel lower than in the uncorrected scenario.

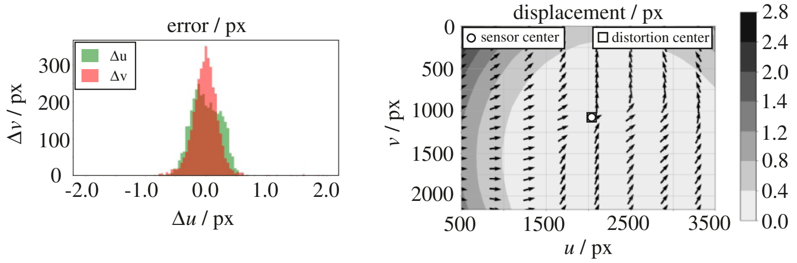
Now not only the absolute values of ${}^{O_1}\mathbf{T}_{O_2,1}$ and ${}^{O_1}\mathbf{T}_{O_2,2}$ are similar but also possess the same signs [cf. to red (dotted underline), blue (wavy underline), and green (dashed underline) boxed values in Fig. 6] even after refinement. (Just the sign of the y value is changing but is very close to zero. This change is not connected to the mirror effect). The same applies to the truncated matrices ${}^{C_1}\tilde{\mathbf{T}}_{O_1,i}$ and ${}^{C_2}\tilde{\mathbf{T}}_{O_1,i}$; the critical signs do not change in the refinement procedure, meaning that the start values of both cameras have been successfully combined.

The error histogram for the second camera is shown in Fig. 7(a). The error is approximately normally distributed for the v direction, whereas the u direction deviates from a normal

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Parameter	Start parameters	Refined parameters
Camera matrix $\mathbf{K}_1$	$\begin{bmatrix} 27.063 & 0 & 2056 \\ 0 & 27.063 & 1088 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 27.062 & -0.0802 & 2056 \\ 0 & 27.061 & 1088 \\ 0 & 0 & 1 \end{bmatrix}$
Distortion coefficients $\mathbf{k}_1$	$\begin{bmatrix} k_1 \\ k_2 \\ p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 0.0 \\ 0.0 \\ 0.0 \\ 0.0 \end{bmatrix}$	$\begin{bmatrix} k_1 \\ k_2 \\ p_1 \\ p_2 \end{bmatrix} = \begin{bmatrix} 9.738 \times 10^{-8} \\ 9.703 \times 10^{-12} \\ -3.097 \times 10^{-6} \\ -2.010 \times 10^{-6} \end{bmatrix}$
Truncated rigid body transformation ${}^{C_1}\tilde{\mathbf{T}}_{O_1}$ for first target pose	$\begin{bmatrix} -0.7017 & 4.079 \times 10^{-2} & -0.7113 & 35.908 \\ -3.826 \times 10^{-5} & -0.9985 & -5.720 \times 10^{-2} & 4.917 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} -0.7019 & 3.795 \times 10^{-2} & -0.711 & 35.915 \\ 1.672 \times 10^{-3} & -0.9985 & -5.492 \times 10^{-2} & 4.905 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Truncated rigid body transformation ${}^{C_2}\tilde{\mathbf{T}}_{O_1}$ for first target pose	$\begin{bmatrix} \boxed{\underline{-1.0001}} & 8.240 \times 10^{-3} & 1.723 \times 10^{-2} & 9.283 \\ -9.063 \times 10^{-3} & -0.9987 & -5.340 \times 10^{-2} & 4.409 \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} \boxed{\underline{-0.9998}} & 6.3380 \times 10^{-3} & 1.854 \times 10^{-2} & 9.287 \\ -7.268 \times 10^{-3} & -0.9987 & -5.055 \times 10^{-2} & 4.395 \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Rigid body transformation ${}^{O_1}\mathbf{T}_{O_2,1}$	$\begin{bmatrix} 0.7339 & 5.567 \times 10^{-3} & \boxed{-0.679 \times 10^{-3}} & 35.001 \\ -1.570 \times 10^{-3} & 0.99998 & \boxed{6.499 \times 10^{-3}} & -7.792 \times 10^{-3} \\ \boxed{0.679} & \boxed{-3.703 \times 10^{-3}} & 0.734 & \boxed{10.800} \\ 0 & 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0.7338 & 5.463 \times 10^{-3} & \boxed{-0.679 \times 10^{-3}} & 35.002 \\ -1.004 \times 10^{-3} & 0.99998 & \boxed{6.956 \times 10^{-3}} & 1.494 \times 10^{-2} \\ \boxed{0.679} & \boxed{-4.422 \times 10^{-3}} & 0.734 & \boxed{10.811} \\ 0 & 0 & 0 & 1 \end{bmatrix}$
Rigid body transformation ${}^{O_1}\mathbf{T}_{O_2,2}$	$\begin{bmatrix} 0.7341 & 5.864 \times 10^{-3} & \boxed{-0.679 \times 10^{-3}} & 34.978 \\ -1.826 \times 10^{-3} & 0.99998 & \boxed{6.663 \times 10^{-3}} & -7.218 \times 10^{-2} \\ \boxed{0.679} & \boxed{-3.651 \times 10^{-3}} & 0.7341 & \boxed{10.804} \\ 0 & 0 & 0 & 1 \end{bmatrix}$	see refined ${}^{O_1}\mathbf{T}_{O_2,1}$
errors $e_{abs,mean,1}$/pixel and $e_{abs,mean,2}$/pixel	–	0.274, 0.287

Fig. 6 Calibration result for exemplary parameters for scenario two. The start values for the first camera are estimated based on a mirrored point cloud but are corrected by $\mathbf{T}_{\text{mir}}$. ${}^{O_1}\mathbf{T}_{O_2,1}$ is used as start value for the stereo optimization.



(a) (b)

Fig. 7 (a) Error histogram for the second camera. (b) Distortion model for the first camera: only the illuminated sensor area is depicted. The corresponding distortion coefficients $\mathbf{k}_1$ are given in Fig. 6.

distribution. The reason for this cannot be assessed conclusively but could be due to a slightly biased target pose distribution.

Altogether, error distribution and mean absolute projection errors are equal to the previous scenario without correction, which is comprehensible due to mathematical equivalence of mirrored and nonmirrored solution. This mathematical equivalence should not be confused with a physical equivalence. The determined parameters in the noncorrected scenario are false in sign.

In addition, the lens distortion for the first camera is shown in Fig. 7(b). The telecentric lens does not allow for a complete sensor illumination, resulting in masked areas near the right and left sensor boundaries. This is why the displayed sensor area is reduced by about 500 pixel from the sides. Altogether, the lens distortion is relatively low, as the distortion model introduces a pixel correction distinctly below 0.4 pixel for the greater part of the sensor. Even lower corrective effect is introduced by the distortion model for the first lens, confirming the assumption of low distortion generated by high quality telecentric lenses.

Noteworthy is also the matrix entry (1,1) of ${}^{C_2}\tilde{\mathbf{R}}_{O_1,1}$ [as part of the start value of ${}^{C_2}\tilde{\mathbf{T}}_{O_1,1}$, indicated by single black (solid underline) box], as it results in a deviation to an orthonormal basis. This might be due to numerical inaccuracies when performing the Euclidean upgrade but apparently had no effect on the parameter refinement in the next step, as the refined matrix represents an orthonormal basis.

The plausibility of the optimized rigid body transformation ${}^{O_1}\mathbf{T}_{O_2}$ (here in terms of ${}^{O_1}\mathbf{T}_{O_2,1}$) is evaluated by analyzing the angles between the axes of the two coordinate systems $\{O_1\}$ and $\{O_2\}$. As the second orthonormal basis of the refined rotation matrix ${}^{O_1}\mathbf{R}_{O_2,1}$ (as part of ${}^{O_1}\mathbf{T}_{O_2,1}$) is nearly $(5.463 \times 10^{-3}, 0.99998, -4.422 \times 10^{-3})^T \approx (0,1,0)^T$, the angle between the y axis of

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

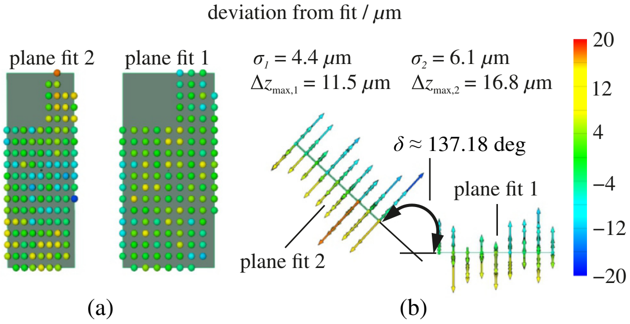


Fig. 8 Triangulated 3-D coordinates of calibration target in coordinate frame $\{O_1\}$ of first target pose: (a) lateral view, *xy* plane and (b) top view, *xz* plane.

$\{O_1\}$ and $\{O_2\}$ is approximately zero. This is in good agreement with the obvious orientation of the two planes in Fig. 4(b), in which the y axes are nearly in parallel. The angles between both x and z axes are approximately equal and about 42.80 deg (obtained by scalar product between corresponding orthonormal bases). This is again in agreement with the previous result, as the rotation in order to transfer $\{O_1\}$ to $\{O_2\}$ is performed around the y axis.

The triangulation angle between the cameras has been manually set to 45 deg and is validated by calculating the angle between the sensors' normal axes based on the calibration result. To this end, the third orthonormal bases of ${}^{C_1}\tilde{\mathbf{T}}_{O_2}$ and ${}^{C_2}\tilde{\mathbf{T}}_{O_2}$ are calculated. The procedure is exemplary given for the first camera. The cross product of ${}_{C_1}\mathbf{r}_1$ and ${}_{C_1}\mathbf{r}_2$ yields ${}_{C_1}\mathbf{r}_3$. The triangulation angle θ is calculated based on the scalar product of ${}_{C_1}\mathbf{r}_3$ and ${}_{C_2}\mathbf{r}_3$, resulting in an angle of 46.425 deg. This is in good agreement with the roughly measured value of 45 deg.

Altogether, postulated model and experiment are in good agreement. If ${}^{O_1}\mathbf{T}_{O_2,2}$ of the second camera is used as start value, the initial mean projection error is even a bit lower (0.72 pixel).

4.3 *Plausibility Test: Triangulation Result*

The 3-D coordinates of the calibration target are triangulated in coordinate system $\{O_1\}$ of the first target pose. An analysis of the camera specific projection errors for this measurement pose helps to eliminate the possibility, that it deviates strongly from the mean errors and represents an outlier pose. As this is not the case, it is used for triangulation. The point correspondences (in pixel) are used to reconstruct the target points according to (e.g., cf. Ref. 6):

$$\begin{bmatrix} {}^{c_1}\mathbf{M}_{O_1} \\ {}^{c_2}\mathbf{M}_{O_1} \end{bmatrix} {}_{O_1}\mathbf{X} = \begin{bmatrix} {}_{c_1}\mathbf{u} - {}_{c_1}\mathbf{p} \\ {}_{c_2}\mathbf{u} - {}_{c_2}\mathbf{p} \end{bmatrix}. \tag{27}$$

The affine projection matrices—for the image pair defining the measurement system—are obtained by combining the camera matrices with the truncated rigid body transformations according to Eq. (3). ${}_{c_1}\mathbf{u}$ and ${}_{c_2}\mathbf{u}$ are the undistorted pixel correspondences of both cameras. ${}_{O_1}\mathbf{X}$ is calculated by the least squares method, as Eq. (27) is overdetermined. The triangulation result is given in Fig. 8. For each plane, the standard deviation σ and the maximum deviation $\Delta z_{\max}$ are given based on an individual plane fitting. The result implies a satisfactory planarity.

Furthermore, the rooftop angle $\delta = 137.18$ deg is depicted, obtained by the angle between the two plane fits, resulting in an angle of $180\ \text{deg} - 137.18\ \text{deg} = 42.82\ \text{deg}$ between the planes' normal vectors. This is in accordance with the previous angle analysis, where an angle of 42.80 deg was calculated between the planes' z axes.

5 Conclusion

In this paper, a robust and direct calibration routine for a structured light sensor with telecentric stereo camera pair is proposed. The routine combines an affine autocalibration approach with a nonlinear parameter refinement based on a Levenberg–Marquardt optimization. The used low-

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

cost 3-D calibration target combines two 2-D planes with metric distance information and makes an additional camera magnification determination dispensable. This reduces the calibration effort. The problem of affine mirror ambiguity is theoretically addressed and solved by analyzing the rigid body transformation between the two 2-D target planes and by introducing a correction matrix. Moreover, radial-tangential lens distortion is considered to allow for a more accurate camera model. A representative data base for optimization is provided by acquiring individual target poses for each camera.

Provided a nondegenerate and sufficient number of target poses is acquired for each camera (here at least three, with at least one image pair defining the measurement coordinate frame), the following general conclusions can be derived: if the start value determination is coincidentally based on a mirrored calibration point cloud for both cameras, the nonlinear optimization will converge robustly, but based on a mirrored sensor arrangement, resulting in mirrored triangulated point clouds. If the start parameters for only one camera are affected by mirror ambiguity, the subsequent nonlinear optimization not necessarily converges (cf. Sec. 4.2.1), as the outcome depends on the selected start value for ${}^{O_1}\mathbf{T}_{O_2}$.

The monitoring of ${}^{O_1}\mathbf{T}_{O_2}$ via yaw–pitch–roll decomposition allows for the detection of a potential point cloud mirroring. The correction of affected start values by the introduced matrix $\mathbf{T}_{\mathrm{mir}}$ guarantees a rapid optimization convergence, independently of the choice of ${}^{O_1}\mathbf{T}_{O_2}$. Moreover, the initial projection error is smaller. In consequence, the triangulated results are always defined accurately and not mirrored. The obtained experimental results verify the effectiveness of the proposed approach.

In the present version of the calibration approach, due to the start value determination by the factorization algorithm, the detected target features must be visible in all views of a single camera. A higher degree of flexibility could be achieved using an affine reconstruction approach, which does not depend on this constraint (cf. Sec. 1.2). Especially, the estimation of the lens distortion parameters could benefit from a higher number of sensor boundary points. In the present routine, such points are more likely to be excluded from affine reconstruction, due to limited visibility. Another approach to provide a wider data basis could be achieved by the re-usage of former excluded points for nonlinear optimization, as the visibility constraint does not apply here.

Moreover, the introduction of an affine analogy to the ideal image plane for perspective cameras could potentially increase numerical stability, when optimizing the distortion parameters of the lenses (cf. Sec. 2). This could become more important, if higher-order distortion coefficients are meant to be introduced.

Also, the practicality of the suggested hardware setup is limited to measurement scenarios, in which the required measurement volume is relatively small. This is due to the camera's restricted DOF, and telecentricity range, resulting in a small cross section in which an object point is in focus, and sharply displayed on both affine sensors. A potential solution is the application of telecentric lenses with Scheimpflug adapters (e.g., Opto Engineering TCSM096 or a comparable product of a different manufacturer).

The telecentricity range for which an object is mapped with constant magnification onto a sensor is smaller than the DOF. In order to use the complete DOF, it could be interesting to introduce slightly different magnification values (and in consequence camera matrices), depending on the distance from object to lens. To this end, an accurate estimate of lens magnification ratios for the target poses in different distances would be needed. This could be achieved by introducing other metric constraints for the Euclidean upgrade, based on the so-called scaled-orthographic model (e.g., as given in Ref. 41, p. 217), instead of the orthographic model. The introduction of additional parameters could affect the stability of the nonlinear optimization routine, which therefore needs to be analyzed. Also, point data triangulation would become more costly, as a first rough point cloud reconstruction would be required, in order to judge which magnification value to use in a second, more accurate triangulation step.

A final remark on the potential of the scaled-orthographic model: The model allows for an image dependent modeling of scaling. It is therefore thinkable to apply the start value determination via factorization algorithm on the complete pose data set captured by both cameras, and still obtain camera specific start values for magnification. The advantage would be a fitting start

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

value data set for nonlinear optimization, as it would either depend on a mirrored or nonmirrored point cloud. Still, a check for affine mirror ambiguity and potential correction would be necessary in order to avoid the optimization of an inverse sensor setup. Also, if more than one stereo image is captured, the errors of camera one and camera two could be further coupled for these specific poses, as in this case the rigid body transformation between the cameras is constant (defining the stereo rig). Hereby, the advantages of the stereo image based approach by Liu et al.[12] could be combined with the presented method.

Acknowledgments

We want to thank the Deutsche Forschungsgemeinschaft (DFG) for funding subproject C5 Multiscale Geometry Inspection of Joining Zones as part of the Collaborative Research Centre (CRC) 1153 Process chain to produce hybrid high performance components by Tailored Forming (252662854). Also, we would like to thank Mr. Töberg for the valuable discussions on affine cameras and autocalibration. The authors declare no conflicts of interest.

References

1. M. Rahlves and J. Seewig, *Optisches Messen technischer Oberflächen*, Messprinzipien und Begriffe, Beuth Verlag (2009).
2. S. Zhang, "High-speed 3D shape measurement with structured light methods: a review," *Opt. Lasers Eng.* **106**, 119–131 (2018).
3. S. V. der Jeught and J. J. Dirckx, "Real-time structured light profilometry: a review," *Opt. Lasers Eng.* **87**, 18–31 (2016).
4. K. Chen et al., "Microscopic three-dimensional measurement based on telecentric stereo and speckle projection methods," *Sensors* **18**, 3882 (2018).
5. Y. Hu et al., "A new microscopic telecentric stereo vision system—calibration, rectification, and three-dimensional reconstruction," *Opt. Lasers Eng.* **113**, 14–22 (2019).
6. Z. Chen, H. Liao, and X. Zhang, "Telecentric stereo micro-vision system: calibration method and experiments," *Opt. Lasers Eng.* **57**, 82–92 (2014).
7. H. Liu et al., "Epipolar rectification method for a stereovision system with telecentric cameras," *Opt. Lasers Eng.* **83**, 99–105 (2016).
8. K. Haskamp, M. Kästner, and E. Reithmeier, "Accurate calibration of a fringe projection system by considering telecentricity," *Proc. SPIE* **8082**, 80821B (2011).
9. B. Li and S. Zhang, "Flexible calibration method for microscopic structured light system using telecentric lens," *Opt. Express* **23**, 25795–25803 (2015).
10. L. Rao et al., "Flexible calibration method for telecentric fringe projection profilometry systems," *Opt. Express* **24**, 1222–1237 (2016).
11. D. Li, C. Liu, and J. Tian, "Telecentric 3D profilometry based on phase-shifting fringe projection," *Opt. Express* **22**, 31826–31835 (2014).
12. H. Liu, H. Lin, and L. Yao, "Calibration method for projector-camera-based telecentric fringe projection profilometry system," *Opt. Express* **25**, 31492–31508 (2017).
13. Q. Mei et al., "Structure light telecentric stereoscopic vision 3D measurement system based on Scheimpflug condition," *Opt. Lasers Eng.* **86**, 83–91 (2016).
14. J. Peng et al., "Distortion correction for microscopic fringe projection system with Scheimpflug telecentric lens," *Appl. Opt.* **54**, 10055–10062 (2015).
15. R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed., Cambridge University Press, Cambridge (2004).
16. Z. Zhang et al., "A simple, flexible and automatic 3D calibration method for a phase calculation-based fringe projection imaging system," *Opt. Express* **21**, 12218–12227 (2013).
17. D. Lanman, D. Hauagge, and G. Taubin, "Shape from depth discontinuities under orthographic projection," in *IEEE 12th Int. Conf. Comput. Vision Workshops*, pp. 1550–1557 (2009).
18. Z. Zhang, "Flexible camera calibration by viewing a plane from unknown orientations," in *Proc. Seventh IEEE Int. Conf. Comput. Vision*, IEEE (1999).

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

19. H. Liao, Z. Chen, and X. Zhang, "Calibration of camera with small FOV and DOF telecentric lens," in *IEEE Int. Conf. Rob. and Biomim.*, pp. 498–503 (2013).
20. D. Li and J. Tian, "An accurate calibration method for a camera with telecentric lenses," *Opt. Lasers Eng.* **51**, 538–541 (2013).
21. L. Yao and H. Liu, "A flexible calibration approach for cameras with double-sided telecentric lenses," *Int. J. Adv. Rob. Syst.* **13**(3), 82 (2016).
22. Y. Hu et al., "Calibration of telecentric cameras with distortion center estimation," *Proc. SPIE* **10827**, 1082720 (2018).
23. S. Zhang and P. S. Huang, "Novel method for structured light system calibration," *Opt. Eng.* **45**(8), 083601 (2006).
24. O. D. Faugeras, Q. T. Luong, and S. J. Maybank, "Camera self-calibration: theory and experiments," *Lect. Notes Comput. Sci.* **588**, 321–334 (1992).
25. J. J. Koenderink and A. J. van Doorn, "Affine structure from motion," *J. Opt. Soc. Am. A* **8**, 377–385 (1991).
26. C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: a factorization method," *Int. J. Comput. Vision* **9**, 137–154 (1992).
27. L. Quan, "Self-calibration of an affine camera from multiple views," *Int. J. Comput. Vision* **19**, 93–105 (1996).
28. S. S. Brandt, "Conditional solutions for the affine reconstruction of N-views," *Image Vision Comput.* **23**(7), 619–630 (2005).
29. S. S. Brandt and K. Palander, "A Bayesian approach for affine auto-calibration," *Lect. Notes Comput. Sci.* **3540**, 577–587 (2005).
30. N. Guilbert, A. Bartoli, and A. Heyden, "Affine approximation for direct batch recovery of Euclidian structure and motion from sparse data," *Int. J. Comput. Vision* **69**, 317–333 (2006).
31. R. Horaud, S. Christy, and R. Mohr, "Euclidean reconstruction and affine camera calibration using controlled robot motions," in *Proc. IEEE/RSJ Int. Conf. Intell. Rob. and Syst. Innovative Rob. Real-World Appl.*, Vol. 3, pp. 1575–1582 (1997).
32. I. Shimshoni, R. Basri, and E. Rivlin, "A geometric interpretation of weak-perspective motion," *IEEE Trans. Pattern Anal. Mach. Intell.* **21**, 252–257 (1999).
33. K. Kanatani, Y. Sugaya, and Y. Kanazawa, *Guide to 3D Vision Computation*, Springer International Publishing, Cham, Switzerland (2016).
34. T. Collins and A. Bartoli, "Planar structure-from-motion with affine camera models: closed-form solutions, ambiguities and degeneracy analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1237–1255 (2017).
35. S. Ullman, "The interpretation of structure from motion," *Proc. R. Soc. London Ser. B* **203**(1153), 405–426 (1979).
36. M. Han and T. Kanade, *Perspective Factorization Methods for Euclidean Reconstruction*, Carnegie Mellon University, The Robotics Institute (2000).
37. S. Garrido-Jurado et al., "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognit.* **47**(6), 2280–2292 (2014).
38. D. C. Brown, "Decentering distortion of lenses," *Photogramm. Eng. Remote Sens.* **23**(3), 444–462 (1966).
39. C. B. Duane, "Close-range camera calibration," *Photogramm. Eng.* **37**(8), 855–866 (1971).
40. J. G. Fryer and D. C. Brown, "Lens distortion for close-range photogrammetry," *Photogramm. Eng. Remote Sens.* **52**(1), 51–58 (1986).
41. C. Poelman and T. Kanade, "A paraperspective factorization method for shape and motion recovery," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 206–218 (1997).
42. Z. Zhang and G. Xu, "A unified theory of uncalibrated stereo for both perspective and affine cameras," *J. Math. Imaging Vision* **9**, 213–229 (1998).
43. D. Eggert, A. Lorusso, and R. Fisher, "Estimating 3-D rigid body transformations: a comparison of four major algorithms," *Mach. Vision Appl.* **9**, 272–290 (1997).
44. K. E. Ozden, K. Schindler, and L. V. Gool, "Multibody structure-from-motion in practice," *IEEE Trans. Pattern Anal. Mach. Intell.* **32**, 1134–1141 (2010).

Rüdiger Beermann is a research associate at the Institute of Measurement and Automatic Control at the Leibniz Universität Hannover. He received his diploma in mechanical engineering

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

from Leibniz Universität Hannover in 2013, and his state examination as a teacher for math and metal technology for vocational schools in 2015. His current research interests include the development of fringe projection systems for high temperature workpieces and thermo-optical simulations.

Lorenz Quentin is a research associate at the Institute of Measurement and Automatic Control at the Leibniz Universität Hannover. He obtained his diploma in mechanical engineering in 2016. His current research interests include the development of fringe projection systems for high temperature workpieces.

Markus Kästner is the head of the Production Metrology research group at the Institute of Measurement and Automatic Control at the Leibniz Universität Hannover. He received his PhD in mechanical engineering in 2008 and his postdoctoral lecturing qualifications in 2016 from the Leibniz Universität Hannover. His current research interests are optical metrology from macro- to nanoscale and optical simulations.

Eduard Reithmeier is a professor at the Leibniz Universität Hannover and head of the Institute of Measurement and Automatic Control. He received his diplomas in mechanical engineering and in math in 1983 and 1985, respectively, and his doctorate degree in mechanical engineering from the Technische Universität München in 1989. His research focuses on system theory and control engineering.

Downloaded From: https://www.spiedigitallibrary.org/journals/Optical-Engineering on 18 Dec 2020
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use