

Calibration of multiple cameras for large-scale experiments using a freely moving calibration target

Muller, K.; Hemelrijk, C. K.; Westerweel, J.; Tam, D. S.W.

DOI

[10.1007/s00348-019-2833-z](https://doi.org/10.1007/s00348-019-2833-z)

Publication date

2020

Document Version

Final published version

Published in

Experiments in Fluids

Citation (APA)

Muller, K., Hemelrijk, C. K., Westerweel, J., & Tam, D. S. W. (2020). Calibration of multiple cameras for large-scale experiments using a freely moving calibration target. *Experiments in Fluids*, 61(1), Article 7. <https://doi.org/10.1007/s00348-019-2833-z>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.



Calibration of multiple cameras for large-scale experiments using a freely moving calibration target

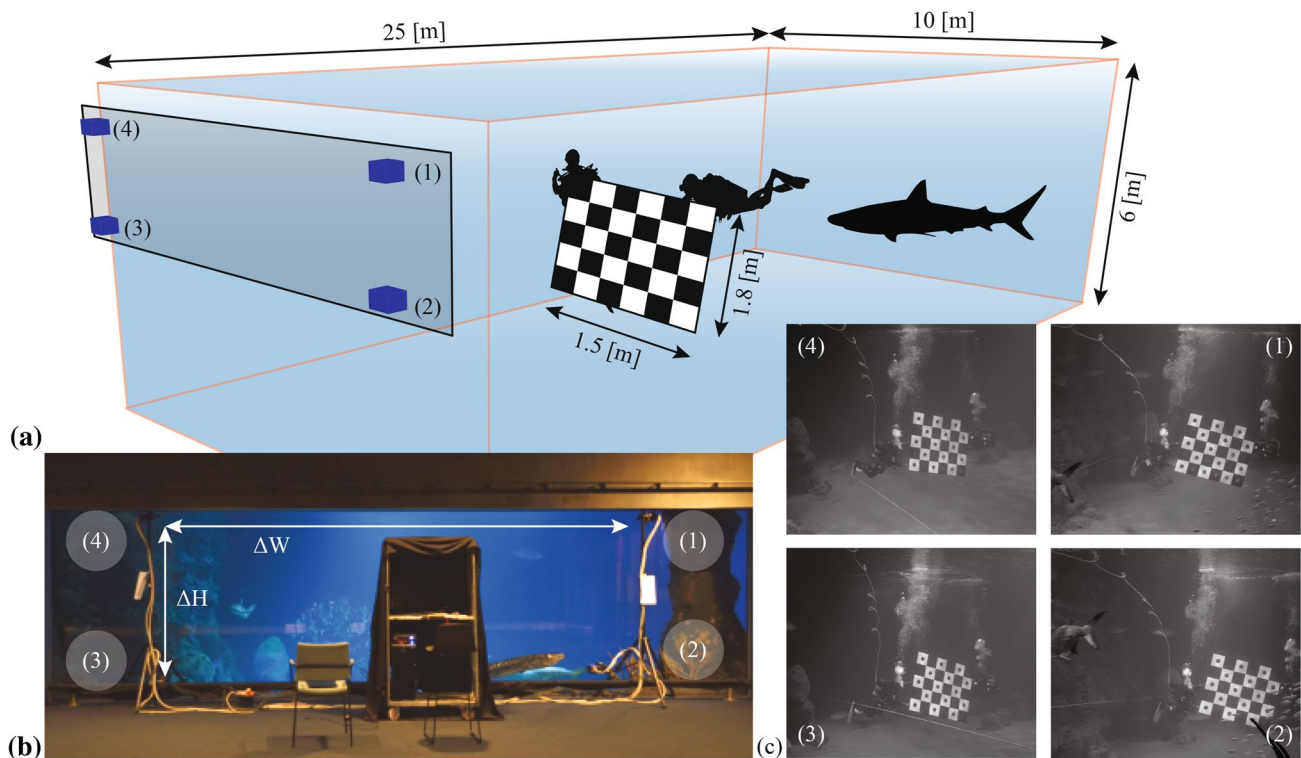
K. Muller¹ · C. K. Hemelrijk² · J. Westerweel¹ · D. S. W. Tam¹

Received: 22 August 2019 / Revised: 18 October 2019 / Accepted: 22 October 2019
© The Author(s) 2019

Abstract

Obtaining accurate experimental data from Lagrangian tracking and tomographic velocimetry requires an accurate camera calibration consistent over multiple views. Established calibration procedures are often challenging to implement when the length scale of the measurement volume exceeds that of a typical laboratory experiment. Here, we combine tools developed in computer vision and non-linear camera mappings used in experimental fluid mechanics, to successfully calibrate a four-camera setup that is imaging inside a large tank of dimensions $\sim 10 \times 25 \times 6 \text{ m}^3$. The calibration procedure uses a planar checkerboard that is arbitrarily positioned at unknown locations and orientations. The method can be applied to any number of cameras. The parameters of the calibration yields direct estimates of the positions and orientations of the four cameras as well as the focal lengths of the lenses. These parameters are used to assess the quality of the calibration. The calibration allows us to perform accurate and consistent linear ray-tracing, which we use to triangulate and track fish inside the large tank. An open-source implementation of the calibration in MATLAB is available.

Graphic abstract



Extended author information available on the last page of the article

1 Introduction

New studies in biophysics and fluid mechanics require the quantitative imaging of large-scale field experiments. Such studies include the large-scale Lagrangian tracking of bats and bird flocks (Theriault et al. 2014; Attanasi et al. 2015), super-large-scale particle image velocimetry measurements using natural snowfall (Toloui et al. 2014) and recent advancements in tomographic-PIV (Jux et al. 2018).

Obtaining reliable two- and three-dimensional imaging data in these large-field experiments is challenging and requires a camera calibration that is accurate down to the smallest physical length scale of interest. Non-linear polynomial camera mappings (Soloff et al. 1997; Wienieke 2005, 2008) are often used in laboratory experiments (Kühn et al. 2010), but their application at length scales beyond that of the laboratory is, in practice, limited. First, the size of the calibration target is limited, such that it only covers a small portion of the measurement. Second, the absence of conventional laboratory equipment, providing access to the measurement volume, does not support the accurate spatial positioning of the target, required in conventional calibration procedures.

In the present work, we combine the pinhole camera model (Tsai 1987) with non-linear polynomial camera mappings used in experimental fluid mechanics (Soloff et al. 1997) to perform a multiple camera calibration over a large-scale measurement volume inside the tank of the aquarium located in the Rotterdam zoo. Our method integrates the use of the pinhole camera model with a non-linear camera mapping to correct for optical distortion across refractive interfaces (Belden 2013). Our approach uses the framework of projective geometry in computer vision (Hartley and Zisserman 2004) and apply advanced self-calibration techniques (Svoboda et al. 2005; Shen et al. 2008).

Here we apply the planar checkerboard calibration technique by Zhang (2000), see also Zhang (1998, 1999), Sturm and Maybank (1999), Menudet et al. (2008) and Bouguet (2015). This approach eliminates the need to accurately position the calibration target, as required in conventional calibration procedures. Instead, the checkerboard calibration target is moved to arbitrary and unknown positions and orientations, here with the help of a team of divers. Second, by sequentially acquiring multiple calibration images while freely moving the calibration target, we achieve a camera calibration that spans over length scales much larger than the calibration target itself. This approach yields an accurate calibration over the measurement volume with a characteristic length scale on the order of several tens of meters.

We process the camera calibration in steps (Heikkilä and Silvén 1997). First, we correct for optical distortions

(Fryer and Brown 1986) by rectifying the curved lines of the checkerboard images (Prescott and McLean 1997; Devernay and Faugeras 2001). Second, we perform a calibration based on a single view for each camera following (Zhang 2000). Finally, we combine the single views and find the positions and orientations of the cameras over multiple views (Geiger et al. 2012) and optimize the calibration for spatial accuracy and consistency between the different views.

The camera calibration yields accurate results. To assess the validity of the camera calibration, we compare the estimated effective focal length obtained from it against the true value of the focal length of the lenses. We quantify the spatial accuracy of the camera calibration, by computing the skewness of optical rays associated with the multiple views. Our calibration allows us to use linear ray-tracing (Hedrick 2008) to track and triangulate multiple fish swimming over the entire visual depth of the tank. The method is versatile and can be implemented in field experiments over large length scales and for measurement volumes that are challenging to access experimentally.

2 Camera setup and calibration procedure

We image inside the large tank of the aquarium in the Rotterdam zoo in a measurement volume of dimensions $\sim 10 \times 25 \times 6 \text{ m}^3$ (Fig. 1). We use four 5.5 megapixel sCMOS cameras with wide-angle lenses of focal length $f_{\text{lens}} = 24 \text{ mm}$ (NIKKOR AF-24 mm) that cause significant variations in magnification over the large depth-of-field of $\text{DOF} \approx 25 \text{ m}$ (Adrian and Westerweel 2011).

The camera setup is positioned behind an acrylic window of thickness $\sim 50 \text{ cm}$. Optical distortions in the image plane are due to refraction at the water ($n_{\text{water}} = 1.363$) /acrylic ($n_{\text{acryl}} = 1.51$) and the acrylic/air ($n_{\text{air}} = 1.0003$) interfaces (Sedlazeck and Koch 2012), where n is the refractive index. The optical access is limited to the acrylic window, which constrains the spacing between the cameras to $\Delta H \approx 1 \text{ m}$ in height and $\Delta W \approx 6 \text{ m}$ in width, and limits the relative angles between the cameras from 5° to 20° (Fig. 1).

To calibrate the camera setup we image a planar checkerboard calibration target of dimensions $1.5 \times 1.8 \text{ m}^2$ with 5×6 tiles (Geiger et al. 2012) that are of area $A_{\text{tile}} = 30 \times 30 \text{ cm}^2$. This calibration target is moved within the aquarium by a team of divers that swim with the checkerboard under arbitrary and unknown positions and orientations throughout the aquarium.

2.1 Image processing

The different images of the calibration target are processed to identify the curves and the nodes corresponding to the

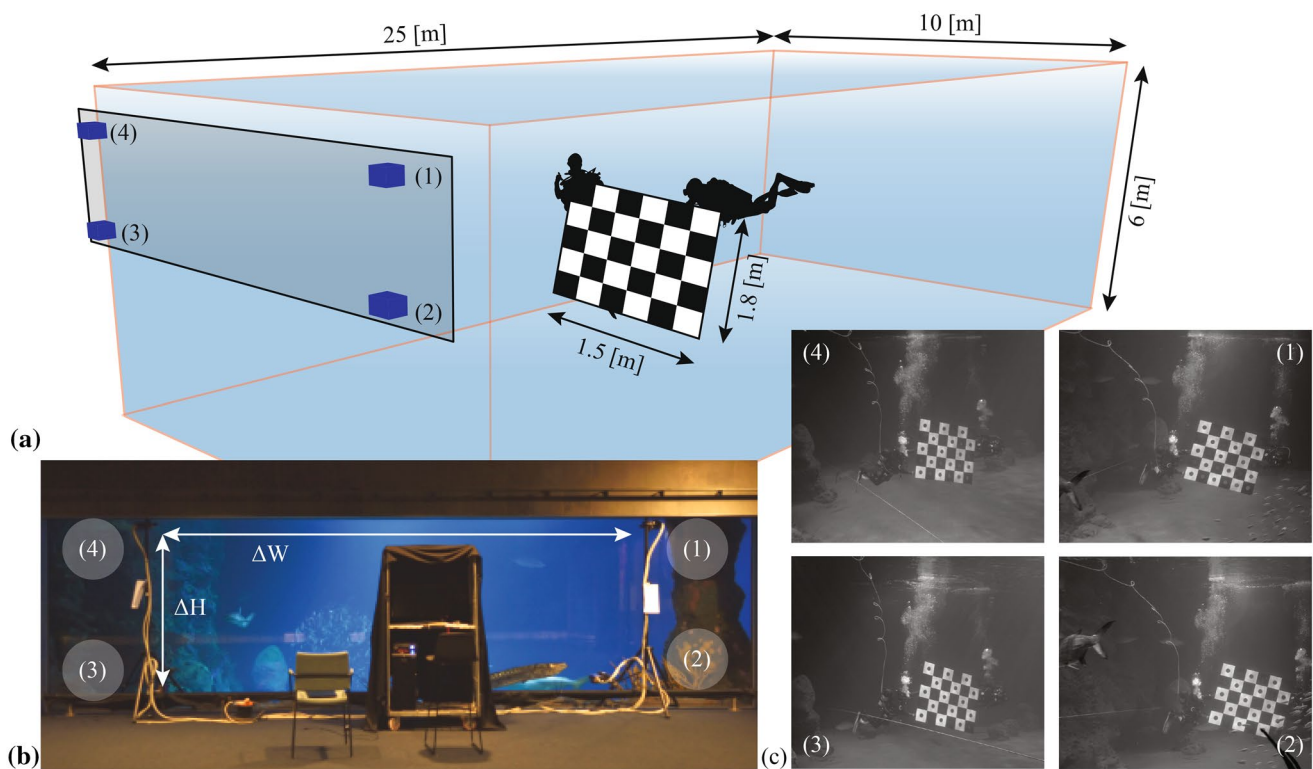


Fig. 1 The four-camera setup at the Rotterdam zoo. **a** Schematic representation of the large measurement volume of $\sim 10 \times 25 \times 6 \text{ m}^3$ along the $\text{DOF} \approx 25 \text{ m}$, including the flat checkerboard of dimensions $1.5 \times 1.8 \text{ m}^2$ that is positioned by a team of divers. **b** Optical

access through the large window spanning $2 \times 8 \text{ m}^2$ including the camera setup at relative spacing of $\Delta H \approx 1 \text{ m}$ by $\Delta W \approx 6 \text{ m}$. **c** An example of calibration images acquired while positioning the checkerboard

gridlines and intersections between the tiles of the checkerboard (Fig. 2). In our application, using a checkerboard calibration target is advantageous over using a pattern of dots. This is because the image gradient obtained from a checkerboard determines grid points more accurately and more robustly over the large depth-of-field of our experiment.

The calibration images are converted to the image gradient using a Savitsky–Golay image differentiation approach (Meer and Weiss 1992) to mark locations on the gridlines between the tiles. For each image, we then fit a set of polynomial curves $\gamma_i(t)$ to the $I = 9$ gridlines between the tiles of the checkerboards using the local intensity values from the image gradient. These fitting curves are written as $\gamma_i(t) = \sum_k \mathbf{a}_i^k t^{k-1}$, where \mathbf{a}_i^k are two-dimensional vectors. Here the parameter t varies within the interval $t \in [0, 1]$ over the checkerboard image, such that $\gamma_i(t = 0)$ and $\gamma_i(t = 1)$ correspond to the beginning and end-points of the gridline of the checkerboard image, see the lines on Fig. 2. At last, we find the $J = 20$ intersections between all the gridlines as a set of nodes \mathbf{x}_j in the image plane of each camera. Here $\mathbf{x}_j = [x_j \ y_j]^T$ with $(\bullet)^T$ the vector transpose, where the numbering $j = 1 \dots J$ is consistent between the different camera views (Geiger et al. 2012) (Fig. 2).

2.2 Distortion correction

Following the path of an optical ray for a single camera in Fig. 2, the linear path is refracted across the air/water interface (Belden 2013). This causes optical distortions in the image plane for each camera and, therefore, we rectify the image plane by dewarping the optical distortion. The coordinates $\mathbf{x} = [x \ y]^T$ in the image plane are mapped to distortion-corrected image coordinates $\hat{\mathbf{x}} = [\hat{x} \ \hat{y}]^T$ by determining a distortion mapping $\hat{\mathbf{x}} = m(\mathbf{x})$. This mapping ensures that collinear points in the object domain are projected as collinear points in the dewarped image plane and, therefore, supports linear ray tracing. For moderate optical distortions, a polynomial distortion map (Soloff et al. 1997) is sufficient. In this study, we write the distortion map as

$$\hat{\mathbf{x}} = \mathbf{x} + \sum_k \mathbf{c}_k \phi^k(\mathbf{x}) = \mathbf{x} + \mathbf{c}_1 x^2 + \mathbf{c}_2 xy + \mathbf{c}_3 y^2 + \mathbf{c}_4 x^3 + \mathbf{c}_4 x^2 y + \mathbf{c}_5 xy^2 + \mathbf{c}_6 y^3. \tag{1}$$

The distortion map is determined by rectifying the curved gridlines in the N original calibration images. We follow Devernay and Faugeras (2001) and minimize the percentage of deflection along the gridlines. We consider the nodes \mathbf{x}_j^n

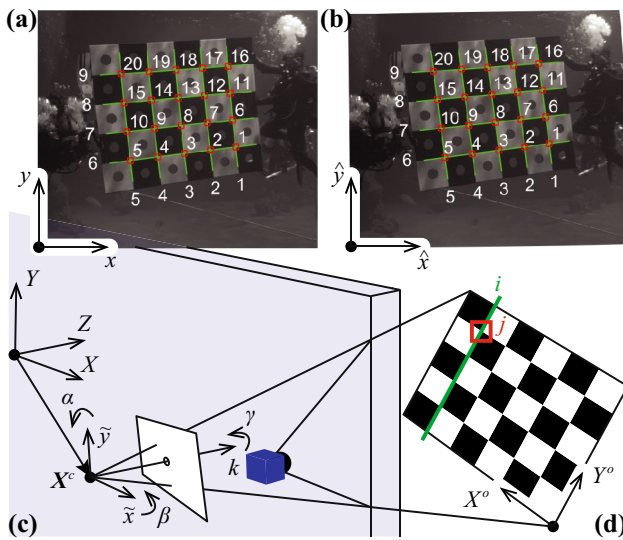


Fig. 2 Image processing and the geometry of the optical path. **a** Processed calibration image with the identified gridlines fitted as second-order polynomial curves (green lines) and the nodes at the gridline intersections (red squares). The gridlines are numbered from $i = 1$ to 9 and the nodes from $j = 1$ to 20. **b** The dwarped calibration image corresponding to **a** in which the gridlines are rectified for minor optical distortion using the mapping of Eq. 1. Supplementary-Fig. 8 provides an example of dewarping for severe optical distortion. **c** The geometry of the optical path where the optical rays are refracted across the air/water interface (neglecting the acrylic window for simplicity) and the positioning of the (virtual-)camera coordinate system $[\tilde{x} \tilde{y} k]^T$ with the origin at the camera center \mathbf{X}^c in world coordinate system $[X Y Z]^T$. **d** The local coordinate system $[X^o Y^o]^T$ attached to the plane of the checkerboard. The indexing i corresponds to the gridlines and j to the locations of the nodes

along a particular gridline $\gamma_i^n(t)$ and their images $\hat{\mathbf{x}}_j^n$ and $\hat{y}_i^n(t)$ in the dwarped image plane through the map $m(\mathbf{x})$. We fit a straight line $\hat{\ell}_i^n$ through the nodes $\hat{\mathbf{x}}_j^n$ and compute the point-line distance $d(\hat{\mathbf{x}}_j^n, \hat{\ell}_i^n)$. The parameters \mathbf{c}_k defining the distortion map are then determined by solving a minimization problem over all gridlines in all calibration images:

$$\min_{\mathbf{c}_k} \sum_{i,n} \sum_{\substack{\text{nodes } j \\ \text{along } \hat{y}_i}} \frac{d(\hat{\mathbf{x}}_j^n, \hat{\ell}_i^n)^2}{\left\| \hat{y}_i^n(1) - \hat{y}_i^n(0) \right\|^2}. \tag{2}$$

This minimization problem can be solved efficiently using a steepest descend algorithm and numeric integration techniques described in Boyd and Vandenberghe (2004). This approach directly extends to larger optical distortions requiring more elaborate distortion models (see Appendix A for the air/water interface and Supplementary-Fig. 8). An example of a dwarped image can be found in Fig. 2.

2.3 Camera calibration and projective geometry

We consider a physical point in the object domain \mathbf{X} of coordinates $\mathbf{X} = [X Y Z]^T$ in a world coordinate system and its projected image $\hat{\mathbf{x}} = [\hat{x} \hat{y}]^T$ in the dwarped image plane of a single camera. The calibration is defined by the mapping function F , such that $\hat{\mathbf{x}} = F(\mathbf{X})$. Our method uses the framework of projective geometry to express the mapping function F and implicitly assumes a pinhole camera model. In the following, we outline the main notations used in projective geometry; a more complete introduction can be found in Hartley and Zisserman (2004).

We make use of augmented vectors to represent points in both the image plane and in the object domain. The coordinates in the dwarped image plane $\hat{\mathbf{x}}$ are augmented to the ray-tracing vector $\tilde{\mathbf{x}}$ such that $\tilde{\mathbf{x}} = [k\hat{x} k\hat{y} k]^T$, where k is a scaling parameter in the direction of the principal optical axis. The associated inverse function that projects $\tilde{\mathbf{x}}$ back to $\hat{\mathbf{x}}$ is defined as the projection $p(\tilde{\mathbf{x}}) = [\tilde{x} \tilde{y}]^T / k = \hat{\mathbf{x}}$. Similarly, the world coordinates \mathbf{X} are augmented to a homogeneous vector as $\tilde{\mathbf{X}} = [X Y Z 1]^T$ (Hartley and Zisserman 2004). Using augmented vectors a geometric transformation, consisting of a rotation and a translation, is simply written as a matrix multiplication $[R \mathbf{t}]\tilde{\mathbf{X}}$, where R is a rotation matrix and \mathbf{t} is a translation vector.

With these notations, the mapping function can be written in the following form that is widely used in projective geometry: (Hartley and Zisserman 2004):

$$\hat{\mathbf{x}} = F(\mathbf{X}) = p(K[R \mathbf{t}]\tilde{\mathbf{X}}). \tag{3}$$

Here K is the 3×3 camera calibration matrix and $[R \mathbf{t}]$ is a 3×4 matrix, with R the 3×3 rotation matrix, and \mathbf{t} the 3×1 translation vector. The matrix K in Eq. 3 has the following form:

$$K = \begin{bmatrix} \alpha_x & s & p_x \\ 0 & \alpha_y & p_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{4}$$

where $\alpha_x = fr_x$ and $\alpha_y = fr_y$ are scale factors, with f the focal length of the lens in mm and r_x and r_y the pixel pitch of the sCMOS sensor in (px/mm). s is the pixel skew, characterizing the angle between the x and the y pixel axes, and $[p_x p_y]^T$ are the coordinates of the principal point at the intersection between the optical axis and the dwarped image plane (Hartley and Zisserman 2004). The elements of K are often referred to as the intrinsic camera parameters, representing the characteristic properties of the camera itself (Hartley and Zisserman 2004), while $[R \mathbf{t}]$ are referred to as the extrinsic parameters representing the position of the camera with respect to the world coordinate system (Fig. 2).

Together, K, R, \mathbf{t} define the mapping function F of Eq. 3 and have to be determined for each of the cameras separately.

In the following, we use the superscript $c = 1 \dots 4$ and the notations K^c , R^c and \mathbf{t}^c , when we distinguish explicitly between the different cameras. We omit the superscript for clarity when no distinction between the cameras is needed; the details of the algorithms can be found in [Appendices](#).

2.4 Single-camera calibration

First, the camera matrix K is determined for each of the four separate cameras by calibrating a single camera using the method developed by Zhang (2000). We consider a local coordinate system $\mathbf{X}^o = [X^o \ Y^o \ Z^o]^T$ attached to the planar checkerboard in the object domain, where $Z^o = 0$ corresponds to the plane of the checkerboard. In this coordinate system, the J nodes at the intersections between the gridlines have known coordinates $\mathbf{X}_j^o = [X_j^o \ Y_j^o \ 0]^T$. The nodes \mathbf{X}_j^o are mapped to their images $\hat{\mathbf{x}}_j^n$ following the formalism used in Eq. 3. This transformation can be written as $p(K[R^n \ \mathbf{t}^n]\tilde{\mathbf{X}}_j^o)$, where the rotation matrix R^n and translation vector \mathbf{t}^n characterize the position of the checkerboard in the object domain. Geometrically, this corresponds to a rotation and translation of the checkerboard plane in the object domain, followed by a projection on the image plane. The camera calibration matrix K and the positioning of the checkerboard, characterized by R^n and \mathbf{t}^n , are determined following Zhang (2000) and are refined by minimizing the following functional:

$$\min_{K, R^n, \mathbf{t}^n} \sum_{j,n} \left\| \hat{\mathbf{x}}_j^n - p(K[R^n \ \mathbf{t}^n]\tilde{\mathbf{X}}_j^o) \right\|^2. \tag{5}$$

2.5 Multiple-camera calibration

To complete the calibration over the multiple cameras, we determine the rotation R^c and the translation \mathbf{t}^c representing the position of each of the cameras in the world coordinate system. R^c and \mathbf{t}^c correspond to the extrinsic camera parameters described in Sect. 2.3 and a first estimate is deduced directly from the calibration of single cameras performed in the previous step. Selecting two different calibrated cameras, we use the Kabsch algorithm (Kabsch 1976) to estimate the relative positions of the two cameras by comparing the positions of the checkerboards, $R^{n,c}$ and $\mathbf{t}^{n,c}$, for these two views. Considering all camera pairs, we determine the relative positions between all the views and deduce a first estimate for R^c and \mathbf{t}^c ; see Appendix B for detail. Next, for each of the N calibration images, we estimate position R^n and \mathbf{t}^n of the checkerboard in the world coordinate system. We do this by averaging the position estimates $R^{n,c}$ and $\mathbf{t}^{n,c}$ obtained from the four separate single-camera calibrations.

In the last step, we compute the final values for K^c , R^c and \mathbf{t}^c by minimizing the reprojection errors from all cameras

and calibration images. For this optimization step, we use as initial conditions, the camera matrices K^c obtained from Sect. 2.4, the positions R^c and \mathbf{t}^c obtained from the Kabsch algorithm and the estimates of the checkerboard positions R^n and \mathbf{t}^n . We define the reprojection error $\epsilon_j^{n,c}$ for each of the N calibration images and in each camera view as

$$\epsilon_j^{n,c} = \frac{1}{\sqrt{\hat{\mathcal{A}}_j^{n,c}}} \left\| \hat{\mathbf{x}}_j^{n,c} - p \left(K^c [R^c \ \mathbf{t}^c] \begin{bmatrix} R^n & \mathbf{t}^n \\ \mathbf{0}^T & 1 \end{bmatrix} \tilde{\mathbf{X}}_j^o \right) \right\|. \tag{6}$$

The reprojection error $\epsilon_j^{n,c}$ is normalized by $\hat{\mathcal{A}}_j^{n,c}$, which is the area in pixels of a single tile on the checkerboard projection in the dewarped image plane (see Appendix C). Hence, $\epsilon_j^{n,c}$ provides a measure of the error relative to the size of the checkerboard. This error is relatively independent of the location of the calibration target within the large depth of field and the positions of the cameras and, therefore, independent of the apparent size of the checkerboard in the image plane. The final parameters for the calibration function F for each view are obtained from the minimization of the following summation:

$$\min_{K^c, R^c, \mathbf{t}^c, R^n, \mathbf{t}^n} \sum_c \sum_{j,n} \left(\epsilon_j^{n,c} \right)^2. \tag{7}$$

The resulting camera calibration is shown in Fig. 3.

3 Assessment of the calibration method

In the following, we evaluate the performance of the calibration. First, we report the extrinsic and intrinsic camera parameters from Table 1 and discuss their physical interpretation. Second, we assess the robustness and convergence of the method as a function of the number of calibration images used. Third, we study the spatial accuracy of the calibration and identify the sources of error.

3.1 Intrinsic and extrinsic camera parameters

First, we consider the numerical values of the extrinsic camera parameters obtained from a calibration, see Table 1. As discussed in Sect. 2.3, these parameters characterize the spatial position and orientation of each camera. The reconstructed positions of the cameras are in agreement with the experimental setup, with cameras 4 and 1 positioned above cameras 3 and 2, and cameras 4 and 3 positioned on the left-hand side while cameras 1 and 2 are located on the right-hand side (as shown in Fig. 1), see Table 1. Furthermore, the relative distances between cameras are also in agreement with the experimental scene as we find the horizontal distance between cameras $\Delta W \approx 5.6$ m and a vertical distance between cameras

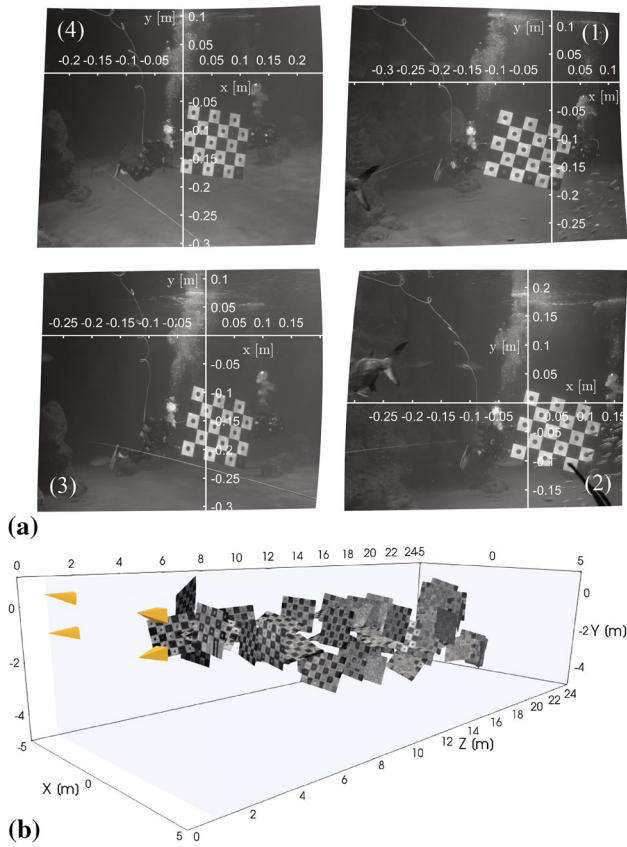


Fig. 3 The resulting camera calibration. **a** The calibrated views that include physical scales on a reference plane at $k = 1$ m in the depth of field for each view. **b** Three-dimensional reconstruction of a random selection of checkerboards used for the calibration, in yellow the reconstructed (virtual) cameras in front of the acrylic window of Fig. 1

$\Delta H \approx 1.2$ m, as deduced from Table 1. Likewise, the reconstructed camera orientations are consistent with the cameras on the right-hand side oriented with a positive angle α , while the cameras on the left-hand side are oriented with a comparable negative angle α .

In addition to reconstructing the position of the camera, the calibration procedure reconstructs the intrinsic camera properties, which we compare to the specification of the instrumentation. The coefficients of the camera calibration matrix K of Eq. 4 are provided for each camera in Table 1. We focus on the values of the focal length of the lenses and deduce an effective focal length f_{eff} directly from the coefficients as

$$f_{\text{eff}} = \sqrt{\frac{\alpha_x \alpha_y \tilde{J} \tilde{n}^2}{r^2}}, \tag{8}$$

where r is the resolution of the camera sensor in px/mm, which is known from the camera specifications, \tilde{J} represents a correction factor for the image expansion due to optical

Table 1 Numerical values of the calibration parameters

Camera		1	2	3	4
\tilde{J}	(-)	0.97	0.98	0.93	0.96
α_x	(px)	5166.3	5245.8	4982.9	5116.0
α_y	(px)	5056.3	5240.5	4941.3	5050.2
s	(px)	- 48.5	- 63.5	- 189.4	- 249.5
p_x	(px)	1888.8	1655.0	1419.6	1237.0
p_y	(px)	1488.0	980.0	1530.5	1552.8
X^c	(m)	2.873	2.878	- 2.872	- 2.87848
Y^c	(m)	0.617	- 0.616	- 0.658	0.657
Z^c	(m)	0.010	- 0.010	0.009	- 0.009
α	(°)	9.02	14.00	- 13.04	- 11.51
β	(°)	- 5.18	- 3.93	2.56	- 5.19
γ	(°)	- 2.45	- 0.81	- 2.29	0.76
f_{eff}	(mm)	23.91	24.66	22.67	23.68
$\epsilon_j^{n,c}$	%	1.84	1.85	2.08	1.95
		± 1.57	± 1.64	± 1.72	± 1.54
$(\epsilon_j^{n,c})^*$	(px)	1.63	1.72	1.85	1.84
		± 1.49	± 1.62	± 1.59	± 1.60
N	(#)	176	153	197	186

From top to bottom: the expansions factor of the distortion mapping \tilde{J} , the intrinsic camera properties from the matrix K , the extrinsic camera positions X , Y and Z and orientations in pitch–yaw–roll angles α , β and γ (Figs. 1, 2). The effective focal lengths f_{eff} deduced from K with Eq. 8, the reprojection errors $\epsilon_j^{n,c}$ in percentage and equivalent pixel-dimensions $(\bullet)^*$, and the total number of calibration images N used for each camera

distortion (see Appendix D) and $\tilde{n} = n_{\text{air}}/n_{\text{water}}$ corrects for the magnification due to refraction at the air/water interface. Our calibration yields values for the effective focal length of the four cameras of $f_{\text{eff}} = 23.73 \pm 0.82$ mm. This reconstruction of the focal length lies within 1–2 % of the actual focal length $f_{\text{lens}} = 24$ mm of the lenses that were used. Hence, we find that both the extrinsic and intrinsic camera parameters deduced from our calibration procedure are in agreement with the dimensions and characteristics of our experimental setup.

3.2 Convergence and robustness

The camera calibration is obtained by minimizing the sum of the squared reprojection errors $\epsilon_j^{n,c}$ over all four cameras, all N calibration images and all nodes on the checkerboard, see Eq. 7. The camera calibration converges to low values for $\epsilon_j^{n,c}$ see Table 1. Here, the average normalized reprojection error is on the order of $\epsilon_j^{n,c} \sim 2\%$ of the size of a checkerboard tile, which corresponds to an error of less than 1 cm.

The camera calibration requires a minimum of two non-coplanar checkerboard images (Zhang 2000). Increasing the number of calibration images increases the sampling

of the measurement volume and, therefore, improves the reliability of the calibration. We further characterize the performance of the method as a function of the number of calibration images used, by randomly selecting different subsets of checkerboard images.

We first consider the effective focal length f_{eff} of Eq. 8 deduced from the matrix K to characterize the quality of the position in space of the checkerboards and the cameras. In Fig. 4, we select different subsets of calibration images ranging from $N = 2$ to $N = 50$ images and compute f_{eff} from the associated calibration. Figure 4a represents the ratio between f_{eff} and f_{lens} as a function of N . With a low number of $N = 2$ to 10 calibration images, the ratio $f_{\text{eff}}/f_{\text{lens}}$ is far from one and the focal length is under- and over-estimated by 50%, indicating an unreliable calibration. Increasing the number of calibration images to $N = 15$ shows that the estimated focal length f_{eff} approaches f_{lens} and represents a clear improvement of the calibration. Further increasing N beyond $N = 15$, the ratio $f_{\text{eff}}/f_{\text{lens}}$ does not significantly converge further (Fig. 4a).

Second, we consider the reprojection errors $\epsilon_j^{n,c}$ in Fig. 4b. For a low number of $N = 2$ to 10 calibration images, the average reprojection error is as high as $\epsilon_j^{n,c} \sim 50$ to 60 %. Increasing the number of calibration images from $N = 15$ to 50 shows an additional decrease of the normalized reprojection error from $\epsilon_j^{n,c} \sim 5$ to 2 % (Fig. 4b) and inset. This shows that 15 calibration images are sufficient to achieve a valid calibration. Further

increasing the number of calibration images improves the convergence for the camera calibration while the ratio $f_{\text{eff}}/f_{\text{lens}}$ remains at a value close to 1.

3.3 Spatial accuracy of the camera calibration

By inverting the camera matrix K , one can directly associate an optical ray in the object domain to a point in the dewarped image plane. For an ideal calibration, the four optical rays associated with the images of the same point on each of the four cameras should intersect at a unique location in the object domain. In practice, the four optical rays are skew lines and do not intersect at a single point. Here, we characterize the spatial accuracy of our camera calibration by estimating the skewness among the four optical rays.

For this, we use the nodes identified at gridline intersections on the N calibration images. We proceed by evaluating the four optical rays associated with each node of each calibration image. We then triangulate the location of each node by finding the point \mathbf{X}_j^n in the object domain that minimizes the sum of the squared distances from the point to the four optical rays. We report for each node the skewness s_j^n as the average distance from the triangulated location \mathbf{X}_j^n to the four optical rays (see Appendix E for more detail).

The calibration images were acquired over the entire depth of the tank and used to characterize the spatial accuracy, by reporting the skewness as a function of the position along the Z -axis of the world coordinate system (Fig. 5a). We find the skewness to be mostly uniform over the large depth of the measurement volume $Z = 5 - 25$ m. Our calibration yields a high spatial accuracy characterized by an average skewness of less than 1 cm. Only a slight increase in the skewness can be observed towards the back of the aquarium although the average skewness still remains below 1 cm. This small increase is due to the decrease in spatial resolution and the decrease in angles between the optical path from the different views.

The variation in skewness over the height and width of the tank are represented in Fig. 5b, c. Figure 5b is a map of the skewness in a XY -plane at the front of the aquarium, while Fig. 5c provides a similar map at the back of the aquarium. Both are qualitatively similar and the skewness remains small over the depth of the tank. For reference, Fig. 5d, e shows the sampling density, which indicates the number of checkerboard images that were used at a location to compute the calibration. It is noteworthy that the center of the tank was better sampled than the sides and the bottom. We find that the skewness, and hence the spatial accuracy, is relatively constant over a large part of the measurement volume, but increases towards the edges of the tank. This is likely a result of the lower sampling away from the camera

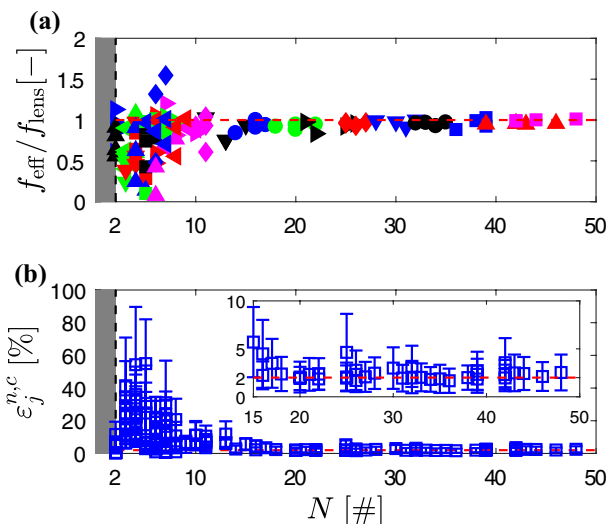


Fig. 4 Quality assessment of the camera calibration procedure for different subsets of randomly selected calibration images. **a** The ratio in estimated focal length and the true focal length of the used lenses $f_{\text{eff}}/f_{\text{lens}}$ as function of the number of calibration images N . Different symbols and colors indicate different selections of images. **b** Averaged reprojection errors $\epsilon_j^{n,c}$ for each camera as a function of N . The error bars indicate the standard deviation of the error per camera

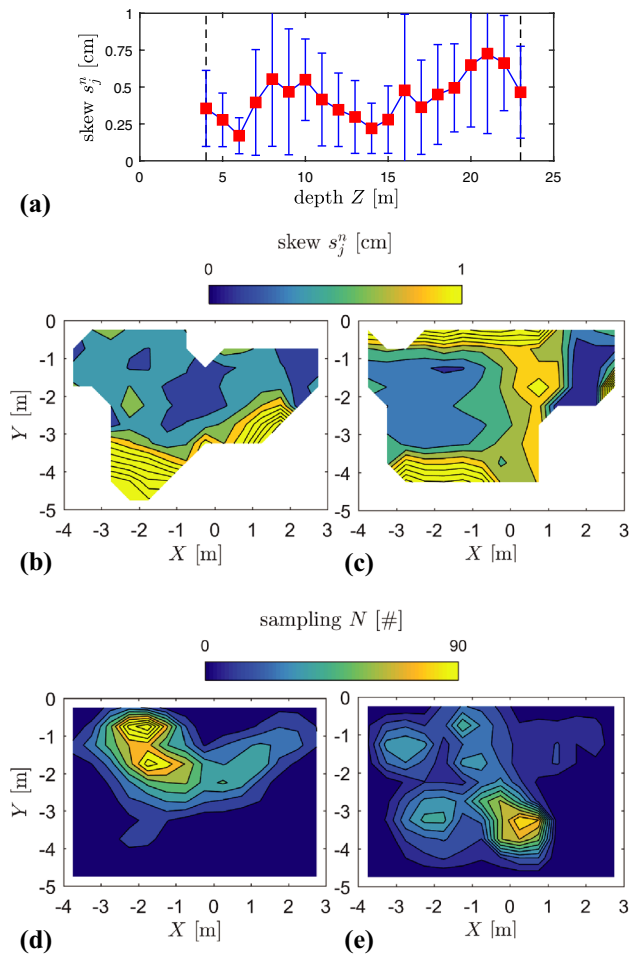


Fig. 5 Spatial accuracy of the camera calibration. **a** The average back-projection error in cm over the depth of the tank Z . **b** The back-projection error over the width and height of the measurement volume at the front of the tank from $Z = 4$ to $Z = 17$ m. **c** The back-projection error at the back of the tank from $Z = 17$ to $Z = 24$ m. **d, e** The sampling density of checkerboards associated with **(b, c)**, respectively

center, as well as unresolved optical distortions at the edges of the image plane.

4 Application to field experiments

We demonstrate the effectiveness of the calibration method by performing three-dimensional measurements in the large aquarium at the Rotterdam Zoo. We begin by focusing on an element which is easily identifiable on each camera view and show that we can accurately triangulate the position. We end our validation by tracking the three-dimensional position of fish of various sizes that are freely swimming over the depth of the tank.

Large predatory fish in the aquarium, such as sharks, swim through the entire aquarium. They provide a good

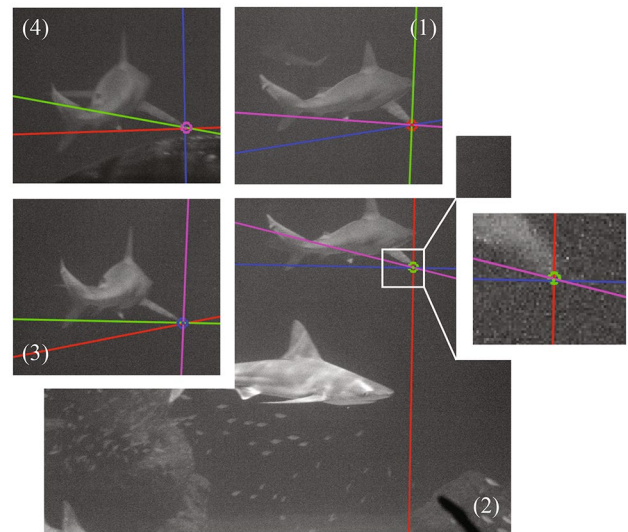


Fig. 6 Triangulation of the vertex of a shark fin. For each camera view: the point corresponding to the vertex of the shark fin is identified with a marker, while the three lines correspond to the epipolar lines associated with the three markers of the remaining three camera views. The color coding is consistent across the multiple views, for example, the vertex on camera 1 is identified with a red marker and the epipolar lines associated with this point in cameras 2, 3, 4 are red. Likewise the vertex in cameras 2, 3, 4 and the corresponding epipolar lines are respectively represented in green, blue and magenta. The inset on camera 2 zooms on the vertex of the fin and shows that the epipolar lines intersect at pixel accuracy

target to evaluate the robustness of the camera calibration as their sharply tipped fins provide easily recognizable and well-defined features. Figure 6 shows how accurately such features can be triangulated with our calibration method. We first identify the vertex of the right-hand side pectoral fin of a shark on each camera view. Similar to Sect. 3.3, we directly associate the vertices, identified on each of the four images, with four optical rays in the object domain. We triangulate the position of the vertex and find a skewness of only 0.35 cm, which demonstrates the accuracy of the method. This small skewness is illustrated in Fig. 6, where, for each camera view, the optical rays associated with the other camera views are projected on the image plane onto the epipolar lines (Hartley and Zisserman 2004). The epipolar lines intersect nearly perfectly at the identified vertex of the shark fin. The inset in Fig. 6 presents a closeup view from which one can infer the reprojection error from the slight skew between the optical rays. This associated reprojection error is of only 1.11 px.

Further, we demonstrate that our calibration supports the tracking of several fish simultaneously over large distances by tracking a small group of six tuna fish (Fig. 7). By triangulation, we reconstruct the three-dimensional time-resolved position of the group (Fig. 7b). Using an in-house automated tracking code, we track the group of six individual fish (tuna)

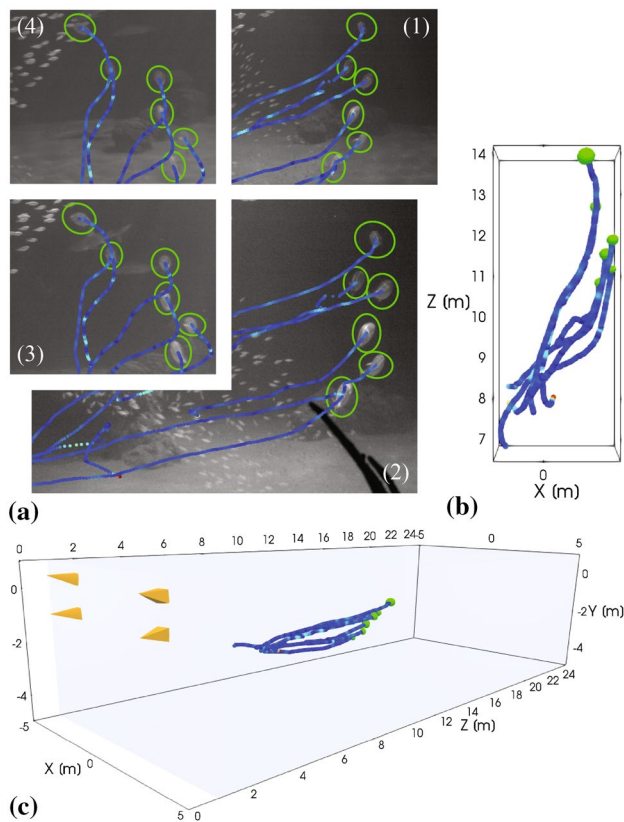


Fig. 7 Tracking and triangulation of a group of tuna fish inside the measurement volume. **a** The group of tracked tuna fish in the image plane. **b** Top view of the tuna fish swimming over a distance of ~ 7 m within the depth of the tank. **c** The three-dimensional reconstruction of the fish swimming in object space including the camera position of Fig. 3. In all visualizations, the color code of the tracks corresponds to the linear velocity in object space

swimming away from the cameras over a large distance of 7 m. Together with the triangulated shark fin, this experiment demonstrates the robustness and accuracy of the calibration and its potential use in large-scale field experiments. The calibration supports accurate triangulation over a large distance along the depth of field. This makes it of interest to further application for the tracking of particles (Schanz et al. 2015), birds (Attanasi et al. 2015), insects (van der Vaart et al. 2019), fish and other animals, and the study of fluid motion using tomographic methods (Elsinga et al. 2006) for large-scale field experiments.

5 Conclusion

Here, we have described and characterized a versatile, accurate and robust calibration method, which supports the three-dimensional tracking and triangulation of multiple fish. Our method is of particular interest to large-scale fields experiments, when spatial access to the

measurement volume is limited and laboratory equipment to precisely position the target cannot be installed. The method does not require a large calibration target to be moved with known displacements. Rather, we use a freely moving checkerboard calibration target, which is much smaller than the measurement volume itself. The calibration mapping uses the framework of projective geometry, which assumes linear ray tracing. It combines a pinhole camera model for the linear ray tracing, with a non-linear camera mappings commonly used in experimental fluid mechanics to correct for optical distortion. All the algorithms necessary for the implementation of the calibration method are described here with details provided in [Appendices](#).

The calibration method has been implemented to obtain an accurate and consistent multiple-view camera calibration in the large-scale aquarium of the Rotterdam Zoo. Here, the calibration target was positioned arbitrarily by a team of divers. We have characterized the spatial accuracy of the calibration to be less than 2% of the size of a checkerboard tile, corresponding to 1 cm over the entire measurement volume that spans several tens of meters. When correcting for the difference in refractive index of air and seawater, we find that our method reconstructs both the camera position and the intrinsic properties of the camera such as the focal length of the lenses. Selecting different subsets of calibration images in a quality assessment, we show that in our case 15 calibration images are sufficient to achieve a valid calibration. Increasing the number of checkerboard images and better sampling the measurement volume further improves the spatial accuracy. The resulting camera calibration allows accurate imaging and three-dimensional tracking over a large measurement volume, here with application to biological fluid mechanics and the tracking of fish.

Acknowledgements We would like to thank the Rotterdam zoo for providing access to the large tank in the Oceanium. In particular, we would like to thank Mark de Boer providing access to the facilities and Martijn van der Veer for coordination of the diving activities. This work was supported by the Netherlands Organisation for Scientific Research (NWO) with project number 824.15.021 and by the H2020 European Research Council (Grant agreement no. 716712).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Availability The implementation of the camera calibration method in MATLAB and an accompanying data set is freely available through Muller et al. (2019).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>)

[mmons.org/licenses/by/4.0/](https://creativecommons.org/licenses/by/4.0/)), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

A: Optical distortion across an interface

Refraction across an interface between two media of different refractive index is governed by Snell’s law:

$$\frac{\sin(\theta_2)}{\sin(\theta_1)} = \tilde{n},$$

where $\tilde{n} = n_1/n_2$ with n the refractive index, θ_1 is the incident angle and θ_2 is the exit angle to the normal vector on the interface. For a flat interface, the image plane of a camera can be warped parallel to the interface by the linear image mapping H :

$$H = \begin{bmatrix} A & \mathbf{p} \\ \mathbf{v}^T & 1 \end{bmatrix},$$

where A is a 2×2 matrix that describes an affine image transformation that changes the aspect ratio and skew of the image, \mathbf{p} centers the image at the principle ray, and \mathbf{v}^T describes a perspective change (Hartley and Zisserman 2004). With these notations, the distortion mapping can be written as

$$\hat{\mathbf{x}} = \frac{A\mathbf{x} + \mathbf{p}}{\sqrt{\lambda\|A\mathbf{x} + \mathbf{p}\|_2^2 + (\mathbf{v}^T\mathbf{x} + 1)^2}},$$

where $\lambda = (1 - \tilde{n}^2)/f$ with f the focal length in pixels dimensions.

B: Relative camera positioning from calibrated views

We determine the rigid body motion from a view c to another view c' , by first finding the best rotation, using the Kabsch algorithm (Kabsch 1976). We first compute the cross-covariance matrix $A = \sum_{n,j} \mathbf{X}_j^{n,c'} (\mathbf{X}_j^{n,c})^T$ using the paired object coordinates $\mathbf{X}_j^{n,c}$ and $\mathbf{X}_j^{n,c'}$ from multiple checkboards. We then compute the singular value decomposition of the cross-covariance matrix $A = USV^T$ and extract the best rotation matrix $R^{c,c'}$ as

$$R^{c,c'} = U \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \det(UV^T) \end{bmatrix} V^T.$$

Knowing $R^{c,c'}$, we use the rigid body motion $\mathbf{X}_j^{n,c'} = R^{c,c'} \mathbf{X}_j^{n,c} + \mathbf{t}^{c,c'}$ to compute the translation vector $\mathbf{t}^{c,c'}$ between the views.

From the relative position between the views we find a unique extrinsic camera positioning R^c, \mathbf{t}^c by taking the first view at reference and solving the minimization problem for the translation \mathbf{t}^c ,

$$\min_{\mathbf{t}^c} \sum_{c \neq c', c''} \left\| R^{c,c'} \mathbf{t}^{c'} - \mathbf{t}^{c,c''} - \mathbf{t}^c \right\|^2$$

subject to $\mathbf{t}^1 = \mathbf{0}$

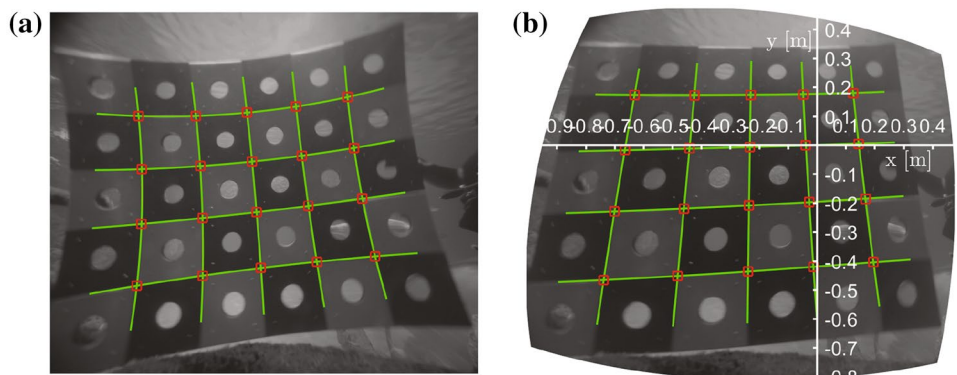
and the rotation R^c ,

$$\min_{R^{c,c'}} \sum_{c \neq c', c''} \left\| R^{c,c'} R^{c''} - R^c \right\|_F^2$$

subject to $R^1 = I$,

where F is the Frobenius norm that sums over all matrix components squared and I is the identity matrix. These linear least squares problems have a direct solution using methods described in Boyd and Vandenberghe (2004) and generalize to any number of views (Fig. 8).

Fig. 8 Supplementary optical distortions for an ultrawide-angle lens [VENUS OPTICS LAOWA 7.5 mm MFT]. **a** Processed calibration image with set of curved gridlines (second-order polynomial curves) and intersections. **b** Calibrated calibration image using the window model of Appendix A



C: Projected area of a planar object

A planar object with area $\mathcal{A}_{\text{plane}}$ and local plane coordinates $\mathbf{X} = [X \ Y \ 0]^T$ is mapped to the dewarped image plane by $\tilde{\mathbf{x}} = p(K[R \ \mathbf{t}]\tilde{\mathbf{X}})$. The projected area in the dewarped image plane $\mathcal{A}_{\text{dewarped}}$ can be computed by integrating the determinant of the Jacobian of the projection map:

$$\mathcal{A}_{\text{dewarped}} = \int_{\mathcal{A}_{\text{plane}}} |\nabla p(X, Y)| \, dX \, dY.$$

This integral can be evaluated using standard numeric integration techniques.

D: Magnification of the distortion map

An image is dewarped according to the distortion map $\hat{\mathbf{x}} = m(\mathbf{x})$. Similar to Appendix C, the area deformation of the distortion map can be computed by integrating the determinant of the Jacobian:

$$\mathcal{A}_{\text{dewarped}} = \int_{\mathcal{A}_{\text{image}}} |\nabla m(\mathbf{x})| \, dx \, dy.$$

This integral can be used to correct the light intensity per pixel area for the varying magnification of the distortion map. The average area expansion of the map is found by integration over the complete image:

$$\tilde{J} = \frac{1}{\mathcal{A}_{\text{image}}} \int_{\mathcal{A}_{\text{image}}} |\nabla m(\mathbf{x})| \, dx \, dy.$$

E: Point triangulation and skewness

A point \mathbf{X} in the object domain is triangulated by minimizing the point-line distance to the optical rays from the different cameras. The optical rays associated with each view are computed by inverting the camera calibration matrix $K^c \setminus \tilde{\mathbf{x}}^c k^c$, where k^c scales the depth of field and $\tilde{\mathbf{x}} = [\hat{x} \ \hat{y} \ 1]^T$. The object location is then triangulated by the linear least squares problem:

$$\min_{\mathbf{X}, k^c} \sum_c \|K^c \setminus \tilde{\mathbf{x}}^c k^c - [R^c \ \mathbf{t}^c]\tilde{\mathbf{X}}\|^2.$$

We then compute the skewness s as the average point-line distance between the optical rays and the object location by

$$s = \frac{1}{C} \sum_c \|K^c \setminus \tilde{\mathbf{x}}^c k^c - [R^c \ \mathbf{t}^c]\tilde{\mathbf{X}}\|^2.$$

References

Adrian RJ, Westerweel J (2011) Particle image velocimetry, 1st edn. Cambridge University Press, Cambridge (ISBN: 978-0-521-44008-0)

Attanasi A, Cavagna A, Castello LD, Giardina I, Jelić A, Melillo S, Parisi L, Pellacini F, Shen E, Silvestri E, Viale M (2015) Greta-a novel global and recursive tracking algorithm in three dimensions. *IEEE Trans Pattern Anal Mach Intell* 37(12):2451–2463. <https://doi.org/10.1109/TPAMI.2015.2414427>

Belden J (2013) Calibration of multi-camera systems with refractive interfaces. *Exp Fluids* 54(2):1463. <https://doi.org/10.1007/s00348-013-1463-0>

Bouguet JY (2015) Camera calibration toolbox for matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/index.html. Accessed Sept 2017

Boyd S, Vandenberghe L (2004) Convex optimization, 2nd edn. Cambridge University Press, Cambridge (ISBN: 9780521833783)

Devernay F, Faugeras O (2001) Straight lines have to be straight. *Mach Vis Appl* 13(1):14–24. <https://doi.org/10.1007/PL00013269>

Elsinga GE, Scarano F, Wieneke B, van Oudheusden B (2006) Tomographic particle image velocimetry. *Exp Fluids* 41(6):933–947. <https://doi.org/10.1007/s00348-006-0212-z>

Fryer JG, Brown DC (1986) Lens distortion for close-range photogrammetry. *Photogramm Eng Remote Sens* 52(1):51–58

Geiger A, Moosmann F, Car O, Schuster B (2012) Automatic camera and range sensor calibration using a single shot. *IEEE Int Conf Robot Autom*. <https://doi.org/10.1109/ICRA.2012.6224570>

Hartley RI, Zisserman A (2004) Multiple view geometry in computer vision, 2nd edn. Cambridge University Press, Cambridge (ISBN: 0521540518)

Hedrick TL (2008) Software techniques for two- and three-dimensional kinematic measurements of biological and biomimetic systems. *Bioinspiration Biomim*. <https://doi.org/10.1088/1748-3182/3/3/034001>

Heikkilä J, Silvén O (1997) A four-step camera calibration procedure with implicit image correction. *IEEE Comput Soc Proc Conf Comput Vis Pattern Recognit*. <https://doi.org/10.1109/CVPR.1997.609468>

Jux C, Sciacchitano A, Schneiders JFG, Scarano F (2018) Robotic volumetric piv of a full-scale cyclist. *Exp Fluids* 59(4):74. <https://doi.org/10.1007/s00348-018-2524-1>

Kabsch W (1976) A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr Sect A* A32:922–923. <https://doi.org/10.1107/S0567739476001873>

Kühn M, Ehrenfried K, Bosbach J, Wagner C (2010) Large-scale tomographic particle image velocimetry using helium-filled soap bubbles. *Exp Fluids* 50(4):929–948. <https://doi.org/10.1007/s00348-010-0947-4>

Meer P, Weiss I (1992) Smoothed differentiation filters for images. *J Vis Commun Image Represent* 3(1):58–72. [https://doi.org/10.1016/1047-3203\(92\)90030-W](https://doi.org/10.1016/1047-3203(92)90030-W)

Menudet J, Becker J, Fournel T, Mennessier C (2008) Plane-based camera self-calibration by metric rectification of images. *Image Vis Comput* 26(7):913–934. <https://doi.org/10.1016/j.imavis.2007.10.005>

Muller K, Tam DSW (2019) Open source package for the calibration of multiple cameras for large-scale experiments using a freely moving calibration target. <https://doi.org/10.4121/uuid:3b0134e7-4436-4c6f-964b-d3dfd4ab7770>


Prescott B, McLean G (1997) Line-based correction of radial lens distortion. *Graph Models Image Process* 59(1):39–47. <https://doi.org/10.1006/gmip.1996.0407>

Schanz D, Schröder A, Gesemann S, Wieneke B (2015) 'shake the box': Lagrangian particle tracking in densely seeded flows at high spatial

- resolution. In: International symposium on turbulence and shear flow phenomena
- Sedlazeck A, Koch R (2012) Perspective and non-perspective camera models in underwater imaging—overview and error analysis. In: Dellaert F, Frahm J, Leal-Taix MPL, Rosenhahn B (eds) *Outdoor and large-scale real-world scene analysis*, vol 37. Springer, Berlin, pp 212–242
- Shen R, Cheng I, Basu A (2008) Multi-camera calibration using a globe. In: 8th Workshop on omnidirectional vision <https://hal.inria.fr/inria-00325386/document>. Accessed Sept 2017
- Soloff SM, Adrian RJ, Liu ZC (1997) Distortion compensation for generalized stereoscopic particle image velocimetry. *Meas Sci Technol* 8(12):1441–1454
- Sturm P, Maybank S (1999) On plane-based camera calibration: a general algorithm, singularities, applications. In: IEEE conference on computer vision and pattern recognition, pp 432–437. <https://doi.org/10.1109/CVPR.1999.786974>
- Svoboda T, Martinec D, Pajdla T (2005) A convenient multicamera self-calibration for virtual environments. *Presence Teleoperators Virtual Environ* 14(4):407–422. <https://doi.org/10.1162/105474605774785325>
- Theriault DH, Fuller NW, Jackson BE, Bluhm E, Evangelista D, Wu Z, Betke M, Hedrick TL (2014) A protocol and calibration method for accurate multi-camera field videography. *J Exp Biol* 217(11):1843–1848. <https://doi.org/10.1242/jeb.100529>
- Toloui M, Riley S, Hong J, Howard K, Chamorro LP, Guala M, Tucker J (2014) Measurement of atmospheric boundary layer based on super-large-scale particle image velocimetry using natural snow-fall. *Exp Fluids*. <https://doi.org/10.1007/s00348-014-1737-1>
- Tsai RY (1987) A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *IEEE J Robot Autom* RA-3(4):321–344. <https://doi.org/10.1109/JRA.1987.1087109>
- van der Vaart K, Sinhuber M, Reynolds AM, Ouellette NT (2019) Mechanical spectroscopy of insect swarms. *Sci Adv*. <https://doi.org/10.1126/sciadv.aaw9305>
- Wieneke B (2005) Stereo-piv using self-calibration on particle images. *Exp Fluids* 39(2):267–280. <https://doi.org/10.1007/s00348-005-0962-z>
- Wieneke B (2008) Volume self-calibration for 3d particle image velocimetry. *Exp Fluids* 45(4):549–556. <https://doi.org/10.1007/s00348-008-0521-5>
- Zhang Z (1998) A flexible new technique for camera calibration. Tech. rep., Microsoft Research, Microsoft Corporation, One Microsoft Way Redmond, WA 98052. <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/tr98-71.pdf>. Accessed Sept 2017
- Zhang Z (1999) Flexible camera calibration by viewing a plane from unknown orientations. In: IEEE proceedings of the seventh IEEE international conference on computer vision. <https://doi.org/10.1109/ICCV.1999.791289>
- Zhang Z (2000) A flexible new technique for camera calibration. *IEEE Trans Pattern Anal Mach Intell* 22(11):1330–1334. <https://doi.org/10.1109/34.888718>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

K. Muller¹  · C. K. Hemelrijk² · J. Westerweel¹ · D. S. W. Tam¹

✉ K. Muller
K.Muller@tudelft.nl

✉ D. S. W. Tam
D.S.W.Tam@tudelft.nl

¹ Laboratory for Aero and Hydrodynamics, Faculty of Mechanical, Materials and Maritime Engineering, Delft University of Technology, Leeghwaterstraat 21, 2628 CA Delft, The Netherlands

² Faculty of Science and Engineering, Groningen Institute for Evolutionary Life Sciences, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands