A Spatial AR System for Axis-aligned Augmentation of Planar Scenes in Industrial Settings

Michael Hornáček*, Hans Küffner-McCauley, Majesa Trimmel, Patrick Rupprecht, Sebastian Schlund

Human Centered Cyber Physical Production and Assembly Systems, Institute for Management Sciences, TU Vienna, Austria

Abstract

Augmented reality (AR) promises to enable use cases in industrial settings that include the embedding of assembly instructions directly into the scene, potentially reducing or altogether obviating the need for workers to refer to instructions in paper form or on a screen. *Spatial* AR, in turn, is a form of AR whereby the augmentation of the scene is carried out using a projector, with the advantage of rendering the augmentation visible to all onlookers simultaneously without calling for each to wear some form of head-mounted display. Care must be taken, however, to distort the images to be projected in a manner that they appear undistorted to the viewer, since the geometry of the scene as it relates to the geometry of the projector plays a role in how the pixels of the projector's image plane map to points in the scene. For planar scene geometry (such as a floor, wall, or table), this can be done in a cumbersome manual process called keystone correction, often using software bundled with the projector.

We propose a spatial AR system for planar scenes that produces the effect of keystone correction analytically, and that facilitates placement by enabling intuitively placing the desired augmentations in a manner aligned with the axes of an image of the scene acquired by a camera facing downwards towards the scene plane. Moreover, our system is able to handle a projector equipped with

*Corresponding author
Email address: michael.hornacek@tuwien.ac.at (Michael Hornáček)

a steerable mirror, enabling factory floor augmentation exceeding the bounds of the projector's own immediate field of view, using only a single projector.

Keywords: Spatial augmented reality (SAR), Industry 4.0, Pilotfabrik

1. Introduction

Augmented reality (AR) [1, 2] promises to enable use cases in industrial settings that include the embedding of assembly instructions directly into the scene [3, 4, 5], potentially reducing or altogether doing away with the need for workers to refer to instructions in paper form or on a screen. Typically, AR works by embedding the augmentation in an image of the scene acquired from the viewpoint of a single individual, with the resulting augmented image in turn displayed using some form of head-mounted display. Reliance on head-mounted displays, however, has two adverse consequences: (i) a head-mounted display must be worn by each individual wishing to view the augmentation, and (ii) such a head-mounted display—in some cases taking the form of a helmet in order to house multiple sensors in support of accurately tracking the viewpoint of the viewer relative to the scene—can be obtrusive. *Spatial* AR is a form of augmented reality carried out not by embedding the augmentation in an image of the scene as with head-mounted displays, but by projection to the scene itself [6], thus eliminating both aforementioned problems. Yet considering for the moment a planar surface to be augmented, unless the projector faces the surface frontally, the bounds of a projected rectangular image will not appear rectangular; more generally, they will instead appear trapezoidal (i.e., the projected image will appear distorted). Such distortions can be eliminated by carrying out a cumbersome manual process called keystone correction to appropriately warp the image to be projected, often using software bundled with the projector.

Our contribution is to propose a spatial AR system that produces the effect of keystone correction analytically, and—using the X- and Y-axes of an image of the scene acquired by a downwards-facing camera as a proxy—in a manner aligning the axes of the augmentation with those of the proxy image. We achieve

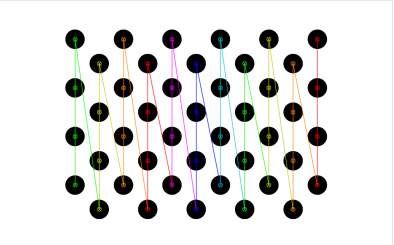
this by distorting the image to be projected using a plane-induced homography computed to produce the effect of projecting the image not from the actual projector viewpoint, but in accordance with the viewpoint of a *virtual* projector (i) facing directly downwards to the scene plane and (ii) rotated to place the axes of the image plane of the virtual projector in line with those of the camera. This facilitates placement of augmentations by intuitively placing them in accordance with the axes of an image of the scene, and eliminates the need for manual keystone correction. Moreover, a consequence of our approach is that our system is able to handle a projector equipped with a steerable mirror (without need for explicitly modeling the action of the steerable mirror on the projector), thereby enabling factory floor applications exceeding the immediate field of view of the projector without needing to rely on multiple projectors.

1.1. Related Work

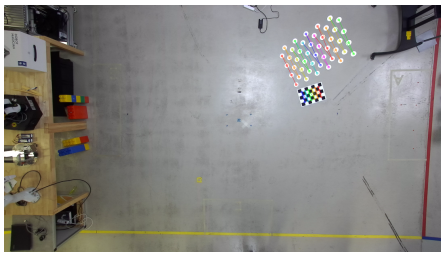
An early spatial AR system for using a projector with a steerable mirror is the IBM Everywhere Displays prototype of Pinhanez [7], yet carry out keystone correction manually. Rather than rely on a steerable mirror to support spatial AR to multiple locations using a single projector, some works mount the projector and a camera on a rigid rig and subject the rig to motion [8, 9, 10]. Steering only a mirror, however, places humbler requirements on the system from a hardware standpoint than if an entire camera-projector rig is to be subject to steering. ...

2. Approach

Correcting for projective distortions of the sort outlined in Section 1 can be achieved by modeling the manner in which the respective rays through the pixels of the projector's image plane fan out into the scene (i.e., by 'calibrating' the projector) and the geometry of the scene itself (i.e., by recovering the scene plane relative to the projector) within at least the projector's field of view. This is because the scene point 'illuminated' by a pixel in the projector's image plane



(a) Circles pattern image, in image plane of projector (detections overlain).



(b) Projector calibration image (one for each target location), in image plane of camera (detections overlain).

Figure 1: Recovering 2D positions in support of projector calibration. (a) 2D positions of the 2D-3D correspondences to be used for calibrating the projector are obtained by detecting—in the image plane of the projector—the circle centers in the circles pattern image, projected by the projector to each of the target locations in the scene plane. (b) For each such target location, an image is acquired from the viewpoint of the camera and the circle centers of the projected circles pattern are detected, in the image plane of the camera. A chessboard pattern to be used for recovering the local scene plane is placed near the projected pattern, whose corners are likewise detected. Detected 2D projected circles pattern center points and chessboard corners overlain for illustration.

can be recovered by intersecting its corresponding ray with the geometry of the scene surface. To model this interaction, we (i) carry out a one-time projector calibration, which in our approach calls for additionally calibrating a camera facing downward to the scene plane and includes recovery of the scene plane as a convenient side effect. Next, we use the relative camera-projector-scene plane geometry to (ii) compute a plane-induced homography that distorts the image to be projected in a manner that it appear undistorted to the viewer, and placed in alignment with the axes of a proxy image of the scene. These two points are treated in Sections 2.1 and 2.2, respectively.

2.1. Recovering Geometry

Calibration of a projector (or camera) in the sense we employ the term here[1] renders one able to project a scene point $\mathbf{X} \in \mathbb{R}^3$ to its corresponding pixel $\mathbf{x} \in \mathbb{R}^2$ in the projector's (or camera's) image plane, or to compute the 'back-projection' of $\mathbf{x}$ (cf. Figure 2), i.e., the ray from the projector's (or camera's) center of projection through $\mathbf{x}$ along which $\mathbf{X}$ must lie. Such a calibration can be expressed in terms of (i) a 3×3 calibration matrix `K` derived from the projector's (or camera's) focal length and principal point [11], and (ii) the coefficients of a lens distortion model used to correct for radial or tangential distortions caused by the lens system [12].

In practice, calibrating a camera relies on (i) establishing 2D-3D correspondences between pixels in the camera's image plane and corresponding points in the scene, and on (ii) using those correspondences as input to an optimization procedure that relies on bundle adjustment [13] to output the calibration matrix `K`, the associated lens distortion model coefficients, and, for each calibration image, the pose (i.e., position and orientation) of the camera relative to the 3D points [11, 14]. To calibrate a camera, a calibration surface such as a chessboard pattern is used to identify the correspondences. Calibration of a projector can be carried out in precisely the same manner insofar as step (ii) is concerned; the major difference in projector calibration relative to the calibrating a camera concerns the manner in which 2D-3D correspondences are identified, i.e., between pixels in the image plane of the projector and the corresponding points in the scene. A convenient consequence of the approach we take to identifying the 3D points of the 2D-3D correspondences needed for projector calibration is, for each target location, recovery of the pose (i.e., position and orientation) of (i) the projector and of (ii) the scene plane, both relative to the coordinate frame of the camera.

[1] We are referring to a geometric calibration; not, e.g., to a color calibration.

Camera calibration. We recover 2D-3D correspondences in support of calibrating the downwards-facing camera is carried out by relying on a planar calibration surface to automatically identify correspondences between the 3D points on the calibration surface and their 2D correspondences in the image plane. The classical calibration surface is a chessboard pattern. The 3D corner points of the chessboard are obtained *a priori* in a coordinate system defined in the plane of the chessboard[2], requiring knowledge only of the dimensions of the chessboard pattern and of the length of a side of a chessboard square. The corresponding 2D points are obtained, in the same order, using a specialized algorithm [15]. A set of calibration images is acquired, each with the calibration pattern visible in a different part of the image plane, and such that the center and all corners and edges of the image plane are covered, the camera's autofocus setting be off, and the camera's zoom factor remain fixed. 2D-3D correspondences are then recovered for each calibration image, and the resulting list is passed on as input to an optimization procedure that relies on bundle adjustment to yield the camera calibration matrix $\mathtt{K}_{\mathrm{cam}}$ and the associated lens distortion model coefficients.

Projector calibration. As with camera calibration, projector calibration relies on 2D-3D correspondences, yet we obtain them in this case by projecting a pattern of circles. We rely on an algorithm to detect the circle pattern center points in the circles pattern image in the image plane of the projector [15], giving the 2D positions of our 2D-3D correspondences for calibrating the projector (cf. Figure 1(a)). We project the circles pattern image to each of the target locations, and use the camera to acquire a projector calibration image for each. Given a projector calibration image acquired using the downwards-facing camera, we detect the circle centers of the *projected* circles pattern (cf. Figure 1(b)); given the scene plane (the recovery of which we shall return to in the paragraph that

[2]E.g., $(0,0,0),(1.5,0,0),(3,0,0),\ldots,(9,7.5,0)$ for a chessboard with 7×6 corners (8×7 squares), with each square of length and width of 1.5 unit, respectively. Note that the units of the chessboard's 3D points give the units of the camera calibration, and—owing to how our projector calibration relies on the camera calibration—of the projector calibration as well.

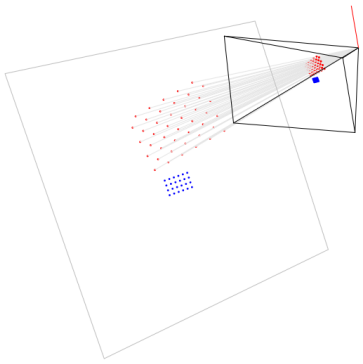


Figure 2: Scene plane (gray) recovered via spatial resection with respect to 2D-3D correspondences obtained using a chessboard pattern (blue); 3D circles pattern points—i.e., the 3D positions of the 2D-3D correspondences to be used for calibrating the projector—obtained by intersection with the scene plane of back-projections (likewise gray) of the 2D circles pattern center points detected in the image plane (red). Note that the points in the figure are the points recovered for the projector calibration image in Figure 1(b), acquired by the downwards-facing camera (frustum of the camera in black, with up vector in red).

follows) and such a 2D circle center $\mathbf{x}$, its 3D correspondence is obtained by intersecting the back-projection of $\mathbf{x}$ with the scene plane (cf. Figure 2). Since the algorithm that yields 2D circle centers does so in a consistent ordering, we thus obtain the 2D-3D correspondences between the projector's image plane and the scene required for projector calibration, yielding the projector calibration matrix $\mathtt{K}_{\mathrm{proj}}$.

We recover the scene plane via spatial resection by applying a PnP algorithm [16] to the 2D-3D correspondences obtained using a chessboard pattern. Note that this step is separate from camera calibration, yet could well be carried out using the same calibration pattern used in the camera calibration step.[3] While a single image of such a chessboard pattern placed on the floor could be sufficient if the floor is even, we place a chessboard pattern in close proximity to the projected circles pattern in each projector calibration image in order to

[3]The critical point is that the pattern should ideally be coplanar with the local scene plane, meaning its height above the scene plane should not exceed a few millimeters.

recover the scene plane locally to each target location, in order to account for the possibility of an uneven floor.

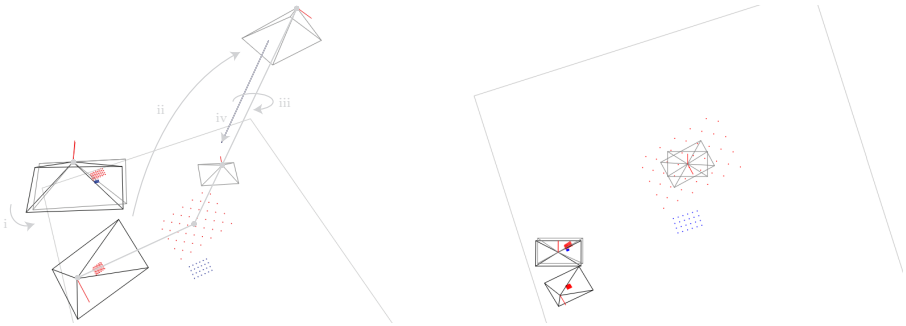
The pose (i.e., position and orientation) of the projector, for each projector calibration image, is provided alongside $\mathtt{K}_{\text{proj}}$ by the aforementioned optimization procedure. Note that for a fixed projector with steerable mirror, given a projector calibration image, the recovered projector's pose is the pose the projector would have to have had to project to the given target location *in the absence of the mirror*. As this is sufficient for our needs in Section 2.2, it is in this sense that our system is able to handle a projector equipped with a steerable mirror, without need for modeling the steerable mirror explicitly.

2.2. Correcting for Projective Distortion

If the projector is calibrated and its pose relative to the scene plane is known, a 'virtual' projector (with the same calibration `K` and lens distortion coefficients) can be placed elsewhere relative to the scene plane. If we for a moment imagine that the projector—at its recovered pose—functions as a camera,[4] then (i) projecting an image to the scene plane *from the viewpoint of the virtual projector* and (ii) acquiring the resulting projected image from the viewpoint of the recovered projector gives the desired corrective warp. Projecting an image warped in this manner to the scene plane *from the viewpoint of the recovered projector* then has the same effect as projecting the unwarped image to the scene plane from the viewpoint of the virtual projector. This warp can computed using a plane-induced homography, computed analytically as a function of the scene plane, the projector, and the virtual projector.

Virtual projector. The placement of the virtual projector determines from which pose the image to be projected is to *appear* to have been projected. This

[4] Recall that the calibration matrix `K` enables computing both (i) the projection of a scene point to the image plane (the function of a camera), or (ii) the back-projection of a pixel in the image plane, giving a ray into the scene (along which a projector illuminates the scene with the given pixel).



(a) Oblique view. (b) Downwards view.

Figure 3: The virtual camera is obtained by (i) rotating the camera (top left, black) about its center of projection such that its optical axis be made parallel with the normal vector of the scene plane. The virtual projector is obtained by (ii) rotating the projector (bottom left, black) about the point of intersection of its optical axis with the scene plane such that the optical axis be made parallel with the scene plane's normal vector, (iii) rotating the X- and Y-axes to align them with those of the virtual camera, and optionally (iv) translating along the normal direction to achieve the desired projected image dimensions.

placement is produced according to a small handful of steps. First, we (i) rotate the camera about its center of projection to align its optical axis with the normal vector of the scene plane, giving a virtual camera likewise facing directly[5] downwards to the scene plane (cf. Figure 3). Next, we (ii) intersect the scene plane with the optical axis (i.e., the ray from the projector's center of projection through the center of the image plane) and rotate the projector's placement about that point of intersection, aligning the optical axis with the scene plane's normal vector and giving an initial virtual projector. Finally, we (iii) align the X- and Y-axes of the initial virtual projector with those of the virtual camera, which gives the virtual projector (cf. again Figure 3). The virtual projector is thus rendered fronto-parallel with the scene plane, enabling projection to the scene plane absent of projective distortions. Optionally, we

[5] A physical camera placed to face downwards is likely to not face downwards precisely; in contrast, the virtual camera's optical axis is aligned exactly with the scene plane's normal vector, rendering it fronto-parallel with respect to the scene plane.

(a) Projecting original image. (b) Projecting our warped image.

Figure 4: Projection from the recovered projector viewpoint (bottom left, black; virtual projector in center, gray) to the scene plane of the original image and of the warped image. (a) Projecting the original image to the scene plane. (b) After warping the original image according to our plane-induced homography for the given target location, the image is projected in a manner that appears free of perspective distortions, aligned with the axes of the virtual camera (via the virtual projector), and to have the desired dimensions in the scene plane, expressed in metric units. Note that background in the projected image is shown set to black.

additionally (iv) adjust the height above the scene plane of the virtual projector, in order to satisfy desired projected image dimensions provided in metric units.

Owing to the manner in which we place the virtual projector, the virtual projector's axes and thus the augmentation are aligned with the axes of the downward-facing camera; the placement of the camera thereby intuitively determines the principal axes according to which augmentations are placed. Note further that a consequence of placing the virtual projector by rotating about the point of intersection of the projector's optical axis with the scene plane is that the center of the projector's image plane remains invariant to the placement of the virtual projector, i.e., a steerable mirror can be aimed with respect to a point projected from the center of the projector's image plane, further facilitating placement.

Plane-induced homography. Let $\mathtt{K}_{\text{proj}}$ express the calibration matrix of the recovered projector and $(\mathtt{R}, \mathbf{t}) \in SE(3)$ the rigid body transformation that transforms points from the coordinate frame of the recovered projector to that of

the virtual projector, for a given target location. Moreover, let $(\mathbf{n}^\top, -d)^\top$ give the scene plane, expressed in the coordinate frame of the recovered projector, where $\mathbf{n} \in \mathbb{R}^3$ is the scene plane's normal vector and $d = \mathbf{n}^\top \mathbf{X}$ for any point $\mathbf{X}$ in the plane, so that $(\mathbf{n}^\top, -d)(\mathbf{X}^\top, 1)^\top = 0$. The transformation that distorts the image to be projected to the scene plane by the recovered projector such that it appear as if were projected to the scene plane by the virtual projector (cf. Figure 4(b)) is given the by the 3×3 matrix

$$\mathtt{H} = \mathtt{K}_{\text{proj}} \left(\mathtt{R} - \frac{\mathbf{t}\mathbf{n}^\top}{d} \right) \mathtt{K}_{\text{proj}}^{-1}, \tag{1}$$

a form of 'plane-induced' homography [11]. For convenience, we enable optional rotation of the image to be projected *before* applying the homography, about the image center; that rotation, parameterized in degrees, is thus in effect likewise carried out intuitively relative to the placement of the camera.

3. Evaluation

We evaluate our approach by augmenting 15 locations across the floorspace at the Pilotfabrik[6] of TU Vienna, a collaborative space for research on Industry 4.0 topics situated in Vienna, Austria. We contrast our approach with a baseline approach involving manual keystone correction, by aiming with both approaches to place the same image for each location aligned with the principal axes of the floorspace, absent of projective distortions, and with the same metric dimensions (50 cm $\times$ 31.25 cm[7]). All experiments were carried out by the same technician, experienced in both approaches.

The hardware setup employed in the evaluation comprised a Panasonic PT-RZ660BE projector with a steerable mirror system—used in our experiments to point the projection to each of the 15 locations—manufactured by Dynamic

[6] `https://www.pilotfabrik.at/`

[7] The dimensions in pixels of the image we project are 960 pixels $\times$ 600 pixels; we chose for our experiments to set the projected metric length of the horizontal axis of the image to 50 cm, which implies 31.25 cm for the vertical axis.

Projection Institute [17]. The steerable mirror system was bundled with the MDC-X software for steering the mirror, loading imagery, and optionally carrying out manual keystone correction, such that each position and (warped) image can be registered as a preset. In addition, we used a downward-facing Zed 2 stereo camera manufactured by Stereolabs, yet relied only on the left view. The floorspace used for our experiments measured dimensions of ca. 6 m $\times$ 4 m; the projector was mounted at approximately the center of this space, at a height of ca. 3.5 m.

Our approach. We began by carrying out a calibration of the camera, acquiring 10 camera calibration images (cf. Section 2.1) of a chessboard calibration pattern with 6×4 corners (7×5 squares) and feeding the images as input to our camera calibration module. Separately, for each of the 15 target locations, we produced a projector calibration image (cf. again Section 2.1) by projecting a 11×4 circles pattern image to the location in question using the steerable mirror, placing a checkerboard pattern beside the projected pattern, and acquiring the image using the downward-facing camera. We then fed these images alongside the output of the camera calibration module to our projector calibration module. For each of the target locations, the steerable mirror was made to point to that location, the circles pattern image was projected to the scene plane, an image was using the camera, and the location was registered in the MDC-X software as a preset. The output of the projector calibration module is a homography per input projector calibration image (cf. Section 2.2). Next, we warped the images to be projected to the respective locations using their corresponding homography, using a third dedicated custom module. These warped images were finally imported into the MDC-X software and associated with their respective location presets.

The total amount of time to carry out all the above steps amounted to ca. 20 min, with ca. 2 min going to acquisition of the camera calibration images, and ca. 5 min going to that of projector calibration images. The remainder of the time was spent running our modules or working with the MDC-X software.

Note that once the camera is calibrated, that calibration can be reused if the camera's intrinsics remain fixed, in particular if no change is made to the zoom factor of the camera.

Baseline. ...

4. Conclusion

We presented a spatial AR system for planar scenes that produces the effect of keystone correction analytically, and that intuitively places the desired augmentations in a manner aligned with the axes of an image of the scene acquired by a camera. Moreover, we showed our system to be able to handle a projector equipped with a steerable mirror, enabling factory floor augmentation exceeding the bounds of the projector's own immediate field of view. Our evaluation demonstrated our approach to produce compelling results at less time than the more cumbersome traditional manual approach to keystone correction.

A subject for a followup paper would be consider in greater detail the causes of and propose a solution to the apparent growth in discrepancy between intended metric projected dimensions and actual metric projected dimensions as the mirror is pointed further away from directly downwards. A further natural extension of this work would be to address non-planar scenes. To handle non-planar scenes would call for a change in how scene geometry is recovered and how distortion of the image to be projected is carried out; projector calibration could, however, be left unchanged.

References

[1] D. Van Krevelen, R. Poelman, A survey of augmented reality technologies, applications and limitations, International Journal of Virtual Reality 9 (2) (2010) 1–20.

[2] F. Zhou, H. B.-L. Duh, M. Billinghurst, Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR, in: 2008 7th

IEEE/ACM International Symposium on Mixed and Augmented Reality, IEEE, 2008, pp. 193–202.

[3] S. Schlund, W. Mayrhofer, P. Rupprecht, Möglichkeiten der Gestaltung individualisierbarer Montagearbeitsplätze vor dem Hintergrund aktueller technologischer Entwicklungen, Zeitschrift für Arbeitswissenschaft 72 (4) (2018) 276–286.

[4] T. Masood, J. Egger, Augmented reality in support of industry 4.0—implementation challenges and success factors, Robotics and Computer-Integrated Manufacturing 58 (2019) 181–195.

[5] M. Gattullo, G. W. Scurati, M. Fiorentino, A. E. Uva, F. Ferrise, M. Bordegoni, Towards augmented reality manuals for industry 4.0: A methodology, Robotics and Computer-Integrated Manufacturing 56 (2019) 276–286.

[6] O. Bimber, R. Raskar, Spatial Augmented Reality: Merging Real and Virtual Worlds, AK Peters/CRC Press, 2019.

[7] C. Pinhanez, The everywhere displays projector: A device to create ubiquitous graphical interfaces, in: International Conference on Ubiquitous Computing, Springer, 2001, pp. 315–331.

[8] J. Ehnes, K. Hirota, M. Hirose, Projected augmentation-augmented reality using rotatable video projectors, in: Third IEEE and ACM International Symposium on Mixed and Augmented Reality, IEEE, 2004, pp. 26–35.

[9] S. Borkowski, J. Letessier, J. L. Crowley, Spatial control of interactive surfaces in an augmented environment, in: IFIP International Conference on Engineering for Human-Computer Interaction, Springer, 2004, pp. 228–244.

[10] A. Butz, A. Kruger, Applying the peephole metaphor in a mixed-reality room, IEEE Computer Graphics and applications 26 (1) (2006) 56–63.

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

[12] D. C. Brown, Close-range camera calibration, Photogrammetric Engineering 37 (8) (1971) 855–866.

[13] B. Triggs, P. F. McLauchlan, R. I. Hartley, A. W. Fitzgibbon, Bundle adjustment—a modern synthesis, in: International Workshop on Vision Algorithms, Springer, 1999, pp. 298–372.

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

[15] G. Bradski, The OpenCV library, Dr. Dobb's Journal of Software Tools 25 (2000) 120–125.

[16] G. Terzakis, M. Lourakis, A consistently fast and globally optimal solution to the perspective-n-point problem, in: European Conference on Computer Vision, Springer, 2020, pp. 478–494.

[17] P. Rupprecht, H. Kueffner-McCauley, S. Schlund, Information provision utilizing a dynamic projection system in industrial site assembly, Procedia CIRP 93 (2020) 1182–1187.