

# A Spatial AR System for Wide-area Axis-aligned Metric Augmentation of Planar Scenes

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

*Human Centered Cyber Physical Production and Assembly Systems, Institute for Management Sciences, TU Wien, 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 such instructions in paper form or on a statically situated 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 any to hold a handheld device such as a tablet or for each to wear some form of head-mounted display. In carrying out spatial AR, however, care must be taken to appropriately warp the image to be projected such that it, when projected, appear free of projective distortions to the viewer. For planar scene geometry (such as a floor, wall, or table), a manual process called keystone correction can be used to carry out an appropriate corrective image warp, a process that can be cumbersome. Another drawback of conventional spatial AR relying only on a single projector is that it is capable of augmenting only the portion of the scene within the projector's static field of view, thereby hindering its applicability to use cases calling for augmentation of wide areas such as a factory floorspace.

We propose a spatial AR system for wide-area metric augmentation of planar scene surfaces that produces the effect of keystone correction analytically as a function of the relative geometry of the projector and scene plane, using a projector equipped with a steerable mirror to direct the projection across varying target locations and a camera facing the scene plane in support of calibrating the system. Our system renders the placement of augmentations in the scene more intuitive than manual keystone correction in two ways. First, (i) the horizontal and vertical axes of the desired augmentations are set in accordance with the horizontal and vertical image axes of the camera, thereby making setting those axes as simple as appropriately rotating the camera relative to the scene plane, and thereby providing a consistent axial reference frame

---

\*Corresponding author

Email address: [michael.hornacek@tuwien.ac.at](mailto:michael.hornacek@tuwien.ac.at) (Michael Hornáček)

across all target locations. Second, (ii) the desired dimensions of the projected augmentations are specified in metric terms, thereby providing for consistent scaling, likewise across all target locations.

*Keywords:* Spatial augmented reality (SAR), computer vision, steerable mirror projector, projector-camera calibration, Industry 4.0

---

## 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, 6, 7, 8, 9, 10], potentially reducing or altogether doing away with the need for workers to refer to such instructions in paper form or on a statically situated screen. Typically, AR works by embedding the augmentation in imagery of the scene acquired from the viewpoint of a single individual, with the resulting augmented image in turn displayed typically using some form of hand-held or head-mounted display. Handheld devices such as tablets do not enable the user to work hands-free while using the device, or force the user to wear additional gear. Reliance on head-mounted displays has two adverse consequences of its own: a head-mounted display must be worn by each individual wishing to partake in the augmentation, and 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. A HoloLens 2 (Microsoft Corporation, USA), for instance, is such a head-mounted display; at a weight of 566 grams [11], it places an additional ergonomic load on the worker, in particular if worn over a longer duration or in bent positions.

In turn, a form of augmented reality referred to as *spatial* AR[12] is carried out not by embedding the augmentation in an image of the scene as with a head-mounted or handheld display, but by projection to the scene itself, thus eliminating the aforementioned problems. Spatial AR thus turns surfaces such as the object being manufactured, a desk, the floor, a wall, or even the ceiling (considering overhead work) into displays. Information provision—of the work plan or of quality-related information—can be carried out by projection to the given workspace, while directly referring to exact positions within that workspace. Moreover, the fact that this information can be viewed by multiple users simultaneously serves to facilitate collaboration [7].

While interest in spatial AR in industrial settings has been growing, its deployment in industrial settings lags well behind that of wearable AR systems such as handhelds or head-mounted displays. One of the main shortcomings of typical projector-based spatial AR relative to AR relying on wearables is the ability to project augmentations only within the projector’s own static field of view. Here, dynamic projection systems—using steerable mirror heads to direct the projection across wider areas—have the potential to bridge the gap to wider industrial penetration. Yet the increase in the area to be covered comes with



Figure 1: Naïve projection to a given target location in the scene plane (e.g., the floor, a wall, or table) if not facing the scene plane directly. (a) Original image, which we wish to project to the scene plane. (b) Projecting the original image to the scene plane from the viewpoint of the projector (shown as a frustum in the bottom left, in black with up vector red) directed to the target location introduces projective distortions if the projector does not face the scene plane directly (as is the case here), with orientation and scale of the augmentation varying in the scene plane as a function of target location.

noteworthy challenges in terms of preparing the augmentations, in particular as the number of target locations in the scene becomes larger than a small handful.

40 Considering a planar surface to be augmented (e.g., a floor, wall, or table), unless the projector faces the surface frontally, the bounds of a projected rectangular image will not appear rectangular, but will instead be subject to projective distortions (cf. Figure 1). More generally, naïvely projecting an image to a planar surface causes not only projective distortions as a function of target location on the surface, but also variation in scale and orientation. Such effects  
45 can be corrected for by carrying out a cumbersome manual process referred to as keystone correction—involving warping an image by perturbing its corners until the desired effect is produced when projected to the scene—to appropriately warp the image to be projected, often using software bundled with the projector. Automating this manual, time-consuming, and error-prone process significantly reduces the time needed for the industrial engineering of a spatial AR application and therefore has potential to promote adoption by industry.  
50 Especially manufacturing processes of large products that remain at a static mounting position during the assembly process—i.e., industrial site assembly—offer a significant application potential for the use of spatial AR as part of worker assistance systems [8].

Our contribution is to propose a spatial AR system for planar scenes that produces the effect of manual keystone correction analytically in a manner for

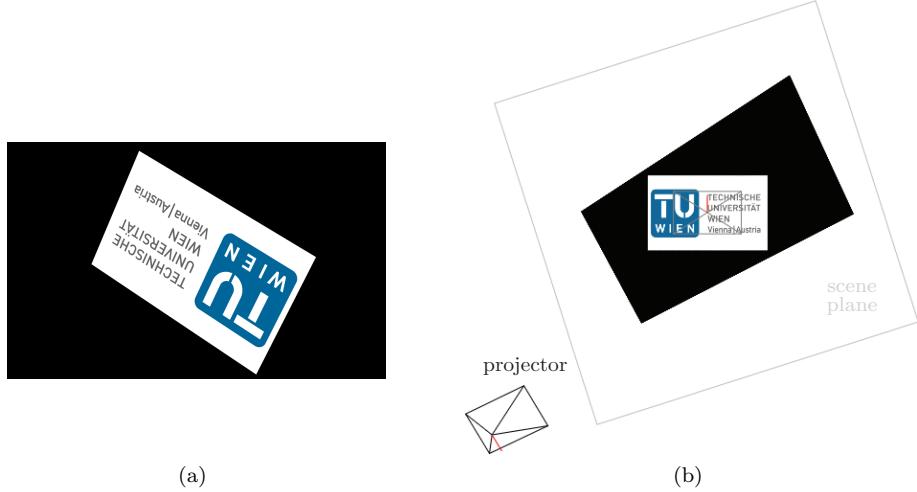


Figure 2: Projection to a given target location in the scene plane, correcting for projective distortions in terms of a *virtual* projector placed and utilized according to our approach. (a) Original image (cf. Figure 1(a)) warped using our approach. (b) Projecting the warped image to the scene plane from the viewpoint of the projector (bottom left, frustum in black with up vector red; same as in Figure 1(b)) directed to the target location has the effect of projecting the *original* image from the viewpoint of a virtual projector (center, frustum in gray with up vector red). This renders the projection free of projective distortions, oriented in accordance with the axes of the ‘downwards-facing’ camera, and at the desired metric scale. Note that black pixels will in effect be ignored by the projector.

use with a projector equipped with a steerable mirror, thereby enabling wide-area applications exceeding the immediate field of view of the projector without needing to rely on multiple projectors. Besides correcting for projective distortions, our system enables placement of augmentations at a desired metric scale. Notably, the system moreover enables setting the  $X$ - and  $Y$ -axes of the augmentation in accordance with the  $X$ - and  $Y$ -axes of a camera placed to face downwards towards the scene plane (cf. Figure 7(a)), which in our case is a factory floor (but could well be generalized to other planar surfaces). We refer to this camera as ‘downwards facing’ for brevity, in quotes since the camera likely to deviate at least gently from facing downwards towards the scene plane directly.<sup>1</sup> Relying on the placement of the ‘downwards-facing’ camera to orient augmentations substantially facilitates placement relative to manual keystone correction, since setting the  $X$ - and  $Y$ -axes of augmentations thus reduces to appropriately rotating the ‘downwards-facing’ camera relative to the scene plane.

---

<sup>1</sup>More precisely, we understand a camera to face downwards towards the scene plane *directly* if the camera’s optical axis is parallel with the scene plane’s normal vector, and correct for possible deviation.

We compute our corrective image warp for a given target location by warping 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 from the viewpoint of a *virtual* projector (cf. Figure 2). We place this virtual projector to (i) face directly downwards to the target location in the scene plane and (ii) rotate its axes in accordance with those of the ‘downwards-facing’ camera. Finally, we (iii) place the virtual projector at the appropriate height above the scene plane to produce the augmentation with the desired metric dimensions in the scene plane.

### 1.1. Related Work

An early spatial AR system explicitly using a projector mounted with a steerable mirror is the IBM Everywhere Displays prototype of Pinharez [13]. For the purposes of the prototype, a projector mounted with a steerable mirror was set up with a camera to demonstrate a variety of use cases, including collaborative assembly encompassing the projection of assembly instructions and an interactive projected user interface [14, 15]. The authors, however, carry out keystone correction manually, by interactively adjusting for the position and orientation of a virtual scene plane and a scaling factor and computing a corresponding 2D homography accordingly.

More generally, keystone correction for planar scenes can be said to reduce to computing a 2D homography [16], an invertible transformation that preserves colinearity and is thus fittingly alternatively referred to as a ‘colineation’; this is the case whether computing the homography relies on manual interaction as in the case of Pinharez, or is obtained free of manual interaction as in ours. The intuition for why it is that a transformation that preserves colinearity—i.e., maps lines to lines—can serve to model an appropriate corrective image warp can be drawn from considering a planar chessboard pattern: looking at a chessboard from one angle, one observes that lines parallel in the chessboard may appear to meet in respective vanishing points, yet always remain lines; looking at the same chessboard from another angle, one observes those same lines remain lines. Accordingly, to warp an image of the chessboard acquired at one angle such that the lines of the chessboard appear as if the chessboard were acquired at another angle accordingly calls for a transformation that maps lines to lines. Since this holds true for any set of lines in a plane, it follows that there exists a line-to-line mapping (i.e., a homography) that models the perspective transformation relating a pair of views of a plane.

One way to compute such a homography is by identifying at least four correspondences between pixel positions in two images of a planar surface [16]. The keystone correction approach of Sukthankar *et al.* [17] reduces to using a segmentation approach to identify the four corners of projection screen in an image acquired by a camera, and computing a homography that maps the resulting four corners of the projection screen to the four corners of the projector’s image plane. Additional sensors can be used to inform the computation of a homography; Raskar and Beardsley [18] use a tilt sensor to recover the projector’s

gravity vector relative to the projector, which they use to carry out a rotational correction in projecting to a wall.

120 Still another way to compute a homography is as a function of a plane and  
a pair of pinhole cameras; the effect produced by the homography is one of  
projecting an image to the plane from the viewpoint of the one camera, and  
acquiring the projected image from the viewpoint of the other. Such a homog-  
raphy is said to be ‘induced’ by a plane [16]. We exploit the fact that cameras  
125 and projectors alike can be modeled using a pinhole camera model [12], proceed  
to recover the relative geometry relating the scene plane and projector, and  
compute a particular form of plane-induced homography to effect the intended  
corrective image warp, additionally taking into account both desired in-plane  
orientation and metric scale.

## 130 2. Approach

Correcting for projective distortions of the sort outlined in Section 1—and  
exemplified in Figure 1—can be achieved by means of an appropriate plane-  
induced homography [16], computed by modeling the manner in which the  
135 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  
140 in the projector’s image plane can be recovered by intersecting its corresponding  
ray with the geometry of the scene plane; modeling the geometry that relates  
projector and scene plane thus enables one to reason about the problem of cor-  
recting for distortions in terms of geometry. To model this interaction, we (i)  
145 additionally calibrating a camera facing downwards to the scene plane and includes  
recovery of the scene plane. Next, we use the relative camera-projector-scene  
plane geometry to (ii) compute a plane-induced homography that warps the im-  
age to be projected in a manner that it appear undistorted to the viewer, placed  
in alignment with the axes of a proxy image of the scene, and at the desired  
metric scale. 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<sup>2</sup> 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  $\mathbf{x}$ , i.e., the ray from the projector’s (or camera’s) center  
of projection through  $\mathbf{x}$  along which a projecting point  $\mathbf{X}$  must lie. Such a  
calibration can be expressed in terms of (i) a  $3 \times 3$  calibration matrix  $\mathbf{K}$  derived  
from the projector’s (or camera’s) focal length  $f$  and principal point  $\mathbf{p}_0 =$

---

<sup>2</sup>We are referring to a geometric calibration; not, e.g., to a color calibration.



Figure 3: Obtaining the 3D positions of the 2D-3D correspondences for use in projector calibration. (a) For each target location, a projector calibration image is acquired from the viewpoint of the ‘downwards-facing’ camera, containing a projected asymmetric circle pattern at the target location and a chessboard pattern placed on the floor nearby. Circle pattern center points and chessboard corners detected automatically (overlaid for illustration). (b) Given a projector calibration image, the scene plane is recovered in the coordinate frame of the ‘downwards-facing’ camera (frustum in black with up vector red) via spatial resection with respect to 2D-3D correspondences obtained using the chessboard pattern (blue), with 3D positions known *a priori*. With the scene plane available, the 3D positions of the projected asymmetric circles pattern to be used in calibrating the projector (red, in scene plane) are obtained by intersection with the scene plane of the back-projections (gray rays) of the detected 2D center points of the circle pattern (red, in image plane of camera).

$(x_0, y_0)^\top \in \mathbb{R}^2$  [16], and (ii) the coefficients of a lens distortion model used to correct for radial or tangential distortions caused by the lens system [19, 20]; the calibration matrix  $K$  is then

$$K = \begin{bmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (1)$$

where  $f$ ,  $x_0$ , and  $y_0$  are each expressed in units of pixels. The calibration matrix  $K$  models a so-called pinhole camera [16]; such a model is applicable to an image acquired using a real-world camera assuming the image has in a prior step been corrected for lens distortion effects.

A camera can be calibrated by (i) establishing 2D-3D correspondences between pixels in the camera’s image plane and the corresponding points in the scene, and on (ii) using those correspondences as input to an optimization procedure that relies on bundle adjustment [21] to output maximum likelihood estimates of the calibration matrix, 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 scene points from among the 2D-3D correspondences [16, 22].

Calibration images containing a calibration surface such as a chessboard pattern of known dimensions and scale are acquired from varying viewpoints using

the camera, to be used to establish the 2D-3D correspondences  $\{\mathbf{x}_{i,j} \leftrightarrow \mathbf{X}_{i,j}\}$ ,  $i \in \{1, \dots, n_{\text{pt}}\}, j \in \{1, \dots, n_{\text{im}}\}$ , where  $n_{\text{pt}}$  gives the number of correspondences obtained from one calibration image of the pattern,  $n_{\text{im}}$  the number of such images, and  $\mathbf{X}_{i,j} = \mathbf{X}_{i,j'}$ , for  $j, j' \in \{1, \dots, n_{\text{im}}\}$ . Bundle adjustment is then used to obtain the maximum likelihood estimates  $(\hat{\omega}, \hat{\mathbf{K}}, \{(\hat{\mathbf{R}}_j, \hat{\mathbf{t}}_j)\})$ ,  $j \in \{1, \dots, n_{\text{im}}\}$  of the lens distortion model coefficients  $\omega$ , calibration matrix  $\mathbf{K}$ , and rigid body transformations  $(\mathbf{R}_j, \mathbf{t}_j)$ ,  $j \in \{1, \dots, n_{\text{im}}\}$ . These are obtained using bundle adjustment by minimizing the cumulative reprojection error

$$\sum_{i=1}^{n_{\text{pt}}} \sum_{j=1}^{n_{\text{im}}} d_{\hat{\omega}}(\mathbf{x}_{i,j}, \hat{\mathbf{x}}'_{i,j}), \quad (2)$$

where  $\hat{\mathbf{x}}'_{i,j} \in \mathbb{R}^2$  gives the projection of the point  $\mathbf{X}_{i,j} \in \mathbb{R}^3$  to the image plane of the camera underlying the  $j^{\text{th}}$  calibration image, subject to the maximum likelihood estimates  $\hat{\mathbf{K}}, (\hat{\mathbf{R}}_j, \hat{\mathbf{t}}_j)$ . The function  $d_{\hat{\omega}}$ , in turn, is a function that first (i) corrects  $\mathbf{x}_{i,j}, \hat{\mathbf{x}}'_{i,j}$  for lens distortions in accordance with the maximum likelihood estimate of the lens distortion model coefficients  $\hat{\omega}$  [19, 20], and then (ii) computes a distance in pixels with respect to two resulting corrected positions. More generally, the projection  $\mathbf{x}' \in \mathbb{R}^2$  of a point  $\mathbf{X} \in \mathbb{R}^3$  to the image plane of a camera given calibration matrix  $\mathbf{K}$  and pose  $(\mathbf{R}, \mathbf{t})$  is given by

$$(\mathbf{x}'^\top, 1)^\top \sim \mathbf{K}(\mathbf{R}\mathbf{X} + \mathbf{t}) \in \mathbb{P}^2, \quad (3)$$

where  $\mathbf{R}\mathbf{X} + \mathbf{t} \in \mathbb{R}^3$  expresses  $\mathbf{X}$  in the coordinate frame of the camera. Inverting the rigid body transformation  $(\mathbf{R}_j, \mathbf{t}_j)$  gives the pose  $(\mathbf{R}_j^{-1}, -\mathbf{R}_j^{-1}\mathbf{t}_j)$  of the  $j^{\text{th}}$  camera relative to the 3D scene points of the calibration pattern, with  $-\mathbf{R}_j^{-1}\mathbf{t}_j \in \mathbb{R}^3$  the corresponding center of projection.

Calibrating a projector [23, 24, 25] can be carried out in precisely the same manner as calibrating a camera insofar as step (ii) is concerned; the major difference in projector calibration relative to the calibrating a camera then 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. What remains of this section is concerned primarily with the recovery of 2D-3D correspondences, first in support of calibrating the ‘downwards-facing’ camera (giving the camera calibration matrix  $\mathbf{K}_{\text{cam}}$ ), then—in a manner that makes use of  $\mathbf{K}_{\text{cam}}$ —for calibrating the projector (giving the projector calibration matrix  $\mathbf{K}_{\text{proj}}$  and the projector poses  $(\hat{\mathbf{R}}_j, \hat{\mathbf{t}}_j)$ , for each projector calibration image).

*2D-3D correspondences for camera calibration.* We recover 2D-3D correspondences in support of calibrating the ‘downwards-facing’ camera 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, and is such a pattern that we use as well. The 3D corner points of the chessboard are

known *a priori* in a coordinate system defined in the plane of the chessboard<sup>3</sup>,  
 requiring knowledge only of the dimensions of the chessboard pattern and of the  
 185 length of a side of a chessboard square. The corresponding 2D points are ob-  
 tained, in the same order, using a specialized algorithm [26]. A set of calibration  
 images is acquired, each with the calibration pattern visible in a different part  
 190 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  
 195 calibration image, respectively, and the resulting list of correspondences is used  
 to recover the camera calibration matrix  $\mathbf{K}_{\text{cam}}$ .

2D-3D correspondences for projector calibration. For projector calibration, we  
 obtain the 2D-3D correspondences by *projecting* a calibration pattern to the  
 195 scene plane, once for each target location. Since we know that the 3D posi-  
 tions of the projected pattern must lie in the scene plane, we can recover the  
 200 3D correspondence of a pattern point detected in the 2D image plane of the  
 ‘downwards-facing’ camera by intersecting the back-projection of the 2D posi-  
 tion with the scene plane. We reuse the chessboard pattern—used above for  
 205 camera calibration—to recover the scene plane locally to each target location  
 via spatial resection [16], by placing the chessboard pattern on the floor near the  
 projected pattern and using the resulting 2D-3D chessboard correspondences as  
 210 input to a PnP algorithm [27]. While a single image of such a chessboard pat-  
 tern placed on the floor could be sufficient if the floor is even, we recover a scene  
 215 plane per projector calibration image to account for the possibility of an uneven  
 floor. Since our projector calibration images thus already contain a chessboard,  
 we choose an alternative calibration pattern as the pattern to project, namely  
 220 an asymmetric circle pattern. We again use a specialized algorithm [26] to ob-  
 tain the circle centers of the asymmetric circle pattern, capable of recovering  
 225 circle centers even under projective distortions.

Given a projector calibration image acquired using the ‘downwards-facing’  
 camera, we detect the circle centers of the *projected* asymmetrical circle pattern  
 (cf. Figure 3(a)); given the scene plane and such a 2D circle center  $\mathbf{x}$ , we obtain  
 230 its 3D correspondence by intersecting the back-projection

$$\mathbf{K}_{\text{cam}}^{-1}(\mathbf{x}^\top, 1)^\top \in \mathbb{P}^2 \quad (4)$$

of  $\mathbf{x}$  with the recovered scene plane (cf. Figure 3(b)). We obtain the 2D positions  
 235 of the 2D-3D correspondences by simply running the algorithm for detecting  
 circle centers on the original asymmetric circles pattern image (i.e., the one we  
 feed to the projector). Since—as in the case of recovering chessboard pattern  
 240 corners—the algorithm that yields 2D circle centers does so in a consistent

---

3E.g.,  $(0, 0, 0), (1.5, 0, 0), (3, 0, 0), \dots, (9, 7.5, 0)$  for a chessboard with  $7 \times 6$  corners ( $8 \times 7$   
 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—since our projector  
 calibration relies on the camera calibration—the units of the projector calibration as well.

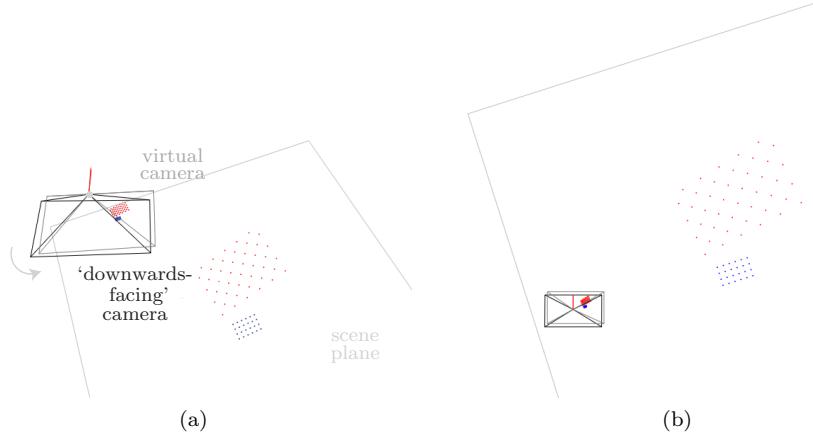


Figure 4: Obtaining the virtual camera. Our physical ‘downwards-facing’ camera—placed to face downwards towards the scene plane, and in terms of whose rotation relative to the normal vector of the scene plane we wish to orient our augmentations—is almost certain to not face downwards towards the scene plane directly. Accordingly, we rotate the recovered ‘downwards-facing’ camera (frustum in black) about its center of projection such that its optical axis be made parallel with the scene plane’s normal, with respect to the minimum arc-length rotation relating the two. It is the  $X$ - and  $Y$ -axes of the virtual camera that we then use to set the  $X$ - and  $Y$ -axes of the virtual projector.

ordering, we thus obtain respective 2D-3D correspondences for each projector calibration image, which we use to recover the projector calibration matrix  $K_{\text{proj}}$ .

Recall that the pose  $(\hat{\mathbf{R}}_j, \hat{\mathbf{t}}_j)$  of the projector, for each projector calibration image  $j$ , is provided alongside  $K_{\text{proj}}$ , relative to the 3D points of the pattern  
220 and thus in the coordinate frame of the ‘downwards-facing’ camera. Note that for a fixed projector with steerable mirror, given a projector calibration image,  
225 the recovered projector’s pose is the pose the projector would have had to project to the given target location *in the absence of the mirror* (i.e., facing directly towards the target location). 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 model coefficients)  
230 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,<sup>4</sup> then

---

<sup>4</sup>Recall that the calibration matrix  $K$  enables computing (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).

(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 (cf. Figure 2). 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 original image to the scene plane from the viewpoint of the virtual projector. This warp can be effected using a plane-induced homography, computed analytically as a function of the scene plane, the projector, and the virtual projector.

<sup>240</sup> *Virtual camera and projector.* The placement of the virtual projector—for a given a target location—determines from which pose the image to be projected is to *appear* to have been projected, with the effect of (i) correcting for projective distortions, (ii) placing the axes of the augmentation in accordance with the axes of the ‘downwards-facing’ camera, and (iii) setting the dimensions of the <sup>245</sup> augmentation in line with the desired target dimensions. We carry out this placement according to a small handful of steps.

<sup>250</sup> It is in accordance with the  $X$ - and  $Y$ -axes of the ‘downwards-facing’ camera that we wish to place the  $X$ - and  $Y$ -axes of the augmentation, since we wish to intuitively be able to use the rotation of the camera relative to the scene <sup>255</sup> plane to set those axes. It is, however, only if the virtual projector is placed to face downwards to the scene plane directly that the projection will appear free of distortions; this is noteworthy, since a physical camera placed to face downwards is almost certain to not face directly downwards to the scene plane. Accordingly, we rotate the recovered camera about its center of projection to align its optical axis with the normal vector of the scene plane by computing and applying the minimum arc-length rotation that relates the two, giving a <sup>255</sup> *virtual* camera facing directly downwards to the scene plane (cf. Figure 4).

To obtain the virtual projector, we (i) intersect the scene plane with the optical axis of the projector and rotate the projector’s placement about that point of intersection, aligning the optical axis with the scene plane’s normal vector, and (ii) align the  $X$ - and  $Y$ -axes with those of the virtual camera. Finally, we (iii) translate along the scene plane’s normal, in order to satisfy the desired metric projected image dimensions; the height  $h$  [m] above the ground plane at which the virtual projector is to be placed if the image—with original image width of  $w_{\text{im}}$  [px]—is to be projected to the scene plane with a metric width of  $w$  [m] is given by

$$h = w \times f_{\text{proj}} / w_{\text{im}}, \quad (5)$$

<sup>260</sup> where  $f_{\text{proj}}$  [px] is the focal length of the projector. These three steps are illustrated in Figure 5. The resulting virtual projector is thus rendered fronto-parallel with the scene plane—enabling projection to the scene plane absent of projective distortions—and projects to the scene plane in the desired scale and with the the  $X$ - and  $Y$ -axes of the augmentation set in accordance with those of the ‘downwards-facing’ camera.

<sup>265</sup> It is thus in the sense that we set the  $X$ - and  $Y$ -axes of the virtual projector to the  $X$ - and  $Y$ -axes of the virtual camera that we place the horizontal and

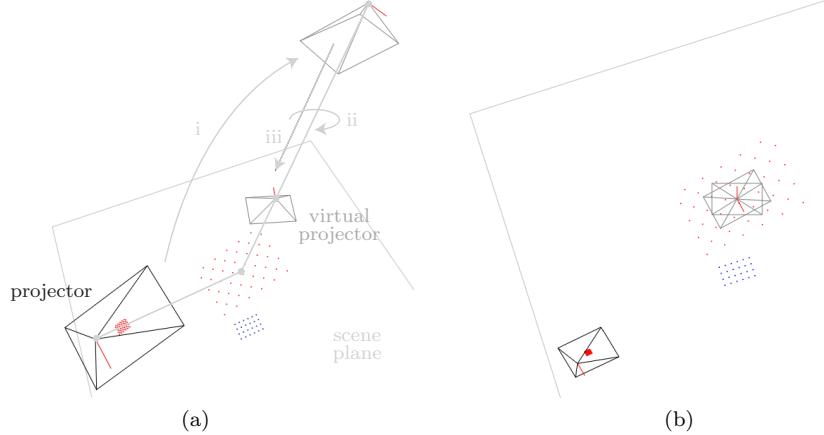


Figure 5: Obtaining the virtual projector, with  $X$ - and  $Y$ -axes aligned with those of the virtual camera (cf. Figure 4). For a given target location, the virtual projector is obtained by (i) rotating the recovered projector (bottom left, frustum in black) about the point of intersection of its optical axis with the scene plane such that the optical axis be made parallel with the scene plane’s normal vector, (ii) rotating the  $X$ - and  $Y$ -axes to align them with those of the virtual camera facing *directly* downwards to the scene plane (cf. Figure 4), and (iii) translating along the normal direction to achieve the desired metric projected image dimensions. Projection from the viewpoint of the virtual projector thus produces an augmentation absent of projective distortions, aligns the augmentation in accordance with the horizontal and vertical image axes of the ‘downwards-facing’ camera, and ensures the augmentation appears in the scene plane with the desired metric dimensions.

vertical axes of the desired augmentations ‘in accordance’ with the horizontal and vertical image axes of the ‘downwards-facing’ camera; the placement of the camera thereby intuitively determines the principal axes according to which augmentations are to be 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. A steerable mirror can be aimed with respect to a point projected from the center of the projector’s image plane, thereby further facilitating placement.

*Plane-induced homography.* Let  $K_{\text{proj}}$  express the calibration matrix of the recovered projector and  $(R, t)$  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  $(n^\top, -d)^\top \in \mathbb{P}^3$  give the scene plane, expressed in the coordinate frame of the recovered projector, where  $n \in \mathbb{R}^3$  is the scene plane’s normal vector and  $d = n^\top \mathbf{X}$  for any point  $\mathbf{X} \in \mathbb{R}^3$  in the plane, so that  $(n^\top, -d)(\mathbf{X}^\top, 1)^\top = 0$ . The transformation that warps the image to be projected—by the recovered projector to the scene plane—such that it appear as if were projected to the scene plane by the virtual projector

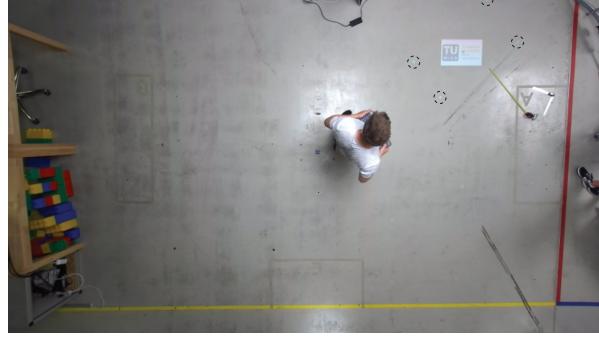


Figure 6: Projection of the warped image from Figure 2(a) to corresponding target location at the upper-right corner of the floorspace, acquired by the ‘downwards-facing’ camera. Note that besides correcting for projective distortions, the warp places the axes of the augmentation in accordance with the axes of the ‘downwards-facing’ camera, and that the dimensions of the augmentation are in line with the desired target dimensions (50 cm × 31.25 cm). To provide context, the corners of the full projected image extent are projected to the floor in light green (emphasized with overlain dashed circles).

(cf. Figures 2 and 6) is given by the  $3 \times 3$  matrix

$$H = K_{\text{proj}} \left( R - \frac{tn^T}{d} \right) K_{\text{proj}}^{-1}, \quad (6)$$

<sup>275</sup> a form of ‘plane-induced’ 2D homography [16]. For convenience, we enable optional rotation of the image to be projected *before* applying  $H$ , about the image center; that rotation, parameterized in degrees, is thus in effect carried out—again intuitively—with respect to the ground plane.

### 3. Evaluation

<sup>280</sup> We evaluate our approach by using a 960 pixel × 600 pixel image of the TU Wien logo (cf. Figure 1(a)) to augment 15 locations across the floorspace at the Pilotfabrik<sup>5</sup> of TU Wien, 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 <sup>285</sup> approaches to place the same image of the TU Wien logo for each of the 15 locations, aligned with the principal axes of the floorspace (i.e., the yellow and red lines in Figure 3(b)), absent of projective distortions, and with the same metric dimensions of 50 cm × 31.25 cm<sup>6</sup>. All experiments were carried out by the same technician, experienced in both approaches.

---

<sup>5</sup><https://www.pilotfabrik.at/>

<sup>6</sup>The dimensions in pixels of the image we project are 960 × 600; 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 if aspect ratio is to be preserved.

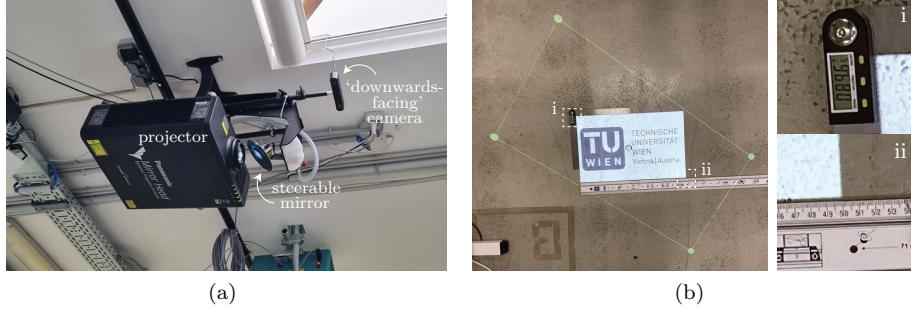


Figure 7: Evaluation scenario. (a) Our hardware setup, comprised of a projector, a steerable mirror system, and a stereo camera, of which we used only the left view. (b) Example augmentation produced by projecting—to one of the 15 target locations across the floorspace considered in our evaluation, to the upper-left corner of our floorspace—an image warped using our approach. The digital angle gauge reads  $89.6^\circ$  for the top-left corner, and the steel rule 49.9 cm for the bottom edge. Corners of the full projected image extent projected to the floor in light green to provide context. Note that blur has a minimal impact on the quality of the augmentation, and hence on the accuracy of the measurements.

290 The hardware setup employed in the evaluation (cf. Figure 7(a)) comprised  
 a Panasonic PT-RZ660BE projector (Panasonic Corporation, Japan) with a  
 steerable mirror system—used in our experiments to point the projection to  
 each of the 15 locations—manufactured by Dynamic Projection Institute (Dy-  
 namic Projection Institute, Austria) [9, 10]. The steerable mirror system was  
 295 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 Stereolabs  
 Zed 2 stereo camera (Stereolabs Inc., France) as our ‘downwards-facing’ camera,  
 making use of only its left view. The floorspace used for our experiments mea-  
 300 sured dimensions of ca.  $6 \text{ m} \times 4 \text{ m}$ ; the projector and camera were mounted at  
 approximately the center of this space, at a height of ca. 3.5 m. At this height,  
 the pixel size (i.e., length or width of a pixel if projected to the scene plane) of  
 the camera was ca. 3.5 mm.

Table 1: Mean absolute error (MAE) and maximum absolute error in units of centimeters, with respect to the four sides (top, bottom, left, right) of the 15 augmentations produced using our approach for 15 target locations spread across the floorspace, respectively. Errors were computed against intended augmentation dimensions of  $50 \text{ cm} \times 31.25 \text{ cm}$  (i.e., 50 cm for top and bottom, 31.25 cm—rounded to 31.3 cm for the purposes of the computing the errors—for left and right).

	Top	Bottom	Left	Right
MAE [cm]	0.23	0.17	0.22	0.2
Max AE [cm]	0.8	0.6	0.5	0.5

*Manual approach.* Having pointed the steerable mirror to the desired location,  
 305 manual keystone correction using the MDC-X software involves warping the  
 image to be projected by manipulating the corners of the image until the de-  
 sired effect is produced on the projection surface. In order to provide a tem-  
 plate for manually carrying out keystone correction, we prepared a rectangular  
 310 50 cm  $\times$  31.25 cm cutout of cardboard. For each target location, we placed this  
 piece of cardboard in a manner that its axes were aligned with the principal axes  
 315 of the floorspace; we then manually warped the image of the TU Wien logo for  
 the given target location so that any corner of the augmentation deviated by at  
 most 1 cm from the corresponding corner of the template. End-to-end, the pro-  
 cess of setting the presets in the MDC-X software having carried out keystone  
 correction manually took ca. 32 min for the 15 target locations.

Table 2: Mean absolute error (MAE) and maximum absolute error in units of degrees, with respect to the top-left and bottom-right corners of the 15 augmentations produced using our approach for 15 target locations spread across the floorspace, respectively. Errors computed in terms of deviation from  $90^\circ$ .

	Top Left	Bottom Right
MAE [deg]	0.17	0.32
Max AE [deg]	0.9	0.5

*Our approach.* We began by carrying out a calibration of the camera, acquiring  
 320 10 camera calibration images (cf. Section 2.1) of a chessboard calibration pattern  
 with  $6 \times 4$  corners ( $7 \times 5$  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  
 325  $11 \times 4$  asymmetrical circle pattern image to the target location in question by  
 directing the projection using the steerable mirror, placing a chessboard pattern  
 beside the projected pattern, and acquiring the image using the ‘downwards-  
 facing’ camera. We then fed these images alongside the output of the camera  
 330 calibration module to our projector calibration module. For each of the target  
 locations, the target 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 image of  
 335 the TU Wien logo to the respective target locations using the corresponding  
 homography, using a third dedicated custom module responsible for warping an  
 image according to a homography. These warped images were finally imported  
 into the MDC-X software and associated with their respective target location  
 presets (cf. Figure 7(b)).

Tables 1 and 2 give the errors in measured lengths and angles, respectively,  
 335 with respect to the intended quantities. Lengths were measured using a steel  
 rule for the top, bottom, left, and right sides of the 15 augmentations; angles  
 were measured using a digital angle gauge at the 15 augmentations’ top-left  
 and bottom-right corners. Note that absolute errors in lengths in no instance

exceeded 1 cm for any side of any of the 15 augmentations, or 1° for any of  
340 the measured corners. The total amount of time expended for carrying 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  
345 or working with the MDC-X software. Note that once the camera is calibrated,  
that calibration can be reused if the camera’s optics remain fixed, notably if no  
change is made to the zoom factor or focus settings of the camera.

#### 4. Conclusion

We presented a spatial AR system for planar scenes that produces the effect  
350 of keystone correction analytically as a function of the geometry relating  
projector and scene plane, exploiting a steerable mirror system to enable augmentation  
exceeding the bounds of the projector’s own immediate field of view.  
Our method produces this effect in a manner that enables intuitive placement  
355 of the augmentations in accordance with the  $X$ - and  $Y$ -axes of a camera placed  
to face downwards towards the scene plane. Moreover, our method allows for  
specifying the desired dimensions of augmentations in metric terms, thus enabling  
consistent scaling. Our evaluation demonstrated our approach to produce compelling  
results at less time than a more cumbersome manual approach to keystone correction.

A natural extension of this work would be to consider the impact of varying  
360 calibration patterns, their dimensions, and their scale relative to the camera,  
or to consider the impact of the ‘downwards-facing’ camera’s pixel size. Still another  
would be to address the augmentation of non-planar scenes. To handle  
non-planar scenes would call for a change in how scene geometry is recovered  
365 and how warping of the image to be projected is carried out; the methodology  
we proposed for projector calibration could, however, be left unchanged.

#### 5. Acknowledgments

This work was supported by the Austrian Research Promotion Agency (FFG)  
through its endowed professorship in Human Centered Cyber Physical Production  
and Assembly Systems at TU Wien (FFG-852789).

#### 370 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.  
375

- 380 [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.
- 385 [4] A. E. Uva, M. Gattullo, V. M. Manghisi, D. Spagnulo, G. L. Casella, M. Fiorentino, Evaluating the effectiveness of spatial augmented reality in smart manufacturing: a solution for manual working stations, *The International Journal of Advanced Manufacturing Technology* 94 (1) (2018) 509–521.
- 390 [5] 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.
- 395 [6] 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.
- [7] D. Aschenbrenner, F. Leutert, A. Çençen, J. Verlinden, K. Schilling, M. Latoschik, S. Lukosch, Comparing human factors for augmented reality supported single-user and collaborative repair operations of industrial robots, *Frontiers in Robotics and AI* 6 (2019) 37.
- 400 [8] W. Mayrhofer, P. Rupprecht, S. Schlund, One-fits-all vs. tailor-made: user-centered workstations for field assembly with an application in aircraft parts manufacturing, *Procedia Manufacturing* 39 (2019) 149–157.
- 405 [9] P. Rupprecht, H. Kueffner-McCauley, S. Schlund, Information provision utilizing a dynamic projection system in industrial site assembly, *Procedia CIRP* 93 (2020) 1182–1187.
- [10] P. Rupprecht, H. Kueffner-McCauley, M. Trimmel, S. Schlund, Adaptive spatial augmented reality for industrial site assembly, *Procedia CIRP*.
- 410 [11] Hololens 2—overview, features, and specs, <https://www.microsoft.com/en-us/hololens/hardware>, [Online; accessed 2-September-2021] (2021).
- [12] O. Bimber, R. Raskar, *Spatial Augmented Reality: Merging Real and Virtual Worlds*, AK Peters/CRC Press, 2019.
- 415 [13] C. Pinhanez, The everywhere displays projector: A device to create ubiquitous graphical interfaces, in: *International Conference on Ubiquitous Computing*, Springer, 2001, pp. 315–331.
- [14] R. Kjeldsen, C. Pinhanez, G. Pingali, J. Hartman, T. Levas, M. Podlaseck, Interacting with steerable projected displays, in: *Proceedings of Fifth IEEE International Conference on Automatic Face Gesture Recognition*, IEEE, 2002, pp. 402–407.

- [15] C. Pinhanez, R. Kjeldsen, A. Levas, G. Pingali, M. Podlaseck, N. Sukaviriya, Applications of steerable projector-camera systems, in: Proceedings of the IEEE International Workshop on Projector-Camera Systems at ICCV 2003, 2003.
- 420 [16] R. I. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd Edition, Cambridge University Press, ISBN: 0521540518, 2004.
- [17] R. Sukthankar, R. G. Stockton, M. D. Mullin, Smarter presentations: Exploiting homography in camera-projector systems, in: Proceedings Eighth IEEE International Conference on Computer Vision (ICCV), Vol. 1, IEEE, 2001, pp. 247–253.
- 425 [18] R. Raskar, P. Beardsley, A self-correcting projector, in: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, IEEE, 2001, pp. II–II.
- 430 [19] D. C. Brown, Close-range camera calibration, *Photogrammetric Engineering* 37 (8) (1971) 855–866.
- [20] J. Weng, P. Cohen, M. Herniou, et al., Camera calibration with distortion models and accuracy evaluation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14 (10) (1992) 965–980.
- 435 [21] 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.
- [22] Z. Zhang, A flexible new technique for camera calibration, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (11) (2000) 1330–1334.
- 440 [23] D. Moreno, G. Taubin, Simple, accurate, and robust projector-camera calibration, in: *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization & Transmission*, IEEE, 2012, pp. 464–471.
- [24] X. Zhang, L. Zhu, Projector calibration from the camera image point of view, *Optical Engineering* 48 (11) (2009) 117208.
- 445 [25] S. Zhang, P. S. Huang, Novel method for structured light system calibration, *Optical Engineering* 45 (8) (2006) 083601.
- [26] G. Bradski, The OpenCV library, *Dr. Dobb's Journal of Software Tools* 25 (2000) 120–125.
- 450 [27] T. Collins, A. Bartoli, Infinitesimal plane-based pose estimation, *International Journal of Computer Vision* 109 (3) (2014) 252–286.