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 an image to be projected in a manner that it, when projected, appear free of 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 is cumbersome and time consuming. Another drawback of conventional projector-based spatial AR is that it is capable of augmenting only the portion of the scene within the projector's own 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

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

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 is carried out 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. Second, (ii) the desired dimensions of the projected augmentations are specified in metric terms, thereby allowing for consistent scaling 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 an image of the scene acquired from the viewpoint of a single individual, with the resulting augmented image in turn displayed typically using some form of handheld 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 product 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 promote collaboration [7].

While interest in spatial AR in industrial settings has been growing, its deployment in industrial settings remains 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 applicability of spatial AR only to augmenting scenes within the projector's own static field of view. Here, dynamic projection systems—using mirror heads to deflect the projection across wider areas—has the potential to bridge the gap to wider industrial penetration. Yet the increase in the area to be covered comes with noteworthy challenges in terms of ensuring that the images to be projected appear free of distortions to the users. 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. Such distortions 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. 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 wide-area spatial AR system for planar scenes that produces the effect of keystone correction analytically, free of any manual interaction. Most importantly, we do this in a manner placing the axes of the augmentation in accordance with the axes of a camera placed to face directly towards the scene plane (a camera we shall in what follows refer to as 'downwards facing', for brevity). This facilitates placement of augmentations, since establishing the principal axes of the augmentation thus reduces to placing the downwards-facing camera in a manner such that the X- and Y-axes of the image align with the intended X- and Y-axes of the augmentation. We achieve this 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 in accordance with the viewpoint of a *virtual* projector. 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 camera. Moreover, our system enables specifying the dimensions of augmentations in metric terms, which we achieve by placing the virtual projector at the appropriate height above the scene plane. Finally, our system set up 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 wide-area 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 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 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 an image of a chessboard acquired from an oblique angle, one observes that lines parallel in the chessboard appear to meet in respective vanishing points; looking at an image of the same chessboard acquired frontally with respect to the plane of the chessboard, lines parallel in the chessboard appear parallel in the image (i.e., they are said to meet 'at infinity'). To warp the former image (where lines parallel in the scene meet in respective vanishing points) such that the lines of the chessboard appear as in the latter image (where lines parallel in the scene meet at infinity) calls for a transformation that maps lines to lines.

One way to compute 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, which they use to carry out a rotational correction in projecting to a wall. 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 homography is said to be 'induced' by a plane [16]. We exploit the fact that cameras 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, taking into account both desired in-plane orientation and metric scale.

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 can be recovered by intersecting its corresponding ray with the geometry of the scene surface; modeling the geometry that relates projector and scene plane thus enables one to reason about the problem of correcting for distortions in terms of geometry. To model this interaction, we (i) carry out a one-time projector calibration, which in our approach calls for additionally calibrating a camera facing downwards 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 warps 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 $\mathbf{x}$, 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 3×3 calibration matrix $\mathtt{K}$ derived from the projector's (or camera's) focal length f and principal point $\mathbf{p}_0 = (x_0, y_0)^\top \in \mathbb{R}^2$ [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 $\mathtt{K}$ is then

$$\mathtt{K} = \begin{bmatrix} f & 0 & x_0 \\ 0 & f & y_0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{1}$$

where f and $(x_0, y_0)^\top$ are both expressed in units of pixels. The calibration matrix $\mathtt{K}$ models a so-called pinhole camera; such a model is applicable to an image acquired using a real-world camera assuming the image has been corrected for lens distortion effects, by having applied the appropriate lens distortion model coefficients.

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 co-

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

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

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

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

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

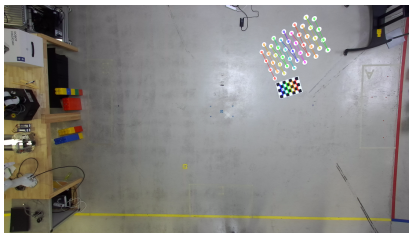
where $\mathtt{R}\mathbf{X} + \mathbf{t} \in \mathbb{R}^3$ expresses the point $\mathbf{X}$ in the coordinate frame of the camera. Inverting the rigid body transformation $(\mathtt{R}_j, \mathbf{t}_j)$ gives the pose $(\mathtt{R}_j^{-1}, -\mathtt{R}_j^{-1}\mathbf{t}_j) \in SE(3)$ of the j^{th} camera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

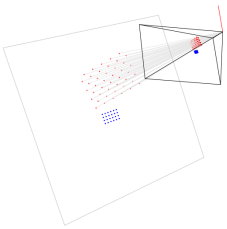
ing accordingly, the major difference in projector calibration relative to the calibrating a camera concerns the manner in which 2D-3D correspondences are identified, i.e., between pixels in the image plane of the *projector* and the corresponding points in the scene. What remains of this section is concerned primarily with the recovery of 2D-3D correspondences, first in support of calibrating the downwards-facing camera (used to obtain the camera calibration matrix $\mathtt{K}_{\text{cam}}$), then—in a manner that makes use of $\mathtt{K}_{\text{cam}}$—for calibrating the projector (used for obtaining the projector calibration matrix $\mathtt{K}_{\text{proj}}$ and the projector poses $(\hat{\mathtt{R}}_j, \hat{\mathbf{t}}_j) \in SE(3)$, 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 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 [26]. 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, respectively, and the resulting list of correspondences is used to recover the camera calibration matrix $\mathtt{K}_{\text{cam}}$.

[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—since our projector calibration relies on the camera calibration—the units of the projector calibration as well.



(a) Projector calibration image (detections overlain).



(b) 3D positions for projector calibration (red, in scene plane).

Figure 1: Obtaining the 3D positions of the 2D-3D correspondences for 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. Automatically detected circle pattern center points and chessboard corners overlain for illustration. (b) Given a projector calibration image, the scene plane (gray) is recovered in the coordinate frame of the camera (black, up vector red) via spatial resection with respect to 2D-3D correspondences obtained using the chessboard pattern (blue); the 3D positions of the 2D-3D correspondences to be used for calibrating the projector (red, in scene plane) are obtained by intersection with the scene plane of back-projections (rays, likewise gray) of the 2D center points of the circle pattern detected in the image plane of the camera.

2D-3D correspondences for projector calibration. For projector calibration, we obtain the 2D-3D correspondences by *projecting* a calibration pattern to the scene plane, once for each target location. Since we know that the 3D positions of the projected pattern must lie in the scene plane, we can recover the 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 position with the scene plane. We reuse the chessboard pattern—used above for 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 input to a PnP algorithm [27]. While a single image of such a chessboard pattern placed on the floor could be sufficient if the floor is even, we recover a scene 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 an asymmetric circle pattern. We again use a specialized algorithm [26] to obtain the circle centers of the asymmetric circle pattern, capable of recovering 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 1(a)); given the scene plane and such a 2D circle center $\mathbf{x}$, we obtain its 3D correspondence by intersecting the back-projection

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

of $\mathbf{x}$ with the recovered scene plane (cf. Figure 1(b)). We obtain the 2D positions 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 corners—the algorithm that yields 2D circle centers does so in a consistent ordering, we thus obtain respective 2D-3D correspondences for each projector calibration image, which we use to recover the projector calibration matrix $\mathtt{K}_{\mathrm{proj}}$.

Recall that the pose $(\hat{\mathtt{R}}_j, \hat{\mathbf{t}}_j) \in SE(3)$ of the projector, for each projector calibration image j, is provided alongside $\mathtt{K}_{\mathrm{proj}}$, relative to the 3D points of the pattern 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, 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* (i.e., facing directly towards the target location). This is sufficient for our needs in Section 2.2, and 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.

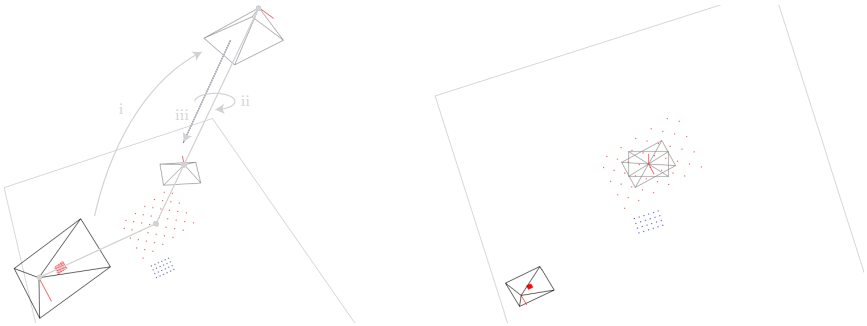
(a) Oblique view (annotated). (b) Downwards view.

Figure 2: Obtaining the virtual camera. Our physical downwards-facing camera—in terms of which we wish to orient our augmentations—is almost certain to not face downwards precisely. Accordingly, we rotate the recovered downwards-facing camera (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. It is then, for each respective target position, the X- and Y-axes of the resulting *virtual* camera (gray) that we subsequently use to orient our virtual projector.

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) 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,[3] 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 original image to the scene plane from the viewpoint of the virtual projector. This warp can be effected

[3] 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).



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

Figure 3: Obtaining the virtual projector. For a given target location, the virtual projector is obtained by (i) rotating the recovered projector (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 (cf. Figure 2), and (iii) translating along the normal direction to achieve the desired metric projected image dimensions.

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. We carry out this placement according to a small handful of steps. First, we rotate the camera about its center of projection to align its optical axis[4] with the normal vector of the scene plane by applying the minimum arc-length rotation that relates the two, giving a *virtual camera* facing directly[5] downwards to the scene plane (cf. Figure 2). It is to the virtual camera's X- and Y-axes that we subsequently set the X- and Y-axes of the virtual projector. To obtain the virtual projector, we (i) intersect the scene plane with the optical axis and rotate the projector's

[4]Strictly speaking, the optical axis is the back-projection of the principal point; in our usage, we understand it to refer to the ray through center of the image plane.

[5]A physical camera placed to face downwards is almost certain 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 genuinely fronto-parallel with respect to the scene plane.

(a) Original image.

(b) Projecting the original image (downwards view).

Figure 4: Projection to the scene plane from the viewpoint of the recovered projector (bottom left, black; virtual projector in center, gray) for a given target location. (a) Original image. (b) Projecting the original image to the scene plane introduces projective distortions if the projector is not fronto-parallel with the scene plane, as is the case here.

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 projected image dimensions provided in metric units. 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 with the desired scale.

It is, strictly speaking, 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 vertical axes of the desired augmentations in accordance with the horizontal and vertical image axes of the 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 facilitating placement of target locations.

(a) Our warped image.

(b) Projecting our warped image (downwards view).

Figure 5: Projection to the scene plane from the viewpoint of the recovered projector (bottom left, black; virtual projector in center, gray) for a given target location. (a) Original image warped using our approach. (b) After warping the original image according to our approach for the given target location, the image is projected with the effect of having the original image from the viewpoint of the virtual projector. This renders the projection free of projective distortions, aligned with the axes of the virtual camera, and with the desired dimensions in the scene plane, expressed in metric units. Note that pixels shown in black will not be projected to the scene.

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 warps 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 5(b)) is given the by the 3×3 matrix

$$\mathtt{H} = \mathtt{K}_{\text{proj}} \left(\mathtt{R} - \frac{\mathbf{t}\mathbf{n}^\top}{d} \right) \mathtt{K}_{\text{proj}}^{-1}, \tag{5}$$

a form of 'plane-induced' 2D homography [16]. For convenience, we enable optional rotation of the image to be projected *before* applying $\mathtt{H}$, about the image center; that rotation, parameterized in degrees, is thus in effect carried

(a) Hardware setup (projector, steerable mirror system, camera).

(b) Example projection of image warped for target location using our approach.

Figure 6: Evaluation scenario. (a) Our hardware setup, comprised of a Panasonic PT-RZ660BE projector, a steerable mirror system manufactured by Dynamic Projection Institute, and a Stereolabs Zed 2 stereo camera (extreme front, black), of which we used only the left view. (b) Example augmentation produced by projecting—to one of our 15 target locations—an image warped using our approach, as seen from the downwards-facing camera. Note that the warp places the axes of the augmentation in accordance with the axes of the camera, and that the dimensions of the augmentation are in line with the desired target dimensions (50 cm × 31.25 cm). Corners of the full projected image extent projected to the floor in light green.

out intuitively with respect to the ground plane.

3. Evaluation

We evaluate our approach by using a 960 pixel × 600 pixel image of the TU Wien logo (cf. Figure 5) to augment 15 locations across the floorspace at the Pilotfabrik[6] 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 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 of 50 cm × 31.25 cm[7]. All experiments were carried out by the same technician,

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

[7] 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

experienced in both approaches.

The hardware setup employed in the evaluation (cf. Figure 6(a)) 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 Projection Institute [9, 10]. 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 downwards-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. 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.

Manual approach. Having pointed the steerable mirror to the desired location, manual keystone correction using the MDC-X software involves warping the image to be projected by manipulating the corners of the image until the desired effect is produced on the projection surface. In order to provide a template for manually carrying out keystone correction, we prepared a rectangular 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 of the scene (cf. the yellow and red lines in Figure 6(b)); we then manually warped the image 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 process of setting the presets in the MDC-X software having carried out keystone correction manually took ca. 32 min for the 15 target locations.

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 pat-

implies 31.25 cm for the vertical axis if aspect ratio is to be preserved.

Table 1: Means and standard deviations (in units of centimeters and pixels, respectively) of deviations in measured lengths from 50 cm for top and bottom and from 21.25 cm for left and right, with respect to the 15 augmentations produced using our approach (intended augmentation dimensions 50 cm × 31.25 cm). Pixel size given the downward-facing camera's height above the ground plane was 3.5 mm; note that in no instance do mean and standard deviation exceed 1 pixel from the viewpoint of the camera.

	Top	Bottom	Left	Right
Mean [cm]	0.03	0.07	0.09	0.06
Stdev [cm]	0.33	0.17	0.17	0.17
Mean [px]	0.1	0.19	0.26	0.18
Stdev [px]	0.95	0.48	0.49	0.49

Table 2: Means and standard deviations of deviations in measured angles from 90° with respect to the 15 augmentations produced using our approach.

	Top Left	Bottom Right
Mean [deg]	-0.03	0.0
Stdev [deg]	0.4	0.34

tern 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 asymmetrical circle pattern image to the location in question 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 calibration module to our projector calibration module. For each of the target locations, the steerable mirror was made to point to that location, the asymmetrical circle pattern image was projected to the scene plane, an image was acquired 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 cali-

bration 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 (cf. Figure 6(b)).

Tables 1 and 2 give the deviations in measured lengths and angles, respectively, from the intended quantities. Lengths were measured using a measuring tape for the top, bottom, left, and right sides of the augmentations; angles were measured using an angle gauge at the augmentations' top-left and bottom-right corners (the measuring tape and angle gauge are both visible in Figure 6(b)). Note that if deviations in measured lengths are expressed in units of pixels of the downwards-facing camera, mean and standard deviation in no instance exceed 1 pixel. 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 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 or focus settings of the camera.

4. Conclusion

We presented a spatial AR system for planar scenes that produces the effect 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 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 com-

pelling results—with mean and standard deviation of errors in measured lengths within the downwards-facing camera's pixel size—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 calibration patterns, their dimensions, and their scale relative to the camera. 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 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).

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] A. E. Uva, M. Gattullo, V. M. Manghisi, D. Spagnulo, G. L. Cascella, 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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[27] T. Collins, A. Bartoli, Infinitesimal plane-based pose estimation, International Journal of Computer Vision 109 (3) (2014) 252–286.