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

Michael Horná\v{c}ek*, Hans Küffner-McCauley, Sebastian Schlund

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

Abstract

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

We propose a system that produces the effect of keystone correction analytically, and that intuitively places the desired augmentations in a manner aligned with the axes of an image of the scene acquired by a camera. Moreover, our system is able to handle a projector equipped with a steerable mirror, enabling wide-area factory floor augmentation exceeding the bounds of the projector's own immediate field of view.

*Corresponding author
Email address: `michael.hornacek@tuwien.ac.at` (Michael Horná\v{c}ek)

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

1. Introduction

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

Using the X- and Y-axes of an image of the scene as a proxy, our contribution is to propose a system that produces the effect of keystone correction analytically, and in a manner aligning the axes of the augmentation with those of the proxy image. We achieve this by distorting the image to be projected using a plane-induced homography computed to produce the effect of projecting the image not from the actual projector viewpoint, but in accordance with the viewpoint of a *virtual* projector (i) facing directly downwards to the scene plane and (ii) rotated to place the axes of the image plane of the virtual projector in line with those of the camera. This enables intuitive placement of augmentations in the scene, and eliminates the need for manual keystone correction. Moreover, a consequence of our approach is that our system is able to handle a projector equipped with a steerable mirror (without need for explicitly modeling the action of the steerable mirror on the projector), thereby enabling wide-area applications.

1.1. Related Work

todo - maybe Majesa would like to help? Sebastian's inputs would be valuable for more general background story.

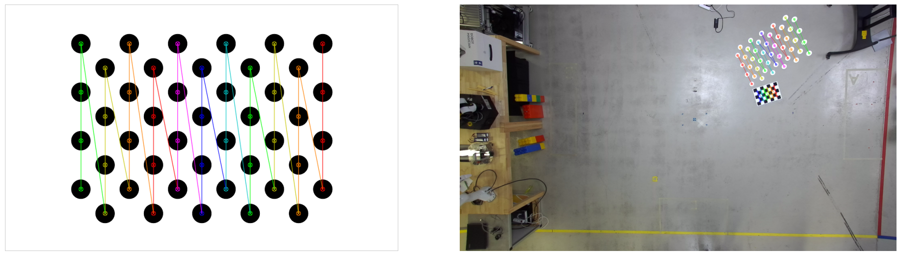
2. Hardware Setup

The hardware setup employed in this work comprised a Panasonic PT-RZ660BE projector with a steerable mirror system manufactured by Dynamic Projection Institute. In addition, we used a Zed 2 stereo camera manufactured by Stereolabs, yet relied only on the left view. The setup was mounted on the ceiling of the Pilotfabrik[1] of TU Vienna, a collaborative space for research on Industry 4.0 topics situated in Vienna, Austria. The floorspace used for our experiments measured dimensions of ca. X m $\times$ Y m; the projector was mounted at approximately the center of this space, at a height of ca. XX m.

3. Approach

Correcting for projective distortions of the sort outlined in Section 1 calls for knowledge of the manner in which the respective rays through the pixels of the projector's image plane fan out into the scene (i.e., we must 'calibrate' the projector) and the geometry of the scene itself (i.e., we must recover the

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



(a) Circles pattern image, in image plane of projector (detected 2D circles pattern center points overlain).

(b) Projector calibration image (one for each target location), in image plane of camera (detected 2D projected circles pattern center points and chessboard corners overlain).

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

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 is given by intersecting its corresponding ray with the geometry of the scene surface. To model this interaction, we (i) carry out a one-time projector calibration, which in our approach calls for additionally calibrating a camera overlooking the scene and includes recovery of the scene plane as a convenient side effect. Next, we use the relative camera-projector-scene plane geometry to (ii) compute a plane-induced homography that distorts the image to be projected in a manner that it appear undistorted to the viewer, and placed in alignment with the axes of a proxy image of the scene. These two points are treated in Sections 3.1 and 3.2, respectively.

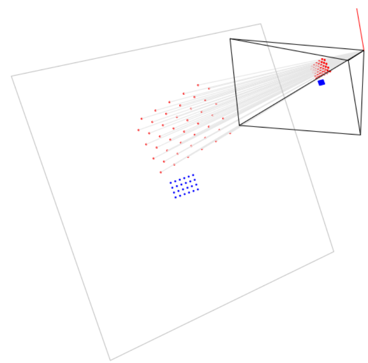


Figure 2: Scene plane (gray) obtained via spatial resection with respect to 2D-3D correspondences obtained using a chessboard pattern (blue); 3D circles pattern points obtained by intersection with the scene plane of back-projections (likewise gray) of circles pattern center points detected in the image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

3.1. Recovering Geometry

Calibration of a projector (or camera) in the sense we employ the term here[2] 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 compute the 'back-projection' of $\mathbf{x}$ (cf. Figure 2), i.e., the ray from the projector's (or camera's) center of projection through $\mathbf{x}$ along which $\mathbf{X}$ must lie. Such a calibration can be expressed in terms of (i) a 3×3 calibration matrix $\mathtt{K}$ derived from the projector's (or camera's) focal length and principal point [7], and (ii) the coefficients of a lens distortion model to correct for radial or tangential distortions caused by the lens system [8].

In practice, calibrating a camera relies on (i) establishing 2D-3D correspondences between pixels in the camera's image plane and corresponding points in the scene, and on (ii) using those correspondences as input to an optimiza-

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

tion procedure that relies on bundle adjustment [9] to output the calibration matrix `K`, the associated lens distortion model coefficients, and, for each calibration image, the pose (i.e., position and orientation) of the camera relative to the 3D points [7, 10]. To calibrate a camera, a calibration surface such as a chessboard pattern is used to identify the correspondences. Calibration of a projector can be carried out in precisely the same manner insofar as step (ii) is concerned; the major difference in projector calibration relative to camera calibration concerns the manner in which 2D-3D correspondences are identified, i.e., between pixels in the image plane of the projector and the corresponding points in the scene. A convenient consequences of the approach we take to identifying the 3D points of the 2D-3D correspondences needed for projector calibration is, for each target location, recovery of the pose (i.e., position and orientation) of (i) the projector and of (ii) the scene plane, both relative to the coordinate frame of the camera.

Camera calibration. The recovery of 2D-3D correspondences in support of camera calibration is carried out by relying on a planar calibration surface to automatically identify correspondences between the 3D points on the calibration surface and their 2D correspondences in the image plane. The classical calibration surface is a chessboard pattern. The 3D corner points of the chessboard are obtained *a priori* in a coordinate system defined in the plane of the chessboard[3], 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 [11]. 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. 2D-3D correspondences are then recov-

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

ered for each calibration image, and the resulting list is passed on as input to an optimization procedure that relies on bundle adjustment to yield the camera calibration matrix $\mathtt{K}_{\mathrm{cam}}$ and the associated lens distortion model coefficients.

Projector calibration. As in the case of camera calibration, projector calibration relies on 2D-3D correspondences, yet we obtain them in this case by projecting a pattern of circles. We rely on an algorithm to detect the circle pattern center points in the circles pattern image in the image plane of the projector, giving the 2D positions of our 2D-3D correspondences (cf. Figure 1(a)). We project the circles pattern image to each of the target locations, and use the camera to acquire a projector calibration image for each. Given a projector calibration image, we again detect the centers of the *projected* circles pattern (cf. Figure 1(b)); given the scene plane (the recovery of which we shall turn to in the paragraph that follows) and such a 2D circle center $\mathbf{x}$, its 3D correspondence is obtained by intersecting the back-projection of $\mathbf{x}$ with the scene plane (cf. Figure 2). Since the algorithm that yields 2D circle centers does so in a consistent ordering, we thus obtain the 2D-3D correspondences between the projector's image plane and the scene required for projector calibration, yielding the projector calibration matrix $\mathtt{K}_{\mathrm{proj}}$.

We recover the scene plane via spatial resection by applying a PnP algorithm [12] to the with respect to 2D-3D correspondences of a chessboard pattern. Note that this step is separate from camera calibration, yet could well be carried out using the same calibration pattern used in the camera calibration step.[4] While a single image of such a chessboard pattern placed on the floor could be sufficient if the floor is even, we place a chessboard pattern in close proximity to the projected circles pattern in each projector calibration image in order to recover the scene plane locally to each target location, in order to account for the possibility of an uneven floor.

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

The pose (i.e., position and orientation) of the projector, for each projector calibration image, is provided alongside $\mathtt{K}_{\text{proj}}$ by the aforementioned optimization procedure. Note that for a fixed projector with steerable mirror, given a projector calibration image, the recovered projector's pose is the pose the projector would have to have had to project to the given target location *in the absence of the mirror*. As this is sufficient for our needs in Section 3.2, it is in this sense that our system is able to handle a projector equipped with a steerable mirror, without need for modeling the steerable mirror explicitly.

3.2. Correcting for Projective Distortion

If the projector is calibrated and the scene plane is known in the coordinate frame of the projector, a 'virtual' projector (with the same calibration $\mathtt{K}$ and lens distortion coefficients) can be placed elsewhere relative to the scene plane. The projection of an image from the virtual projector to the scene plane can in turn be projected to the image plane of the original projector;

Virtual projector.

Plane-induced homography. Let $\mathtt{K}_{\text{proj}}$ express the 3×3 calibration matrix of the projector and $(\mathtt{R}', \mathbf{t}') \in SE(3)$ the rigid body transformation that transforms points from the coordinate frame of the projector to that of the virtual projector.

$$\mathtt{H} = \mathtt{K}_{\text{proj}} \left(\mathtt{R}' - \frac{\mathbf{t}'\mathbf{n}^\top}{d} \right) \mathtt{K}_{\text{proj}}^{-1}, \tag{1}$$

where $(\mathbf{n}^\top, -d)^\top$ gives the scene plane and $d = \mathbf{n}^\top \mathbf{X}$, for any point $\mathbf{X}$ in the plane.

4. Evaluation

todo - maybe Hans would like to help?

5. Conclusion

We presented a system that produces the effect of keystone correction analytically, and that intuitively places the desired augmentations in a manner aligned with the axes of an image of the scene acquired by a camera. Moreover, we showed our system to be able to handle a projector equipped with a steerable mirror, enabling wide-area factory floor augmentation exceeding the bounds of the projector's own immediate field of view. Our evaluation demonstrated ...

References

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

[2] F. Zhou, H. B.-L. Duh, M. Billinghurst, Trends in augmented reality tracking, interaction and display: A review of ten years of ismar, in: 2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, IEEE, 2008, pp. 193–202.

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

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

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

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

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

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

[9] 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.

[10] Z. Zhang, A flexible new technique for camera calibration, IEEE Transactions on pattern analysis and machine intelligence 22 (11) (2000) 1330–1334.

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

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