

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

Michael Hornáček, 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 work step 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 AR glasses. 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 projector's own immediate field of view.

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

1. Introduction

Augmented reality (AR) promises to enable use cases in industrial settings that include the embedding of work step instructions directly into the scene, 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 AR glasses. Reliance on AR glasses has two adverse consequences: (i) AR glasses must be worn by each individual wishing to view the augmentation, and (ii) such glasses—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 AR glasses, but by projection to the scene itself, 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.

35 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-
40 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 Pilotfabrik¹ of TU Vienna, a collaborative space for research on Industry 4.0 topics situated in Vienna, Austria. The floorspace used for our
45 experiments measured dimensions of $X \text{ m} \times Y \text{ m}$; the projector was mounted at approximately the center of this space, at a height of ca. $XX \text{ m}$.

3. Approach

Correcting for projective distortions of the sort outlined in Section 1 requires knowledge of the manner in which the respective rays through the pixels of the
50 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 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
55 surface. To model this interaction calls for (i) a one-time projector calibration,

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 of several), 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. Since the projected circles pattern lies in this plane, we are in a subsequent step able to rely on that plane to derive the 3D positions corresponding to the 2D projected circles pattern center points detected in the image.

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, (ii) the relative camera-projector-scene plane geometry is used to compute a plane-induced homography that distorts the image to be projected in a manner
60 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.

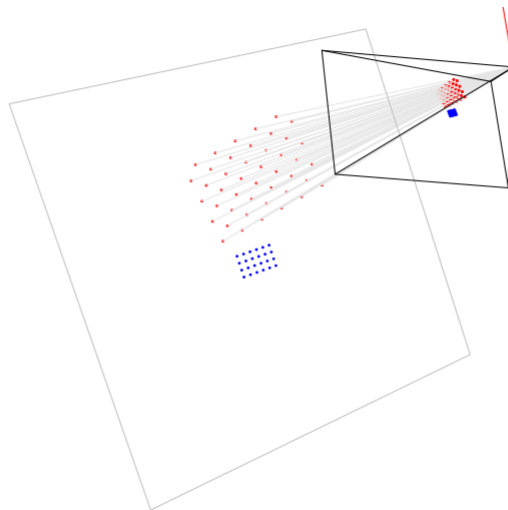


Figure 2: Scene plane via camera resection with respect to chessboard correspondences (blue); 3D circles pattern points via intersection of back-projections of detected circles pattern points with scene plane (red).

3.1. Recovering Geometry

Calibration of a projector (or camera) in the sense we employ the term here² entails that one become able to project a scene point \mathbf{X} to its corresponding pixel \mathbf{x} in the projector’s (or camera’s) image plane, or compute the ‘back-projection’ of \mathbf{x} , the ray from the projector’s (or camera’s) center of projection \mathbf{C} through \mathbf{x} along which \mathbf{X} must lie. Such a calibration can be expressed in terms of (i) a 3×3 calibration matrix \mathbf{K} derived from the projector’s (or camera’s) focal length and principal point, and (ii) the coefficients of a lens distortion model to account for radial or tangential distortions caused by the lens system.

In practice, calibrating a camera relies on (i) establishing 2D-3D correspondences between pixels in the camera’s image plane and corresponding points in the scene, and on (ii) using those correspondences as input to an optimization procedure that outputs the calibration matrix \mathbf{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. 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 consequence of the approach we take to identifying the 3D points of the 2D-3D correspondences needed for projector calibration is, for each target location, recovery of the pose (i.e., position and orientation) of (i) the projector and of (ii) the scene plane, both relative to the coordinate frame of the camera.

Camera calibration. The recovery of 2D-3D correspondences in support of camera calibration is carried out, conventionally, by relying on a planar calibration surface of known scale to automatically identify correspondences between the

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

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 chessboard³, requiring knowledge only of the dimensions
95 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. 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
100 are then recovered for each calibration image, and the resulting list is passed on as input to an optimization procedure that relies on bundle adjustment to yield the camera calibration matrix K_{cam} and the associated lens distortion model coefficients.

Projector calibration. As in the case of camera calibration, projector calibration
105 relies on 2D-3D correspondences, yet in this case we obtain them 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. We project the circles pattern image to each of the target locations, yielding a projector calibration
110 image for each. Given a calibration image, we can detect the centers of the *projected* circles pattern; if we know the scene plane, then given such a 2D circle center \mathbf{x} , its 3D correspondence is obtained by intersecting the back-projection of \mathbf{x} with the scene plane. Since the algorithm that yields 2D circle centers does so in a consistent ordering, we thus obtain the 2D-3D correspondences required
115 for projector calibration, yielding the projector calibration matrix K_{proj} .

We recover the scene plane via camera resection by applying a PnP algorithm

³E.g., $(0, 0, 0), (1.5, 0, 0), (3, 0, 0), \dots, (9, 7.5, 0)$ 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.

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 pattern used in the camera calibration step.⁴ While a single
120 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.

125 The pose (i.e., position and orientation) of the projector, for each projector calibration image, is provided alongside K_{proj} by the aforementioned optimization procedure based on bundle adjustment. 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 to project to the given target
130 location *in the absence of the mirror*.

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 K and lens distortion coefficients) can be placed elsewhere relative to the scene plane.
135 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 K_{proj} express the 3×3 calibration matrix of the projector and $(R', \mathbf{t}') \in SE(3)$ the rigid body transformation that transforms
140 points from the coordinate frame of the projector to that of the virtual projector.

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

⁴The 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.

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?

145 5. Conclusion

References