

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, intuitively placing 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.

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.

1.1. Related Work

35 todo - maybe Majesa would like to help?

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
40 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 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. $XX \text{ m}$.

45 3. Approach

The problem of correcting for distortions as outlined in Section 1 requires 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
50 scene plane) 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 calls for (i) a one-time projector calibration, which in our approach calls for additionally calibrating a camera overlooking the scene
55 and includes recovery of the scene plane as a convenient side effect. Next, (ii)

¹<https://www.pilotfabrik.at/>

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 that it appear undistorted to the viewer, and placed in alignment with the axes of a proxy image of the scene. These to points are treated in Sections 3.1 and 3.2, respectively.

3.1. Recovering Geometry

The ‘output’ of projector (or camera) calibration is a mathematical model that determines, for each pixel x_{proj} in the projector’s (or camera’s) image plane $proj$, the ray from the projector’s center of projection C_{proj} through x_{proj} , along which the pixel is projected to the scene. Inversely, the model also determines the projection \mathbf{x}'_{proj} of any point \mathbf{X}'_{proj} in the scene to the projector’s (or camera’s) image plane π_{proj} , obtained by intersecting π_{proj} with the ray from \mathbf{C}_{proj} through \mathbf{X}'_{proj} . Such a model is said to function in accordance with the principle of central projection.

Notably, central projection is the same principle that underlies the modeling of conventional cameras. A camera modeled in accordance with the principle of central projection is called a pinhole camera. A pinhole camera can be expressed in terms of a 3×3 camera calibration matrix K , derived from the camera’s focal length and principal point. In practice, a pinhole camera model works well in modeling how the respective rays through the pixels of the image plane fan out into the scene, on the condition that images to which the model is applied have been corrected for lens distortions caused by the camera’s lens system. The ‘output’ of camera calibration is accordingly not only the matrix K , but the associated lens distortions model coefficients as well. The lens distortion model coefficients enable applying a lens distortion model to undistort any image acquired by the camera.

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 K and the associated distortion

model coefficients. 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
90 projector and the corresponding points in the scene.

Camera calibration.

Scene plane recovery.

Projector calibration. The pose of the projector is given in the coordinate frame of the camera.

95 3.2. Correcting for Distortion

Virtual projector.

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

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

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

4. Evaluation

todo - maybe Hans would like to help?

5. Conclusion

105 References