MEASURING IN IMAGES WITH PROJECTIVE GEOMETRY

Bastian Erdnüß[1,2]

[1] Institute of Photogrammetry and Remote Sensing, KIT, Karlsruhe, Germany - bastian.erdnuess@kit.edu
[2] Fraunhofer Institute of Optronics, System Technologies and Image Exploitation, Karlsruhe, Germany

Commission I, WG I/9

KEY WORDS: projective geometry, affine geometry, barycentric coordinates, barycentric ratio, cross-ratio, camera pose estimation, 3D point reconstruction

ABSTRACT:

There is a fundamental relationship between projective geometry and the perspective imaging geometry of a pinhole camera. Projective scales have been used to measure within images from the beginnings of photogrammetry, mostly the cross-ratio on a straight line. However, there are also projective frames in the plane with interesting connections to affine and projective geometry in three dimensional space that can be utilized for photogrammetry. This article introduces an invariant on the projective plane, describes its relation to affine geometry, and how to use it to reduce the complexity of projective transformations. It describes how the invariant can be use to measure on projectively distorted planes in images and shows applications to this in 3D reconstruction. The article follows two central ideas. One is to measure coordinates in an image relatively to each other to gain as much invariance of the viewport as possible. The other is to use the remaining variance to determine the 3D structure of the scene and to locate the camera centers. For this, the images are projected onto a common plane in the scene. 3D structure not on the plane occludes different parts of the plane in the images. From this, the position of the cameras and the 3D structure are obtained.

1. INTRODUCTION

Photogrammetry deals with images and an image can be considered as a recording of spectral properties of rays. In the simplest case (pinhole camera) the rays meet in a point (camera center), are equally distributed (no distortion), straight (no atmospheric or relativistic distraction), and not subject to optical issues like out-of-focus blur. Projective geometry deals with the one dimensional subspaces (lines through the origin) of a vector space. It appropriately describes the imaging geometry of the pinhole camera by identifying the camera center with the origin of the vector space and the rays of light with the corresponding lines through the origin.

In order to reconstruct the depicted scene, the usually way is to calibrate the camera and to find its position and orientation relative to the scene. This way, a ray of light through the scene can be associated to each image pixel. New scene points can be triangulated by intersecting the rays of light from different images. This approach is especially suited when the imaging geometry of a camera does not change and only one set of parameters has to be estimated for the whole sequence of images. But often, the imaging geometry of the camera changes significantly between frames. This is the case e. g. for variable zoom cameras or for cameras with a software image stabilization that warps each frame to better align with the previous frame. In this case it might be more suitable to project the images onto a common plane in the scene instead of their own sensor planes. The advantage is that the common plane has to be "calibrated" just once for all images.

The first part of the article describes the connection between affine and projective geometry. Applying affine geometry to projective geometry shows that there is just a very small delta between them. In some sense, projective transformations diagonalize under some kind of affine transformation. A projective

invariant will be developed based on this observation. The invariant is called *barycentric ratio*. The barycentric ratio will be used to measure projectively invariant in images. This article mainly concerns measurements on a plane. But it will also be shown how the cross-ratio on a line relates to the barycentric ratio. Moreover, there are also applications to the barycentric ratio in 3D space (Erdnüß, 2017).

The second part of the article explains how projections on common planes of a scene can be used to reconstruct the depicted scene. The main idea is that a plane in a scene can conceptually be extended to cover the whole image plane. Therefore, each scene point in the scene intersects the plane and the location of the intersection on the plane can be determined. The camera center, the scene point, and its known projection point onto the plane are collinear. The article shows how this information can be used for the triangulation of the camera centers and scene points. The resulting equation system will be linear and provides a solution without requiring an initial estimate.

2. PROJECTIVE SCALES

A projective scale refers to a method to measure within an image. Conceptually, the scale is projected into the image and measurements are read from it there. Figure 1 shows a ruler at the bottom that is projectively distorted. Its scale is not longer affine like the scale of the ruler at the top where same distances equal the same number of ticks on the scale. The first part of this article deals with the question how to measure on a projective scale by just using an affine scale (like the top ruler in Figure 1).

2.1 Homogeneous Coordinates

It is common in computer vision to represent a point $(x, y)^t$ in homogeneous coordinates $(x, y, 1)^t$ by appending a one as ad-

ditional component (Hartley and Zisserman, 2004, p. 2). Scaling a vector of homogeneous coordinates by a real number does not change the point it refers to. This is motivated by the idea that a pinhole camera located at the origin projects a point $P = (X, Y, Z)^t$ in space along the line through the origin and P onto a screen at $z = 1$. On the screen the projected point has coordinates $(X/Z, Y/Z, 1)^t$. Conversely, a point measured at the coordinates $(x, y, 1)^t$ on the screen corresponds to a point $(kx, ky, k)^t$ in space with some non-zero distance factor k.

The choice to put the screen at position $z = 1$ is arbitrary but conventional. The line corresponding to a homogeneous vector with 0 as the last component will never intersect the screen at $z = 1$. In projective geometry one interprets this as an intersection at infinity and calls such an object a *direction* instead of a *point*. E. g. the homogeneous vector $(1, 0, 0)^t$ corresponds to the direction of the x-axis, the homogeneus vector $(0, 1, 0)^t$ corresponds to the direction of the y-axis and the homogeneous vector $(0, 0, 1)^t$ is a point on the screen and corresponds to its origin.

The mapping from $(x, y)^t$ to $(x, y, 1)^t$ is called homogenization and the mapping from $(X, Y, Z)^t$ to the point $(X/Z, Y/Z)^t$ is called dehomogenization.

2.2 Barycentric Coordinates

In the vector space $\mathbb{R}^2$ each vector x can be decomposed into a linear combination of two basis vectors a and b:

$$x = \alpha \cdot a + \beta \cdot b. \tag{1}$$

Let $B = (a, b)$ denote the matrix composed of the two basis vectors. One can find the coordinates $(\alpha, \beta)^t$ of a vector x by the inverse of B:

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} = B^{-1} x. \tag{2}$$

The affine plane is closely related to the vector space $\mathbb{R}^2$. Additionally, it has an explicit origin O. The point $X = O + x$ can be rewritten in terms of the points $A = O + a$ and $B = O + b$ using Equation (1) as (cf. Figure 2):

$$\begin{aligned} X &= O + \alpha \cdot (A - O) + \beta \cdot (B - O) \\ &= \alpha \cdot A + \beta \cdot B + (1 - \alpha - \beta) \cdot O. \end{aligned} \tag{3}$$

The coefficients $\boldsymbol{\xi} = (\alpha, \beta, 1 - \alpha - \beta)^t$ of the basis points A, B, and O are the *barycentric coordinates* of X with respect to the triangle A, B, O.
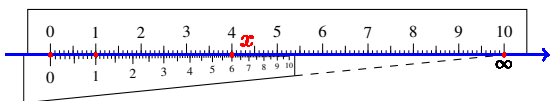

Figure 1. An affine scale (top) and a projective scale (bottom). Both scales agree at 0 and 1. The bottom ruler would already show infinity where the top rule is just at 10. In this case the bottom ruler shows $\frac{9x}{10-x}$ when the top ruler shows x, e. g. the bottom ruler shows $6 = \frac{9 \cdot 4}{10 - 4}$ when the top ruler shows $x = 4$. In general, the value of a projective scale on a line can be found by measurements of an affine scale using the cross-ratio (cf. Section 2.5).

Barycentric coordinates usually sum up to 1, but this requirement can be relaxed by understanding them as homogeneous coordinates. They allow the same interpretation as the homogeneous coordinates in Section 2.1 as lines through the origin. But they do not project onto the screen at $z = 1$. Instead they project onto the screen at $x + y + z = 1$. This screen contains all three unit vectors $(1, 0, 0)^t$, $(0, 1, 0)^t$, and $(0, 0, 1)^t$. The first two unit vectors correspond to the unit point of the x- and y-axis, and the last one corresponds to the origin. The choice to put the coefficient of the origin in the last position is arbitrary but complies with the convention in Section 2.1.

The homogeneous coordinates in Section 2.1 could be dehomgenized by dividing them trough their last component. In opposite, barycentric coordinates have to be divided by their sum for dehomogenization. The last component can than be stripped since it will always be 1 minus the sum of the remaining components.

2.3 The Barycentric Ratio

A projective transformation, or homography, is a linear transformation H of homogeneous coordinate vectors. It is uniquely defined by four point correspondences where no three points are collinear. Let

$$\begin{aligned} A &\leftrightarrow A', & B &\leftrightarrow B', \\ C &\leftrightarrow C', & D &\leftrightarrow D' \end{aligned} \tag{4}$$

be such correspondences. Consider the perspective transformation that maps barycentric coordinates with respect to the triangle A, B, C to barycentric coordinates with respect to the triangle A', B', C'. Since $(1, 0, 0)^t$ are the barycentric coordinates of A in the first coordinate system and also the barycentric coordinates of A' in the second coordinate system, H must map $(1, 0, 0)^t$ to a multiple of itself. In other words, $(1, 0, 0)^t$ is an eigenvector of H. Analogously, $(0, 1, 0)^t$ and $(0, 0, 1)^t$ are eigenvectors of H. Therefore, H must be diagonal.

Let $(\alpha, \beta, 1 - \alpha - \beta)^t$ be the barycentric coordinates of D with respect to A, B, C and $(\alpha', \beta', 1 - \alpha' - \beta')^t$ be the barycentric coordinates of D' with respect to A', B', C'. H has to map the first set of coordinates to a multiple of the second one, therefore H must itself be a multiple of the diagonal matrix

$$H \sim \text{diag}\left(\frac{\alpha'}{\alpha}, \frac{\beta'}{\beta}, \frac{1 - \alpha' - \beta'}{1 - \alpha - \beta} \right). \tag{5}$$

This formulation is also used in computer graphics for efficient texture mapping (Blinn, 2003, chap. 13).

Let $X \leftrightarrow X'$ be another correspondence with barycentric coordinates $(\xi, \eta, 1 - \xi - \eta)^t$ and $(\xi', \eta', 1 - \xi' - \eta')^t$, respectively.

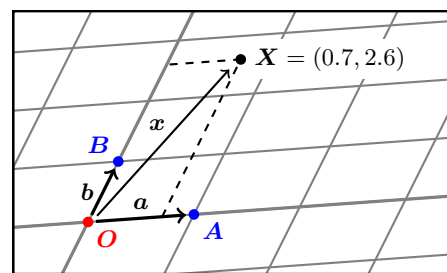Figure 2. Affine grid. Vector x can be written as linear combination $x = 0.7 \cdot a + 2.6 \cdot b$ of the vectors a and b. Moreover, $(0.7, 2.6, 1 - 0.7 - 2.6)^t$ are the barycentric coordinates of X with respect to the triangle A, B, O.
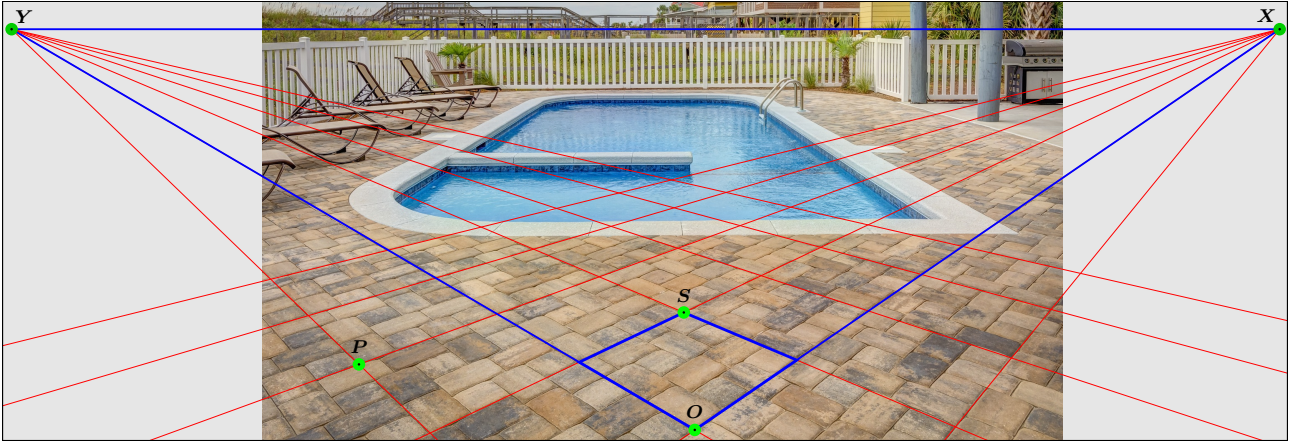
Figure 3. Pool and Pavement (Photo by GregoryButler, 2013). Vanishing points X and Y are estimated by the joints in the pavement. O and S are arbitrary chosen and define the origin and unit square. The position of an arbitrary point P on the ground plane can be found by the barycentric ratio $(X, Y, O; S, P)$.

Applying H from Equation (5) to the barycentric coordinates of X and rearranging α', β', and $1 - \alpha' - \beta'$ to the left-hand side yields

$$\left(\frac{\xi'}{\alpha'}, \frac{\eta'}{\beta'}, \frac{1 - \xi' - \eta'}{1 - \alpha' - \beta'} \right)^t \sim \left(\frac{\xi}{\alpha}, \frac{\eta}{\beta}, \frac{1 - \xi - \eta}{1 - \alpha - \beta} \right)^t . \quad (6)$$

The value of the equation is independent of H since the right-hand side depends only on A, B, C, D, and X. Therefore, it provides a projective invariant. The numerators of Equation (6) contain the barycentric coordinates of X and the denumerators contain the barycentric coordinates of D. This motivates the name *barycentric ratio* of X to D with respect to the triangle A, B, C for the term in Equation (6) (Erdnüß, 2017). In the following, it will be denoted as

$$(A, B, C; D, X), \quad (7)$$

with the barycentric basis A, B, C in front of the semicolon and the reference point and target point following. Barycentric coordinates and the barycentric ratio are not limited to the plane but directly extend to arbitrary dimension (Erdnüß, 2017).

2.4 An Interpretation of the Barycentric Ratio

Figure 3 shows a pool in the backyard. The image is dominated by the ground plane. Two of its vanishing points X and Y can be estimated by the joints in the paving. The point O marks the (arbitrarily) chosen origin of a perspective grid on the ground plane with axes through X and Y. The point S defines the scale on the plane and is located diagonally opposite to the origin on the unit square (drawn in blue). The point P is another (arbitrary chosen) point. It happens to have the coordinates $(-1, 2)^t$ in the projective coordinate system defined by X, Y, O, and S, i.e. starting from the origin O, the point P is two units towards the vanishing point Y and one unit away from the vanishing point X.

The position of P on the perspective grid can be calculated by the barycentric ratio of P to S with respect to the triangle X, Y, O. In Figure 3 the barycentric coordinates of S with respect to the triangle X, Y, O are $(0.149, 0.143, 0.707)^t$ and those of P are

$(-0.176, 0.340, 0.836)$. Their ratio is given by

$$(X, Y, O; S, P) = \left(\frac{-0.176}{0.149}, \frac{0.340}{0.143}, \frac{0.836}{0.707} \right)^t \quad (8)$$
$$= (-1.18, 2.36, 1.18)^t .$$

To better understand how to interpret this result, consider the following barycentric ratios

$$\begin{aligned}
(X, Y, O; S, X) &\sim (1, 0, 0)^t \\
(X, Y, O; S, Y) &\sim (0, 1, 0)^t \\
(X, Y, O; S, O) &\sim (0, 0, 1)^t \\
(X, Y, O; S, S) &\sim (1, 1, 1)^t .
\end{aligned} \quad (9)$$

As stated in Section 2.1, in the usual interpretation of homogeneous coordinates as projected onto the screen $z = 1$, the homogeneous vector $(1, 0, 0)^t$ corresponds to the direction of the x-axis. That corresponds to the interpretation that X is the vanishing point of the x-axis in Figure 3. The same holds true for Y corresponding to the vanishing point of the y-axis, and for O corresponding to the origin. The homogeneous vector $(1, 1, 1)^t$ corresponds to the point $(1, 1)^t$ that is located diagonally opposite to the origin on the unit square. That corresponds to the interpretation of S as well. Therefore, dehomogenizing the homogeneous vector in Equation (8) by dividing all coordinates by the last component yields the coordinates $(-1, 2)^t$ of the point P on the distorted projective grid defined by X, Y, O, and S.

2.5 An Application of the Cross-Ratio

The cross-ratio can be used to measure in a perspectively distorted image. Equally spaced dots on a line in space captured by a photo appear closer together the further they are away from the camera.

Figure 4 shows a street scene in Paris. Some parallel lines on the facade of the left building were identified, and their corresponding vanishing point was estimated. The borders of the balcony railings on the second floor are marked with green dots and the length of the railing on the balcony above the main door is set to unit length.

The lines on the facade appear perspectively distorted in the image, similar to the ruler on the bottom of Figure 1. The rectified
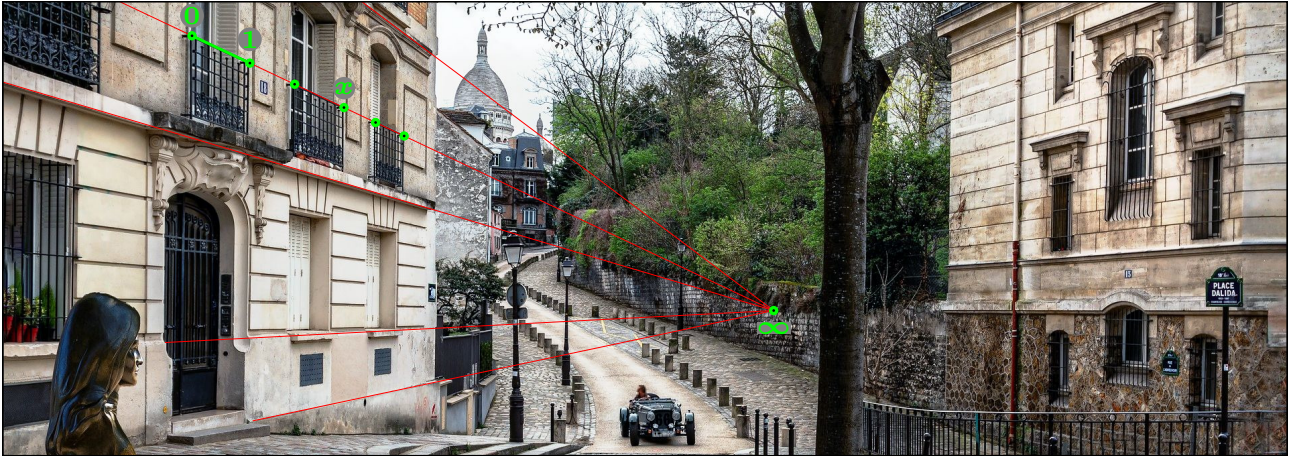
Figure 4. Place Dalida, Paris (Photo by endmondlafoto, 2018). Red lines show parallel lines found on the facade of the left building meeting in their vanishing point (∞). The other greed dots are placed on a line and mark the edges of the ralings on the second floor of the building. The length of the first railing is defined as unit length. The rectified position of an arbitrary other point $\boldsymbol{x}$ on that line can be found by the cross-ratio $(\infty, \boldsymbol{0}; \boldsymbol{1}, \boldsymbol{x})$.

position ξ of an arbitrary point $\boldsymbol{x}$ on the line with the green dots is given by the cross-ratio

$$\xi = (\infty, \boldsymbol{0}; \boldsymbol{1}, \boldsymbol{x}) = \frac{\|\boldsymbol{x} - \boldsymbol{0}\|}{\|\boldsymbol{1} - \boldsymbol{0}\|} : \frac{\|\infty - \boldsymbol{x}\|}{\|\infty - \boldsymbol{1}\|}. \qquad (10)$$

Here, the bold letters $\boldsymbol{0}, \boldsymbol{1}, \boldsymbol{x}$ and ∞ denote the vectors with the image coordinates of the corresponding points in an arbitrary (affine) grid (e. g. in the pixel grid of the image).

Consider the line segment between $\boldsymbol{0}$ and ∞ and the ratios in which the segment is divided by the two points $\boldsymbol{x}$ and $\boldsymbol{1}$, respectively. Using a straight edge on the print of Figure 4, the division ratios are roughly

$$\begin{aligned} \|\boldsymbol{x} - \boldsymbol{0}\| : \|\infty - \boldsymbol{x}\| &\approx 0.26 : 0.74 \\ \|\boldsymbol{1} - \boldsymbol{0}\| : \|\infty - \boldsymbol{1}\| &\approx 0.10 : 0.90 \,. \end{aligned} \qquad (11)$$

Note that the cross-ratio in Equation (10) relates these two ratios to each other:

$$\xi = \frac{0.26}{0.1} : \frac{0.74}{0.9} \approx 2.6 : 0.82 \approx 3.17 \,. \qquad (12)$$

The point $\boldsymbol{1}$ divides the line segment between $\boldsymbol{0}$ and ∞ in the ratio $0.1 : 0.9$ and the barycentric coordinates of $\boldsymbol{1}$ with respect to ∞ and $\boldsymbol{0}$ are $(0.1, 0.9)^t$. The two concepts are very closely related. Moreover, the barycentric ratio of $\boldsymbol{x}$ to $\boldsymbol{1}$ with respect to ∞ and $\boldsymbol{0}$ is

$$(\infty, \boldsymbol{0}; \boldsymbol{1}, \boldsymbol{x}) = \left(\frac{0.26}{0.1}, \frac{0.74}{0.9} \right)^t. \qquad (13)$$

Dehomogenizing this vector by dividing the first value by the second value leads to the same result as in Equation (12). That justifies the use of the same notation for the cross-ratio and the barycentric ratio, and allows to understand the barycentric ratio as a generalization of the cross-ratio.

2.6 Other Interpretations of the Barycentric Ratio

The previous section showed a typical application of the cross-ratio. Given a line and four points on the line, the cross-ratio is invariant under perspective transformations. Given a certain interpretation of the four points on the line, one can interpret the cross-ratio as the rectified position of one point with respect to a scale defined by the other three points. Here, the scale was defined by the vanishing point, the origin, and the unit point. The unit point together with the origin fixed the scale on projectively distorted line.

Section 2.4 showed a similar interpretation for a plane with five points on it, where four of them defined the scale; two vanishing points, the origin, and one unit point defining the size of the unit square. The barycentric ratio of the fifth point and the four points fixing the scale could be interpreted as the rectified coordinates of the fifth point on that scale. But that's not the only interpretation possible.

Equation (9) states that the four basis points of the projective grid correspond to the three unit vectors and the vector with all components equal to 1. The interpretation stated in the last paragraph is based on the mental image that the homogeneous vectors would be projected onto the screen $z = 1$. Different interpretations are possible when projecting on other screens.

Section 2.2 already introduced the screen $x + y + z = 1$ where the first two basis points correspond to the unit points on the axes instead of their vanishing points. However, the last basis point corresponded to the barycenter of the unit triangle (bounded by the origin and the other two basis points). To project $(1, 1, 1)^t$ on the screen $x + y + z = 1$ one has to divide the vector by the sum of its components $1 + 1 + 1 = 3$. The first two components of the resulting vector $(1/3, 1/3)^t$ are then the affine coordinates of the projected point.

A very handy screen is given by the equation $x + y - z = 1$, where the basis points correspond to the unit square. As with the screen $x + y + z = 1$, the first three basis points correspond to the unit point on the x-axis, the unit point on the y-axis, and the origin, respectively. However, the last unit point is not the barycenter of the unit triangle, but instead completes the unit square. The point $(1, 1, 1)^t$ is already on the screen since $1 + 1 - 1 = 1$. Stripping off the last component of the projected point leads to the coordinates $(1, 1)^t$ of the last point on the unit square.

Hence, let $\boldsymbol{O}, \boldsymbol{X}, \boldsymbol{S}, \boldsymbol{Y}$ be a projectively distorted square (cf. Figure 5) and $\boldsymbol{P}$ some point on it (or outside of it), when
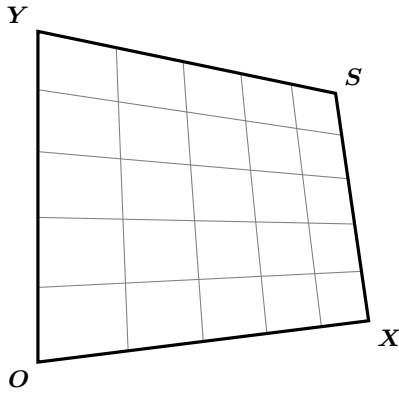
Figure 5. Perspectively distorted square.

$(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{O}; \boldsymbol{S}, \boldsymbol{P}) = (\xi, \eta, \zeta)^t$ is their barycentric ratio then

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{\xi + \eta - \zeta} \begin{pmatrix} \xi \\ \eta \end{pmatrix} \qquad (14)$$

are the rectified coordinates of $\boldsymbol{P}$.

3. PLANES IN AN IMAGE

The previous chapter demonstrated how to use a projective invariant to measure on a projective grid defined by some control points in an image. This can be used to construct grids on planes in images that are invariant of the camera pose. A structure in an image defining a plane might only cover a small region of the image; however, the plane can conceptionally always be extended to cover the whole image plane. This turns out to be very valuable, since any scene point seen in the image projects on that plane. I. e. one can exactly determine the coordinate on the plane, in front of which the scene point appears in the image. The camera center, the scene point, and the projected point onto the plane are collinear. This is a very precise information about the positions of the camera and the scene points. It is comparable to aiming a riffle with notch and bead sights, that have to be aligned with the target to make sure the barrel points in the right direction.

3.1 The Projection onto the Plane at Infinity

Consider a scene in front of the night sky (cf. Figure 6). The night sky can be understood as the plane at infinity and the position of a scene point relative to the night sky holds some useful information. E. g. the green point in Figure 6 (corner of the building) is located at the right ascension of about $11^h 30^m$ and declination of about $40°$ in the celestial coordinate system in which astronomer report the positions of the stars. This is a spherical coordinate system and the direction of $(11^h 30^m, 40°)$ corresponds to the unit vector

$$\boldsymbol{v} = \begin{pmatrix} \cos(11^h 30^m)\cos(40°) \\ \sin(11^h 30^m)\cos(40°) \\ \sin(40°) \end{pmatrix} \approx \begin{pmatrix} -0.76 \\ 0.01 \\ 0.64 \end{pmatrix}. \qquad (15)$$

The direction vector $\boldsymbol{X} - \boldsymbol{C}$ from the camera center $\boldsymbol{C}$ to the corner of the building $\boldsymbol{X}$ is a multiple of $\boldsymbol{v}$. While the points $\boldsymbol{C}$ and $\boldsymbol{X}$ are unknown, the direction vector $\boldsymbol{v}$ can be calculated directly from the projection of the scene point onto the plane at infinity in the image.

A reconstruction of the scene can be done by just the direction vectors $\boldsymbol{v}$ of the projections onto the plane at infinity. Consider

several images of the scene taken at the same time (i. e. the scene did not move with respect to the night sky). Number the images by $n = 1, \ldots, N$ and denote the (unknown) projection center of the respective camera by $\boldsymbol{C}_n$. Find some point correspondences in the images, number them by $m = 1, \ldots, M$, and denote the corresponding (unknown) scene points by $\boldsymbol{X}_m$. Let $\boldsymbol{v}_{mn}$ be the direction vector between camera center $\boldsymbol{C}_n$ and scene point $\boldsymbol{X}_m$. I. e. calculate $\boldsymbol{v}_{mn}$ by Equation (15) from where the scene point $\boldsymbol{X}_m$ appears to be in image n, relatively to the night sky. As stated before, $\boldsymbol{X}_m - \boldsymbol{C}_n$ is a multiple of $\boldsymbol{v}_{mn}$, i. e.

$$\boldsymbol{X}_m - \boldsymbol{C}_n = \lambda_{mn} \boldsymbol{v}_{mn} \qquad (16)$$

for some real valued scaling factor λ_{mn}. This type of equation system can be solved by a direct linear transformation (DLT). Let

$$P_{mn}^{\perp} = I - \frac{\boldsymbol{v}_{mn} \boldsymbol{v}_{mn}^t}{\|\boldsymbol{v}_{mn}\|^2} \qquad (17)$$

be the orthogonal projection onto the plane through the origin with normal vector $\boldsymbol{v}_{mn}$. Here, I denotes the 3-by-3 identity matrix . The optimal solution to Equation (16) is then given by

$$\begin{aligned} \left(\sum_{m=1}^{M} P_{mn}^{\perp} \right) \boldsymbol{C}_n &= \sum_{m=1}^{M} P_{mn}^{\perp} \boldsymbol{X}_m \quad \text{for } n = 1, \ldots, N\,, \\ \left(\sum_{n=1}^{N} P_{mn}^{\perp} \right) \boldsymbol{X}_m &= \sum_{n=1}^{N} P_{mn}^{\perp} \boldsymbol{C}_n \quad \text{for } m = 1, \ldots, M\,. \end{aligned} \qquad (18)$$

This is an homogeneous system of linear equations in the camera centers $\boldsymbol{C}_n$ and the scene points $\boldsymbol{X}_m$. Its nullity is at least four. This reflects the fact that the origin of the reconstruction (three degrees of freedom) and the overall scale (one degree of freedom) cannot be determined by the data. Choosing an arbitrary origin (e. g. by requiring $\sum_{m=1}^{M} \boldsymbol{X}_m = \boldsymbol{0}$ as additional linear equations) usually leads to a nullity of one (for non-singular data), and choosing any non-zero solution gives a reconstruction at an arbitrary scale.

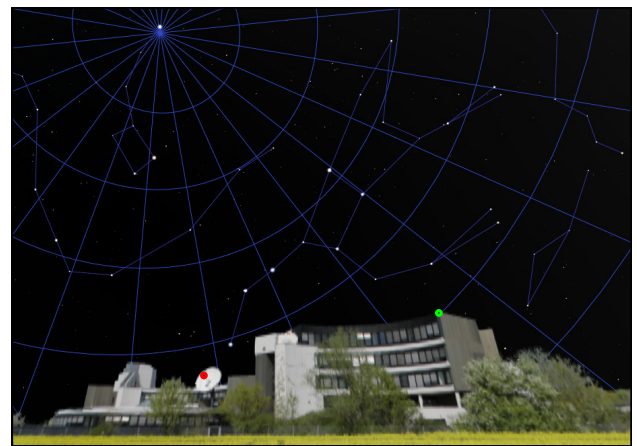It's remarkable that in this setting it's possible to perform a full

Figure 6. Nightsky over Munich (Image by Stellarium, 2018). A corner of the building (green) and the LNB of the satellite dish (red) are marked with dots. From the perspective of the image the corner of the building (green dot) is seen in front of the night sky at the right ascension of about $11^h 30^m$ and declination of about $40°$ on the celestial coordinate system, and the LNB (red dot) is seen at the right ascension of about $14^h 10^m$ and declination of probably $47°$.

multi-view reconstruction by solving a single system of linear equations. Usually, a reconstruction has seven degrees of freedom, the location of the origin, the rotation, and the overall scale. In this case the degree of freedom is just four, for the rotation has already been fixed by the celestial coordinate system on the plane at infinity. Instead of calibrating the camera and finding its rotation, the plane at infinity was "calibrated" into the camera image and only the position of the camera has still to be found. And finding the camera positions and scene points together is possible with a single system of linear equations. Note also, that by shifting the camera rotation problem into the plane at infinity, Equation (18) is fully symmetric in the camera centers and scene points.

Moreover, in this case the reconstruction is Euclidean, since the celestial coordinate system is based on an orthogonal coordinate system. Furthermore, measurement errors are therefore well balanced in Equation (18). However, usually one does not have access to the plane at infinity and an orthogonal coordinate system on it in real-world images. But most of the method can be generalized to more realistic scenarios.

3.2 The Projection onto an Arbitrary Plane

Often, photos show a dominating plane like the ground plane or the facade of a building. Figure 7 shows a tennis court. The original photo (Photo by HeungSoon, 2016) showed the game in oblique view. However, Figure 7 was perspectively rectified to simulate an orthogonal view from above (i.e. the markings are perpendicular and form factor corresponds to that of a standard tennis court). But because of the rectification of the ground plane of the tennis court, everything above the ground plane was perspectively distorted, like the players and the net. They now look like shadows of themselves, just in color. Indeed, suppose the camera would have been replaced by a strong spotlight and the scene would have been captured from above, the shadows from the additional spotlight would have exactly the shape of the now distorted players and net.

In Figure 7 the tennis ball appears to be about one feet left and one feet above the very middle of the tennis court. In space coordinates this could be expressed as $\boldsymbol{P} = (-1, 1, 0)^t$ where the first axis goes to the right, the second axis goes up in the image plane and the third axis goes up in space perpendicular to the tennis courts ground plane and the origin is in the middle of the tennis court.

Let $\boldsymbol{C}$ be the position of the camera in this fixed coordinate system and $\boldsymbol{X}$ be the position of the tennis ball. The three points $\boldsymbol{C}, \boldsymbol{X}$ and $\boldsymbol{P}$ must be collinear since $\boldsymbol{P}$ is the projection of $\boldsymbol{X}$ through $\boldsymbol{C}$ onto the ground plane. This information is similar to that of Equation (16). But there, the direction vector $\boldsymbol{v}$ between the camera center $\boldsymbol{C}$ and a scene point $\boldsymbol{X}$ could immediately be calculated by Equation (15). Unfortunately, some more work has to be done in this case.

By writing

$$\boldsymbol{C} = (C_x, C_y, C_z, 1)^t$$
$$\boldsymbol{X} = (X_x, X_y, X_z, 1)^t \quad\quad (19)$$
$$\boldsymbol{P} = (P_x, P_y, 0, 1)^t$$

as homogeneous coordinates and applying the homography that exchanges the last two components, the ground plane will be

exchanged with the plane at infinity. This homography maps $\boldsymbol{C} = (C_x, C_y, C_z, 1)^t$ to

$$\boldsymbol{C'} = (C_x, C_y, 1, C_z)^t$$
$$\sim \left(\frac{C_x}{C_z}, \frac{C_y}{C_z}, \frac{1}{C_z}, 1 \right)^t \quad\quad (20)$$
$$= (C_x', C_y', C_z', 1)^t$$

and for $\boldsymbol{X}$ analogously. $\boldsymbol{P}$ will be mapped to the direction vector

$$\boldsymbol{v'} = (P_x, P_y, 1, 0)^t . \quad\quad (21)$$

In this setting, Equation (18) can be applied to $\boldsymbol{C'}$, $\boldsymbol{X'}$, and $\boldsymbol{v'}$. After solving for $\boldsymbol{C'}$ and $\boldsymbol{X'}$ the original coordinates can be retrieved by

$$\boldsymbol{C} = \left(\frac{C_x'}{C_z'}, \frac{C_y'}{C_z'}, \frac{1}{C_z'} \right)^t \quad\quad (22)$$

and for $\boldsymbol{X}$ analogously.

The ambiguity in scale and origin in the solution of Equation (18) has stronger influence in this case, and only projective reconstruction can be expected. The ambiguity can be resolved and a metric reconstruction can be obtained given the position of one more point outside the ground plane. Here the top tip of the middle of the net is appropriate. That is supposed to be exactly three feet above the middle of the court. Unfortunately, measurement errors in the directions $\boldsymbol{v'}$ in this projectively distorted space will not balance as well as in the original Equation (18), but this could probably be improved by a meaningful reweighing of the sum.

It is worth noting that in the case of Figure 7, a reconstruction is even possible from the single image. Similar as the camera center $\boldsymbol{C}$, a scene point $\boldsymbol{X}$ (e.g. the tennis ball) and it's projected point $\boldsymbol{P}$ on the ground are collinear, also the sun $\boldsymbol{S}$, the scene point $\boldsymbol{X}$ and its shadow $\boldsymbol{Q}$ on the ground plane are collinear (e.g. the shadow of the tennis ball is located about at $\boldsymbol{Q} = (-2.5, 4, 0)^t$ on the ground plane). Therefore, the shadows in the image are equally well suited for reconstruction. However, it might be useful to enforce that the sun is (virtually) infinitely far away from the scene. In homogeneous coordinates, $\boldsymbol{S} = (S_x, S_y, S_z, 0)^t$ must therefore be mapped to a point $\boldsymbol{S'} = (S_x/S_z, S_y/S_z, 0, 1)^t$ on the ground plane of the distorted projective space. Then, after the reconstruction of $\boldsymbol{S'}$ the direction to the sun is given by $\boldsymbol{S} = (S_x', S_y', 1)^t$.

Figure 7. Rectified tennis court (Photo by HeungSoon, 2016). The original photo shows a scene of a tennis game in oblique perspective. The ground plane was rectified assuming a standard 78 ft by 36 ft tennis court. Objects not on the ground plane (like the players and the net) appear distorted. They are projected through the camera center onto the ground plane.
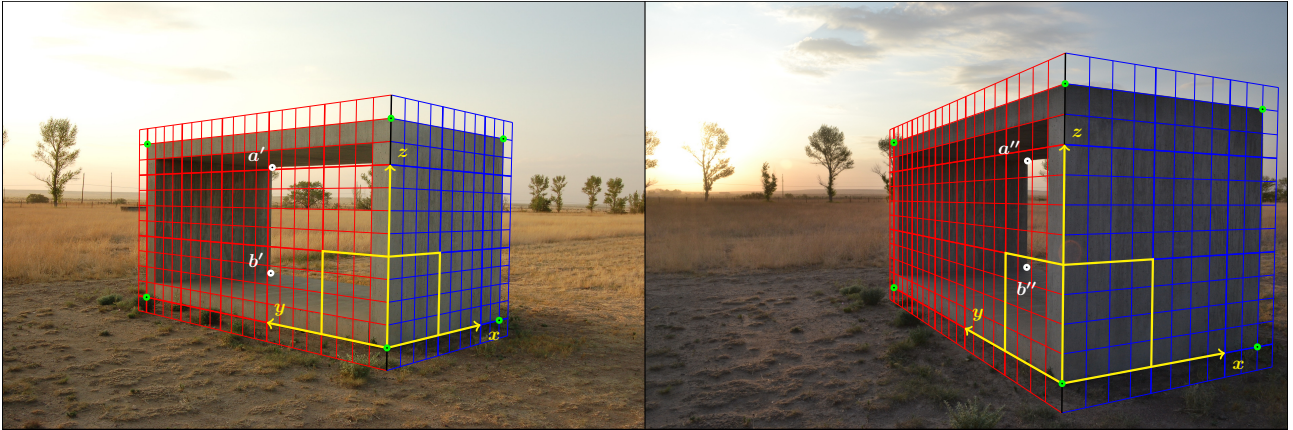
Figure 8. One of Donald Judd's concrete blocks in Marfa, Texas (Photo by Emanue3, 2011, left) and (Photo by Kbennett0016, 2011, right). The green dots are measured on the outer corners of the $2.5\,\mathrm{m} \times 2.5\,\mathrm{m} \times 5\,\mathrm{m}$ concrete block. The blue and red plane are defined by the four corresponding green outer corners of the corresponding sides of the concrete block. The yellow squares show the unit squares on each plane ($1\,\mathrm{m} \times 1\,\mathrm{m}$). The red and blue plane could principally be arbitrarily extended over the whole image plane (in the image they reach about 25 cm over the dimensions of the concrete block). The white dots show two more homologue points in the two images. It is easy to see that e. g. $\boldsymbol{a}'$ projects on the red plane at y-z-coordinates of about $(2, 2)$. However, $\boldsymbol{a}'$ also projects on the blue plane at x-z-coordinate of roughly $(-2, 2)$. The image does not show the extension of the blue plane over the white dots since it would mess up the drawing too much.

3.3 Two planes in an image

In Section 3.1 and Section 3.2 the projections of scene points through the camera center onto a plane in the scene was used for reconstruction. The reconstruction problem is reduced to a single homogeneous system of linear equations in the positions of the scene points and camera centers. However, this system of linear equations is still large and the position of all the scene points and all the camera centers couple together. This can be further simplified by a second known plane in the image.

Figure 8 shows two photos of a concrete sculpture by Donald Judd in Marfa. The sculpture is a rectangular prism and as such well suited to define some planes in the image. Two planes are indicated in the image, the red plane on the long 2.5 m by 5 m side and the blue plane on the short squared side. A coordinate systems is drawn on each plane, the x-z-coordinate system on the blue plane and the y-z-coordinate system on the red plane. Here, the z axis is common to both planes and their origins agree. Moreover, they are placed in a way such that the x, y and z axes form also an orthogonal coordinate system in the Euclidean space. Despite the fact that the red and blue plane are only shown to some extent, they virtually cover the whole image plane. Thus each point in the image has coordinates on the red plane, as well as on the blue plane.

Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be the two inner corners of the backside of the sculpture denoted $\boldsymbol{a}'$ and $\boldsymbol{a}''$, respectively $\boldsymbol{b}'$ and $\boldsymbol{b}''$ in the images. It is easy to estimate their coordinates on the red plane. Roughly they are

$$\begin{aligned}
\boldsymbol{a}'_{\mathrm{red}} &= (1.9, 2.0)^t, & \boldsymbol{a}''_{\mathrm{red}} &= (0.6, 1.9)^t, \\
\boldsymbol{b}'_{\mathrm{red}} &= (1.9, 0.7)^t, & \boldsymbol{b}''_{\mathrm{red}} &= (0.6, 0.9)^t.
\end{aligned} \quad (23)$$

It's harder to estimate their coordinates on the blue plane by just looking at the images but using the pixel coordinates of the defining greed corner points of the planes, by Equation (14) they are

roughly

$$\begin{aligned}
\boldsymbol{a}'_{\mathrm{blue}} &= (-1.7, 1.9)^t, & \boldsymbol{a}''_{\mathrm{blue}} &= (-0.4, 1.9)^t, \\
\boldsymbol{b}'_{\mathrm{blue}} &= (-1.7, 0.9)^t, & \boldsymbol{b}''_{\mathrm{blue}} &= (-0.4, 1.0)^t.
\end{aligned} \quad (24)$$

Furthermore, $\boldsymbol{a}'$ is a point on the red plane with coordinates $\boldsymbol{A}'_{\mathrm{red}} = (0, 1.9, 2.0)^t$ in space and can be understood as the projection of the point $\boldsymbol{A}$ through the camera center $\boldsymbol{C}'$ of the left image onto the red plane in space. On the other hand the projection of $\boldsymbol{A}$ onto the blue plane in space is at $\boldsymbol{A}'_{\mathrm{blue}} = (-1.7, 0, 1.9)^t$ and the four points $\boldsymbol{C}'$, $\boldsymbol{A}$, $\boldsymbol{A}'_{\mathrm{red}}$, and $\boldsymbol{A}'_{\mathrm{blue}}$ are collinear.

Because there are four collinear points, one can decompose the reconstruction in two independent problems. The camera center $\boldsymbol{C}'$ corresponding to the left image is the intersection of the line through $\boldsymbol{A}'_{\mathrm{red}}$ and $\boldsymbol{A}'_{\mathrm{blue}}$ with the line through $\boldsymbol{B}'_{\mathrm{red}}$ and $\boldsymbol{B}'_{\mathrm{blue}}$, both determined in the left image. The scene point $\boldsymbol{A}$ is the intersection of the line through $\boldsymbol{A}'_{\mathrm{red}}$ and $\boldsymbol{A}'_{\mathrm{blue}}$ determined in the left image with the line through $\boldsymbol{A}''_{\mathrm{red}}$ and $\boldsymbol{A}''_{\mathrm{blue}}$ determined in the right image. This decomposes the large system of linear equations in Equation (18) into independent small systems; one for each scene point and one for each camera center.

4. DISCUSSION

The previous chapter showed some applications to 3D reconstruction using planes in the scene. The main idea behind the approach is to separate the reconstruction problem into two distinct parts. One is the aspect of the position of the camera, it's perspective. The other is the imaging geometry of the camera.

The first aspect, the position or perspective of the camera, is actually the three dimensional problem. It decides about which parts of the scene are visible and which parts are occluded. The occlusion structure is key to the question where the camera was, even for more complex camera models as e. g. fish eye cameras; given just, there still exists something like a camera center in which all lines of sight meet (in opposite to e. g. a push broom camera).

The approach presented in this article was to use two consecutive scene points mapped exactly behind one another in the image to determine a line in space where the camera center must be located on. Ideally, one would like to use two feature points in the image that are mapped exactly on the same place. However, this rarely happens and if it happens, one likely would not recognize this double feature as corresponding to any of the individual features anymore. Therefore, another method was chosen. One target point was compared to a reference plane in the scene. The reference plane has to be established by four feature points, but then, it's virtually omnipresent in the image, and the intersection with the target scene point can be exactly determined. This yields two consecutively mapped scene points in the image, the target scene point, and it's projection through the camera center onto the reference plane.

The second aspect is that of the imaging geometry. It's a purely two dimensional problem. Here, a pinhole camera model was actually assumed, that does project the scene through the camera center onto a plane with an affine grid on it (i. e. without e. g. radial distortion or the like). For this reason, affine geometry could be used to defer the camera calibration problem into the scene. By choosing three point correspondences as basis for an affine coordinate system (in form of barycentric coordinates), the camera calibration becomes irrelevant, since the barycentric coordinates are already invariant to the affine grid on the sensor plane. Moreover, the corresponding three points in the scene already define a unique reference plane, where the images will be commonly projected onto.

The remaining problem is to find a fourth point correspondence on that plane to establish a projective grid on it. The problem is equivalent to determining the direction of the principal axis of the camera with respect to the reference plane in the scene. It comprehends the projective component of the camera projection, as the rotation around the principal axis already entered the affine part of the transformation. If the scene does not offer an obvious choice like markings of a court or tie point markers arranged into the scene, a robust homography based image registration algorithm like (Hartley and Zisserman, 2004, Algorithm 4.3) might reveal a dominant plane from which four inlier can be selected. However, if prior knowledge of the scene is not available it might be difficult to find an affine grid on the reference plane such that measurement errors balance well in Equation (18).

REFERENCES

Blinn, J., 2003. *Jim Blinn's Corner: Notation, Notation, Notation*. Morgan Kaufmann Publishing.

Erdnüß, B., 2017. Das baryzentrische Verhältnis als Invariante der projektiven Geometrie. *tm – Technisches Messen* 84(7-8), pp. 479–492.

Hartley, R. I. and Zisserman, A., 2004. *Multiple View Geometry in Computer Vision*. Second edn, Cambridge University Press.

Image by Stellarium, 2018. Landscape "Garching" by Fabien Chéreau: Max Plank Institute for Astrophysics in Munich. Stellarium: http://stellarium.org.

Photo by Emanue3, 2011. Concrete sculpture by Donald Judd, Marfa, Texas. Wikimedia Commons: https://commons. wikimedia.org/wiki/File:Donald_Judd_Concrete_Blocks.jpg. License: CC-BY-SA-3.0.

Photo by endmondlafoto, 2018. Place Dalida, Paris. Pixabay: https://pixabay.com/de/paris-tourismus-perspektive-3310967/. License: CC0.

Photo by GregoryButler, 2013. Pool and pavement. Pixabay: https://pixabay.com/de/schwimmbad-fliese-schwimmen-389261/. License: CC0.

Photo by HeungSoon, 2016. Tennis. Pixabay: https://pixabay.com/de/tennis-%C3%BCbung-doppel-spiel-sport-2557074/. License: CC0.

Photo by Kbennett0016, 2011. One of Donald Judd's concrete pieces at the Chinati foundation, Marfa, Texas. Wikimedia Commons: https://commons.wikimedia.org/wiki/File:Marfa_Chinati_Bennett.JPG. License: Public Domain.