Introduction to Geometric Computer Vision

Prakash Chockalingam

I. Introduction

We have always wondered how two railways lines that are parallel to each other meet at a point far away from us and how the ocean and sky meet at a distant line. These parallel lines and planes meet because the capturing device maps the 3D world coordinates to 2D coordinates using a perspective projection. There are two kinds of projection - orthographic and perspective. Orthographic projection corresponds to perspective projection in the hypothetical case where the capturing device is at an infinite distance with infinite focal length.

There are different transformations possible (*e.g.* translation, rotation, shear, etc.) within the realm of perspective projection leading to a hierarchy of projective transformations and the study of all the different geometric properties (*e.g.* area, volume, distance, angles, etc.) that are invariant under these transformations have given rise to the field of projective geometry. The knowledge of projective geometry, that captures the geometric relation between the 3D scene and a 2D image, can be extended to multiple 2D images capturing the same 3D scene from different viewpoints. The study of geometric relations between the structures in the multiple 2D views of the 3D scene forms the crux of multiple view geometry. Epipolar geometry is a specialization of multiple view geometry where the the number of views are restricted to two, similar to the two eyes capturing the 3D world.

Many applications have been developed based on these geometric aspects of computer vision:

- Building 3D models of a scene from images that captures multiple views of the scene is one of the classic applications of geometric computer vision. 3D modeling has gained more attention after Google Earth introduced 3D models of city generated based on the images captured. Using motions of the object over time to analyze the 3D structure using factorization methods have given impressive results.
- Automotobile industries use the concepts of stereo vision, a sub-field of computer vision, which deals with the epipolar geometry of two images capturing the scene, similar to the human eyes. Stereo vision is used to analyze the structure of the scene which helps in automatic locomotion by estimating the distance of the obstacles in front of the vehicle.
- Image registration is one of the popular techniques of computer vision that is employed in a wide gamut of applications and products ranging from medical imaging to comercial digital cameras. Registration is the process of aligning the images taken from different views into a single coordinate system. This helps in many applications like image stabilization to remove camera jitter and image mosaicking to build panoramic views of the scene by stitching multiple images together.
- Super-resolution techniques, based on multiple frames or images, are used to enhance the resolution of imaging systems.
- Multiple view object recognition algorithms have shown significant performance gain over their single view counterparts in clutter and occlusion using the extra information from different views.

Given the significance of the geometric concepts, this report discusses the projective geometry and epipolar geometry and their relation to comptuer vision. Section II and III give an overview about the 2D projective plane and the 2D transformations respectively. IV details out the properties of the 3D projective space. Image formation and application of the projective geometry to computer vision is discussed next in section V. A description on epipolar geometry, properties of fundamental and essential matrix is included in the final section.

II. Projective Plane

A. Homogenous Coordinates

Lets begin with the representation of lines and points and extend that understanding to analyze the projective plane which forms the basis of the multiple view geometry. A line can be represented using the equation $y = mx+c$, where m and c are the slope and the intercept of the line respectively. The problem with this representation is that vertical lines, whose slope tends to ∞, cannot be represented. A more better representation would be using the equation $ax + by + c = 0$, where different choices of a, b, and c gives rise to different lines. Hence each line can be represented using a 3-vector (a, b, c). Another neat advantage of this representation is that, for any non-zero constant k, the lines $ax + by + c = 0$ and $kax + kby + kc = 0$ are same and the vectors (a, b, c) and (ka, kb, kc) are considered to be equivalent. All vectors which are equivalent are known as homogenous vectors.

A point in $\mathfrak{R}^2$ is represented as (x, y). We can easily extend this representation to $\mathfrak{R}^3$ by adding another dimension and fixing the value as 1. Hence each point is now represented as $(x, y, 1)$. As explained above, $(x, y, 1)$ and (kx, ky, k) are said to be equivalent. Hence it is obvious that the set of vectors (kx, ky, k), for different values of k, are the representation of the point (x, y) in $\mathfrak{R}^2$. The inhomogeneous representation of any homogenous point (x, y, w) is given by $(x/w, y/w)$.

B. Ideal Points and Lines

The most obvious and intriguing question is the necessity for this homogenous representation of a point in higher dimension. This can be best illustrated by analyzing intersection of two parallel lines l and l' represented by $ax + by + c = 0$ and $ax + by + c' = 0$ respectively. Intersection of these lines is given by their cross products:

$$l \times l' = \begin{vmatrix} i & j & k \\ a & b & c \\ a & b & c' \end{vmatrix} = \begin{vmatrix} bc' - bc \\ -(ac' - ac) \\ ab - ab \end{vmatrix} = \begin{vmatrix} b(c' - c) \\ -a(c' - c) \\ 0 \end{vmatrix} \tag{1}$$

Leaving out the scaling term $(c' - c)$, we get $(b, -a, 0)$. This is the homogeneous point where the parallel lines meet. The inhomogeneous representation of this point is $(b/0, -a/0)$, which results in infinitely large coordinates. Now we have a mechanism to represent points at infinity and all points with homogeneous coordinates $(x, y, 0)$ represent points at infinity. These points are termed as *ideal points*. All the ideal points lie on the line $(0, 0, 1)$ which forms the *line at infinity*. The nice thing about this representation is that the first two coordinates gives the direction of this infinity point. For example, the point $(1, 0, 0)$ is a point at infinity in the direction of x-axis. Now the projective space, P^2, can be visualized as the regular $\mathfrak{R}^2$ with affine transformations and a huge circle encapsulating $\mathfrak{R}^2$ [1]. The circle represents the line at infinity and all the points on this circle form the ideal points.

Alternatively, it can also be visualized as rays in $\mathfrak{R}^3$. The set of vectors $k(x, y, w)$ forms a ray emanating from origin as w varies. Hence a line in $\mathfrak{R}^3$ corresponds to a point in P^2 and a plane in $\mathfrak{R}^3$ corresponds to a line in P^2. All the ideal points and the line at infinity lie on the plane $w = 0$ and intersecting the lines and planes at $w = 1$ gives the inhomogenous representation of the points and lines respectively.

C. Duality Principle

It can be noted that the representation for lines and points are similar in the projective plane. The intersection of two lines is given by $x = l \times l'$ and similarly the intersection of two points is given by $l = x \times x'$. The incidence of a point on line is given by the relation $x^T l = 0$ and since it is symmetric, it is equivalent to $l^T x = 0$. In all these equations and representations, it is clearly seen that swapping lines and points still holds all the equations true. This principle where lines and points can be swapped in any statement or equation describing the properties of the projective plane is known as the duality principle.

D. Cross-Ratios

Cross ratio is a ratio of ratios of distances. Given four collinear points p_1, p_2, p_3, and p_4 in P^2 and denoting the Euclidean distance between two points p_i and p_j as Δ_{ij}, cross ratio can be defined as shown in figure 1,

$$\tau_{p_1 p_2 p_3 p_4} = \frac{\Delta_{13}\Delta_{24}}{\Delta_{14}\Delta_{23}} \tag{2}$$

Depending on the order in which the four points are chosen for calculating the cross-ratios, there are 24 possible values. However, only six distint values are produced which are related to each other as:

$$\{\tau, \frac{1}{\tau}, 1 - \tau, \frac{1}{1 - \tau}, \frac{\tau - 1}{\tau}, \frac{\tau}{\tau - 1}\}. \tag{3}$$

E. Conics

Thus far, we considered lines and planes represented using first-degree equations. In Euclidean geometry, the family of second-degree equations give rise to three main geometrical figures - ellipse, parabola and hyperbola. In projective geometry, the discrimination between the three types is lost and they can be converted from one form to another. The inhomogenous representation is given as

$$ax^2 + bxy + cy^2 + dx + ey + f = 0 \tag{4}$$

The homogenous representation of the above second-degree equation would be:

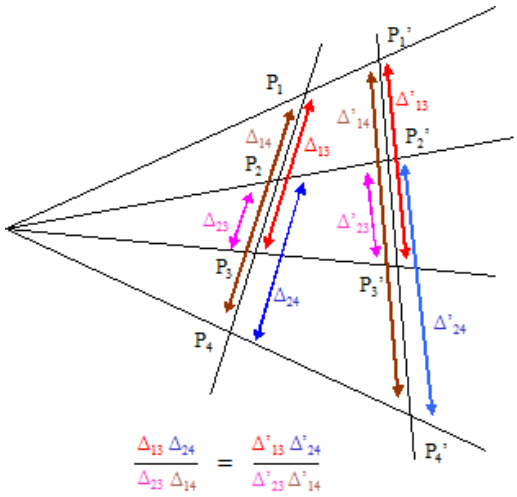

Fig. 1. Invariant nature of cross ratio for four collinear points subjected to a projective transformation

$$a(\frac{x}{w})^2 + b\frac{x}{w}\frac{y}{w} + c(\frac{y}{w})^2 + d\frac{x}{w} + e\frac{y}{w} + f = 0$$
$$ax^2 + bxy + cy^2 + dxw + eyw + fw^2 = 0 \tag{5}$$

In matrix form,

$$\mathbf{x}^T C\mathbf{x} = 0 \tag{6}$$

where C is given by

$$C = \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix} \tag{7}$$

This matrix is known as the conic coefficient matrix. It has five degrees of freedom since there are six unique elements which can be defined using five independent ratios.

III. 2D Projective Transformations

A. Homography

Projective transformations is the mapping of points in P^2 to points in P^2 that preservers collinearity of any given set of points. The new point in P^2 represented by the vector $(a', b', c')^T$ can be obtained by multiplying the point, represented by the vector $(a, b, c)^T$, with a non-singular 3×3 matrix H as:

$$\begin{pmatrix} a' \\ b' \\ c' \end{pmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} \tag{8}$$

Since there are eight independent ratios out of the nine elements in H, a projective transformation has eight degrees of freedom. Such transformations are useful to extract information about the position of the observed objects when the point of view of the observer (camera) changes. Projective transformation is also known as projectivity, homography, or collineation.

B. Group of Transformations

1) Euclidean Transformation: The Euclidean transformation is a composition of translation and rotation. A translation can be represented as:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} 0 & 0 & t_x \\ 0 & 0 & t_y \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{9}$$

where t_x and t_y are the translation in the two directions. A rotation is represented as:

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{10}$$

The above two representations can be combined together to form the Euclidean transformation which is given by:

$$H_{Euclidean} = \begin{bmatrix} \cos\theta & -\sin\theta & t_x \\ \sin\theta & \cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \tag{11}$$

Euclidean transformation can be used to model the motion of a rigid object and has three degrees of freedom, one for rotation (θ) and two for translation in x and y direction. Since euclidean transformation allows only translation and rotation of objects, it preserves distance, angles and area.

2) Similarity Transformation: To model scale changes of an object, the above transformation should be modified such as to incorporate isotropic scaling, *i.e.*, uniform scaling in x and y direction. Such a transformation which is a composition of translation, rotation and isotropic scaling is known as Similarity transformation and has the following representation:

$$H_{Similarity} = \begin{bmatrix} s\cos\theta & -s\sin\theta & t_x \\ s\sin\theta & s\cos\theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \tag{12}$$

where s represents the isotropic scaling parameter. A simililarity transformation has four degrees of freedom, three from Euclidean transformation and one more from the isotropic scaling. Since scaling are now allowed, lengths and area are no longer preserved. However, ratios of lengths and area are still preserved and angle measurements are also not affected by the isotropic scaling.

3) Affine Transformation: Generalizing the above Similarity transformation by allowing non-isotropic scaling will result in the affine representation:

$$H_{Affine} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \tag{13}$$

where a_{11}, a_{12}, a_{21}, and a_{22} are the affine parameters that allow rotation and non-isotropic scaling. The affine transformation has six degrees of freedom (four affine parameters and two translation parameters) and can be computed using three point correspondences. Since affine transformations allow non-isotropic scaling, angle measurements are not preserved. Ratios of areas are invariant as area is affected by the scaling in the two directions and this scaling cancels out in the ratio of areas. Parallel lines are also preserved as parallel lines meet at some point $(x, y, 0)$. This point under affine transformation is still mapped to another point at infinity and hence parallel lines are invariant. Ratios of lengths on a line is also invariant under affine transformation.

4) Projective Transformation: Affine transformations can be generalized to the form the superset projective transformation by allowing the first two entries of the last column to be variable.

$$H_{Projective} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ p_1 & p_2 & 1 \end{bmatrix} \tag{14}$$

It can be seen that $H_{Projective}$ is a homogenous matrix with eight degrees of freedom. Unlike affine transformation, this representation will map a point at infinity to some other point in the projective plane as:

$$\begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ p_1 & p_2 & 1 \end{bmatrix} \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} = \begin{pmatrix} a_{11}x + a_{12}y \\ a_{21}x + a_{22}y \\ p_1x + p_2y \end{pmatrix} \tag{15}$$

Hence parallel lines and ratio of lengths on a line are not preserved. However, ratio of ratios of lengths, known as cross ratios, are preserved and this forms one of the significant and fundamental properties of projective transformation.

C. Estimation of the Homography

Having learnt all the invariant properties of the homography, the next obvious thing is to devise an algorithm to estimate the homography given the point correspondences in P^2. The homogenous matrix has eight degrees of freedom and each point is represented using two parameters giving rise to two equations for every point correspondence. Hence atleast four point correspondences are required to compute the homography.

Let $\mathbf{x}_i = [x_i \ y_i \ w_i]^T$ and $\mathbf{x}'_i = [x'_i \ y'_i \ w'_i]^T$ be the given point correspondence. We know that the transformation is given by $\mathbf{x}'_i = H\mathbf{x}_i$. Since the vectors are represented in the homogenous representation, the vectors $\mathbf{x}'_i$ and $H\mathbf{x}_i$ have only the same direction but not the magnitude. The magnitude of the cross product of two parallel vectors are zero. This property can be employed here to derive a simple linear solution for H.

$$\begin{pmatrix} x_i \\ y_i \\ w_i \end{pmatrix} \times \begin{pmatrix} h_{11}x_i + h_{12}y_i + h_{13}w_i \\ h_{21}x_i + h_{22}y_i + h_{23}w_i \\ h_{31}x_i + h_{32}y_i + h_{33}w_i \end{pmatrix} = 0 \tag{16}$$

$$\begin{pmatrix} y'_i x_i h_{31} + y'_i y_i h_{32} + y'_i w_i h_{33} - w'_i x_i h_{21} - w'_i y_i h_{22} - w'_i w_i h_{33} \\ w'_i x_i h_{11} + w'_i y_i h_{12} + w'_i w_i h_{13} - x'_i x_i h_{31} - x'_i y_i h_{32} - x'_i w_i h_{33} \\ x'_i x_i h_{21} + x'_i y_i h_{22} + x'_i w_i h_{23} - y'_i x_i h_{11} - y'_i y_i h_{12} - y'_i w_i h_{13} \end{pmatrix} = 0 \tag{17}$$

The above equation can be re-arranged and written as:

$$\begin{bmatrix} 0 & 0 & 0 & -w'_i x_i & -w'_i y_i & -w'_i w_i & y'_i x_i & y'_i y_i & y'_i w_i \\ w'_i x_i & w'_i y_i & w'_i w_i & 0 & 0 & 0 & -x'_i x_i & -x'_i y_i & -x'_i w_i \\ -y'_i x_i & -y'_i y_i & -y'_i w_i & x'_i x_i & x'_i y_i & x'_i w_i & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{bmatrix} = 0 \tag{18}$$

Though the matrix on the left has three rows, the third row is a linear combination of the first two rows and it can be omitted while solving for H. Hence the above equation becomes,

$$\begin{bmatrix} 0 & 0 & 0 & -w'_i x_i & -w'_i y_i & -w'_i w_i & y'_i x_i & y'_i y_i & y'_i w_i \\ w'_i x_i & w'_i y_i & w'_i w_i & 0 & 0 & 0 & -x'_i x_i & -x'_i y_i & -x'_i w_i \end{bmatrix} \begin{bmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{bmatrix} = 0 \tag{19}$$

This equation is of the form $A_i h = 0$, where A_i has the dimensions 2×9. To solve for H, a matrix A is constructed by stacking the two rows of A_i for each point correspondence given. Thus A has a dimension of $2n \times 9$, where n is the number of given point correspondences. If all the given point correspondences are correct, then an exact solution for h can be obtained with a simple constraint $\|h\| = 1$. However, if they are not exact, due to noisy measurements, then an approximate solution has to be obtained by minimizing the norm $\|Ah\|$ which is equivalent to minimizing $\|Ah\|/\|h\|$. The solution to such a system is given by the eigenvector of A^TA corresponding to the least eigenvalue.

If there are large number of outliers in the given measurement, then a robust estimator like RANdom SAmple Consensus (RANSAC) can be used to determine a set of inliers from the given set of correspondences so that the homography can be estimated reliably using techniques like DLT.

IV. 3D Projective Space

Projective transformations of 3-space can be understood as direct generalizations of 2-space. In the projective 3-space P^3, the ideal points form a plane at infinity π_∞ instead of a line at infinity, l_∞, as in P^2. Parallel lines and parallel planes intersect at π_∞.

A. Point and plane representation

Each point (x, y, z) in the 3-space is homogenously represented using a 4D vector (x', y', z', w'). The inhomogenous coordinates can be obtained from the homogenous representation as:

$$x = \frac{x'}{w'}, y = \frac{y'}{w'}, z = \frac{z'}{w'}. \tag{20}$$

A plane in 3-space can be represented as $\pi_1 x + \pi_2 y + \pi_3 z + \pi_4 = 0$, where the parameters π_1, π_2, π_3 define the orientation of the plane and π_4 represent the distance of the plane from the origin. Homogenizing this plane representation using (20) yields:

$$\pi_1 x' + \pi_2 y' + \pi_3 z' + \pi_4 w' = 0 \tag{21}$$

The above representation clearly shows that the point X lies on plane Π and also illustrates the duality between points and planes in P^3

B. Line representation using Plücker coordinates

A line is defined by the intersection of two planes or the join of two points. A homogenous representation of line in 3-space requires a 5-vector as a line has four degrees of freedom. Using such a 5-vector with a line or plane representation that has 4-vector makes all the mathematical equations more complex to express. To overcome this issue, different representations have been used. The most elegant and commonly used is the Plücker coordinates which is obtained from a 4×4 Plücker matrix which is defined as

$$L = X_1 X_2^T - X_2 X_1^T \tag{22}$$

where X_1 and X_2 are the two homogenous points joining the line. Similarly, a Plücker matrix for a line formed by the intersection of the planes Π_1 and Π_2 can be written as

$$L^* = \Pi_1 \Pi_2^T - \Pi_2 \Pi_1^T. \tag{23}$$

The Plücker matrix representation helps in representing the join and incidence properties:

- The plane defined by the join of the point X and line L is given by

$$\Pi = L^* X \tag{24}$$

- The point defined by the intersection of the line L with the plane Π is given by

$$X = L\Pi \tag{25}$$

To get a more clear understanding between the Plücker matrix and the line representation, the elements of a Plücker matrix for a line joining two points represented by (x_1, y_1, z_1, w_1) and (x_2, y_2, z_2, w_2) can be written as:

$$L = \begin{bmatrix} x_1x_2 - x_2x_1 & x_1y_2 - x_2y_1 & x_1z_2 - x_2z_1 & x_1w_2 - x_2w_1 \\ y_1x_2 - y_2x_1 & y_1y_2 - y_2y_1 & y_1z_2 - y_2z_1 & y_1w_2 - y_2w_1 \\ z_1x_2 - x_1z_2 & z_1y_2 - y_1z_2 & z_1z_2 - z_2z_1 & z_1w_2 - z_2w_1 \\ w_1x_2 - x_1w_2 & w_1y_2 - y_1w_2 & w_1z_2 - z_1w_2 & w_1w_2 - w_2w_1 \end{bmatrix} \tag{26}$$

It can be noted that L is a skew-symmetric matrix, *i.e.*, $L^T = -L$. Out of the twelve non-zero elements, a line can be represented using six non-zero elements. The elements that are chosen to represent the line are:
$(l_{12}, l_{31}, l_{14}, l_{23}, l_{24}, l_{34})$, where

$$\begin{aligned} l_{12} &= x_1y_2 - x_2y_1, \\ l_{31} &= x_2z_1 - z_2x_1, \\ l_{14} &= w_2x_1 - x_2w_1, \\ l_{23} &= y_1z_2 - y_2z_1, \\ l_{24} &= y_1w_2 - y_2w_1, \\ l_{34} &= z_1w_2 - z_2w_1 \end{aligned} \tag{27}$$

The above six coordinates are known as Plücker coordinates. There are also alternate choices for the Plücker coordinates from the elements of the Plücker matrix to represent the line. The choice of the above coordinates helps in Euclidean representation as:

$$X_1' - X_2' = (x_1w_2 - x_2w_1 \quad y_1w_2 - y_2w_1 \quad z_1w_2 - z_2w_1)^T = (l_{14} \; l_{24} \; l_{34})^T \tag{28}$$

$$X_1' \times X_2' = \frac{1}{w_1w_2}(x_1y_2 - x_2y_1 \quad x_2z_1 - x_1z_2 \quad y_1z_2 - y_2z_1)^T = (l_{12} \; l_{31} \; l_{23})^T \tag{29}$$

where X_1' and X_2' are the inhomogenous counterparts of X_1 and X_2 respectively. Since $\det(L) = 0$, it follows that

$$l_{12}l_{34} + l_{31}l_{24} + l_{14}l_{23} = 0 \tag{30}$$

A 6-vector that satisfies the above equation will form a line in 3-space.

C. Group of Transformations

The hierarchy of projective transformations of a 3-space is similar to that of 2-space. The simplest specialization of the projective transformation is the Euclidean transformation which has six degrees of freedom, three from the translation in three directions and three from the rotation about the three axis. Similarity transformation has seven degrees of freedom, one extra for the isotropic scaling. Affine transformation has twelve degrees of freedom and preserves the parallelism of planes and volume ratios. The projective transformation has 15 degrees of freedom and is represented as:

$$H = \begin{bmatrix} A & \mathbf{t} \\ \mathbf{v}^T & v \end{bmatrix} \tag{31}$$

where A is the 3×3 invertible affine matrix, $\mathbf{t}$ is a 3×1 vector representing the 3D translation, $\mathbf{v}$ is a general 3-vector and v is a scalar.

V. Application of projective geometry in computer vision

This section deals with the geometry of a single perspective camera and the formation of images. Image formation using a simple camera model is discussed and the next subsection gives the camera matrix required to convert the 3D world coordinate into 2D image coordinates.

A. Image Formation

Image formation involves the mapping of world coordinates $(X, Y, Z)^T$ in P^3 to image coordinates $(x, y)^T$in P^2. Let the image plane be $Z = f$, then the 2D image coordinates is given by

$$x = -f\frac{X}{Z} \tag{32}$$

$$y = -f\frac{Y}{Z} \tag{33}$$

As shown in figure 2, the line perpendicular to the image plane joining the center of the camera C is the principal axis and the point P at which the principal axis intersects the image plane is called the principal point. In this figure, it is assumed that the camera is at the origin of the world coordinate.

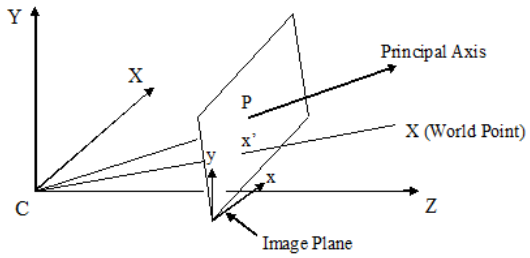

Fig. 2. Image Formation using pinhole camera

B. Camera Matrix

If the world and image coordinates are represented as homogenous vectors, then the above equation can be written in terms of matrix multiplication as

$$\begin{pmatrix} x \\ y \\ w \end{pmatrix} = \begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} \tag{34}$$

The above equation can be compactly written as

$$\mathbf{x} = \widetilde{P}\mathbf{X} \tag{35}$$

where $\mathbf{x}$ and $\mathbf{X}$ are the homogenous points in 2D image coordinates and 3D world coordinates respectively, and $\widetilde{P}$ is the 3×4 homogenous projection matrix. In the above representation, it was assumed that the principal point is at the origin of the image coordinates and the image pixels produced by the camera are square pixels of unit length. Allowing an offset (u_0, v_0) for the principal point from the origin of the image coordinate and different scale factors k_u and k_v in the x and y directions respectively with a skew parameter k_s, the above equation can be written as

$$\begin{pmatrix} x \\ y \\ w \end{pmatrix} = \begin{bmatrix} k_u & k_s & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} \tag{36}$$

The only assumption made in the above model is that the camera center is in the origin of the world coordinate. To relax that assumption, an Euclidean transformation in the 3-space is required. Hence the final representation is given by

$$\begin{pmatrix} x \\ y \\ w \end{pmatrix} = \begin{bmatrix} k_u & k_s & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -f & 0 & 0 \\ 0 & -f & 0 \\ 0 & 0 & 1 \end{bmatrix} [R \quad \mathbf{t}] \begin{pmatrix} X \\ Y \\ Z \\ W \end{pmatrix} \tag{37}$$

where R is a 3×3 rotation matrix and $\mathbf{t}$ is a 3×1 translation vector. Hence the homogenous 3×4 camera projection matrix is given by the product of 3 matrices,

$$\mathbf{x} = P_{internal} P_{projection} P_{external} \mathbf{X} \tag{38}$$

where $P_{internal}$ is the 3×3 matrix that captures the camera intrinsics, $P_{projection}$ is the perspective projection matrix and $P_{external}$ is the matrix that captures the camera extrinsics. The final camera projection matrix, $\widetilde{P}$, has 11 degrees of freedom: 6 from the $P_{external}$, 4 from $P_{internal}$ and 1 from $P_{projection}$. Typically $P_{internal}$ and $P_{projection}$ are combined together and known as internal camera calibration matrix,

$$P_{calibration} = \begin{bmatrix} \alpha_u & k_s & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{39}$$

where $\alpha_u = -k_u \times f$ and $\alpha_v = -k_v \times f$.

C. Vanishing points and lines

As explained in the introduction section, when parallel railway track lines are observed standing on the railway track, it can be seen that the two parallel lines meet at a point. These ideal points that map to the 2D image plane are known as vanishing points. There are some ideal points that does not map to the image plane. For example, the ideal point formed by two lines that are parallel to each other and to the the image plane does not map to the image plane and remains an ideal point. This concept can be extended to planes. Parallel planes meet to form a line at infinity and these ideal lines that get projected to the image plane are known as vanishing lines.

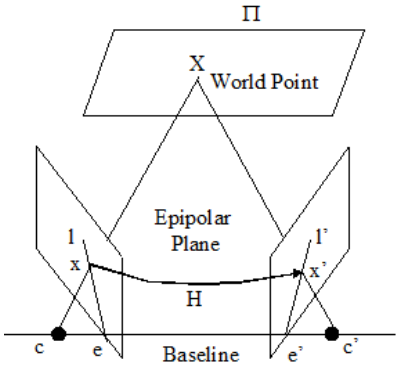

Fig. 3. Epipolar geometry

VI. Two View Geometry

A. Epipolar Geometry

Figure 3 shows two pin-hole cameras looking at a 3D world point X located at the plane Π. The points c and c' form the center of projection or focal points of the left and right camera respectively. Since the two focal points of the cameras are distinct, each focal point projects onto a distinct point into the other camera's image plane. The line connecting the focal points of the camera cc' form the baseline. The two points on the image plane e and e' where the baseline intersects the image plane are called the epipoles.

The perspective projection of the line cX on the left image plane forms a point x in the left image as the world point X is directly in line with the left camera's focal point. However the right camera sees this line as the line l' which joins the points x' and e'. This line is called the epipolar line. Similary the line $c'X$ is seen as the line l joining the points x and e in the left image which forms the epipolar line for x'. Here, the world point X and the two center of projections c and c' form a plane called epipolar plane which intersects the image plane as the epipolar line. All epipolar lines intersect at the epipoles irrespective of where the world point is located in the epipolar plane. The correspondence between the points x and x' is given by the homography H induced by the plane Π.

B. Fundamental Matrix

Homographies establish the correspondence between the two points in the images but they are dependent on the plane Π on which the world coordinate X, that gave rise to the two points, lies. Hence all the points in the two images cannot be related using a single homography as the points might lie on different planes in the 3D world. Fundamental matrix is a 3×3 matrix that establishes a relationship for point correspondences in the two images, irrespective of the planes in which the 3D points are located. In the following subsections, we derive the expression for the fundamental matrix, analyze the relationship of fundamental matrix with the homographies and detail out the methods for computing the fundamental matrix.

1) Point Correspondence relation: We need to establish a geometric relationship between corresponding pairs of points in the images formed from the two cameras. Let $\widetilde{x}$ and $\widetilde{x'}$ be the normalized coordinates in the two images, *i.e.*, coordinates with respect to their camera coordinate frame and x and x' be the pixel coordinates in the two images. Let c and c' be the focal point of the cameras. Since $c\widetilde{x}$, $c\widetilde{x'}$ and cc' are all coplanar,

$$\widetilde{x'}^T(t \times R\widetilde{x}) = 0 \tag{40}$$

where t and R define the translation and rotation between the two camera coordinate frames respectively. We know that the camera calibration matrix is used to convert the normalized coordinates to pixel coordinates as:

$$\begin{aligned} x &= K\widetilde{x} \\ x' &= K'\widetilde{x'} \end{aligned} \tag{41}$$

where K and K' are the calibration matrix as defined in equation (39). Substituting (41) in (40) yields,

$$\begin{aligned} (K'^{-1}x')^T(t \times R)(K^{-1}x) = 0 \\ x'^T K'^{-T}(t \times R)(K^{-1}x) = 0 \end{aligned}$$

$$x'^T F x = 0 \tag{42}$$

where $F = K'^{-T}(t \times R)K^{-1}$. Thus, a fundamental matrix can be computed from the camera intrinsics and the relative camera extrinsics. The fundamental matrix has seven parameters and its rank is two [2].

2) Relation with Homography: In the previous section, we saw the role of fundamental matrix in defining the geometric relationship between points in the left and right images. We can also deduce a relationship between fundamental matrix and homographies. We know that the points x and x' are related by the homography as $x' = Hx$. The epipolar line l' can be expressed in terms of the two points e' and x' as $l' = e' \times x'$. Using the cross product notation, we have, $l' = [e']_\times x'$. Combining these two information, we get

$$\begin{aligned} l' &= [e']_\times x' = [e']_\times Hx \\ l' &= Fx \end{aligned} \tag{43}$$

where $F = [e']_\times H$. The fundamental matrix can also be expressed in terms of the left epipole using the facts $l = [e]_\times x$ and $x = H^{-1}x'$. From these equations, we have

$$\begin{aligned} l &= [e]_\times H^{-1} x' \\ l &= Fx' \end{aligned}$$

where $F = [e]_\times H^{-1}$. Note that the homogrpaphy H is induced by the plane Π and homographies can be uniquely determined using planes.

Establishing the relationship between homographies and planes will give a more deeper understanding of the fundamental matrix. To analyze this relationship, we start by considering the plane Π with the coordinates $(\pi^T, d)^T$, where π represents the orientation of the plane and d represents the distance of the plane from the world origin. The world point X can be parameterized as $(\widetilde{x}^T, \lambda)^T$, where λ represents the position of the world point on the line joining the normalized image point $\widetilde{x}$ and world point X. Since the point X lies on the plane Π, we have

$$\Pi^T X = 0 \tag{44}$$

Using the new parameterizations for the plane and the world point in the above equation, we can derive a relationship for λ as

$$\begin{aligned} \pi^T \widetilde{x} + \lambda d &= 0 \\ \lambda &= \frac{-\pi^T \widetilde{x}}{d} \end{aligned} \tag{45}$$

Now, $X = (\widetilde{x}^T, -\pi^T \widetilde{x}/d)^T$. The normalized image point $\widetilde{x'}$ and world point X is related by $\widetilde{x'} = [R\ t]X$, where $[R\ t]$ is the projection matrix of the second camera represented in terms of the relative camera extrinsics with respect to the first camera. Using the new notation for X, the relationship between the two image points can be established as

$$\begin{aligned} \widetilde{x'} &= R\widetilde{x} - \frac{t\pi^T \widetilde{x}}{d} \\ \widetilde{x'} &= (R - \frac{t\pi^T}{d})\widetilde{x} \end{aligned} \tag{46}$$

Hence, the homography is given by

$$H = R - \frac{t\pi^T}{d} \tag{47}$$

For uncalibrated cameras, the normalized image coordinates are converted to pixel coordinates using equation (41). Substituting (41) in (46), the homography for uncalibrated cameras can be obtained as

$$H = K'(R - \frac{t\pi^T}{d})K^{-1} \tag{48}$$

Given the camera intrinsics, relative extrinsics and the plane coordinates, the homography induced by the plane can be uniquely determined using the above equation.

3) Computing fundamental matrix: There are many ways to compute the fundamental matrix. A simple point correspondences based approach is given here. Equation (42) can be written in matrix form as:

$$\begin{bmatrix} x' & y' & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0 \tag{49}$$

$$f_{11}xx' + f_{12}xy' + f_{13}x + f_{21}x'y + f_{22}yy' + f_{23}y + f_{31}x' + f_{32}y' + f_{33} = 0 \tag{50}$$

The above equation can be re-arranged and written as

$$\begin{bmatrix} xx' & xy' & x & x'y & yy' & y & x' & y' & 1 \end{bmatrix} \begin{bmatrix} f_{11} \\ f_{12} \\ f_{13} \\ f_{21} \\ f_{22} \\ f_{23} \\ f_{31} \\ f_{32} \\ f_{33} \end{bmatrix} = 0 \tag{51}$$

Since the number of degrees of freedom for a fundamental matrix is seven, given seven point correspondences, an exact solution to the above equation can be found. If more than seven point correspondences are provided, then an approximate solution has to be obtained by minimizing the norm $\|Af\|$ where A is a matrix that can be constructed from the point correspondences. The solution is given by the eigenvector of A^TA corresponding to the least eigenvalue.

Alternatively, the fundamental matrix can also be found from homographies. Homographies can be estimated using equation (48) or using four coplanar point correspondences using DLT. If two more point correspondences are available, then the epipole e' can be found using the intersection of lines $Hx_1 \times x_1'$ and $Hx_2 \times x_2'$. From the epipole and the homography, the fundamental matrix can be determined using $F = [e']_\times H$

4) Computing epipolar lines using F: Epipolar lines are very useful in constraining the search space of a point in the other image. Given a point in the left image, we know that it will lie on the epipolar line associated with this point in the right image. Hence the search space is greatly reduced to a one-dimensional line. This epipolar line can be found using the fundamental matrix. The epipolar line in the left image associated with a point (x', y') in the right image can be computed as:

$$\begin{bmatrix} x' & y' & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0 \tag{52}$$

$$\begin{bmatrix} f_{11}x' + f_{21}y' + f_{31} & f_{12}x' + f_{22}y' + f_{32} & f_{13}x' + f_{23}y' + f_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = 0 \tag{53}$$

$$ax + by + c = 0 \tag{54}$$

The above equation gives the epipolar line in the left image associated with the point (x', y') in the right image. Similarly the epipolar line in the right image associated with (x, y) can be obtained by:

$$\begin{bmatrix} x' & y' & 1 \end{bmatrix} \begin{bmatrix} f_{11}x + f_{12}y + f_{13} \\ f_{21}x + f_{22}y + f_{23} \\ f_{31}x + f_{32}y + f_{33} \end{bmatrix} = 0 \tag{55}$$

$$ax' + by' + c = 0 \tag{56}$$

C. Essential Matrix

Essential matrix was introduced before fundamental matrix and it is a specialization of the fundamental matrix where the cameras are calibrated beforehand, *i.e.*, $K = K' = I$. The matrix that defines the geometric relationship between the two normalized points when the cameras are calibrated apriori is known as essential matrix and from equation (40), we have

$$E = [t]_{\times}R \tag{57}$$

The translation and rotation matrix contribute three degrees of freedom each. Since the essential matrix is homogenous, the total number of degrees of freedom is five [2].

1) Relation with fundamental matrix: From equations (42) and (57), the relation between fundamental and essential matrix can be established as:

$$F = K'^{-T}EK^{-1} \tag{58}$$

$$E = K'^{T}FK \tag{59}$$

2) Anatomy of essential matrix: The essential matrix is given by $E = [t]_{\times}R = SR$, where $S = [t]_{\times}$ is a skew-symmetric matrix, *i.e.*, $S^T = -S$ and R is an orthogonal matrix, *i.e.*, $R^TR = I$. The eigenvalue decomposition of a skew-symmetric matrix is given by:

$$S = kUZU^T \tag{60}$$

where U is orthogonal and Z is of the form:

$$Z = \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \tag{61}$$

The proof for the decomposition of a skew-symmetric matrix is given in [3]. It can be noted that Z is also skew-symmetric. Now let us consider an orthogonal matrix

$$W = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{62}$$

such that $Z = \Sigma W$, where $\Sigma = diag(1, 1, 0)$, ignoring signs. Since, W, U and R are orthogonal, the eigenvalue decomposition of E can be written as

$$E = SR = UZU^TR = U\Sigma(WU^TR) \tag{63}$$

From the above decomposition, it follows that a 3×3 matrix is an essential matrix if and only if two of its singular values are equal and the third is zero. This forms the internal constraints of the essential matrix and accounts for the reduced number of degrees of freedom.

This property is also very useful in deducing the projection matrix of the two cameras without the overall scale. Let the first camera be at the world origin and $P' = [R\ \mathbf{t}]$ be the projection matrix of the other camera. By assuming $\Sigma = diag(1, 1, 0)$, we ignore the overall scale, and the SVD of E is given by

$$E = U\Sigma V^T \tag{64}$$

We know that $E = SR$ and the decomposition of S is given by equation (60). Let R be decomposed as $R = UAV^T$, where A is some rotation matrix. Hence E can be factorized as

$$E = SR = UZU^TUAV^T = UZAV^T \tag{65}$$

From (64) and (65), we have $ZA = \Sigma$. Since A is an orthogonal rotation matrix, it follows that $A = W$ or $A = W^T$. Hence R can be decomposed in two possible ways:

$$R = UWV^T \text{ or } R = UW^TV^T \tag{66}$$

The translation part of the projective matrix is related to S. Let $\mathbf{t}$ be any vector $(a, b, c)^T$. Then we have

$$S\mathbf{t} = [t]_\times \mathbf{t} = \begin{bmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \mathbf{0} \tag{67}$$

Since $S\mathbf{t} = (UZU^T)\mathbf{t} = \mathbf{0}$ and the last column of Z is $\mathbf{0}$, we get $t = U(0, 0, \pm k)^T = \pm k u_3$ which corresponds to the last column of U. But the scale and sign of the translation component cannot be determined.

To summarize, ignoring the overall scale and assuming that the first camera is at the world origin, the four possible solutions for the projection matrix of the second camera based on the two possible decompositions of R and two possible signs of t are

$$\begin{aligned} P' &= [UWV^T \mid +u_3] \\ P' &= [UWV^T \mid -u_3] \\ P' &= [UW^TV^T \mid +u_3] \\ P' &= [UW^TV^T \mid -u_3] \end{aligned} \tag{68}$$

The sign change in t reverses the baseline and the alternate choice of decomposition of R rotates the camera 180 degrees about the baseline. Interpretation of the four possible solutions are given in figure 4. It can be seen that only one solution has the world coordinate in front of both the cameras. Hence, the solution which gives the world coordinate in front of both the cameras can be used for practical purposes.

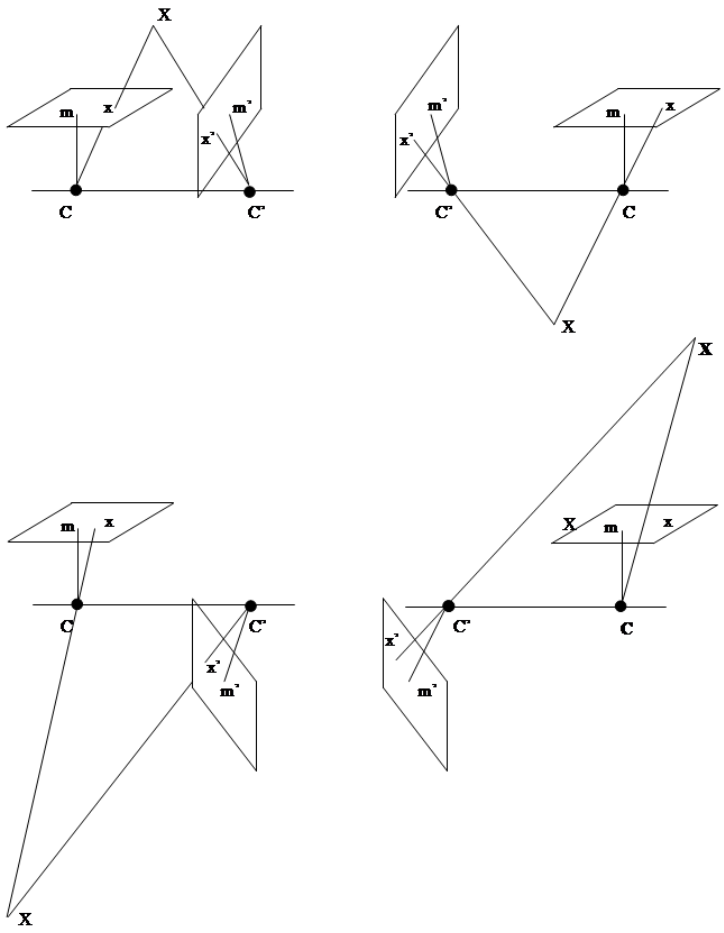

Fig. 4. The four possible solutions for the projective matrix of the camera

References

[1] Stan Birchfield, *An Introduction to Projective Geometry*, http://www.ces.clemson.edu/ stb/projective/, 1998.

[2] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[3] G. H. Golub and C. F. Van Loan, *Matrrix Computations*, The Johns Hopkins University Press, Baltimore, MD, second edition, 1989.