

Discrete Epipolar Geometry

Masatoshi Hamanaka¹, Yukiko Kenmochi², and Akihiro Sugimoto³

¹ Department of Information Technology, Okayama University, Japan

² UMR 8049 - IGM, CNRS/University of Marne-la-Vallée/ESIEE, France

³ National Institute of Informatics, Japan

y.kenmochi@esiee.fr

sugimoto@nii.ac.jp

Abstract. The epipolar geometry, which lies in the basis of 3D reconstruction techniques in the field of computer vision, is formulated in continuous spaces and gives geometric relationships between different views of a point in space. In applications, however, we cannot deal with points themselves in digital images. This is because digital images involve some digitization process and the smallest unit in digital images is a pixel. In this paper, we propose a discrete version of the epipolar geometry, called the discrete epipolar geometry, that gives geometric relationships between pixels rather than points. We then apply this discrete epipolar geometry to 3D reconstruction.

1 Introduction

The problem of inferring 3D information of a scene from a set of its images has a long history in computer vision [3, 4, 7]. In particular, the epipolar geometry [7], which is the intrinsic projective geometry between two views independent of scene structure, has played a central role in reconstructing 3D information.

An intuitive interpretation of the epipolar geometry can be summarized as follows. Suppose a point in the 3D space is imaged in two views observed from two given different viewpoints. Each image point together with its viewpoint defines a straight line (a line in short below) in space, which is known as a ray of sight. Since we consider two viewpoints, we have two rays of sight with respect to a spatial point. In general, two lines in space are in a twisted position and do not meet. In our situation, however, two rays of sight do have the intersection, i.e., the point in space in observation, since we observe the point from two different viewpoints. This results in a constraint on the coordinates of two image points. The geometrical description of this constraint is the epipolar geometry and its algebraic description is the epipolar equation. As we see, the epipolar geometry is on the framework of projections between continuous spaces.

In applications, however, we cannot deal with points themselves in digital images. This is because digital images involve some digitization process and the smallest unit in digital images is not a point but a pixel. In fact, in the presence of noise including digitization errors, two rays of sight do not generally meet, which causes the problem of finding the best point of intersection [6].

A tremendous amount of efforts have been made to deal with noise and to show the robustness of developed methods (see [5, 8, 9, 12], for example). Some introduce a noise model and statistically analyze the stableness and robustness from the theoretical point of view. Some methods are experiment-oriented: they show evidences of the robustness by their intensive experiments. In such studies, however, errors incurred in the digitization process are not directly treated¹ and they are regarded as portions of observation errors. This indicates that few attempts have been made to discriminate digitization errors from observation errors.

Digitization errors and observation errors are originally different from each other and they have to be discriminated. This is because these two kinds of errors are generated in different mechanisms. This paper attempts to discriminate the two kinds of errors, focusing on pixels as the smallest unit of digital images. Namely, we propose the discrete epipolar geometry, i.e., a discrete version of the conventional epipolar geometry. The discrete epipolar geometry is formulated through projections from the continuous 3D space to discrete image planes. We introduce the notions of a pyramidal ray of sight and a discrete epipolar line, and describe them in the mathematically well-defined form. We also present some experiments of 3D reconstruction using the discrete epipolar geometry.

2 Epipolar Geometry

We review here the conventional epipolar geometry and give a mathematically rigorous definition to an epipolar line.

We assume that the position and orientation of a camera with the focal length f are given. We denote by \mathcal{R}^n the n -dimensional space over the real number field \mathcal{R} . We also denote by $\mathbf{I} (\subset \mathcal{R}^2)$ a finite plane representing an image plane. Let C be the camera coordinate system, as shown in Fig. 1, where the origin coincides with the viewpoint, the xy -plane is parallel to the image plane \mathbf{I} , and the z -coordinate is parallel to the optical axis and toward to the principal point of the image plane. We also introduce a coordinate system to \mathbf{I} so that its origin is the principal point and its x - and y -coordinates are respectively identical with those of C .

Let² $\mathbf{P} = (X, Y, Z)^\top$ denote a point in $\mathbf{V} (\subset \mathcal{R}^3)$ where \mathbf{V} is the 3D space observable by the camera. We then have its image point $\mathbf{p} = (x, y)^\top$:

$$x = f \frac{X}{Z}, \quad y = f \frac{Y}{Z}. \quad (2.1)$$

The viewpoint, i.e., the origin of C , and $\tilde{\mathbf{p}} = (x, y, f)^\top$ define a line in the 3D space. This line is known as the ray of sight of an image point \mathbf{p} .

¹ In the latter half of 1980s, some discussion on digitization errors was reported [2, 11]. Such discussion, however, seems to be limited with the knowledge available at that time.

² We use column vectors and denote by \mathbf{x}^\top the transposition of a vector \mathbf{x} .

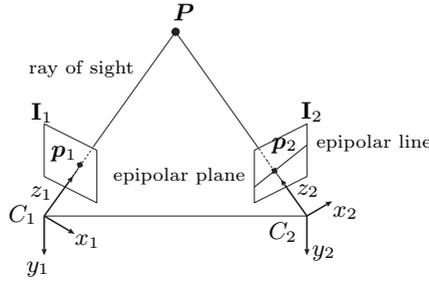


Fig. 1. The conventional epipolar geometry

We now discuss the case of two cameras. Letting two calibrated cameras be available (Fig. 1), we denote by C_1 and C_2 their camera coordinate systems, and by I_1 and I_2 the finite planes representing their image planes. We also denote their focal lengths by f_1 and f_2 .

When we observe a point P in the 3D space, we obtain its two image points p_1 and p_2 on I_1 and I_2 respectively. We call them corresponding points with respect to P . P and two viewpoints define a plane in the 3D space, called an epipolar plane of P . As shown in Fig. 1, two image points p_1 and p_2 are also on the epipolar plane of P . The intersection of the epipolar plane and each image plane forms a line, called an epipolar line. This epipolar line provides us with a constraint on the corresponding points independent of the position of a point in the 3D space. Namely, for a given point in an image plane, the location of its corresponding point in the other image plane is bounded on the epipolar line.

This geometric configuration can be algebraically described with the essential matrix. Essential matrix E , which is defined by³ $E = [t]_{\times}R$ where R and t are a rotation matrix and a translation vector relating the coordinates in C_1 and C_2 [7], thus plays a central role in the conventional epipolar geometry. Namely, for a pair of corresponding points, $p_1 = (x_1, y_2)^T \in I_1$ and $p_2 = (x_2, y_2)^T \in I_2$, we have

$$\tilde{p}_1^T E \tilde{p}_2 = 0, \tag{2.2}$$

where $\tilde{p}_1 = (x_1, y_1, f_1)^T$ and $\tilde{p}_2 = (x_2, y_2, f_2)^T$. (2.2) is called the epipolar equation, providing that two cameras are calibrated.

The epipolar equation gives the constraint on pairs of corresponding points across two views. To be more specific, for a given point $p_1 \in I_1$, the location of its corresponding point in I_2 is restricted into the line satisfying (2.2). This line is identical with the epipolar line. Mathematically, the epipolar line $EL(p_1)$ for $p_1 \in I_1$ is defined by

$$EL(p_1) = \{p_2 \in I_2 \mid \tilde{p}_1^T E \tilde{p}_2 = 0\}. \tag{2.3}$$

We note that $EL(p_1)$ is a set of points in I_2 that satisfy the epipolar equation.

³ $[t]_{\times}$ is the 3×3 skew-symmetric matrix constructed from a 3D vector t : $[t]_{\times}x = t \times x$ for any 3D vector x .

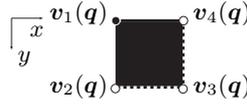


Fig. 2. The region $\Omega(\mathbf{q}) \subset \mathbf{I}$ dominated by pixel $\mathbf{q} = (i, j)^\top \in \mathbf{K}$ and its four vertices

3 Digitization of an Image Plane

The smallest unit of images is not a point but a pixel. This indicates that an image plane is not continuous but digitized. We here introduce a digitization into an image plane. We denote by $1/r_x$ and $1/r_y$ the resolutions of x - and y -coordinates. We also denote by \mathcal{Z} the set of all integers.

For a given point $\mathbf{p} = (x, y)^\top \in \mathbf{I}$, we define

$$i = \lfloor \frac{x}{r_x} + \frac{1}{2} \rfloor, \quad j = \lfloor \frac{y}{r_y} + \frac{1}{2} \rfloor. \tag{3.1}$$

$\mathbf{q} = (i, j)^\top \in \mathcal{Z}^2$ is a digitization of \mathbf{p} and called a pixel representing \mathbf{p} . Applying this digitization to all the points in an image plane \mathbf{I} leads to the digitization of the image plane \mathbf{I} :

$$\mathbf{K} = \left\{ \begin{pmatrix} i \\ j \end{pmatrix} \in \mathcal{Z}^2 \mid r_x \left(i - \frac{1}{2} \right) \leq x < r_x \left(i + \frac{1}{2} \right), \right. \\ \left. r_y \left(j - \frac{1}{2} \right) \leq y < r_y \left(j + \frac{1}{2} \right), \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbf{I} \right\}. \tag{3.2}$$

We note that this digitization ensures a surjective mapping from points in \mathbf{I} to pixels in \mathbf{K} .

Conversely, the region $\Omega(\mathbf{q}) \subset \mathbf{I}$ dominated by pixel $\mathbf{q} = (i, j)^\top \in \mathbf{K}$ is given by

$$\Omega(\mathbf{q}) = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbf{I} \mid r_x \left(i - \frac{1}{2} \right) \leq x < r_x \left(i + \frac{1}{2} \right), \right. \\ \left. r_y \left(j - \frac{1}{2} \right) \leq y < r_y \left(j + \frac{1}{2} \right) \right\}, \tag{3.3}$$

which is illustrated in Fig. 2. Note that the solid lines and the black circles are included in $\Omega(\mathbf{q})$ while the dotted lines and the white circles are not. We see in Fig. 2 that $\Omega(\mathbf{q})$ forms a rectangle.

A point in space is imaged in a view as given in (2.1). Combining (2.1) and (3.1) enables us to uniquely determine the pixel representing its image point for a given point in space. Namely, for a point $\mathbf{P} = (X, Y, Z)^\top$, the pixel $\mathbf{q} = (i, j)^\top$ representing its image point is written by

$$i = \lfloor \frac{fX}{r_x Z} + \frac{1}{2} \rfloor, \quad j = \lfloor \frac{fY}{r_y Z} + \frac{1}{2} \rfloor. \tag{3.4}$$

We now have the following proposition.

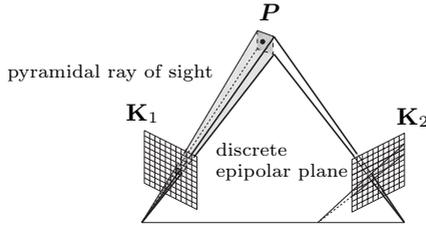


Fig. 3. Discrete epipolar geometry: a pyramidal ray of sight for a pixel in the first image plane and its projection onto the second image plane



Fig. 4. Discrete epipolar line obtained from epipolar inequalities: two examples without (left) and with (right) an epipole

Proposition 1. Any point P in V has the unique pixel in K that represents an image point of P .

Various digitization schemes including (3.1) have been already proposed to discuss geometric properties of digitized objects such as connectivities, bubble-freeness and topologies (see [1, 10], for example). In our case, we employ (3.1) to obtain Proposition 1 that plays an important role for building up our discrete epipolar geometry.

4 Discrete Epipolar Geometry

In the conventional framework, for a point p_1 in the image plane I_1 , a ray of sight is defined as the line going through p_1 and the viewpoint. Projecting this ray of sight onto the image plane I_2 observed from another viewpoint forms an epipolar line in I_2 (see (2.3) and Fig. 1). In our framework, which we call the *discrete epipolar geometry*, on the other hand, for a pixel q_1 in K_1 , i.e., the digitization of I_1 , the quadrangular prism is defined by the pixel q_1 and the viewpoint as shown in Fig. 3. In this paper, this quadrangular prism is called a *pyramidal ray of sight*. Projecting the pyramidal ray of sight onto the other image plane I_2 forms a region that is identified by two inequalities, called *epipolar inequalities*. We digitize the region to identify it as a set of pixels among the digitization of I_2 , i.e., K_2 (Fig. 4). We call the set a *discrete epipolar line*.

The above mentioned geometric configuration is algebraically captured by the following steps.

1. Define a pyramidal ray of sight for a given pixel $\mathbf{q}_1 \in \mathbf{K}_1$.
2. Define the epipolar inequalities as the projection of the pyramidal ray of sight onto \mathbf{I}_2 .
3. Obtain the discrete epipolar line in \mathbf{K}_2 .

4.1 Pyramidal Ray of Sight

We here give the algebraic description to a pyramidal ray of sight (Fig. 3).

In our digitization of an image plane into pixels, any pixel forms a rectangle of r_x by r_y , and thus has four vertices. Given a pixel $\mathbf{q} = (i, j)^\top \in \mathbf{K}$, we define four vertices $\mathbf{v}_k(\mathbf{q}) \in \mathbf{I}$ ($k = 1, 2, 3, 4$) as illustrated in Fig. 2:

$$\begin{aligned} \mathbf{v}_1(\mathbf{q}) &= \left(r_x \left(i - \frac{1}{2} \right), r_y \left(j - \frac{1}{2} \right) \right)^\top, & \mathbf{v}_2(\mathbf{q}) &= \left(r_x \left(i - \frac{1}{2} \right), r_y \left(j + \frac{1}{2} \right) \right)^\top, \\ \mathbf{v}_3(\mathbf{q}) &= \left(r_x \left(i + \frac{1}{2} \right), r_y \left(j + \frac{1}{2} \right) \right)^\top, & \mathbf{v}_4(\mathbf{q}) &= \left(r_x \left(i + \frac{1}{2} \right), r_y \left(j - \frac{1}{2} \right) \right)^\top. \end{aligned}$$

$\mathbf{q} \in \mathbf{K}$ together with a viewpoint defines the quadrangular prism:

$$\mathbf{S}(\mathbf{q}) = \{ \mathbf{X} \in \mathbf{V} \mid (\tilde{\mathbf{v}}_1(\mathbf{q}) \times \tilde{\mathbf{v}}_2(\mathbf{q}))^\top \mathbf{X} \geq 0, (\tilde{\mathbf{v}}_2(\mathbf{q}) \times \tilde{\mathbf{v}}_3(\mathbf{q}))^\top \mathbf{X} > 0, (\tilde{\mathbf{v}}_3(\mathbf{q}) \times \tilde{\mathbf{v}}_4(\mathbf{q}))^\top \mathbf{X} > 0, (\tilde{\mathbf{v}}_4(\mathbf{q}) \times \tilde{\mathbf{v}}_1(\mathbf{q}))^\top \mathbf{X} \geq 0 \}, \quad (4.1)$$

where \mathbf{X} is the coordinates of a 3D point represented in the camera coordinate system C , \mathbf{V} is the 3D space observable from the viewpoint, and $\tilde{\mathbf{v}}_k(\mathbf{q}) = (\mathbf{v}_k(\mathbf{q})^\top, f)^\top$. $\mathbf{S}(\mathbf{q})$ is called the *pyramidal ray of sight* of \mathbf{q} .

4.2 Epipolar Inequalities

For a given pixel $\mathbf{q}_1 \in \mathbf{K}_1$, we have four vertices, each of which together with two given viewpoints defines its epipolar plane. We see that the pyramidal ray of sight $\mathbf{S}(\mathbf{q}_1)$ of \mathbf{q}_1 is located between two of the four epipolar planes (cf. Fig. 3). We call the region between the two epipolar planes that include $\mathbf{S}(\mathbf{q}_1)$ the *discrete epipolar plane* of \mathbf{q}_1 . The discrete epipolar plane is represented by a pair of half-space inequalities. The intersection of the discrete epipolar plane and the image plane \mathbf{I}_2 thus forms a region bounded by the two epipolar lines in \mathbf{I}_2 . The region is represented by two inequalities in \mathbf{I}_2 , and the two inequalities are called *epipolar inequalities*.

The following steps identify the epipolar inequalities for a given $\mathbf{q}_1 \in \mathbf{K}_1$.

1. Consider the epipolar planes of four vertices $\mathbf{v}_k(\mathbf{q}_1)$ ($k = 1, 2, 3, 4$), and select two vertices $\boldsymbol{\mu}(\mathbf{q}_1)$ and $\boldsymbol{\nu}(\mathbf{q}_1)$ among the four vertices so that the pyramidal ray of sight $\mathbf{S}(\mathbf{q}_1)$ is located between the epipolar planes of $\boldsymbol{\mu}(\mathbf{q}_1)$ and $\boldsymbol{\nu}(\mathbf{q}_1)$.
2. Project $\mathbf{S}(\mathbf{q}_1)$ onto \mathbf{I}_2 , and obtain two inequalities from the epipolar lines $\mathbf{EL}(\boldsymbol{\mu}(\mathbf{q}_1))$ and $\mathbf{EL}(\boldsymbol{\nu}(\mathbf{q}_1))$.

For a pixel $\mathbf{q}_1 \in \mathbf{K}_1$, let $\tilde{\mathbf{v}}_k(\mathbf{q}_1) = (\mathbf{v}_k(\mathbf{q}_1)^\top, f_1)^\top$ ($k = 1, 2, 3, 4$). We now represent epipolar planes corresponding to the vertices by

$$(\mathbf{t} \times \tilde{\mathbf{v}}_k(\mathbf{q}_1))^\top \mathbf{X} = 0, \quad (4.2)$$

where \mathbf{t} is the translation vector from the origin of C_1 to that of C_2 , and \mathbf{X} is the coordinates of a 3D point represented in C_1 . We then select two of the four planes between which the pyramidal ray of sight $\mathbf{S}(\mathbf{q}_1)$ is located.

We select a vertex $\mathbf{v}_k(\mathbf{q}_1)$ so that the other three vertices $\mathbf{v}_{k'}(\mathbf{q}_1)$'s ($k' = 1, 2, 3, 4; k' \neq k$) are all in one side of the epipolar plane of $\mathbf{v}_k(\mathbf{q}_1)$. More specifically, we select $\mathbf{v}_k(\mathbf{q}_1)$ such that all the $\mathbf{v}_{k'}(\mathbf{q}_1)$'s except for $\mathbf{v}_k(\mathbf{q}_1)$ ($k' = 1, 2, 3, 4; k' \neq k$) satisfy either (4.3) or (4.4):

$$(\mathbf{t} \times \tilde{\mathbf{v}}_k(\mathbf{q}_1))^\top \tilde{\mathbf{v}}_{k'}(\mathbf{q}_1) \geq 0, \tag{4.3}$$

$$(\mathbf{t} \times \tilde{\mathbf{v}}_k(\mathbf{q}_1))^\top \tilde{\mathbf{v}}_{k'}(\mathbf{q}_1) \leq 0. \tag{4.4}$$

We note that the number of selected vertices is at least two and that it is two if and only if the four epipolar planes are different from each other (this occurs in most cases). In the case where more than two vertices are selected, namely, the case where two epipolar planes coincide with each other, only two vertices having different epipolar planes from each other are used. When $\mathbf{v}_1(\mathbf{q}_1)$ is one of the selected vertices, we always use $\mathbf{v}_1(\mathbf{q}_1)$ due to the speciality of the vertex in our digitization of image planes (see Fig. 2).

Let $\boldsymbol{\mu}(\mathbf{q}_1)$ and $\boldsymbol{\nu}(\mathbf{q}_1)$ be the two selected vertices $\mathbf{v}_k(\mathbf{q}_1)$'s where $\boldsymbol{\mu}(\mathbf{q}_1)$ is the smaller than $\boldsymbol{\nu}(\mathbf{q}_1)$ with respect to k . For $\boldsymbol{\mu}(\mathbf{q}_1)$, we define a function with respect to $\tilde{\mathbf{p}}_2 = (\mathbf{p}_2^\top, f_2)^\top$, i.e., the coordinates of $\mathbf{p}_2 \in \mathbf{I}_2$ represented in C_2 .

$$h_{\boldsymbol{\mu}(\mathbf{q}_1)}(\tilde{\mathbf{p}}_2) = \begin{cases} \tilde{\boldsymbol{\mu}}(\mathbf{q}_1)^\top E\tilde{\mathbf{p}}_2 & \text{if } \boldsymbol{\mu}(\mathbf{q}_1) \text{ is selected by (4.3),} \\ -\tilde{\boldsymbol{\mu}}(\mathbf{q}_1)^\top E\tilde{\mathbf{p}}_2 & \text{if } \boldsymbol{\mu}(\mathbf{q}_1) \text{ is selected by (4.4),} \end{cases}$$

where $\tilde{\boldsymbol{\mu}}(\mathbf{q}_1) = (\boldsymbol{\mu}(\mathbf{q}_1)^\top, f_1)^\top$. In the similar way, we define $h_{\boldsymbol{\nu}(\mathbf{q}_1)}(\tilde{\mathbf{p}}_2)$ for $\boldsymbol{\nu}(\mathbf{q}_1)$. We then have two inequalities that represent the intersection of \mathbf{I}_2 and the discrete epipolar plane of \mathbf{q}_1 :

$$\begin{aligned} h_{\boldsymbol{\mu}(\mathbf{q}_1)}(\tilde{\mathbf{p}}_2) &> 0, \\ h_{\boldsymbol{\nu}(\mathbf{q}_1)}(\tilde{\mathbf{p}}_2) &> 0. \end{aligned} \tag{4.5}$$

We call the above two inequalities *epipolar inequalities*. We remark that if $\boldsymbol{\mu}(\mathbf{q}_1) = \mathbf{v}_1(\mathbf{q}_1)$, (4.5) is replaced by

$$h_{\boldsymbol{\mu}(\mathbf{q}_1)}(\tilde{\mathbf{p}}_2) \geq 0.$$

4.3 Discrete Epipolar Line

For a given pixel $\mathbf{q}_1 \in \mathbf{K}_1$, the region $\mathbf{H}(\mathbf{q}_1) \subset \mathbf{I}_2$ satisfying the epipolar inequalities is given by

$$\mathbf{H}(\mathbf{q}_1) = \begin{cases} \{\mathbf{p}_2 \in \mathbf{I}_2 \mid h_{\boldsymbol{\mu}(\mathbf{q}_1)}(\tilde{\mathbf{p}}_2) > 0, h_{\boldsymbol{\nu}(\mathbf{q}_1)}(\tilde{\mathbf{p}}_2) > 0\} & \text{if } \boldsymbol{\mu}(\mathbf{q}_1) \neq \mathbf{v}_1(\mathbf{q}_1), \\ \{\mathbf{p}_2 \in \mathbf{I}_2 \mid h_{\boldsymbol{\mu}(\mathbf{q}_1)}(\tilde{\mathbf{p}}_2) \geq 0, h_{\boldsymbol{\nu}(\mathbf{q}_1)}(\tilde{\mathbf{p}}_2) > 0\} & \text{otherwise.} \end{cases} \tag{4.6}$$

We digitize the image plane \mathbf{I}_2 in the same way as (3.2) and obtain \mathbf{K}_2 . We also define $\boldsymbol{\Omega}(\mathbf{q}_2) \subset \mathbf{I}_2$ for a pixel $\mathbf{q}_2 \in \mathbf{K}_2$ (cf. (3.3)). Then, $\mathbf{H}(\mathbf{q}_1)$ and $\boldsymbol{\Omega}(\mathbf{q}_2)$ enable us to represent a set of pixels satisfying the epipolar inequalities in \mathbf{K}_2 :

$$\text{DEL}(\mathbf{q}_1) = \{\mathbf{q}_2 \in \mathbf{K}_2 \mid \mathbf{H}(\mathbf{q}_1) \cap \boldsymbol{\Omega}(\mathbf{q}_2) \neq \emptyset\}.$$

We call $\mathbf{DEL}(\mathbf{q}_1)$ the *discrete epipolar line* for a pixel $\mathbf{q}_1 \in \mathbf{K}_1$. As we see the discrete epipolar line can be interpreted as the digitization of $\mathbf{H}(\mathbf{q}_1)$.

Though the discrete epipolar line $\mathbf{DEL}(\mathbf{q}_1)$ for a given pixel \mathbf{q}_1 is defined as a set of pixels in \mathbf{K}_2 , in its computation we require a processing to be done in the continuous space \mathbf{I}_2 because of $\mathbf{H}(\mathbf{q}_1) \cap \Omega(\mathbf{q}_2)$. This indicates that straightforward computing of $\mathbf{DEL}(\mathbf{q}_1)$ is not practical. We present a computationally feasible algorithm below to compute $\mathbf{DEL}(\mathbf{q}_1)$ from $\mathbf{H}(\mathbf{q}_1)$ where each pixel in \mathbf{K}_2 is checked whether it is included in $\mathbf{DEL}(\mathbf{q}_1)$ using epipolar inequalities.

Algorithm 1 [Computing a discrete epipolar line]

Input: $\mathbf{H}(\mathbf{q}_1)$ computed from a given pixel $\mathbf{q}_1 \in \mathbf{K}_1$.

Output: A discrete epipolar line $\mathbf{DEL}(\mathbf{q}_1) \subset \mathbf{K}_2$.

begin

1. Set $\mathbf{DEL}(\mathbf{q}_1)$ to be empty.
2. **For** each $\mathbf{q}_2 \in \mathbf{K}_2$,
 - 2.1. compute four vertices $\mathbf{v}_i(\mathbf{q}_2) \in \mathbf{I}_2$ ($i = 1, 2, 3, 4$);
 - 2.2. **switch** $\mu(\mathbf{q}_1)$
 - case** $\mu(\mathbf{q}_1) \neq \mathbf{v}_1(\mathbf{q}_1)$: **if** at least one $\mathbf{v}_i(\mathbf{q}_2)$ exists that satisfies $h_{\mu(\mathbf{q}_1)}(\tilde{\mathbf{v}}_i(\mathbf{q}_2)) > 0$, where $\tilde{\mathbf{v}}_i(\mathbf{q}_2) = (\mathbf{v}_i(\mathbf{q}_2)^\top, f_2)^\top$ ($i = 1, 2, 3, 4$);
 - case** $\mu(\mathbf{q}_1) = \mathbf{v}_1(\mathbf{q}_1)$: **if** at least one $\mathbf{v}_i(\mathbf{q}_2)$ exists that satisfies $h_{\mu(\mathbf{q}_1)}(\tilde{\mathbf{v}}_i(\mathbf{q}_2)) > 0$ ($i = 2, 3, 4$) or $h_{\mu(\mathbf{q}_1)}(\tilde{\mathbf{v}}_1(\mathbf{q}_2)) \geq 0$;
 - 2.2.1. **if** at least one $\mathbf{v}_i(\mathbf{q}_2)$ exists that satisfies $h_{\nu(\mathbf{q}_1)}(\tilde{\mathbf{v}}_i(\mathbf{q}_2)) > 0$ **then** put \mathbf{q}_2 into $\mathbf{DEL}(\mathbf{q}_1)$.
3. Compute⁴ \mathbf{e}_2 such that $h_{\mu(\mathbf{q}_1)}(\tilde{\mathbf{e}}_2) = 0$ and $h_{\nu(\mathbf{q}_1)}(\tilde{\mathbf{e}}_2) = 0$, where $\tilde{\mathbf{e}}_2 = (\mathbf{e}_2^\top, f_2)^\top$.
4. **If** $\mathbf{e}_2 \in \mathbf{I}_2$ **then**
 - 4.1. obtain the digitization of \mathbf{e}_2 ; let the digitization be $\boldsymbol{\epsilon}_2 \in \mathbf{K}_2$;
 - 4.2. **if** we obtain two connected components \mathbf{D}_1 and \mathbf{D}_2 in $\mathbf{DEL}(\mathbf{q}_1)$ when we remove $\boldsymbol{\epsilon}_2$ from $\mathbf{DEL}(\mathbf{q}_1)$ **then**
 - 4.2.1. $\mathbf{q}_2 = \boldsymbol{\epsilon}_2$;
 - 4.2.2. **while** $\mathbf{q}_2 = \boldsymbol{\epsilon}_2$ **do for** $s \geq 1$
 - let a 3D point be $\mathbf{P} = s(\mathbf{v}_1(\mathbf{q}_1)^\top, f_1)^\top$; represent \mathbf{P} in the second camera coordinate system C_2 and obtain the pixel \mathbf{q}_2 that represents its image point in \mathbf{I}_2 (cf. (3.4));
 - 4.2.3. select a \mathbf{D}_k ($k = 1, 2$) that includes \mathbf{q}_2 , and set $\mathbf{DEL}(\mathbf{q}_1) = \mathbf{D}_k \cup \{\boldsymbol{\epsilon}_2\}$.
5. Return $\mathbf{DEL}(\mathbf{q}_1)$.

end

We see that discrete epipolar lines enjoy the following properties.

Property 1. For a pixel nearby an epipole, its discrete epipolar line dominates a wider region, while it does a small region for a pixel far from an epipole.

⁴ \mathbf{e}_2 is known as an epipole.

Property 2. For the pixel including an epipole in an image, its discrete epipolar line in the other image expands into the whole image.

We denote by \mathbf{V}_1 and \mathbf{V}_2 the 3D spaces in \mathcal{R}^3 observable by two given cameras with their positions and orientations. For a 3D point \mathbf{P} in $\mathbf{V}_1 \cap \mathbf{V}_2$, we let \mathbf{q}_1 and \mathbf{q}_2 be the pixels in \mathbf{K}_1 and \mathbf{K}_2 representing its image points. We call such a pair of pixels \mathbf{q}_1 and \mathbf{q}_2 *corresponding pixels* with respect to \mathbf{P} as a counterpart of corresponding points in the conventional framework. We see that the discrete epipolar line gives a constraint on pairs of corresponding pixels across \mathbf{K}_1 and \mathbf{K}_2 .

Proposition 2. For a pixel $\mathbf{q}_1 \in \mathbf{K}_1$, its corresponding pixel in \mathbf{K}_2 , if exists, is included in $\text{DEL}(\mathbf{q}_1)$.

Corollary 1. For any pair of corresponding pixels, their pyramidal rays of sight always have a non-empty intersection in $\mathbf{V}_1 \cap \mathbf{V}_2$.

In the conventional framework, for a point $\mathbf{p}_1 \in \mathbf{I}_1$, a point $\mathbf{p}_2 \in \mathbf{I}_2$ satisfying the epipolar equation (2.2) is selected as its corresponding point. In practice, however, what we can deal with are not the points \mathbf{p}_1 and \mathbf{p}_2 but the pixels \mathbf{q}_1 and \mathbf{q}_2 , digitization of \mathbf{p}_1 and \mathbf{p}_2 . We thus have to select one point from \mathbf{q}_1 and one point from \mathbf{q}_2 , and then apply triangulation to the selected points to reconstruct a 3D point. The rays of sight obtained in such a way do not always meet in the 3D space [6]. In that case, we cannot reconstruct any point in space as the intersection of the rays of sight. In our framework, on the other hand, the pyramidal rays of sight for any pair of corresponding pixels do have their non-empty intersection in the 3D space even if digitization errors exist. That is, in the framework of the discrete epipolar geometry, a non-empty region in the 3D space does exist that is reconstructed from any pair of corresponding pixels.

5 3D Reconstruction and Its Experiment

5.1 3D Reconstruction from Corresponding Pixels

In the conventional epipolar geometry, we reconstruct a point in space from a pair of corresponding points as the intersection of their rays of sight. In our discrete epipolar geometry, however, a pair of corresponding pixels provides a polyhedron as the intersection of their pyramidal rays of sight.

Let $\mathbf{q}_1 \in \mathbf{K}_1$ and $\mathbf{q}_2 \in \mathbf{K}_2$ be a pair of corresponding pixels, and \mathbf{V}_1 and \mathbf{V}_2 be the 3D spaces observable from the first and second cameras, respectively. Their pyramidal rays of sight $\mathbf{S}_1(\mathbf{q}_1) \subset \mathbf{V}_1$ and $\mathbf{S}_2(\mathbf{q}_2) \subset \mathbf{V}_2$ are then given, similarly to (4.1), by

$$\begin{aligned} \mathbf{S}_1(\mathbf{q}_1) = \{ \mathbf{X}_1 \in \mathbf{V}_1 \mid \\ (\tilde{\mathbf{v}}_1(\mathbf{q}_1) \times \tilde{\mathbf{v}}_2(\mathbf{q}_1))^\top \mathbf{X}_1 \geq 0, (\tilde{\mathbf{v}}_2(\mathbf{q}_1) \times \tilde{\mathbf{v}}_3(\mathbf{q}_1))^\top \mathbf{X}_1 > 0, \\ (\tilde{\mathbf{v}}_3(\mathbf{q}_1) \times \tilde{\mathbf{v}}_4(\mathbf{q}_1))^\top \mathbf{X}_1 > 0, (\tilde{\mathbf{v}}_4(\mathbf{q}_1) \times \tilde{\mathbf{v}}_1(\mathbf{q}_1))^\top \mathbf{X}_1 \geq 0 \}, \quad (5.1) \end{aligned}$$

$$\mathbf{S}_2(\mathbf{q}_2) = \{ \mathbf{X}_1 \in \mathbf{V}_2 \mid \mathbf{X}_1 = R\mathbf{X}_2 + \mathbf{t}, \\ (\tilde{\mathbf{v}}_1(\mathbf{q}_2) \times \tilde{\mathbf{v}}_2(\mathbf{q}_2))^\top \mathbf{X}_2 \geq 0, (\tilde{\mathbf{v}}_2(\mathbf{q}_2) \times \tilde{\mathbf{v}}_3(\mathbf{q}_2))^\top \mathbf{X}_2 > 0, \\ (\tilde{\mathbf{v}}_3(\mathbf{q}_2) \times \tilde{\mathbf{v}}_4(\mathbf{q}_2))^\top \mathbf{X}_2 > 0, (\tilde{\mathbf{v}}_4(\mathbf{q}_2) \times \tilde{\mathbf{v}}_1(\mathbf{q}_2))^\top \mathbf{X}_2 \geq 0 \}, (5.2)$$

where $\tilde{\mathbf{v}}_i(\mathbf{q}_1) = (\mathbf{v}_i(\mathbf{q}_1)^\top, f_1)^\top$, $\tilde{\mathbf{v}}_i(\mathbf{q}_2) = (\mathbf{v}_i(\mathbf{q}_2)^\top, f_2)^\top$ ($i = 1, 2, 3, 4$), and R and \mathbf{t} are a rotation matrix and a translation vector relating the coordinates in C_1 and C_2 . We remark that \mathbf{X}_1 in (5.2) is the coordinates of a 3D point represented in C_1 while \mathbf{X}_2 is the coordinates of the same point represented in C_2 . This is because the coordinate transformation between C_1 and C_2 is required to obtain the intersection of $\mathbf{S}_1(\mathbf{q}_1)$ and $\mathbf{S}_2(\mathbf{q}_2)$ as the sets of points in space represented in the same coordinate system C_1 .

For a pair of corresponding pixels, $\mathbf{q}_1 \in \mathbf{K}_1$ and $\mathbf{q}_2 \in \mathbf{K}_2$, the triangulation reconstructs the intersection of their pyramidal rays of sight:

$$\mathbf{F} = \mathbf{S}_1(\mathbf{q}_1) \cap \mathbf{S}_2(\mathbf{q}_2). \tag{5.3}$$

(5.3) indicates that in our formulation, reconstructed is the polyhedron that consists of all possible points in space though which rays of sight of any points in the two corresponding pixels pass.

Proposition 3. From any pair of corresponding pixels, a polyhedron bounded by at most eight planes is reconstructed. The polyhedron is a set of points in space satisfying the eight inequalities (5.1) and (5.2).

5.2 Experiments

We assume that the following parameters of two cameras are known: the rotation matrix R and the translation vector \mathbf{t} relating the coordinates in C_1 and C_2 , the focal lengths f_1 and f_2 of the two cameras, and the resolutions $1/r_{1x}$ and $1/r_{1y}$ of the x - and y -coordinates on \mathbf{I}_1 , and $1/r_{2x}$ and $1/r_{2y}$ of the x - and y -coordinates on \mathbf{I}_2 . For a given point $\mathbf{P} \in \mathbf{V}_1 \cap \mathbf{V}_2$, we have the unique pair of corresponding pixels $\mathbf{q}_1 \in \mathbf{K}_1$ and $\mathbf{q}_2 \in \mathbf{K}_2$, as given in (3.4). We here reconstruct a polyhedron $\mathbf{F}(\mathbf{P})$ from (5.3) using our discrete epipolar geometry.

To the 3D point $\mathbf{P}^* = (0, 0, 100)^\top$ in C_1 , we reconstructed $\mathbf{F}(\mathbf{P}^*)$ under three different parameters given below.

Parallel camera position:

$$R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \mathbf{t} = \begin{bmatrix} 0 \\ 0 \\ 100 \end{bmatrix}, f_1 = f_2 = 50, \frac{1}{r_{1x}} = \frac{1}{r_{1y}} = \frac{1}{r_{2x}} = \frac{1}{r_{2y}} = 1.$$

Orthogonal camera position:

$$R = \begin{bmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \mathbf{t} = \begin{bmatrix} 100 \\ 0 \\ 100 \end{bmatrix}, f_1 = f_2 = 50, \frac{1}{r_{1x}} = \frac{1}{r_{1y}} = \frac{1}{r_{2x}} = \frac{1}{r_{2y}} = 1.$$

Orthogonal camera position with low image resolution:

the parameters are the same as the above orthogonal camera position except for the image resolutions: $\frac{1}{r_{1x}} = \frac{1}{r_{1y}} = \frac{1}{r_{2x}} = \frac{1}{r_{2y}} = 0.5$.

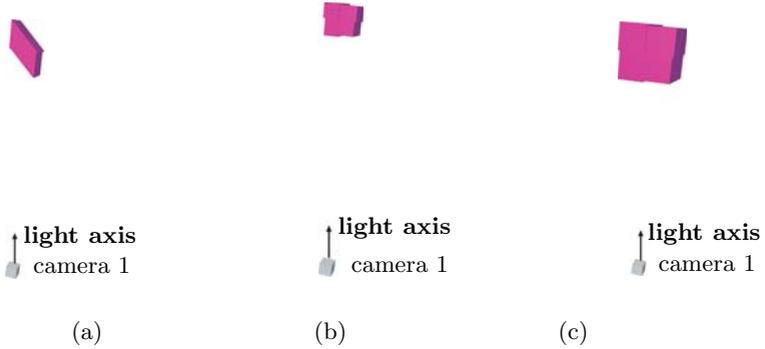


Fig. 5. Reconstructed polyhedra $\mathbf{F}(\mathbf{P}^*)$ for the point $\mathbf{P}^* = (0, 0, 100)^\top$ under the parallel (a) and the orthogonal (b,c) camera positions. The image resolution of (c) is lower than that of (b)

Figure 5 shows the reconstructed polyhedra $\mathbf{F}(\mathbf{P}^*)$ under each camera position. Note that we digitized the 3D space by voxels with sufficiently higher resolution⁵ than the image resolutions and that $\mathbf{F}(\mathbf{P}^*)$ is represented by a set of voxels whose centroids satisfy eight inequalities of (5.1) and (5.2). Voxels observed in Fig. 5 are, therefore, due to the 3D-space digitization.

We see that the shape and size of $\mathbf{F}(\mathbf{P}^*)$ depend on the camera parameters.

Reconstructed $\mathbf{F}(\mathbf{P}^*)$'s in (a) and (b) in Fig. 5 are almost the same in size while they are quite different in shape. When two cameras are with the parallel position where two principal axes are parallel with each other (Fig. 5 (a)), the shape of $\mathbf{F}(\mathbf{P}^*)$ expands to the direction of the principal axes. That is, larger ambiguity in reconstruction exists in the direction of the principal axes than that in the baseline direction. When two cameras are with the orthogonal position where two principal axes are orthogonal with each other, on the other hand, the shape of $\mathbf{F}(\mathbf{P}^*)$ does not expand to the direction of the principal axes. This implies that relative position and orientation between two cameras plays an important role in ambiguity in reconstruction.

When we compare $\mathbf{F}(\mathbf{P}^*)$'s in (b) and (c) in Fig. 5, we see that even with the same position and orientation in setting up two cameras, image resolution affects preciseness in reconstruction. The higher the image resolution becomes, the better preciseness in reconstruction becomes. We note that a smaller size of $\mathbf{F}(\mathbf{P}^*)$ indicates that \mathbf{P}^* is more precisely reconstructed. This is because $\mathbf{F}(\mathbf{P}^*)$ provides us with the possible locations for \mathbf{P}^* reconstructed from \mathbf{q}_1 and \mathbf{q}_2 .

We finally remark that similar discussion on error analysis in 3D reconstruction can be found, for example, in [2, 11], where statistical analysis based on the conventional epipolar geometry is used. Employed statistical models, however, can be applied to only the case of the parallel camera position above. In contrast, the proposed discrete epipolar geometry enables us to investigate errors incurred in 3D reconstruction under general camera parameter setting.

⁵ In our experiments, 3D space resolution is 10 times as high as image resolutions.

6 Conclusions

We focused, in this paper, on pixels as the smallest unit of images rather than points and established the discrete epipolar geometry, i.e., a discrete version of the conventional epipolar geometry. To be more specific, we introduced a pyramidal ray of sight and a discrete epipolar line as counterparts of a ray of sight and an epipolar line, and gave their mathematical descriptions. We also presented some experiments of 3D reconstruction using the discrete epipolar geometry.

The conventional epipolar geometry, which is widely used so far, is formulated in continuous image planes, while the discrete epipolar geometry is in discrete image planes. The smallest unit of digital images is not a point but a pixel. This motivated us to restructure geometric relationships on points between multiple views so that they hold in terms of pixels. The discrete epipolar geometry introduced in this paper is a first step to such a direction and opens a new framework that is capable to discriminate digitization errors from observation errors in handling digital images.

Acknowledgements. This work is in part supported by Grant-in-Aid for Scientific Research of the Ministry of Education, Culture, Sports, Science and Technology of Japan under the contract of 14380161 and 16650040.

References

1. E. Andres: Discrete Linear Objects in Dimension n :the Standard Model, *Graphical Models*, Vol. 65, pp. 92–111, 2003.
2. S. D. Blostein and T. S. Huang: Error Analysis in Stereo Determination of 3-D Point Positions, *IEEE Trans. on PAMI*, Vol. 9, No. 6, pp. 752-765, 1987.
3. O. Faugeras: *Three-Dimensional Computer Vision: A Geometric Viewpoint*, MIT Press, 1993.
4. D. Forsyth and J. Ponce: *Computer Vision: A Modern Approach*, Prentice Hall, 2002.
5. R. Furukawa and H. Kawasaki: Error Analysis of 3D Measurement System Based on Sensitivity of Measuring Conditions, *Proc. of ACCV*, Vol. 2, pp. 664-669, 2004.
6. R. Hartley and P. Sturm: Triangulation, *CVIU*, Vol. 68, No. 2, pp. 146-157, 1997.
7. R. Hartley and A. Zisserman: *Multiple View Geometry in Computer Vision*, Cambridge Univ. Press, 2000.
8. K. Kanatani: *Statistical Optimization for Geometric Computation: Theory and Practice*, Elsevier, 1996.
9. P. Meer, D. Mintz, D. Y. Kim and A. Rosenfeld: Robust Regression Methods in Computer Vision, *Int. J. of Computer Vision*, Vol. 6, pp. 59–70, 1991.
10. U. Montanari: On Limit Properties in Digitization Schemes, *Journal of ACM*, Vol. 17, No. 2, pp. 348–360, 1970.
11. J. J. Rodriguez and J. K. Aggarwal: Stochastic Analysis of Stereo Quantization Error, *IEEE Trans. on PAMI*, Vol. 12, No. 5, pp. 467-470, 1990.
12. C. V. Stewart: Robust Parameter Estimation in Computer Vision, *SIAM Reviews*, Vol. 41, No. 3, pp. 513-537, 1999.