コンピュータビジョン 47－7
（1987. 3. 17）

長さと角度の拘束条件

金谷 健一
群馬大学工学部情報工学科

空間中に存在する線分の長さや直線のなす角度が既知のとき、これらを写したカメラ画像からそれら線分や直線の空間的位置や方向を復元する手法を示す。基本的な考え方は、カメラをレンズの中心のまわりに回転すると同等な画像面の変換を適用して、注目している部分を画像面の中央（"正準位置"）に移動することである。これにより、角度情報に関しては中心投影と平行投影との差がなくなるから、平行投影の場合のみを考えればよい。直交する3辺の場合と、2直角と1既知角度の3辺の場合について、これを実際の画像を用いた例を示す。

CONSTRAINTS ON LENGTH AND ANGLE

Ken-ichi Kanatani*

Department of Computer Science, Gunam University, Kiryu, Gunma 376, Japan

Given a perspective projection of line segments on the image plane, the constraints on their 3D positions and orientations are derived on the assumption that their true lengths or the true angles they make are known. The approach here is first to transform images of line segments to the center of the image plane as if the camera were rotated to aim at them. The 3D information extracted in this *canonical position* is then transformed back to the original configuration. Examples are given, by using real images, for 3D recovery of a rectangular corner and a corner with two right angles.

1. INTRODUCTION

Humans can easily estimate the 3D position and orientation of an object in a scene by vision alone. The most fundamental assumption tacitly made by humans seems to be the *constancy of size*: we know the true shape and size of many familiar objects such as a man, a car and a house, and, seeing these familiar objects, we can easily and fairly accurately reconstruct the 3D world around us from our 2D visual perception.

The same principle applies to computer vision. If the true shape and size of an object are known and its projection image is given, the geometry of projection gives rise to mathematical relations or *constraints* on the 3D position and orientation of the object. The 3D position and orientation can be uniquely determined if a sufficient number of constraints are available from various sources of information.

However, these constraints often have very complicated forms if the projection is perspective, even if the object is a very simple one such as a line segment and a planar face. This is due to the *geometrical inhomogeneity* of the image plane: the extent of perspective distortion is different from position to position. Under orthographic projection, the image plane is geometrically homogeneous and we can freely translate a projected image to an arbitrary position on the image plane. The process of 3D recovery is not affected except for the corresponding translation of the object in the scene. Under perspective projection, however, we cannot arbitrarily translate the projected image.

But must we always analyze a perspectively projected image in that position? Can we not move it, in some way, to another position on the image plane so that analysis becomes easy? These questions lead us to the following observations of human perception. When a human finds a familiar object in the field of view, he rotates his eye or head so that the image of the object in question comes to the center of the field of view. Invoking the knowledge about the true shape and size of the object, and applying the assumption of constancy of size, he estimates the 3D position and orientation of the object. Then, recalling the angle of eye or head rotation, he interprets the 3D information in reference to his body.

This human reaction can be simulated by camera image analysis in the following way. Suppose the camera is rotated around an arbitrary axis by an arbitrary angle with the center of its lens fixed. As a result, a different image is seen on the image plane. However, since a point on the image plane actually corresponds to a *ray* starting from the lens center, occlusion is not affected. If the angle of camera rotation is known, the original image can be recovered as long as the effect of the image boundary is not involved. Thus, the image transformation due to camera rotation *does not require any knowledge about the 3D scene* and

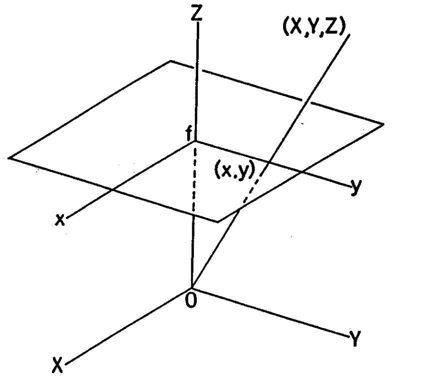

Fig. 1 Perspective projection as a camera model.

hence can be *computed*; the camera need not be actually rotated.

The above consideration implies the following fact: An object image can be *moved* into an arbitrary position on the image plane by applying the transformation corresponding to camera rotation. The geometrical properties of this transformation, especially its *invariants*, were extensively studied by Kanatani <4,6,7> from the viewpoint of group representation theory, especially irreducible representations of the 3D rotation group $SO(3)$.

Consequently, we can move an observed image in to a *canonical position* where analysis becomes easy. The 3D constraints obtained there are then transformed back into the original configuration. This technique was also applied to the analysis of shape from texture by Kanatani and Chou <8>. Evidently, the image origin is a prime candidate for the canonical position. We will show that for angle clues *we only need to consider orthographic projection* if the vertex is located at the image origin.

Even if the object image is moved into the canonical position, the 3D interpretation may not be unique. In such a case, humans invoke an appropriate *hypothesis* and solve the problem uniquely. The underlying mechanism of human hypothesizing is under study by many researchers, and although a definite conclusion has not yet been reached, it is observed that humans assume the "simplest" configuration in some sense. This process is also simulated for geometrical reasoning of computer vision.

In this paper, we study the constraints on the 3D positions and orientations of line segments, assuming that their true lengths and the angles they make are known. The use of "simplifying hypothesis" to restrict the ambiguity are also discussed.

The constraints involving angles have been studied by many researchers. The solution is not unique in general, and a frequently assumed "simplifying hypothesis" is what is called the *rectangularity hypothesis*. Many man-made objects such as buildings, machine parts and furniture have rectangular corners. Besides, the assumption of rectangularity is regarded as very natural from the viewpoint of human perception (cf. Barnard <1>).

Kanade <3> analyzed rectangular corner images with regard to interpretation of polyhedron drawing under orthographic projection. However, since he chose the *gradient* components p, q of the face defined by two edges (probably motivated by the *gradient space* of Huffman <1>, the resulting equations were very complicated, and the solution was obtained only by a numerical or graphical scheme. Later, Kanatani <5> chose the orientation angles of edges as unknowns and derived explicit analytical formulae.

Attempts to handle perspective projection was made by Barnard <1>. His approach is very straightforward, but the solution can be obtained only by numerical iterations even for a rectangular corner. Shakunaga and Kaneko <9> also analyzed angle clues under perspective projection, following the formulation of Huffman <2> and Kanade <3>. Although these approaches can treat a wider class of problems, e.g., lines that do not necessarily meet in the scene, the formulations are very much complicated.

In this paper, we will show that the solution for a rectangular corner can be obtained in very simple analytical terms if we use the image transformation corresponding to camera rotation. Examples are shown by using real images.

2. CAMERA ROTATION TRANSFORMATION

The camera image can be thought of as the projection onto an image plane located at distance f from the viewpoint O; a point P in the scene is projected onto the intersection of the image plane with the *ray* connecting the point P and the viewpoint O. The viewpoint O corresponds to the center of the camera lens, and the distance f equals the focal length of the camera lens if the object is very far away, so that for simplicity we call f the *focal length* although correction is necessary if the object is near the camera.

Let us choose an XYZ-coordinate system such that the viewpoint O is at the origin and the Z-axis coincides with the camera optical axis. Let $Z = f$ be the image plane, and take an xy-coordinate system so that the x- and y-axes are parallel to the X- and Y-axes (Fig. 1). A point (X, Y, Z) in the scene is projected onto point (x, y) on the image plane whose image coordinates x, y are given by

$$x=fX/Z, \quad y=fY/Z. \tag{2.1}$$

Consider a camera rotation around the viewpoint O (i.e., the center of the camera lens) and the induced transformation of the image. Suppose the camera is rotated by rotation matrix R (orthogonal matrix with determinant 1). As a result, the point seen at (x, y) now moves to point (x', y') given by the following theorem.

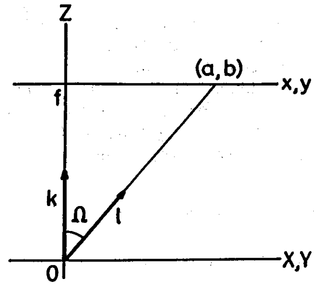

Fig. 2 Point (a, b) on the image defines unit vector l which starts from the viewpoint O and points toward it.

Theorem 1. *The image transformation induced by camera rotation* $R = (r_{ij})$ *is given by*

$$x'=f\frac{r_{11}x+r_{21}y+r_{31}f}{r_{13}x+r_{23}y+r_{33}f}, \quad y'=\frac{r_{12}x+r_{22}y+r_{32}f}{r_{13}x+r_{23}y+r_{33}f}. \tag{2.2}$$

Proof. A rotation of the camera by R is equivalent to the rotation of the scene in the opposite sense. If the scene is rotated by R^{-1} $(= R^T)$, where ^{T} denotes transpose, point (X, Y, Z) moves to point (X', Y', Z') given by

$$\begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix} = \begin{bmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}. \tag{2.3}$$

This point is projected to (x', y') on the image plane, where $x' = fX'/Z'$ and $y' = fY'/Z'$. Combining this with eqs. (2.1), we obtain eqs. (2.2).

It should be emphasized that the image transformation due to camera rotation *does not require any knowledge about the scene*. The transformation has an inverse, which is obtained by interchanging R and R^T. The transformations of the form of eq. (2.2) form a subgroup of the 2D *projective transformation group*. In the following, we assume that the image plane is sufficiently large compared with the projected image of the object we are viewing.[1]

3. STANDARD ROTATION AND TRANSFORMATION

---

1 Strictly speaking, as the camera rotates, a new part comes into view and some part goes out of view even if the image plane is infinitely large. In this paper, we do not consider this effect, assuming that the angle of camera rotation is not so large so that the object we are viewing is always in the field of view. For a mathematically rigorous treatment, see Kanatani <7>.

Consider a camera rotation which maps point (a, b) to the origin $(0, 0)$ on the image plane. The rotation is not unique; we can add a rotation of an arbitrary angle (the *swing*) around the Z-axis. The 3D unit vector starting from the viewpoint O pointing toward the point (a, b) on the image plane is given by

$$l=(\frac{a}{\sqrt{a^2+b^2+f^2}},\frac{b}{\sqrt{a^2+b^2+f^2}},\frac{f}{\sqrt{a^2+b^2+f^2}}). \tag{3.1}$$

(Fig. 2) This vector makes angle

$$\Omega=\tan^{-1}(\sqrt{a^2+b^2}/f) \tag{3.2}$$

with the unit vector $k = (0, 0, 1)$ along the Z-axis. The unit vector normal to both l and k is given by

$$n=\frac{k\times l}{\|k\times l\|}=(-\frac{b}{\sqrt{a^2+b^2}},\frac{a}{\sqrt{a^2+b^2}},0). \tag{3.3}$$

If the camera is rotated around vector $n=(n_1, n_2, n_3)$ by angle Ω screwwise, the point (a, b) is mapped to $(0, 0)$ on the image plane. The corresponding rotation matrix is given by

$$R(a,b)\equiv\begin{bmatrix} E & F & l_1 \\ F & G & l_2 \\ -l_1 & -l_2 & l_3 \end{bmatrix} \tag{3.4}$$

(cf. Kanatani <7>), where we put $l = (l_1, l_2, l_3)$ and

$$E\equiv\frac{a^2 l_3+b^2}{a^2+b^2},\quad F\equiv\frac{ab(l_3-1)}{a^2+b^2},\quad G\equiv\frac{b^2 l_3+a^2}{a^2+b^2}. \tag{3.5}$$

Hence, from Theorem 1, the transformation induced on the image plane is given by

$$x'=f\frac{Ex+Fy-l_1 f}{l_1x+l_2y+l_3f},\quad y'=f\frac{Fx+Gy-l_2f}{l_1x+l_2y+l_3f}. \tag{3.6}$$

We call the rotation $R(a,b)$ the *standard rotation* to map point (a, b) onto the image origin $(0, 0)$, and the transformation of eqs. (3.6), which we denote by $T_{(a,b)}$, the *standard transformation* with respect to point (a, b). Its inverse $T_{(a,b)}^{-1}$ is given by

$$x=-f\frac{Ex'+Fy'+l_1f}{l_1x'+l_2y'-l_3f},\quad y=-f\frac{Fx'+Gy'+l_2f}{l_1x'+l_2y'-l_3f}. \tag{3.7}$$

The standard rotation can be regarded as a rotation which does not contain rotations around the Z-axis (i.e., the *swing* is zero). This is similar to the rotations of the eye or the head: they rotate upward, downward, rightward and leftward, but not around the line of sight.

If we take the limit $f\to\infty$ of infinitely large focal length, i.e., in the limit of *orthographic projection*, we simply obtain $x' = x - a$, $y' = y - b$, namely the *translation* to move point (a, b) onto the image origin $(0, 0)$. Thus, the standard transformation $T_{(a,b)}$ of eqs. (3.6) is a natural extension of image translation under orthographic projection, and hence it can play the role of image translation under perspective projection.

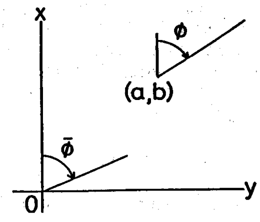

Fig. 3 A half line starting from point (a, b) having orientation φ is mapped by the standard transformation $T_{(a,b)}$ onto a half line starting from the image origin having orientation $\bar{\varphi}$.

4. TRANSFORMATION OF LINES

A line on the image plane is written in the form

$$Ax+By+C=0. \tag{4.1}$$

Here, the ratio of A, B, C alone has a geometrical meaning; A, B, C and cA, cB, cC for a non-zero scalar c define one and the same line.[2] In order to emphasize this fact, let us write $A{:}B{:}C$ to express a line.

If transformation (2.2) is applied, line (4.1) is mapped into another line

$$A'x+B'y+C'=0. \tag{4.2}$$

The line $A'{:}B'{:}C'$ is given by the following theorem.

Theorem 2. *A line $A{:}B{:}C$ on the image plane is transformed by camera rotation R into line*

$$r_{11}A+r_{21}B+r_{31}C/f \;:\; r_{12}A+r_{22}B+r_{32}C/f \;:\; f(r_{13}A+r_{23}B)+r_{33}C. \tag{4.3}$$

Proof. In view of eq. (2.1), eq. (4.1) is written as $A(fX/Z) + B(fY/Z) + C = 0$, or

$$\begin{bmatrix} A & B & C/f \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}=0. \tag{4.4}$$

From eq. (2.3), we find that A, B, C/f are transformed as a vector, i.e.,

$$\begin{bmatrix} A' \\ B' \\ C'/f \end{bmatrix}=\begin{bmatrix} r_{11} & r_{21} & r_{31} \\ r_{12} & r_{22} & r_{32} \\ r_{13} & r_{23} & r_{33} \end{bmatrix}\begin{bmatrix} A \\ B \\ C/f \end{bmatrix}, \tag{4.5}$$

from which eq. (4.3) is obtained.

---

2 This means that A, B, C are the *homogeneous coordinates* of the line of eq. (4.1). If we regard the xy-image plane with the *line of infinity* added as a 2D *projective space*, and use *homogeneous coordinates* to describe points on it, treatment of points becomes completely *dual* to treatment of lines. However, we do not use this *projective geometry* because we are interested in applications to real images; in practice the xy-*inhomogeneous coordinates* is most convenient.

A line passing through point (a, b) is written as

$$A(x-a)+B(y-b)=0, \tag{4.6}$$

or $A{:}B{:}-(Aa+Bb)$. If the camera rotation R is the standard rotation $R(a,b)$, the corresponding standard transformation $T_{(a,b)}$ on the image plane maps this line into a line of the form $A'x + B'y = 0$ or $A'{:}B'{:}0$. From eq. (4.3), we obtain

$$\frac{A'}{B'}=\frac{(fE+al_1)A+(fF+bl_1)B}{(fF+al_2)A+(fG+bl_2)B}. \tag{4.7}$$

Consider a half-line starting from a point on the image plane. Define its *orientation* to be the angle φ $(0 \leqq \varphi < 2\pi)$ made from the positive direction of the x-axis measured in the positive sense (i.e., toward the positive direction of the y-axis) (Fig. 3). From the relation

$$A/B=-\tan\varphi, \tag{4.8}$$

and eq. (4.7), we obtain the following result:

Theorem 3. *A half-line of orientation φ starting from point (a, b) is mapped by the standard transformation $T_{(a,b)}$ into a line starting from the image origin whose orientation $\bar{\varphi}$ is given as follows.*[3]:

$$\bar{\varphi}=-\tan^{-1}\frac{(fE+al_1)\tan\varphi-(fF+bl_1)}{(fF+al_2)\tan\varphi-(fG+bl_2)}. \tag{4.9}$$

Since $\tan^{-1}$ is a two-valued function, there are two values for $\bar{\varphi}$, and the one nearer to φ is chosen.[4]

Corollary. *A half-line of orientation $\bar{\varphi}$ starting from the image origin is mapped by the inverse standard transformation $T_{(a,b)}^{-1}$ into a line starting from point (a, b) whose orientation φ is given by*

$$\varphi=\tan^{-1}\frac{(fG+bl_2)\tan\bar{\varphi}+(fF+bl_1)}{(fF+al_2)\tan\bar{\varphi}+(fE+al_1)}. \tag{4.10}$$

Again, the one nearer to $\bar{\varphi}$ is chosen.[4]

5. CONSTRAINT ON LENGTH

Consider a line segment with endpoints (a_0, b_0), (a_1, b_1) on the image plane, and let P_0, P_1 be the corresponding endpoints in the scene. Assuming that the true 3D length of line segment P_0P_1 is known to be l, consider the resulting constraint. If the standard transformation $T_{(a_0,b_0)}$ is applied, point (a_0, b_0) is mapped onto the image origin. Let (a_1', b_1') be the point onto which point (a_1, b_1) is mapped by eq. (3.6). Let φ, $0 \leqq \varphi < 2\pi$, be the orientation of the line segment starting from the image origin (Fig. 4(a)).

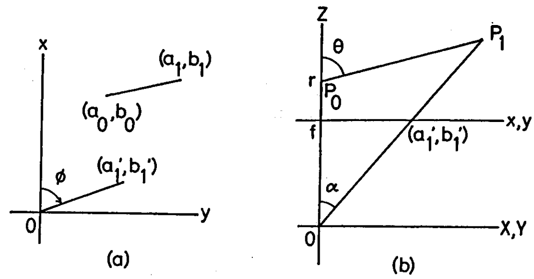

Fig. 4 The projection of line segment P_0P_1 and the mapping by the standard rotation $R(a_0b_0)$.

Let r be the distance of point P_0 from the viewpoint O, and let θ, $0 \leqq \theta < \pi/2$, be the angle of the line segment P_0P_1 measured from the line of sight. The standard transformation $T_{(a_0,b_0)}$ maps point P_0 onto point $(0, 0, r)$ and point P_1 onto

$$(l\sin\theta\cos\varphi,\ l\sin\theta\sin\varphi,\ r+l\cos\theta) \tag{5.1}$$

(Fig. 4(b)). Let α, $0 \leqq \alpha < \pi/2$, be the angle of OP_1 measured from the Z-axis. From Fig. 4(b), angle α is

$$\alpha=\tan^{-1}\sqrt{a_1'^2+b_1'^2}/f. \tag{5.2}$$

By the *law of sines* of trigonometry, the distance r is expressed in terms of θ by

$$r=l\sin(\theta-\alpha)/\sin\alpha, \tag{5.3}$$

and hence one degree of freedom is constrained about the positions of these two points; they are expressed in terms of one parameter θ.

Consider to constrain the remaining degree of freedom by invoking a simplifying hypothesis. A reasonable one may be that the line segment is perpendicular to the ray connecting the viewpoint and the point in question. In the canonical position, this means $\theta = \pi/2$. Under this hypothesis, a unique value for r is given from eq. (5.3) in the form

$$r=fl/\sqrt{a_1'^2+b_1'^2}. \tag{5.4}$$

6. CONSTRAINT ON ANGLE

Suppose we are viewing, on the image plane, two half lines starting from point (a, b), and let φ_1, φ_2, $0 \leqq \varphi_1, \varphi_2 < 2\pi$ be their orientations. Assume that the angle made by the corresponding half lines L_1, L_2 in the scene is known to be α. If the standard transformation $T_{(a,b)}$ is applied, the images of L_1, L_2 start from the image origin, having orientations $\bar{\varphi}_1$,

3 Although eq. (4.9) is sufficient for theoretical purposes, it is not desirable for actual numerical computation, since we have $\tan\varphi \to \infty$ when $\varphi \to \pi/2$. One way to avoid this is to use eq. (4.9) for $0 \leqq \varphi < \pi/4$, $3\pi/4 \leqq \varphi < 5\pi/4$, $7\pi/4 \leqq \varphi < 2\pi$, and to use otherwise

$$\bar{\varphi}=-\cot^{-1}\frac{(fF+al_2)-(fG+bl_2)\cot\varphi}{(fE+al_1)-(fF+bl_1)\cot\varphi},$$

which is equivalent to eq. (4.9). Similar consideration applies to eq. (4.10) as well.

4 Recall that we assume the rotation is not very large.

$\bar{\varphi}_2$ given by eq. (4.9) of Theorem 3 (Fig. 5(b)).

Let θ_1, θ_2, $0 \leq \theta_1, \theta_2 < \pi$, be the unknown angle of L_1, L_2 measured from the Z-axis. Then, the unit vector along them are given by

$$\bar{n}_i = (\sin\theta_i\cos\bar{\varphi}_i,\ \sin\theta_i\sin\bar{\varphi}_i,\ \cos\theta_i), \tag{6.1}$$

for i = 1, 2 (Fig. 5(b)). The condition that they make angle α is $\bar{n}_1 \cdot \bar{n}_2 = \cos\alpha$, or

$$\sin\theta_1\sin\theta_2\cos(\bar{\varphi}_1-\bar{\varphi}_2)+\cos\theta_1\cos\theta_2=\cos\alpha. \tag{6.2}$$

Hence, one degree of freedom is constrained. For example, θ_1 can be expressed in terms of θ_2 and vice versa. The orientations of L_1, L_2 in the original position are prescribed by unit vectors $n_1 = R(a,b)\bar{n}_1$, $n_2 = R(a,b)\bar{n}_2$, respectively.

If we want to constrain the remaining one degree of freedom by invoking a simplifying hypothesis, a natural one is $\theta_1 = \theta_2$. Under this hypothesis, angle θ_1 (= θ_2) is either θ_0 or $\pi - \theta_0$, where

$$\theta_0=\cos^{-1}\sqrt{\frac{\cos\alpha-\cos(\bar{\varphi}_1-\bar{\varphi}_2)}{1-\cos(\bar{\varphi}_1-\bar{\varphi}_2)}}. \tag{6.3}$$

The two solutions are *mirror images* of each other with respect to a mirror perpendicular to the line of sight.

An important fact about angle constraints is that in the canonical position *distinction between perspective and orthographic disappears*; the interpretation of the 3D line orientation does not involve *depth* or the distance from the viewer at all. However, this fact does not seem to have been widely recognized and utilized in image understanding.

7. INTERPRETATION OF A RECTANGULAR CORNER

Consider a rectangular corner having three mutually perpendicular edges. Many familiar objects, especially manufactured objects such as buildings, furniture and machine parts, have rectangular corners. Hence, the study of the rectangularity constraint is of practical importance. In addition, it is often argued that humans invoke this *rectangularity hypothesis* when no prior knowledge about the true angle is obtained (cf. Barnard <1>).

To this problem, Kanatani <5> gave an analytical solution under orthographic projection. Since that perspective projection reduces to orthographic projection in the canonical position as far as orientation is concerned, Kanatani's solution can be directly applied to perspectively projected images as well.

Consider three edges starting from the image origin, having orientations $\bar{\varphi}_i$, i = 1, 2, 3. Let θ_i, i = 1, 2, 3, be the angles of the corresponding edges in the scene measured from the Z-axis. From equations of the form of eq. (6.2) with α = 0, we obtain the condition of rectangularity in the form

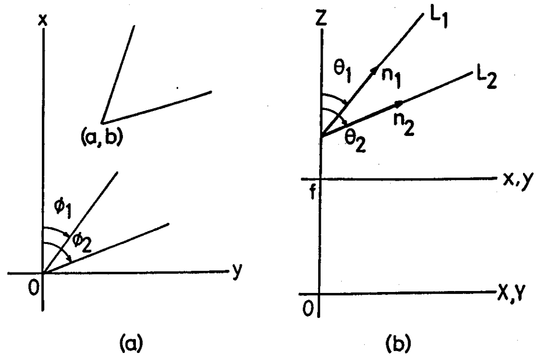

Fig. 5 The projection of two half lines L_1, L_2 and the mapping by the standard transformation $T_{(a,b)}$.

$$\tan\theta_i\tan\theta_j=-1/\cos(\bar{\varphi}_i-\bar{\varphi}_j), \tag{7.1}$$

where (i,j) = (1,2), (2,3), (3.1). If all three edges are assumed to go away from the viewer, i.e., $0 \leq \theta_1, \theta_2, \theta_3 < \pi/2$, we obtain, multiplication of both sides of the three equations (7.1) yields

$$\tan\theta_1\tan\theta_2\tan\theta_3$$

$$=\sqrt{-1/\cos(\bar{\varphi}_1-\bar{\varphi}_2)\cos(\bar{\varphi}_2-\bar{\varphi}_3)\cos(\bar{\varphi}_3-\bar{\varphi}_1)}. \tag{7.2}$$

From eqs. (7.1) and (7.2), we obtain

$$\theta_1=\tan^{-1}\sqrt{\frac{-\cos(\bar{\varphi}_2-\bar{\varphi}_3)}{\cos(\bar{\varphi}_1-\bar{\varphi}_2)\cos(\bar{\varphi}_3-\bar{\varphi}_1)}},$$

$$\theta_2=\tan^{-1}\sqrt{\frac{-\cos(\bar{\varphi}_3-\bar{\varphi}_1)}{\cos(\bar{\varphi}_2-\bar{\varphi}_3)\cos(\bar{\varphi}_1-\bar{\varphi}_2)}}, \tag{7.3}$$

$$\theta_3=\tan^{-1}\sqrt{\frac{-\cos(\bar{\varphi}_1-\bar{\varphi}_2)}{\cos(\bar{\varphi}_3-\bar{\varphi}_1)\cos(\bar{\varphi}_2-\bar{\varphi}_3)}},$$

If edge i goes toward the viewer, i.e., $\pi/2 < \theta_i < \pi$, then θ_i computed above is replaced by $\pi - \theta_i$, i.e., by the *mirror image*.

In deciding which edges go away from or toward the viewer, we must distinguish two configurations. One is the *fork* (or 'Y'), where all pair of edges make angles larger than $\pi/2$ on the image plane (Fig. 6(a)). In this case, we can check that the three edges either all go away from the viewer or all come toward the viewer, and these two interpretations are the mirror images of each other. The other configuration is the *arrow*, where one pair of edges makes an angle larger than $\pi/2$ and the other pairs make angles less than $\pi/2$ (Fig. 6(b)). Then, it can be checked that either the side edges go toward the viewer and the central edge away from the viewer, or the side edges go away from the viewer and the central edge toward the viewer, and the two interpretations are the mirror images of each other. It can also be checked that these two configurations, i.e., the fork and the arrow, exhaust images of rec-

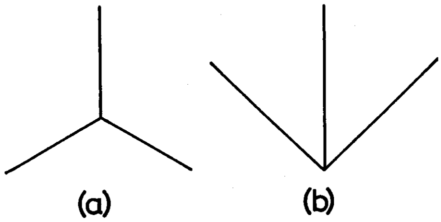

Fig. 6 (a) A *fork* and (b) an *arrow*.

tangular corner.[5]

Once the orientations $\bar{n}_i$, i = 1, 2, 3, of the three edges are determined in this canonical position, their orientations in the original position are given by $n_i = R(a,b)\bar{n}_i$, i = 1, 2, 3. Thus, we can conclude

Theorem 4. *Under perspective projection, the 3D orientation of a rectangular corner can be determined uniquely from its projection except for the mirror image with respect to a mirror perpendicular to the ray connecting the viewpoint and the vertex.*

8. EXAMPLES

Consider the building of Fig. 7. The focal length is f = 28mm. The image coordinates of the upper-right vertex are (10.0mm, 7.9mm), and the orientations of the three edges are φ_1 = 110°, φ_2 = 168°, φ_3 = 224°. If we apply the standard transformation given by eq. (4.9) of Theorem 3, we obtain $\bar{\varphi}_1$ = 11.5°, $\bar{\varphi}_2$ = 165.4°, $\bar{\varphi}_3$ = 224.6° (Fig. 8).

Suppose we know that the three edges are mutually perpendicular. The configuration is an arrow. Applying eqs. (7.3), we obtain θ_1 = 56.1°, θ_2 = 131.3°, θ_3 = 59.8°, if we assume that edge 2 goes away from the viewer and edges 1 and 3 comes toward the viewer. (Otherwise, we obtain the mirror image as well.)

From eq. (6.1), the corresponding unit vectors become

$$\bar{n}_1=\begin{bmatrix} -0.305 \\ 0.772 \\ 0.557 \end{bmatrix}, \quad \bar{n}_2=\begin{bmatrix} -0.727 \\ 0.190 \\ -0.660 \end{bmatrix}, \quad \bar{n}_3=\begin{bmatrix} -0.616 \\ -0.606 \\ 0.503 \end{bmatrix}.$$

From $n_i = R(10.0,7.9)\bar{n}_i$, i = 1, 2, 3, the orientations in the original position are given by

$$n_1=\begin{bmatrix} -0.141 \\ 0.902 \\ 0.408 \end{bmatrix}, \quad n_2=\begin{bmatrix} -0.909 \\ 0.045 \\ -0.414 \end{bmatrix}, \quad n_3=\begin{bmatrix} -0.392 \\ -0.429 \\ 0.814 \end{bmatrix}.$$

Fig. 9 shows the "top view" (orthographic projection onto the *YZ*-plane) and the "side view" (orthographic projection onto the *ZX*-plane).[6]

5 Here, we do no consider the degenerate case where two edges are projected onto the same line (i.e., 'L' or 'T'), assuming that the object is in *general position*.

6 The position of the vertex is taken arbitrarily.

Fig. 7 Image of a building. The upper-right corner has three mutually perpendicular edges.

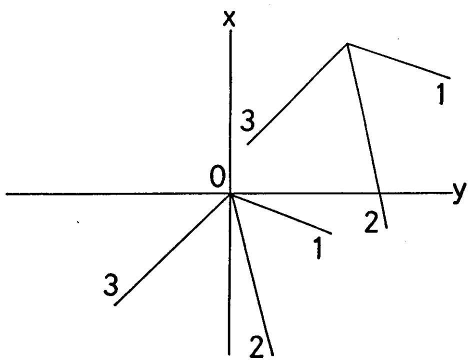

Fig. 8 The standard transformation applied to the three edges in Fig. 7.

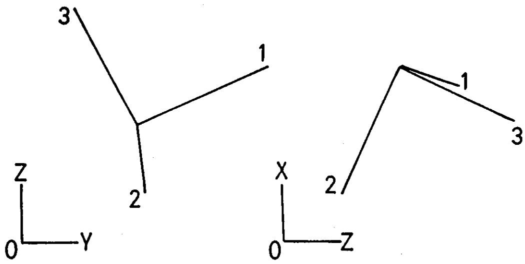

Fig. 9 The *top view* and the *side view* of Fig. 7.

Consider the object in Fig. 10. The focal length is f = 28mm. The image coordinates of the upper-right vertex are (9.0mm, 11.1mm), and the orientations of the three edges are φ_1 = 163°, φ_2 = 193°, φ_3 = 257°. If we apply the standard transformation given by eq. (4.9) of Theorem 3, we obtain $\bar{\varphi}_1$ = 160.8°, $\bar{\varphi}_2$ = 189.7°, $\bar{\varphi}_3$ = 259.7° (Fig. 11).

Suppose we know that edges 1 and 2 make angle 60°, edges 2 and 3 make angle 90°, and edges 3 and 1 make angle 90°. Then, we obtain three equations of the form of eq. (6.2). If

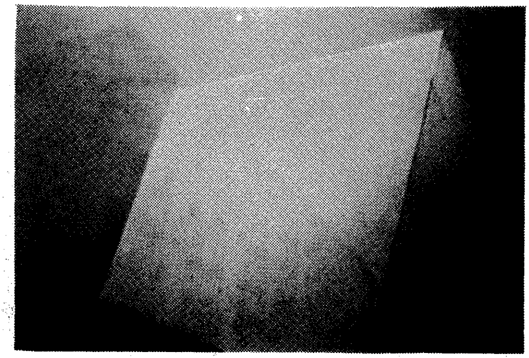

Fig. 10 Object image. The three edges of the upper-right corner make angles of 60°, 90°, and 90°.

two angles are 90°, analytical solutions are obtained.[7] We obtain $\theta_1 = 72.1°$, $\theta_2 = 125.5°$, $\theta_3 = 64.3°$. if we assume that edge 2 comes toward the viewer and edges 1 and 3 go away from the viewer. (Otherwise, we obtain the mirror image as well.) This is the only existing solution (except for mirror image).

From eq. (6.1), the corresponding unit vectors become

$$\bar{n}_1=\begin{bmatrix} -0.899 \\ 0.313 \\ 0.307 \end{bmatrix},\ \bar{n}_2=\begin{bmatrix} -0.803 \\ -0.138 \\ -0.580 \end{bmatrix},\ \bar{n}_3=\begin{bmatrix} -0.161 \\ -0.887 \\ 0.433 \end{bmatrix}.$$

From $n_i = R(9.0, 11.1)\bar{n}_i$, $i = 1, 2, 3$, the orientations in the original position are given by

$$n_1=\begin{bmatrix} -0.789 \\ 0.449 \\ 0.420 \end{bmatrix},\ n_2=\begin{bmatrix} -0.927 \\ -0.291 \\ -0.239 \end{bmatrix},\ n_3=\begin{bmatrix} 0.017 \\ -0.667 \\ 0.745 \end{bmatrix}.$$

Fig. 12 shows the "top view" (orthographic projection onto the YZ-plane) and the "side view" (orthographic projection onto the ZX-plane).[6]

REFERENCES

1. S. T. Barnard, Choosing a basis for perceptual space, *Comput. Vision Graphics Image Process.*, 29 (1985), 87 - 99.
2. D. A. Huffman, A duality concept for the analysis of polyhedral scenes, *Machine Intelligence* (eds. E. Elcock and D. Michie), Vol. 8, Ellis Horwood, Chichester, U.K., pp. 475 - 429.
3. T. Kanade, Recovery of the three-dimensional shape of an object from a single view, *Artif. Intell.*, 17 (1981), pp. 409 - 460.

7 $c_1 \equiv \cos(\bar{\varphi}_2-\bar{\varphi}_3)$, $c_2 \equiv \cos(\bar{\varphi}_3-\bar{\varphi}_1)$, $c_3 \equiv \cos(\bar{\varphi}_1-\bar{\varphi}_2)$,

$$A \equiv (c_2/c_1)^2(c_3^2-\cos^2\alpha),$$
$$B \equiv 2c_2c_3/c_1-(1+(c_2/c_1)^2)\cos^2\alpha,$$
$$X \equiv \sqrt{(-B \pm \sqrt{B^2-4A\sin^2\alpha})/2A},$$
$$\theta_1 = \tan^{-1}X,\ \pi-\tan^{-1}X,$$
$$\theta_2 = \tan^{-1}((c_2/c_1)\tan\theta_1),$$
$$\theta_3 = \tan^{-1}(-1/c_2\tan\theta_1).$$

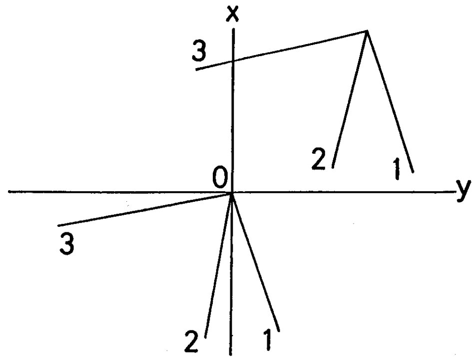

Fig. 11 The standard transformation applied to the three edges in Fig. 10.

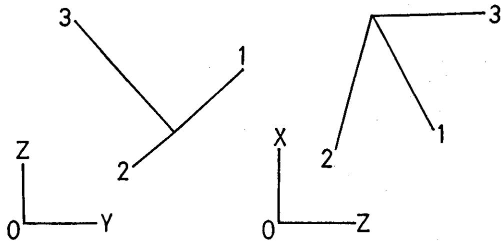

Fig. 12 The *top view* and the *side view* of Fig. 10.

4. K. Kanatani, Transformation of optical flow by camera rotation, *Proc. IEEE Workshop on Motion: Representation and Analysis*, Charleston, SC, May 1986, pp. 113 - 118.
5. K. Kanatani, The constraints on images of rectangular polyhedra, *IEEE Trans. Pattern Anal. Machine Intell.*, PAMI-8(4) (1986), pp. 456 - 463.
6. K. Kanatani, Camera rotation invariance of image characteristics, *Comput. Vision Graphics Image Process.* (to appear).
7. K. Kanatani, *Group Theoretical Methods on Image Understanding*, Springer, Berlin (to appear).
8. K. Kanatani and T.-C. Chou, Shape from texture: General principle, *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, Miami Beach, FL, June 1986, pp. 578 - 583.
9. T. Shakunaga and H. Kaneko, Perspective angle transform and its application to 3-D configuration recovery, *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, Miami Beach, FL, June 1986, pp. 594 - 601.

Acknowledgments. This paper is partly based on the work the author did at the University of Maryland in 1985 - 1986. He thanks Azriel Rosenfeld and Larry Davis for the support and discussions. He also thanks Kinue Akiyama of Gunma University for doing experiments. He also received comments from Kokichi Sugihara and Shun-ich Amari of the University of Tokyo and Takeshi Shakunaga of NTT.