Systems and Computers in Japan, Vol. 19, No. 3, 1988
Translated from Denshi Tsushin Gakkai Ronbunshi, Vol. 70-D, No. 5, May 1987, pp. 937-945

Coordinate Rotation Invariance of Image Characteristics for 3D Shape and Motion Recovery

Ken-ichi Kanatani, Member

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

SUMMARY

In order to determine the shape and the motion of the object from its projection, it suffices to derive and solve the equation relating the parameters specifying the shape and the motion of the object with the features measured on the image. The equation is usually nonlinear, and is difficult to solve in the analytic way. This paper notes that there is no inherent coordinate system on the image, and any two coordinate systems are equivalent. By combining the features of the image, coordinate-rotation-invariant parameters (invariances) are derived. By this scheme, the geometrical meaning of the image characteristics is made clear, and the analytic solution is obtained in a natural way. The idea is demonstrated for the cases of the optical flow analysis and the surface recovery by texture, as examples.

1. Introduction

The goal of the area called image understanding, image recognition and computer vision is to recover the 3D shape and motion of the object from the 2D projection image. The problem can be formulated in general as follows.*

A model is specified for the object, assuming that its shape and motion are specified by a small number of parameters $\alpha_1, \cdots, \alpha_n$. If the object is a plane, for example, the position is specified by the coordinate of a point on the plane and the direction of the normal. The motion can be represented by the translation speed of the reference point and the rotational velocity around the reference. The parameters, which specify the shape and the motion of the object in this way, are called the object parameters.

Let, on the other hand, the data measured on the image be $c_1, \cdots, c_m$, which are called the characteristics of the image. The characteristics may directly be obtained from the gray-level of the image such as the reflectivity of the object surface, or can be obtained as features after image processing, such as boundary, texture and optical flow. Depending on what characteristics are considered, it is often written as "shape from ...," where ... stands for shading, texture, motion, etc.

Applying the model to the object with assumed parameter values for the object and specifying the camera model for the projection, the image characteristics can theoretically be determined. In other words, the characteristics $c_1, \cdots, c_m$ are determined as functions of the object parameters $\alpha_1, \cdots, \alpha_n$:

$$c_i = F_i(\alpha_1, \cdots, \alpha_n), \quad i=1, \cdots, m \tag{1}$$

Consequently, applying the measured characteristics $c_1, \cdots, c_m$, and solving Eq. (1) as a system of equations for the unknowns $\alpha_1, \cdots, \alpha_n$, the object parameters are determined. In this sense, Eq. (1) is called recovery equation.

In most cases, the recovery equation is a system of nonlinear equations, and is difficult to solve in an analytic way. In the following, a powerful, although not always valid, method is proposed to determine the analytic solution for the recovery equation. Its principle is as follows. Although there does not exist a general method for

*This is not the only possible formulation. See Kanatani [4]. In that literature, the formulation in this paper is called "2D formulation."

ISSN0882-1666/88/0003-0087$7.50/0

solving the nonlinear equation, Eq. (1) has a structure which is a reflection of the geometrical property of the problem.

In the following, a property is noted that there does not exist a particular coordinate system which is inherent to the image, and any two coordinate systems are equivalent. From such a viewpoint, the characteristics of the image are combined to compose a property invariant to the rotation of the coordinate. By representing the recovery equation using such invariances, the analytic solution may be obtained in a natural way.

It should also be noted that the invariance corresponds to a certain meaningful geometrical property (Weyl's thesis, which is discussed later). This kind of representation makes it easy to interpret the image characteristics and the geometrical meanings of the recovery equation. This situation is demonstrated using the cases of optical flow analysis and surface recovery by texture, as examples.

2. Rotation of Coordinate and Irreducible Representation

Assume that characteristics $c_1, \cdots, c_m$ are measured in regard to the coordinate on the image plane. It is assumed that the origin of coordinate system is at the same position as the center of the image, i.e., the optical axis of the camera. In this situation, x and y axis do not have a particular meaning, and any direction can be chosen as the axis.

Let the xy-coordinate system be rotated around the origin by angle θ (the anti-clockwise direction is taken as positive) to produce $x'y'$ coordinate system. Let the same characteristics then be measured as $c_1', \cdots, c_m'$. Consider the case where the new characteristics are represented as linear combinations of the old characteristics $c_1, \cdots, c_m$. In other words, let

$$\begin{bmatrix} c_1' \\ \vdots \\ c_m' \end{bmatrix} = \begin{bmatrix} & T(\theta) & \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} \qquad (2)$$

or abbreviated as $c' = T(\theta)c$.

Obviously, the coefficient matrix $T(\theta)$ is a representation of the 2D rotation group SO(2), i.e., is a homeomorphism from SO(2) to the group of the matrix multiplication. In fact, consider $x''y''$ coordinate system, which is obtained by further rotating $x'y'$ coordinate system by angle θ'. Then $c'' = T(\theta')c'$. Consequently, $c'' = T(\theta')T(\theta)c$. On the other hand, $x''y''$ coordinate system is nothing but the one obtained by rotating xy coordinate system by the angle $\theta' + \theta$. Consequently, $c'' = T(\theta' + \theta)c$ and $T(\theta')T(\theta) = T(\theta' + \theta)$.

The individual characteristics of the image are obtained, depending on the measurement method, and they do not always have an essential meaning. Instead of the characteristics c_1, c_2, for example, one may use $C_1 = c_1 + c_2$ and $C_2 = c_1 - c_2$. The characteristics $\{c_1, c_2\}$ are equivalent to the characteristics $\{C_1, C_2\}$, and describe the same property.

Based on such an idea, assume that a new characteristics $c_1, \cdots, c_m$ are formed by linear combinations of characteristics $C_1, \cdots, C_m$.* Let the coordinate system be rotated by angle θ. Then, assume that

$$\begin{bmatrix} C_1' \\ \vdots \\ C_l' \\ C_{l+1}' \\ \vdots \\ C_m' \end{bmatrix} = \left[\begin{array}{c|c} * & 0 \\ \hline 0 & * \end{array}\right] \begin{bmatrix} C_1 \\ \vdots \\ C_l \\ C_{l+1} \\ \vdots \\ C_m \end{bmatrix} \qquad (3)$$

The above relation indicates that the characteristics $\{C_1, \cdots, C_l\}$ and the characteristics $\{C_{l+1}, \cdots, C_m\}$ are transformed independently.

This can be interpreted that the characteristics $\{C_1, \cdots, C_l\}$ and the characteristics $\{C_{l+1}, \cdots, C_m\}$ describe independent properties of the image. In mathematical terminology, this is a reduction of the representation $T(\theta)$ of the 2D rotating group SO(2) into a direct sum of two representations. Consequently, the characteristics with representation which is reducible, i.e., reducible representation, are describing

*At the first step, the characteristics are assumed as real, but complex coefficients are permitted in the linear combination. Consequently, $c_1, \cdots, c_m$ may be complex. It is assumed that two characteristics are equivalent, i.e., the coefficient matrix of the transformation is regular.

simultaneously independent properties of the image.

A similar procedure may be applied to $\{C_1, \cdots, C_l\}$ and $\{C_{l+1}, \cdots, C_m\}$ to reduce the representations successively. Let the final form be

$$\begin{bmatrix} C_1' \\ \vdots \\ C_m' \end{bmatrix} = \begin{bmatrix} * & & \\ & \boxed{*} & \\ & & \ddots \end{bmatrix} \begin{bmatrix} C_1 \\ \vdots \\ C_m \end{bmatrix} \qquad (4)$$

where the small sections form a representation which cannot further be reduced, i.e., irreducible representation. A representation, which is reduced to the direct sum of irreducible representations, is called completely reducible.

The characteristics composing an irreducible representation cannot be separated, since they remain mixed by any transformation. Consequently, they can be interpreted as describing essentially a single property. In contrast, characteristics composing a reducible representation are considered as describing more than one property simultaneously. By this idea, the intuitive and ambiguous notion that a property can be separated into several component properties, is placed in correspondence to the mathematically well-established notion that a representation can be reduced. This is the reasoning presented by Weyl [14].

He presented an assertion that, in order for an observed variable to be called a physical property, it must correspond to an irreducible representation of a group of transformations which does not affect the essence of the phenomenon. Based on such an idea, he introduced the theory of groups into the quantum mechanics [15]. Such an idea is called in the following Weyl's thesis.

3. Invariance and Weight

As is well-known, any representation for the 2D rotation group SO(2) is completely reducible, and any irreducible representations is one-dimensional. Consequently, given characteristics $c_1, \cdots, c_m$ composing the representation of SO(2), new independent characteristics $C_1, \cdots, C_m$, $C_i' = T_i(\theta) C_i$, $i = 1, \cdots, m$ can be formed by a certain linear transformation. Since each $T_i(\theta)$ is a 1D representation, it must satisfy $T_i(\theta')T_i(\theta) = T_i(\theta' + \theta)$. It must also be a periodic function of period 2π, satisfying $T_i(0) = 1$ and $T_i(\theta + 2\pi) = T_i(\theta)$.

It is seen immediately from the above property that there is an integer n such that $T_i(\theta) = e^{-in\theta}$ (i is the imaginary unit). The integer n is called the weight of the characteristic.* The characteristic of weight 0, i.e., the identity characteristic, is called the absolute invariance. The characteristic with nonzero weight n is called a relative invariance of weight n. Both combined are simply called invariance.

Summarizing the above description, the following statement is made. For the characteristics $C_1, \cdots, C_m$ with the transformation of Eq. (2), a certain linear tranformation is applied to form new characteristics $C_1, \cdots, C_m$ as

$$\begin{bmatrix} C_1 \\ \vdots \\ C_m \end{bmatrix} = \begin{bmatrix} & & \\ & P & \\ & & \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_m \end{bmatrix} \qquad (5)$$

(P is a regular matrix), so that the transformations for the new characteristics $C_1, \cdots, C_m$ can be decomposed as follows.

$$\begin{bmatrix} C_1' \\ \vdots \\ C_m' \end{bmatrix} = \begin{bmatrix} P \end{bmatrix} \begin{bmatrix} T(\theta) \end{bmatrix} \begin{bmatrix} P^{-1} \end{bmatrix} \begin{bmatrix} C_1 \\ \vdots \\ C_m \end{bmatrix} = \begin{bmatrix} e^{-in_1\theta} & & \\ & \ddots & \\ & & e^{-in_m\theta} \end{bmatrix} \begin{bmatrix} C_1 \\ \vdots \\ C_m \end{bmatrix} \qquad (6)$$

In other words, the coefficient matrix $T(\theta)$ can simultaneously be diagonalized by a constant matrix P, independently of the value of θ. The reason for this is the 2D rotation group SO(2) is compact and is completely reducible as well as commutative. By Schur's lemma, a reducible representation is always one-dimensional [1, 8]. Since each invariance corresponds to an irreducible

*By writing $e^{-in\theta}$ instead of $e^{in\theta}$, n is sometimes called weight. When the rotation of the object is considered, this notation is more convenient. Since, however, the coordinate system is rotated in the following, the notation $e^{-in\theta}$ is more convenient.

representation, it represents a certain geometrical property.*

4. Scalar, Vector and Tensor

A characteristic, which is invariant by the rotation of the coordinate system:

$$c' = c \tag{7}$$

is called scalar. Equation (7) defines a trivial representation, i.e., the identity representation. A scalar is itself an absolute invariance.

When characteristics a and b are transformed by the rotation θ of the coordinate as

$$\begin{bmatrix} a' \\ b' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \tag{8}$$

a and b are called vector. Equation (8) defines a faithful representation of SO(2), i.e., vector representation. This representation is not irreducible. The reason for this is that when linear combinations $a + ib$ and $a - ib$ are formed, it follows that**

$$\begin{bmatrix} a' + ib' \\ a' - ib' \end{bmatrix} = \begin{bmatrix} e^{-i\theta} & \\ & e^{i\theta} \end{bmatrix} \begin{bmatrix} a + ib \\ a - ib \end{bmatrix} \tag{9}$$

Thus, $z = a + ib$ and $z^* = a - ib$ (the asterisk indicates the complex conjugate) are relative invariances of weights 1 and -1, respectively:

$$z' = e^{-i\theta} z, \quad z^{*\prime} = e^{i\theta} z^* \tag{10}$$

When characteristics A, B, C and D are transformed by the rotation of coordinate by angle θ as

$$\begin{bmatrix} A' & B' \\ C' & D' \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} \cdot \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \tag{11}$$

A, B, C and D are called (second-order) tensor. Equation (11) defines the following linear mapping from A, B, C and D to A', B', C' and D':

$$\begin{bmatrix} A' \\ B' \\ C' \\ D' \end{bmatrix} = \begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta & \cos\theta\sin\theta & \sin^2\theta \\ -\cos\theta\sin\theta & \cos^2\theta & -\sin^2\theta & \cos\theta\sin\theta \\ -\cos\theta\sin\theta & -\sin^2\theta & \cos^2\theta & \cos\theta\sin\theta \\ \sin^2\theta & -\cos\theta\sin\theta & \cos\theta\sin\theta & \cos^2\theta \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \end{bmatrix} \tag{12}$$

The above relation defines the (second-order) tensor representation of SO(2). Neither this representation is irreducible. As the first step, the matrix formed by A, B, C and D can uniquely be decomposed as follows for the symmetrical and antisymmetrical parts:

$$\begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} A & (B+C)/2 \\ (B+C)/2 & D \end{bmatrix} + \begin{bmatrix} 0 & -(C-B)/2 \\ (C-B)/2 & 0 \end{bmatrix} \tag{13}$$

This decomposition can be verified as invariant. In other words, the symmetrical and antisymmetrical parts of the result of transformation of the left-hand side by Eq. (10) are the same as those, respectively, of the result of separate transformations of the symmetrical and the antisymmetrical parts of the right-hand side. (Those properties apply to tensor of any order.) Since the antisymmetrical part of Eq. (13) contains only one independent element, $A - B$ must be an absolute invariance.

The symmetrical part can further be decomposed uniquely as follows into the scalar part (constant multiple of unit matrix) and the deviation part (symmetrical matrix with zero trace).

$$\begin{bmatrix} A & (B+C)/2 \\ (B+C)/2 & D \end{bmatrix} = \frac{A+D}{2} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} (A-D)/2 & (B+C)/2 \\ (B+C)/2 & -(A-D)/2 \end{bmatrix} \tag{14}$$

It is again verified that this decomposition is invariant. In other words, the scalar and deviation parts of the result of the transformation of the left-hand side are

---

*If there exist two or more invariances with the same weight, there is left a freedom for their linear combination. Then, the reduction is not unique (called degenerate). In this case, another group may be operated on to determine the reduction (perturbation in quantum mechanics).

**It is irreducible in the range of real numbers. Schur's lemma, however, does not apply unless complex numbers are considered. The 1D property of the irreducible representation for the commutative group is a direct consequence of Schur's lemma [1, 8].

the same as the results, respectively, of the separate transformations of the scalar and the deviation parts of the right-hand side. (This property applies to a tensor of any order.) Since the scalar part contains only one independent element, $A + D$ must be an absolute invariance.

The deviation part contains two independent elements. Forming $(A - D) + i(B + C)$ and $(A - D) - i(B + C)$ it is verified that they are relative invariances with weight 2 and -2, respectively.* Thus, Eq. (12) can be rewritten as follows:**

$$\begin{bmatrix} A'+B' \\ B'-C' \\ (A'-D')+i(B'+C') \\ (A'-D')-i(B'+C') \end{bmatrix} = \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & e^{-2i\theta} & \\ & & & e^{2i\theta} \end{bmatrix} \begin{bmatrix} A+B \\ B-C \\ (A-D)+i(B+C) \\ (A-D)-i(B+C) \end{bmatrix} \quad (15)$$

In other words, letting

$$\left.\begin{array}{l} T=A+D,\ R=B-C \\ S=(A-D)+i(B+C) \end{array}\right\} \quad (16)$$

T and R are absolute invariances, and S is a relative invariance of weight 1. They are transformed by the coordinate rotation of angle θ as

$$T'=T,\ R'=R,\ S'=e^{-2i\theta}S \quad (17)$$

They have particular geometrical meanings according to Weyl's thesis (see next section).

5. Analysis of Optical Flow

Let xy coordinate system be fixed in the space, and regard xy plane as the image plane. Let the point $(0, 0, -f)$ be the viewpoint, which is at the distance f along z

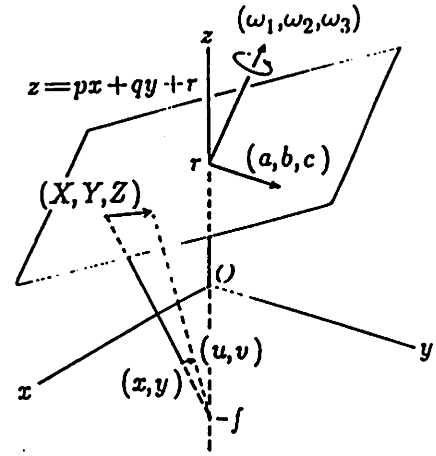

Fig. 1. A plane having equation $z = px + qy + r$ is moving with translation velocity (a, b, c) at $(0, 0, r)$ and rotation velocity $(\omega_1, \omega_2, \omega_3)$ around it. An optical flow is induced on the xy-plane by perspective projection, $(0, 0, -f)$ being the viewpoint.

axis in the negative direction. The object in the space is perspectively projected on xy plane (Fig. 1). In the actual camera, one can consider that a perspective projection is made, where the viewpoint is the center of the lens and f is the focal length.* A point (X, Y, Z) in the space is projected on the point of intersection between the straight-line connecting that point and the viewpoint with xy plane.

Letting the xy coordinate of the projected point be (x, y),

$$x=fX/(f+Z),\ y=fY/(f+Z) \quad (18)$$

The limit of making $f \to \infty$ is the parallel projection.

Consider the case where a plane $z = px + qy + r$ is performing a rigid motion in the space; p and q represent the gradient of the plane, and r represents the distance to the plane along z axis (absolute distance). The rigid motion can be specified by the speed (a, b, c) of the reference point (translation), and the rotation $(\omega_1, \omega_2, \omega_3)$ around the reference point (i.e., clockwise

---

*If the original characteristics are real, the relative invariances appear in a pair with positive and negative values.

**In general, the representation of order r for the n-dimensional rotation group $SO(n)$ can systematically be reduced to irreducible representations based only on the symmetry of (the index of) tensor and the trace (of the index) (Weyl's theorem) [1, 4].

*More strictly, a correction must be made using the lens formula $1/a + 1/b = 1/f$ (where a is the distance from the center of the lens to the object, b is the distance from the center to the film plane, and f is the focal length. If $a \gg f$, however, one can assume that $b \simeq f$.

rotation with angular velocity $\sqrt{\omega_1^2 + \omega_2^2 + \omega_3^2}$ (rad/s) around the axis $(\omega_1, \omega_2, \omega_3)$.

In the following, the intersection of z axis with the plane, $(0, 0, r)$ is used as the reference point. Then, the object parameter is 9, which are p, q, r, a, b, c, ω_1, ω_2 and ω_3.

For such a motion in the space, the motion of the image observed on the image plane $\dot{x} = u(x, y)$, $\dot{y} = v(x, y)$ is represented as follows:

$$\left.\begin{aligned} u(x, y) &= u_0 + Ax + By + (Ex + Fy)x \\ v(x, y) &= v_0 + Cx + Dy + (Ex + Fy)y \end{aligned}\right\} \quad (19)$$

where the coefficients u_0, v_0, A, B, C and D are given as follows:

$$\left.\begin{aligned} & u_0 = \frac{fa}{f+r}, \quad v_0 = \frac{fb}{f+r} \\ & A = p\omega_2 - \frac{pa+c}{f+r} \\ & B = q\omega_2 - \omega_3 - \frac{qa}{f+r} \\ & C = -p\omega_1 + \omega_3 - \frac{pb}{f+r} \\ & D = -q\omega_1 - \frac{qb+c}{f+r} \\ & E = \frac{1}{f}\left(\omega_2 + \frac{pc}{f+r}\right) \\ & F = \frac{1}{f}\left(-\omega_1 + \frac{qc}{f+r}\right) \end{aligned}\right\} \quad (20)$$

Assume that the coefficients u_0, v_0, A, B, C, D and F are estimated, by applying the least-square method to approximate the observed speed on the image by Eq. (19), for example.* In this case, the coefficients are the characteristics of the image, and Eq. (20) is the recovery equation to determine the object parameters p, q, r, a, b, c, ω_1, ω_2 and ω_3. The solution to this equation is given by Longuet-Higgins [10] and Subbarao and Waxman [12]. The following analysis is made by Kanatani [6].

*Kanatani [3] has shown a method, which calculates the coefficient only by measuring the image characteristics, without calculating the optical flow.

Applying the coordinate rotation to Eq. (19), it is seen that u_0, v_0 form a vector; A, B, C, D form a tensor; and E, F form a vector [8]. Consequently, the invariances are constructed as follows:

$$\left.\begin{aligned} & U_0 = u_0 + iv_0 \\ & T = A + D, \quad R = C - B \\ & S = (A - D) + i(B + C), \quad K = E + iF \end{aligned}\right\} \quad (21)$$

T, R are absolute invariances; U_0, K are relative invariances of weight 1; and S is a relative invariance of weight 2.

According to Weyl's thesis, they should represent particular geometrical meanings. As is seen by illustrating the corresponding flows, U_0 represents the translation, T the divergence, R the rotation, S the shear and K the fanning (Fig. 2). Similarly, by applying the coordinate rotation, the following properties are seen for the object parameters: p, q; a, b; ω_1, ω_2 form vectors; and r, c, ω_3 are scalars [8]. Consequently, the following relative invariances of weight 1 can be constructed:

$$P = p + iq, \quad V = a + ib, \quad W = \omega_1 + i\omega_2 \quad (22)$$

Using the above invariances, the recovery Eq. (20) can be written as follows:

$$\left.\begin{aligned} & U_0 = \frac{f}{f+r} V \\ & PW'^{*} = (2\omega_3 - R) - i(2c' + T) \\ & PW' = iS \\ & c'P - iW' = L \end{aligned}\right\} \quad (23)$$

where $c' = c/(f + r)$, $W' = W - iU_0/f$ and $L = fK - U_0/f$. By using the invariances, the equation thus takes a very simple form. Assume that it is verified that $c \neq 0$ [6]. The following result is then obtained. For the case of $c = 0$, see Kanatani [6].

Theorem 1. The third-order equation

$$X^3 + TX^2 + \frac{1}{4}(T^2 - |S|^2 - |L|^2)X + \frac{1}{8}(\mathrm{Re}[L^2 S] - T|L|^2) = 0 \quad (24)$$

has three real roots. Let the root with the middle value be c'. Then, the solution of the recovery Eq. (23) is given as follows:

$$V = \frac{f+r}{f} V_0, \quad c = (f+r)c'$$

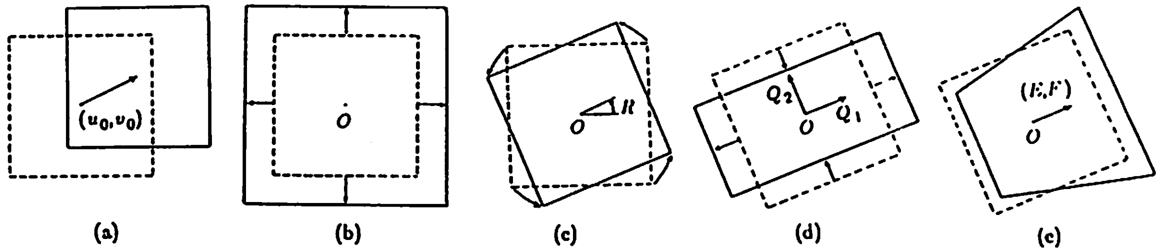

Fig. 2. (a) Translation by (u_0, v_0). (b) Divergence by T. (c) Rotation by R. (d) Shearing with Q_1 and Q_2 as the axes of maximum extension and maximum compression. (e) Fanning along (E, F).

$$\left.\begin{aligned} P &= \frac{1}{2c'}(L \pm \sqrt{L^2 - 4c'S}) \\ W &= \frac{i}{2}(L \mp \sqrt{L^2 - 4c'S}) + \frac{i}{f}U_0 \\ \omega_3 &= \frac{1}{2}R \pm \mathrm{Im}[L^* \sqrt{L^2 - 4c'S}] \end{aligned}\right\} \quad (25)$$

In above, Re and Im represent, respectively, the real and imaginary parts, and the asterisk represents the complex conjugate. Thus, the following properties are obtained.

(i) The absolute distance r is indefinite.

(ii) $a/(f + r)$, $b/(f + r)$ and $c/(f + r)$ are uniquely determined.

(iii) There can be two sets of solutions for the gradient p, q and the rotational velocity ω_1, ω_2 and ω_3.

By definition, L is a relative invariance of weight 1. The product of invariances is an invariance with the weight which is the sum of the weights of the original invariances. By forming the complex conjugate, the weight changes its sign. In the expression represented by invariances, such as Eqs. (23) to (25), the right- and left-hand sides have the same weight. The addition or subtraction can be performed only for the invariances with the same weight. This is very convenient in checking the validity. Or, by utilizing the property, the form of the solution may be predicted. This is a great merit in introducing the invariance.

In addition to the above, Kanatani [6] has presented applications and numerical examples, such as the interpretation of geometrical solution and the connection condition between surfaces.

Taking the limit of $f \to \infty$, the following recovery equation is obtained for translation:

$$\left.\begin{aligned} &u_0 = a, \quad v_0 = b \\ &A = p\omega_2, \quad B = q\omega_2 - \omega_3 \\ &C = -p\omega_1 + \omega_3, \quad D = -q\omega_1 \end{aligned}\right\} \quad (26)$$

The first two equations determine a and b. The rest is represented as follows using the invariances:

$$PW^* = 2\omega_3 - (R + iT), \quad PW = iS \qquad (27)$$

The solution is given as follows [5].

Theorem 2.

$$\left.\begin{aligned} \omega_3 &= \frac{1}{2}(R \pm \sqrt{SS^* - T^2}) \\ W &= k \exp i\left(\frac{\Pi}{4} + \frac{1}{2}\arg(S) - \frac{1}{2}\arg(2\omega_3 - (R + iT))\right) \\ P &= \frac{S}{k} \exp i\left(\frac{\Pi}{4} - \frac{1}{2}\arg(S) + \frac{1}{2}\arg(2\omega_3 - (R + iT))\right) \end{aligned}\right\} \quad (28)$$

where k is an undetermined shape factor. There are two sets of solutions according to the value of k.

For the above solution also, Kanatani [5] has given the connection condition for the surface and numerical examples. He also compared the solution with that of Sugihara and Sugie [13], which is obtained without using the invariance for essentially the same problem. In the latter method, the solution is not obtained in an analytic form, and it is impossible to eliminate completely the physically impossible solution.

Deleting only the term of $O(1/f^2)$ from the recovery equation, leaving the term of $O(1/f)$, the expressions for E and F are obtained as

$$E=\omega_2/f,\quad F=-\omega_1/f \tag{29}$$

Kanatani [6] called this pseudo-parallel projection. The solution then is given as follows.

Theorem 3.

$$\left.\begin{aligned} &V=\frac{f+r}{f}V_0\\ &P=S/L,\quad W=if/K\\ &\omega_3=\frac{1}{2}(R+\mathrm{Im}[Se^{-2i\alpha}])\\ &c=\frac{f+r}{2}(T-\mathrm{Re}[Se^{-2i\alpha}]) \end{aligned}\right\} \tag{30}$$

where $\alpha \equiv \arg(L)$. From above expression, the following properties are seen.

(i) The absolute distance r is indeterminate.

(ii) $a/(f+r)$, $b/(f+r)$ and $c/(f+r)$ are uniquely determined.

(iii) The gradients p and q and rotation velocity ω_1, ω_2 and ω_3 are uniquely determined. Consequently, there does not exist a pseudo-solution. The geometrical meaning of the above solution is discussed by Kanatani [6].

Up to this point, the motion of a plane has been considered. For the case of general curved surface, the optical flow is represented as follows, instead of Eq. (19) Subbarao [11]:

$$\left.\begin{aligned} u(x,\ y)&=u_0+Ax+By+Ex^2+2Fxy+Gy^2+\cdots\\ v(x,\ y)&=v_0+Cx+Dy+Kx^2+2Lxy+My^2+\cdots \end{aligned}\right\} \tag{31}$$

where $\cdots$ indicates higher terms of x and y. In this case also, the invariances are constructed as follows [8]:

$$\left.\begin{aligned} &U_0=u_0+iv_0,\quad T=A+D\\ &R=C-B,\quad S=(A-D)+i(B+C)\\ &H=(E+2L-G)+i(M+2F-K)\\ &I=(E-2L+3G)+i(M-2F+3K)\\ &J=(E-2L-G)-i(M-2F-K) \end{aligned}\right\} \tag{32}$$

T, R are absolute invariances; U_0, H and I are relative invariances of weight 1; and S is a relative invariance of weight 2, and J is a relative invariance of weight 3. In the case of the second-order surface, the recovery equation can be represented by invariances, and the solution can be determined in an analytic form Subbarao [11].

6. Recovery of Surface by Texture

Consider the case where a surface $z = z(x, y)$ with a texture is parallel-projected along z axis on xy plane (Fig. 3). It is assumed that the texture is uniformly distributed on the surface.* Assume that the density $\Gamma(x, y)$ of the texture on the projected image is measured.** It is seen from Fig. 3 that, if the true texture density is ρ, the texture density on each point on the image is given as follows:

$$\Gamma(x,\ y)=\rho\sqrt{\left(\frac{\partial z}{\partial x}\right)^2+\left(\frac{\partial z}{\partial y}\right)^2} \tag{33}$$

Consider the second-order surface

$$z=r+px+qy+ax^2+2bxy+cy^2 \tag{34}$$

The object parameters are then p, q, r, a, b, c and ρ. Substituting Eq. (34) into (33),

$$\Gamma(x,\ y)=A_0\sqrt{1+A_1x+A_2y+A_3x^2+2A_4xy+A_5y^2} \tag{35}$$

The coefficients are given as follows:

$$\left.\begin{aligned} &A_0=\rho\sqrt{1+p^2+q^2},\quad A_1=\frac{4(ap+bq)}{1+p^2+q^2}\\ &A_2=\frac{4(bp+cq)}{1+p^2+q^2},\quad A_3=\frac{4(a^2+b^2)}{1+p^2+q^2}\\ &A_4=\frac{4b(a+c)}{1+p^2+q^2},\quad A_5=\frac{4(b^2+c^2)}{1+p^2+q^2} \end{aligned}\right\} \tag{36}$$

Consider the estimation of the coefficients A_0 ~ A_5 by applying the expression of the form of Eq. (35), for example, to the observed texture density (Kanatani and Chou [9] used a more elaborate technique). Then, A_0 ~ A_5 are the image characteristics, and Eq. (35) is the recovery equation to determine the object parameters p, q, r, a, b, c and ρ.

---

*The uniformity should more strictly be defined, which, however, is omitted. For the details, see Kanatani and Chou [9].

**The density of the discrete texture must be defined, together with the measurement procedure, which, however, is omitted. Kanatani and Chou [9] defined those from the viewpoint of the distribution theory of and presented a measurement procedure based on that definition.

By applying the coordinate transformation to Eq. (34), it is seen that A_0 is a scalar; A_1, A_2 form a vector and A_3, A_4, A_4, A_5 form a tensor [8]. Consequently, the following invariances are constructed:

$$\left.\begin{aligned} &V=\frac{A_1+iA_2}{4},\quad T=\frac{A_3+A_5}{8} \\ &S=\frac{A_3-A_5}{8}+i\frac{A_4}{4} \end{aligned}\right\} \qquad (37)$$

A_0 and T are absolute invariances, V is a relative invariance of weight 1, and S is a relative invariance of weight 2.

Similarly, by applying the coordinate rotation to Eq. (34), it is seen that r is a scalar, p, q form a vector, and a, b, b, c form a tensor. Consequently, the following invariances are constructed:

$$\left.\begin{aligned} &k=\sqrt{1+p^2+q^2},\quad v=\frac{p+iq}{k} \\ &t=\frac{a+c}{2k},\quad s=\frac{a-c}{2k}+i\frac{b}{k} \end{aligned}\right\} \qquad (38)$$

ρ and k, t are absolute invariances, v is a relative invariance of weight 1, and s is a relative invariance of weight 2.

Using those invariances, the recovery Eq. (36) can be written as follows:

$$\left.\begin{aligned} &\rho k=A_0,\quad tv+sv^*=V \\ &t^2+ss^*=T,\quad ts=S/2 \end{aligned}\right\} \qquad (39)$$

When $t \neq 0$ and $t^2 - ss^* \neq 0$, the solution is given as follows [8]:

Theorem 4.

$$\left.\begin{aligned} &t=\pm\sqrt{\frac{T\pm\sqrt{T^2-SS^*}}{2}},\quad s=\frac{S}{2t} \\ &v=\frac{tV-sV^*}{t^2-ss^*},\quad k=\frac{1}{\sqrt{1-vv^*}} \\ &\rho=A_0/k \end{aligned}\right\} \qquad (40)$$

The original object parameters are given as follows:

$$\left.\begin{aligned} &\rho=A_0/k,\quad p=k\mathrm{Re}[v] \\ &q=k\mathrm{Im}[v],\quad a=k(t+\mathrm{Re}[s]) \\ &b=k\mathrm{Im}[s],\quad c=k(t-\mathrm{Re}[s]) \end{aligned}\right\} \qquad (41)$$

As is seen from the first of Eq. (40), there are four sets of solutions. Since the projection is parallel, the mirror image in regard to xy plane cannot be discriminated. Except for the mirror image relation, there are two sets of solutions. The two solutions exist because the rate of texture density change is determined only by the slant of the surface, independently of the tilt.

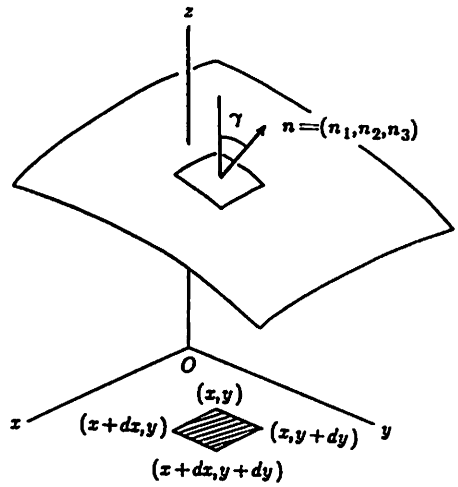

Fig. 3. Under the orthographic projection along the z-axis, the area of the region on the surface which corresponds to the infinitesimal square on the xy-plane defined by four points (x, y), $(x + dx, y)$, $(x + dx, y + dy)$, $(x, y + dy)$ is $dxdy/\cos\gamma$, where γ is the angle made by the unit normal n to the surface and the z-axis.

$t = 0$ applies when the two major curvatures have the same absolute value. It is impossible in this case to discriminate whether the surface is elliptic (with positive Gaussian curvature) or hyperbolic (with negative Gaussian curvature) [8]; $t^2 - ss^* = 0$ applies when the surface is hyperbolic (with zero Gaussian curvature). In this case, the asymptotic direction (direction of the peak) is determined, but the gradient along that direction is indeterminate.

7. Conclusions

This paper considered the 3D recovery problem, which is important in the computer vision, and proposed a method utilizing the invariance providing an irreducible representation of the 2D rotation group SO(2), which corresponds to the rotation of the coordinate system on the image plane. By this method, the recovery equation is simplified. Not only the analytic solution is easily derived, but also the geometrical interpretations of the equation as well as the variables are made clear (Weyl's thesis), making it easy to transform the equations. This situation is demonstrated for the case of the analysis of the optical flow and the recovery of a surface by texture.

When Weyl's thesis is to be used for 3D recovery problem, there can be various groups of transformations, other than the coordinate transformation on the image plane, that do not affect the essential aspect of the phenomenon. When, for example, the camera is rotated around the lens axis, essentially the same information is observed, since the ray is kept as the same. From such a viewpoint, the invariance composing the irreducible representation for the 3D rotation group SO(3), corresponding to the rotation of the camera, can be constructed. Since SO(3) is not commutative, the irreducible representation is not in general one-dimensional. By forming not only the linear combinations of the observed characteristics, but also using an algebraic nonlinear transformation, the problem can be reduced to 1D invariances (for details, see [2], [7], and [8]).

This kind of reasoning can be applied to various problems, and is one of the indispensable techniques in the computer vision. It is expected that this paper will provide a basis for such an idea.

Acknowledgement. This study is a part of the author's study when invited to Maryland Univ., U.S.A. He appreciates the discussions made by Azriel Rosenfeld, Larry S. Davis, John Aloimonos, Murali Subbarao and Tsai-Chia Chou. He also acknowledges the useful comments made by Prof. S. Amari, Assoc. Prof. K. Sugihara, Dept. Phys. Instr., Fac. Eng., Univ. Tokyo.

REFERENCES

1. M. Hammermesh. Group Theory and Its Application to Physical Problems, Addison-Wesley, Reading, MA (1962).
2. K. Kanatani. Transformation rule and invariant decomposition of optical flow, Tech. Rep. I.E.C.E., Japan, PRL85-3 (May 1985).
3. K. Kanatani. Theory of object motion recognition without using correspondence point, Trans. Inf. Proc. Soc. Jap., 27, 3, pp. 373-382 (March 1986).
4. K. Kanatani. 3D Euclid space or 2D non-Euclid space?—Methodology for image recognition, Trans. (D), I.E.C.E., Japan, J70-D, 5, pp. 1035-1038 (May 1987).
5. K. Kanatani. Structure and motion from optical flow under orthographic projection, Comput. Vision Graphics Image Process., 35, pp. 181-199 (1986).
6. K. Kanatani. Structure and motion from optical flow under perspective projection, Comput. Vision Graphics Image Process, 37 (1987).
7. K. Kanatani. Camera rotation invariance of image characteristics, Comput. Vision Graphics Image Process, 38 (1987).
8. K. Kanatani. Group Theoretical Methods on Image Understanding, Springer, Berlin (1987).
9. K. Kantani and T.-C. Chou. Shape from texture: General principle, Proc. 17th Image Eng. Conf., pp. 129-132 (Dec. 1986).
10. H.C. Longuet-Higgins. The visual ambiguity of a moving plane, Proc. Roy. Soc. Lond., B-223, pp. 165-175 (1984).
11. M. Subbarao. Interpretation of Image Motion Fields: A Computational Study, Technical Report CAR-TR-221, Center for Automation Research, University of Maryland (Sept. 1986).
12. M. Subbarao and A.M. Waxman. Closed form solution to image flow equations for planar surfaces in motion, Comput. Vision Graphics Image Process., 36, pp. 208-228 (1986).
13. K. Sugihara and N. Sugie. Recovery of rigid structure from orthographically projected optical flow, Comput. Vision Graphics Image Process., 27, pp. 309-320 (1984).
14. H. Weyl. The Classical Groups, Their Invariants and Representations, 2nd ed., Princeton Univ. Press, Princeton, NJ (1946).
15. H. Weyl. The Theory of Groups and Quantum Mechanics, Dover, New York (1950).

AUTHOR

Ken-ichi Kanatani graduated 1972 Dept. Phys. Instr., Fac. Eng., Univ. Tokyo. Affiliated with Case-Western Reserve Univ., U.S.A. 1969-70. Completed Doctor's Program 1979 Grad. School, Univ. Tokyo (Phys. Instr.). Doctor of Eng. Assistant 1979 and Assoc. Prof. 1983, Dept. Inf. Eng., Gumma Univ. Invited Researcher, 1985-1986, Maryland Univ., U.S.A. Engaged in researches in applied mechanics, particle engineering, soil engineering, and presently, in computer vision, image interpretation and image recognition.