コンピュータビジョン 45－1
（1986. 11. 19）

画像特徴のカメラ回転不変性

金谷健一

群馬大学工学部情報工学科

静止したシーンに対してカメラを回転することによって画像面に生じる変換群を解析し、画面上の重みづけ平均（フィルタ）によって得られる特徴量の変換法則を研究する。これはカメラの回転に対応する３次元回転群のリー代数に基づいた画像の無限小生成作用素の代数構造を調べることによって得られる。それをもとにして種々の不変量や不変概念を導入し、画像の同値性の判定法や、カメラ回転量の対応点を用いない計算法を導く。そして、形状認識やコンピュータビジョンへの応用を論じ、数値計算例を示す。

CAMERA ROTATION INVARIANCE OF IMAGE CHARACTERISTICS

Ken-ichi Kanatani

Department of Computer Science, Gunma University

The image transformation due to camera rotation relative to a stationary scene is analyzed, and the associated transformation rules of features given by weighted averaging of the image are derived by considering infinitesimal generators and projective geometry. The result is applied to shape recognition and computer vision problems where camera rotation is involved. Some numerical results are given.

1. INTRODUCTION

The problem we consider is as follows. Suppose the camera is rotated by a certain angle around its focus relative to a stationary scene. As a result, a different image is seen on the image plane. If the amount of camera rotation is known, the original image can be recovered. (Here, we do not consider the effect of the image boundary. We assume that the image plane is sufficiently large and that the object or scene of interest is always included in the field of view.) Suppose the viewed image is characterized by a finite number of parameters or *features*. If the camera is rotated, the image is accordingly changed so that the features also change their values. If the set of features is *invariant* in the sense that these new values are completely determined by the original values and the amount of the camera rotation, we can predict the values of the features which would be obtained if the camera were rotated by a given amount, or conversely we can compute the amount of camera rotation which would transform the values of the features to prescribed values. These consideration are very important in many problems of computer vision and pattern recognition when the camera orientation is controlled by a computer. We will discuss, as a typical example, the centroid and principal axes of a given region to see how the formulation presented here works for actual problems. Some numerical examples are also given.

2. CAMERA ROTATION AND FEATURES

Let f be the *focal length* of the camera. The camera image is thought of as the projection onto an image plane located at distance f from the *focus F*; a point P in the scene is projected onto the intersection of the image plane with the "ray" connecting point P and the focus F. Let us choose an XYZ-coordinate system such that the focus F is at the origin and the Z-axis coincides with the camera optical axis (Fig. 1). Choose an xy-coordinate system in such a way that the x- and y-axes are parallel to the X- and Y-axes with origin $(0,0,f)$. This xy-plane plays the role of the image plane. A point (X,Y,Z) in the scene is projected onto (x,y) on the image plane, where

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

Consider a camera rotation around its focus F and the induced transformation of the image. Suppose the camera is rotated by rotation matrix R, which is an orthogonal matrix; $RR^T=I$, where ^{T} stands for transpose. As a result, the point in the scene which was seen at (x,y) now moves to another point (x',y'), which is given by

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

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

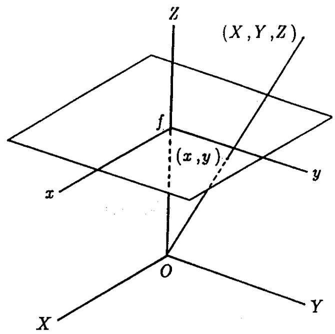

Fig. 1 The XYZ-coordinate system is fixed to the camera, the origin O being the camera focus. The image plane is taken to be $Z=f$, where f is the camera focal length. A point (X,Y,Z) in the scene is projected onto point (x,y) on the image plane.

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

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

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

Suppose the image is characterized by a finite number of parameters J_i, $i=1,2,...,N$, which we call *features* of the image [1,2]. If the image is transformed by eqns (2.2) as a result of camera rotation R, these features take different values J_i', $i=1,...,N$. We say that a set of features J_i, $i=1,...,N$ is *invariant* if the values J_i', $i=1,...,N$, are determined by the original values J_i, $i=1,...,N$ and the amount of camera rotation R alone. Let J_i, $i=1,...,N$, be an invariant set of features. We say that the set is *reducible* if it splits, after an appropriate rearrangement, into two or more sets, each of which is itself invariant. If no further reduction is possible, we say that the set of features is *irreducible*.

If a quantity c does not change its value by transformation (2.2) under camera rotation R, it is called a *scalar*. Obviously, a scalar is itself an invariant and is irreducible. If a pair a, b of numbers is transformed as x, y of eqns (2.2), we call it a *point*. Note that any pair of numbers can be interpreted as a position "on the image plane". However, it is interpreted as indicating a position "in the scene" if and only if it is transformed as a point. Evidently, a point is an invariant set of features and is irreducible.

A line on the image plane is expressed in the form

$$Ax+By+C=0. \tag{2.4}$$

Here, the ratio $A:B:C$ alone has a geometrical meaning; A, B, C and cA, cB, cC for a non-zero scalar c define one and the same line. Hence, we also call the line (2.4) "line $A:B:C$". If transformation (2.2) is applied, line (2.4) is mapped into

$$A'x'+B'y'+C'=0, \tag{2.5}$$

where ratio $A':B':C'$ is given by

Theorem 2. A line $A:B:C$ is transformed by camera rotation $R=(r_{ij})$ into the line

$$(r_{11}A+r_{21}B+r_{31}C/f):(r_{12}A+r_{22}B+r_{32}C/f)$$
$$:(f(r_{13}A+r_{23}B)+r_{33}C). \tag{2.6}$$

Proof. In view of eqns (2.1), eqn (2.4) is written as $Af(X/Z)+B(fY/Z)+C=0$, or

$$[A\ B\ C/f]\begin{bmatrix}X\\Y\\Z\end{bmatrix}=0. \tag{2.7}$$

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

$$\begin{bmatrix}A'\\B'\\C'/f\end{bmatrix}=R^T\begin{bmatrix}A\\B\\C/f\end{bmatrix}, \tag{2.8}$$

from which eqn (2.6) is obtained.

If the ratio of three quantities A, B, C is transformed by eqn (2.6) under camera rotation R, we call it a *line* and write it as $A:B:C$. It is an invariant set of features and is evidently irreducible. As in the case of a point, any triplet of numbers can be interpreted as a line "on the image plane", but it is interpreted as a line "in the scene" if and only if it is transformed as a line.

3. 3D VECTORS AND TENSORS

Consider three quantities a, b, c which are transformed as a *3D vector*, i.e.,

$$\begin{bmatrix}a'\\b'\\c'\end{bmatrix}=R^T\begin{bmatrix}a\\b\\c\end{bmatrix}, \tag{3.1}$$

for camera rotation R. This is, by definition, an invariant set of features but is not irreducible because the "length" $\sqrt{a^2+b^2+c^2}$ is a scalar. We can easily check the following results:

Lemma 1. If a, b, c are transformed as a 3D vector, then fa/c, fb/c are transformed as a point.

Lemma 2. If a, b, c are transformed as a 3D vector, then $a:b:fc$ is transformed as a line.

Theorem 3. A 3D vector is an invariant feature set. It can be irreducibly reduced into a point and a scalar or into a line and a scalar.

Next, consider nine elements A_{ij}, $i,j=1,2,3$, which are transformed by camera rotation R as a *3D tensor*, i.e.,

$$\begin{bmatrix}A_{11}' & A_{12}' & A_{13}'\\A_{21}' & A_{22}' & A_{23}'\\A_{31}' & A_{32}' & A_{33}'\end{bmatrix}=R^T\begin{bmatrix}A_{11} & A_{12} & A_{13}\\A_{21} & A_{22} & A_{23}\\A_{31} & A_{32} & A_{33}\end{bmatrix}R. \tag{3.2}$$

This is, by definition, an invariant set of features. However, it is not irreducible. First, it can be decomposed into a *symmetric part* and an *antisymmetric part* (or *skew part*):

$$\begin{bmatrix}A_{11} & A_{12} & A_{13}\\A_{21} & A_{22} & A_{23}\\A_{31} & A_{32} & A_{33}\end{bmatrix}$$
$$=\begin{bmatrix}A_{11} & (A_{12}+A_{21})/2 & (A_{13}+A_{31})/2\\(A_{21}+A_{12})/2 & A_{22} & (A_{23}+A_{32})/2\\(A_{31}+A_{13})/2 & (A_{32}+A_{23})/2 & A_{33}\end{bmatrix}$$
$$+\begin{bmatrix}0 & (A_{12}-A_{21})/2 & -(A_{31}-A_{13})/2\\-(A_{12}-A_{21})/2 & 0 & (A_{23}-A_{32})/2\\(A_{31}-A_{13})/2 & -(A_{23}-A_{32})/2 & 0\end{bmatrix}, \tag{3.3}$$

and each part is transformed as a 3D tensor like eqn (3.2) separately. Moreover, it can be checked easily that the three independent elements $(A_{23}-A_{32})/2$, $(A_{31}-A_{13})/2$, $(A_{12}-A_{21})/2$ of the antisymmetric part are transformed as a 3D vector. Hence, the antisymmetric part is irreducibly reduced into a point and a scalar or into a line and a scalar.

Suppose $A=(A_{ij})$ is already a symmetric 3D tensor. Such a tensor can be viewed as three mutually perpendicular unit vectors e_1, e_2, e_3 indicating the principal axes and the corresponding principal values σ_1, σ_2, σ_3 in the form

$$A=\sigma_1 e_1 e_1^T+\sigma_2 e_2 e_2^T+\sigma_3 e_3 e_3^T. \tag{3.4}$$

Here, this representation does not change if e_1 (or e_2 or e_3) is replaced by $-e_1$ (or $-e_2$ or $-e_3$). The three principal values are scalars, each of which is an invariant irreducible feature. On the other hand, if we determine the orientations of two of them, say e_1 and e_2, the orientation of the remaining one is uniquely determined. (e_3 and $-e_3$ indicate the same orientation.) As is shown in Theorem 3, the orientations of e_1 and e_2 are represented by two points on the image plane. (If we replace e_1 (or e_2) by $-e_1$ (or $-e_2$), the corresponding points are unchanged as desired.) However, since e_1 and e_2 are perpendicular, one of the two points and the line connecting the two points are sufficient; if one point on the image plane and a line passing through it are given, the three orientations are determined. Thus, we obtain

Theorem 4. A 3D tensor is invariantly reduced to its symmetric part and its antisymmetric part. The antisymmetric part is irreducibly reduced into a point and a scalar or a line and a scalar. The symmetric part is irreducibly reduced to three scalars, a point and a line passing through it.

4. INFINITESIMAL GENERATORS

Let $F(x,y)$ represent an observed image. This may be the intensity of the gray-level or a vectorvalued function corresponding to R, B and G. Here, the value of $F(x,y)$ is assumed to be "inherent to the scene" and independent of the viewing orientation. Chromaticity, for example, has this property. Furthermore, $F(x,y)$ is assumed to be "of finite support", i.e., $F(x,y)$ is zero at a sufficiently large distance from the origin of the image plane.

A 3D rotation is specified by the *rotation axis* (n_1,n_2,n_3), which is taken to be a unit vector, and the *rotation angle* Ω (rad) screwwise around it. If the rotation is *infinitesimal*, i.e., Ω is infinitesimally small, the rotation matrix takes the form $R=I+\delta R+o(\Omega)$, where I is the unit matrix, δR is the matrix given by

$$\delta R=\begin{bmatrix} 0 & -\Omega_3 & \Omega_2 \\ \Omega_3 & 0 & -\Omega_1 \\ -\Omega_2 & \Omega_1 & 0 \end{bmatrix}, \tag{4.1}$$

and $o(\Omega)$ denotes higher order terms of Ω. (We let the context indicate whether these terms are scalars, vectors or tensors.) Here, we put $\Omega_1=\Omega n_1$, $\Omega_2=\Omega n_2$ and $\Omega_3=\Omega n_3$, and they are interpreted as infinitesimal rotation angles around the X-, Y- and Z-axes, respectively.

If the camera rotation is infinitesimal, transformation (2.2) becomes $x'=x+\delta x+o(\Omega)$ and $y'=y+\delta y+o(\Omega)$, where

$$\begin{aligned} \delta x&=-f\Omega_2+\Omega_3 y+\frac{1}{f}(-\Omega_2 x+\Omega_1 y)x, \\ \delta y&= f\Omega_1-\Omega_3 x+\frac{1}{f}(-\Omega_2 x+\Omega_1 y)y. \end{aligned} \tag{4.2}$$

As a result, the image $F(x,y)$ also undergoes an infinitesimal change and becomes

$$F(x-\delta x,y-\delta y)=F(x,y)+\delta F(x,y)+o(\Omega), \tag{4.3}$$

and the first variation $\delta F(x,y)$ is given by

$$\begin{aligned} \delta F(x,y)&=-\frac{\partial F}{\partial x}\delta x-\frac{\partial F}{\partial y}\delta y \\ &=-(\Omega_1 D_1+\Omega_2 D_2+\Omega_3 D_3)F(x,y), \end{aligned} \tag{4.4}$$

where the *infinitesimal generators* are defined by

$$\begin{aligned} D_1&=\frac{xy}{f}\frac{\partial}{\partial x}+(f+\frac{y^2}{f})\frac{\partial}{\partial y}, \\ D_2&=-(f+\frac{x^2}{f})\frac{\partial}{\partial x}-\frac{xy}{f}\frac{\partial}{\partial y}, \\ D_3&=y\frac{\partial}{\partial x}-x\frac{\partial}{\partial y}. \end{aligned} \tag{4.5}$$

It can be checked easily that these infinitesimal generators satisfy the *commutator relations*

$$[D_1,D_2]=D_3,\ [D_2,D_3]=D_1,\ [D_3,D_1]=D_2, \tag{4.6}$$

where the *commutator* is defined by $[A,B]\equiv AB-BA$. Hence, a set of functions can be found which induces a representation of the 3D rotation group $SO(3)$ [3,4].

As is well known, a set of functions which induces an *irreducible representation* is obtained as eigenfunctions of the *Casimir operator*

$$H\equiv-(D_1{}^2+D_2{}^2+D_3{}^2). \tag{4.7}$$

The eigenvalue is $l(l+1)$ and the eigenspace is $2l+1$ dimensional, where l is an integer or half-integer called the *weight* of the irreducible representation. In other words, the differential equation

$$(D_1{}^2+D_2{}^2+D_3{}^2)F+l(l+1)F=0, \tag{4.13}$$

has $2l+1$ independent solutions, which become the basis of the irreducible representation D_l of weight l.

5. ADJOINT FEATURE TRANSFORMATION

Consider a feature obtained by weighted averaging or "filtering":

$$J=\int m(x,y)F(x,y)dxdy. \tag{5.1}$$

Here, $m(x,y)$ is the filter weight function and integration is performed over the entire image plane. (Recall our assumption of finite support of $F(x,y)$.)

If the camera is rotated by R, the image changes and consequently the value of feature J changes into J'. We define the *adjoint transformation operator* T_R^* by

$$J'=\int T_R^* m(x,y)F(x,y)dxdy. \tag{5.2}$$

Once how this adjoint transformation operator T_R^* acts on functions is known, the transformation of features is immediately computed for any given image. This is done by just considering infinitesimal transformations as follows.

If the image is infinitesimally changed as in eqn (4.3), feature J also undergoes an infinitesimally small change $J\rightarrow J+\delta J+o(\Omega)$. Substitution of eqn (4.4) and integration by parts yield

$$\delta J=\int(\Omega_1 D_1^*+\Omega_2 D_2^*+\Omega_3 D_3^*)m(x,y)F(x,y)dxdy, \tag{5.3}$$

where D_1^*, D_2^* and D_3^* are the *adjoint infinitesimal generators* defined by

$$\begin{aligned} D_1^*&=\frac{3y}{f}+\frac{xy}{f}\frac{\partial}{\partial x}+(f+\frac{y^2}{f})\frac{\partial}{\partial y}, \\ D_2^*&=-\frac{3x}{f}-(f+\frac{x^2}{f})\frac{\partial}{\partial x}-\frac{xy}{f}\frac{\partial}{\partial y}, \\ D_3^*&=y\frac{\partial}{\partial x}-x\frac{\partial}{\partial y}. \end{aligned} \tag{5.4}$$

In eqn (5.3), no boundary terms appear due to our assumption of finite support for $F(x,y)$. It can be checked easily that these adjoint infinitesimal generators satisfy the commutator relations

$$[D_1^*,D_2^*]=D_3^*,\ [D_2^*,D_3^*]=D_1^*,\ [D_3^*,D_1^*]=D_2^*. \tag{5.5}$$

Hence, we can find a set of functions which induces a representation of the 3D rotation group SO(3). As before, a basis of the irreducible representation of weight l is obtained as $2l+1$ eigenfunctions of the (adjoint) Casimir operator

$$H^*\equiv-(D_1^{*2}+D_2^{*2}+D_3^{*2}), \tag{5.6}$$

i.e., as $2l+1$ independent solutions of the differential equation

$$(D_1^{*2}+D_2^{*2}+D_3^{*2})m+l(l+1)m=0. \qquad (5.7)$$

We are not interested in representations of half integer weights, because they change sign for 2π rotation around an axis and necessarily involve complex expressions. Skipping the details of calculation, we show only the final results. First, we find that $1/\sqrt{(x^2+y^2+f^2)^3}$ is the basis of a one-dimensional representation ($l=1$). Hence,

$$J=\int \frac{F(x,y)dxdy}{\sqrt{(x^2+y^2+f^2)^3}} \qquad (5.8)$$

is an invariant (i.e., it is transformed as a scalar).

Similarly, computing the basis of a three-dimensional representation ($l=2$), we also see that

$$J_1=\int \frac{xF(x,y)dxdy}{(x^2+y^2+f^2)^2},\quad J_2=\int \frac{yF(x,y)dxdy}{(x^2+y^2+f^2)^2},$$

$$J_3=f\int \frac{F(x,y)dxdy}{(x^2+y^2+f^2)^2} \qquad (5.9)$$

are transformed as a 3D vector. Hence, they are irreducibly reduced to a scalar $\sqrt{(J_1)^2+(J_2)^2+(J_3)^2}$ and a point fJ_1/J_3, fJ_2/J_3 (or a line $J_1{:}J_2{:}fJ_3$).

Computing the basis of five-dimensional representation ($l=3$), we also find that

$$J_{11}=\int \frac{x^2F(x,y)dxdy}{\sqrt{(x^2+y^2+f^2)^5}},\quad J_{12}=\int \frac{xyF(x,y)dxdy}{\sqrt{(x^2+y^2+f^2)^5}},$$

$$J_{21}=\int \frac{xyF(x,y)dxdy}{\sqrt{(x^2+y^2+f^2)^5}},\quad J_{22}=\int \frac{y^2F(x,y)dxdy}{\sqrt{(x^2+y^2+f^2)^5}},$$

$$J_{31}=f\int \frac{xF(x,y)dxdy}{\sqrt{(x^2+y^2+f^2)^5}},\quad J_{32}=f\int \frac{yF(x,y)dxdy}{\sqrt{(x^2+y^2+f^2)^5}},$$

$$J_{13}=f\int \frac{xF(x,y)dxdy}{\sqrt{(x^2+y^2+f^2)^5}},\quad J_{23}=f\int \frac{yF(x,y)dxdy}{\sqrt{(x^2+y^2+f^2)^5}},$$

$$J_{33}=f^2\int \frac{F(x,y)dxdy}{\sqrt{(x^2+y^2+f^2)^5}} \qquad (5.10)$$

are transformed as a 3D (symmetric) tensor. Hence, they are irreducibly reduced to three scalars, a point and a line passing through it.

6. INVARIANT SHAPE CHARACTERIZATION

As an application of the results in the previous sections, let us consider the characterization of a shape on the image plane. Consider a region S on the image plane. Its characteristic function

$$F(x,y)=\begin{cases}1 & (x,y)\in S\\ 0 & \text{otherwise}\end{cases} \qquad (6.1)$$

is taken as the image function $F(x,y)$.

The simplest characteristic of the region S may be its area

$$\bar{S}=\int_S dxdy\left(=\int F(x,y)dxdy\right). \qquad (6.2)$$

However, this area is not invariant with respect to camera rotation. The area of eqn (6.2) changes if the camera is rotated. Hence, eqn (6.2) cannot be considered a characteristic inherent to the scene. In short, *eqn (6.2) is not a scalar.* On the other hand, if eqn (6.2) is replaced by

$$C=f^3\int_S \frac{dxdy}{\sqrt{(x^2+y^2+f^2)^3}}, \qquad (6.3)$$

this is a scalar as was shown in the previous section. If S is a small region located around the image origin, i.e., $x\approx 0$ and $y\approx 0$ in S, then C is approximately equal to its area. We call C the *invariant area* of region S. It is interpreted as the area the region would have if the region were moved to the center of the image plane by changing the camera orientation.

Another simple but important characteristic is the center of gravity of the region S:

$$\bar{x}=\int_S xdxdy\Big/\int_S dxdy,\quad \bar{y}=\int_S ydxdy\Big/\int_S dxdy. \qquad (6.4)$$

Again these quantities do not have invariant meanings. Namely, if region S is moved to another region by camera rotation and $(\bar{x}',\bar{y}')$ is its center of gravity, then $(\bar{x},\bar{y})$ is not necessarily mapped onto $(\bar{x}',\bar{y}')$ by the same camera rotation. In short, *the set of features $\bar{x}$, $\bar{y}$ is not a point.* On the other hand, we know from the previous section that

$$a_1=f\int_S \frac{xdxdy}{(x^2+y^2+f^2)^2},\quad a_2=f\int_S \frac{ydxdy}{(x^2+y^2+f^2)^2},$$

$$a_3=f^2\int_S \frac{dxdy}{(x^2+y^2+f^2)^2}, \qquad (6.5)$$

are transformed as a 3D vector. Hence, fa_1/a_3, fa_2/a_3 are transformed as a point. If the region S is a small region located around the image origin and $x\approx 0$, $y\approx 0$ in S, then $(fa_1/a_3, fa_2/a_3)$ is approximately the center of gravity of the region. We call $(fa_1/a_3, fa_2/a_3)$ the *invariant center of gravity* of region S. It is interpreted as the point which would be mapped into the center of gravity if the region were moved to the center of the image plane by changing the camera orientation.

Another useful characteristic is the moment tensor (M_{ij}), $i,j=1,2$, defined by

$$M_{11}=\int_S (x-\bar{x})^2dxdy,\quad M_{22}=\int_S (y-\bar{y})^2dxdy.$$

$$M_{12}=\int_S (x-\bar{x})(y-\bar{y})dxdy=M_{21}, \qquad (6.6)$$

Its principal values indicate the amount of elongation of the region S along the corresponding principal axes. However, as described above, this tensor does not have invariant meanings. Namely, *the principal values of (M_{ij}) are not scalars, and its principal axes are not lines.*

On the other hand, we know, from the previous section, that B_{ij}, $i,j=1,2,3$, defined by

$$B_{11}=\int_S \frac{x^2dxdy}{\sqrt{(x^2+y^2+f^2)^5}},\quad B_{12}=\int_S \frac{xydxdy}{\sqrt{(x^2+y^2+f^2)^5}},$$

$$B_{21}=\int_S \frac{xydxdy}{\sqrt{(x^2+y^2+f^2)^5}},\quad B_{22}=\int_S \frac{y^2dxdy}{\sqrt{(x^2+y^2+f^2)^5}},$$

$$B_{31}=f\int_S \frac{xdxdy}{\sqrt{(x^2+y^2+f^2)^5}},\quad B_{32}=f\int_S \frac{ydxdy}{\sqrt{(x^2+y^2+f^2)^5}},$$

$$B_{13}=f\int_S \frac{xdxdy}{\sqrt{(x^2+y^2+f^2)^5}},\quad B_{23}=f\int_S \frac{ydxdy}{\sqrt{(x^2+y^2+f^2)^5}},$$

$$B_{33}=f^2\int_S \frac{dxdy}{\sqrt{(x^2+y^2+f^2)^5}}, \tag{6.7}$$

are transformed as a 3D (symmetric) tensor. Since this tensor is positive definite as long as region S is not empty, it has three positive principal values σ_1, σ_2, σ_3. Let σ_3 be the maximum principal value. Let e_1, e_2, e_3 be the corresponding unit eigenvectors (determined up to sign). Let (g_1,g_2) be the point representing vector e_3. Let l_1 be the line passing through point (g_1,g_2) and the point representing vector e_1 (or the line representing vector e_2). Similarly, let l_2 be the line passing through point (g_1,g_2) and the point representing vector e_2 (or the line representing vector e_1). By our method of construction, scalars σ_1, σ_2, point (g_1,g_2) and lines l_1, l_2 are all invariant quantities. It can be checked that point (g_1,g_2) is the center of inertia, l_1 and l_2 are the principal axes, and σ_1, σ_2 are the corresponding principal values if S is a sufficiently small region around the origin. Hence, scalars σ_1 and σ_2 are the principal values the region would have if it were moved to the center of the image plane by camera rotation. Similarly, point (g_1,g_2) and lines l_1 and l_2 would be mapped into the center of inertia and principal axes by the same camera rotation. We call point (g_1,g_2) the *invariant center of inertia*, lines l_1, l_2 the *invariant principal axes*, and σ_1, σ_2 the corresponding *invariant principal values*.

7. CAMERA ROTATION RECONSTRUCTION

In the preceding section, we defined a scalar C by eqn (6.3), a 3D vector $a=(a_i)$ by eqns (6.5) and a 3D tensor $B=(B_{ij})$ by eqns (6.7). We decomposed them into scalars, points and lines. Now, we try to exhaust all *invariants*, i.e., expressions that do not change values under camera rotation.

Evidently, the scalars we have already found are invariants: C, $\|a\|$ or equivalently a^Ta, and the three principal values σ_1, σ_2 and σ_3 or equivalently any three independent algebraic expressions formed from them such as $\sigma_1+\sigma_2+\sigma_3$, $\sigma_1^2+\sigma_2^2+\sigma_3^2$ and $\sigma_1^3+\sigma_2^3+\sigma_3^3$, which are equal to $\mathrm{Tr}B$, $\mathrm{Tr}B^2$ and $\mathrm{Tr}B^3$, respectively.

There are other invariants describing the relationship between 3D vector a and 3D tensor B. A 3D vector is geometrically thought of as a directed axis to which its length is attached and a 3D symmetric tensor as three mutually perpendicular (undirected) axes to which their respective principal values are attached. From this picture, it is easily understood that the remaining invariants are those specifying the orientation of the vector relative to the three mutually perpendicular axes and that two invariants exist. We can use, say, a^TBa and a^TB^2a [5-9]. Of course, the choice is not unique as stated above, and other choices are also possible However, there exist only seven invariants that are algebraically independent.

We say that two regions S and S' on the image plane are *equivalent* if one region can be transformed into the other by camera rotation. If the two regions are equivalent, the above invariants must have identical values. If they have different values, the two regions cannot be equivalent. If the two regions are known to be equivalent, the amount of camera rotation which would take one region into the other can be reconstructed as follows.

Suppose we observe a and B for region S and a' and B' for region S'. Assume that B (hence B' as well) has three distinct eigenvalues and $a\neq 0$. Let e_1, e_2, e_3 be the associated eigenvectors of B. Since the eigenvectors are determined up to sign and magnitude, choose one set such that e_1, e_2, e_3 are mutually perpendicular unit vectors forming a right-hand system in that order. Construct a matrix R_1 having e_1, e_2, e_3 as its columns in that order. Let e_1', e_2', e_3' be the corresponding unit eigenvectors of B' forming a right-hand system. Since the signs of the eigenvectors are arbitrary, there are eight possibilities. For each case, construct the corresponding matrix R_2. Then, the rotation matrix which transforms B' to B is given by

$$R=R_2R_1^T. \tag{7.1}$$

(Matrix B is first transformed by $R_1^{-1}(=R_1^T)$ into a diagonal matrix, which in turn is transformed to B' by R_2.) Finally, choose one among those eight possible R's that transforms a to a'.

If B (hence B' as well) has only two distinct eigenvalues (a single root and a pair of multiple roots), let e_1 be the eigenvector associated with the single root. Suppose a is neither parallel nor perpendicular to e_1. Since the sign of e_1 is arbitrary, choose it so that a and e_1 make an acute angle. Then, we can construct three mutually orthogonal vectors forming a right-hand system e_1, $e_2=e_1\times a/\|e_1\times a\|$, $e_3=e_1\times e_2$. We can form R_1 and R_2 as described above, and the desired rotation is given by eqn (7.1). If a is perpendicular to e_1, there exist two solutions. If a is parallel to e_1, or if B (hence B' as well) has one eigenvalue (i.e., $B(=B')$ is a multiple of I), R is any rotation that maps a onto a' and we can add any rotation around a'. The case where $a=a'=0$ is treated similarly. In sum,

Theorem 5. C, a^Ta, $\mathrm{Tr}B$, $\mathrm{Tr}B^2$, $\mathrm{Tr}B^3$, a^TBa, a^TB^2a exhaust all the invariants constructed from C, a and B. If two regions are equivalent, the amount of camera rotation which takes one region into the other can be reconstructed from a and B.

8. APPLICATIONS AND EXAMPLES

Consider the problem of shape recognition. Suppose we have a reference image obtained from a certain camera orientation. If a test image is obtained from a different camera orientation, the two images cannot be compared directly due to projective distortion. However, the invariants provide an easy test for their equivalence; if the invariants of Theorem 5 have different values, the two regions cannot be equivalent and the test shape is rejected.

If C, $\boldsymbol{a}$ and B alone are sufficient to characterize the set of test shapes in question completely, the equivalence is already determined at this stage. Otherwise, we can move the test shape into the position of the reference shape in such a way that both have the same $\boldsymbol{a}$ and B. Then, the rest of the shape characteristics are compared to test for the equivalence. The necessary camera rotation is reconstructed as mentioned above, and the corresponding image transformation is performed either by actually moving the camera or by numerically computing the camera rotation transformation by eqn (7.1)

We say that a region on the image plane is in the *standard position*, if the invariant center of inertia (g_1, g_2) coincides with the origin of the image plane and the invariant principal axes coincide with the x- and y-axes. Any region on the image plane can be moved into the standard position by camera rotation R such that (i) B is diagonalized in the form

$$R^T B R = \begin{bmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{bmatrix}, \tag{8.1}$$

where σ_3 is the largest principal value, and (ii) if

$$R^T \boldsymbol{a} = \begin{bmatrix} a_1' \\ a_2' \\ a_3' \end{bmatrix}$$

then $a_3' > 0$.

Evidently, shape recognition becomes easier if the test shapes are always moved into the standard position (either by actually rotating the camera or by computation). However, this technique is not restricted to shape recognition. If a camera is tracking a moving object while the camera position is fixed, or if a camera attached to a robot or an autonomous vehicle is aiming at a fixed object in the stationary scene, the technique described above can be used so that the object in question is always seen in the standard position.

On the other hand, testing the equivalence is also viewed as detecting ***active motion***. When an object image moves on the image plane, we call the motion ***passive*** if that motion is induced by camera rotation alone and ***active*** otherwise. When the camera orientation is changed, object images move on the image plane, but those objects may also have moved in the scene independently of the camera. According to the procedure described above, we can detect active motion even if the angle and orientation of camera rotation are not known. If the corresponding two object images are not equivalent, the object must have moved actively. If they are equivalent, the object has not moved in the scene, although motion is observed on the image plane.

Another possible application is camera orientation registration. Even if the camera is rotated by an unknown angle around an unknown axis, the camera orientation can be determined as long as one particular region corresponding to a stationary object is identified on the image plane before and after the camera rotation. An important fact is, as was mentioned above, that knowledge of point-to-point correspondence is not required. Thus, the principle we have described has a wide range of applications to many problems.

Example Consider the three regions S_0, S_1, S_2 on the image plane (Fig. 2(a)). We use a scaling such that the focal length f is unity. Computing the integrations of eqns (6.5) and (6.7), we find their invariant centers of gravity (Fig. 2(b)) and principal axes (Fig. 2(c)) as follows:

S_0	S_1	S_2
(–0.08,–0.20)	(0.46,0.08)	(–0.47,0.35)
$y=-2.81x-0.43$	$y=1.67x-0.70$	$y=-0.08x+0.31$
$y=0.38x-0.17$	$y=-0.48x+0.30$	$y=-16.52x-7.42$

The invariants of (8.12) become as follows:

	S_0	S_1	S_2
C	0.1440	0.1440	0.1121
$\boldsymbol{a}^T\boldsymbol{a}$	0.0202	0.0202	0.0123
$\mathrm{Tr}(B)$	0.1440	0.1440	0.1121
$\mathrm{Tr}(B^2)$	0.0197	0.0197	0.0121
$\mathrm{Tr}(B^3)$	0.0028	0.0028	0.0013
$\boldsymbol{a}^T B\boldsymbol{a}$	0.0028	0.0028	0.0014
$\boldsymbol{a}^T B^2\boldsymbol{a}$	0.0004	0.0004	0.0001

From this result, we can conclude that regions S_0 and S_1 can be equivalent but region S_2 is not equivalent to either. (Here, the data are exact up to rounding. If the data are affected by a large amount of error, a statistical method such as hypothesis testing becomes necessary.) If three regions S_0, S_1, S_2 are known to be images of the same object, we can conclude that an active motion took place between S_0 (or S_1) and S_2 while no such motion took place between S_0 and S_1. By the procedure described in Section 3.8, the camera rotation which maps region S_1 onto region S_0 is reconstructed to be

$$R = \begin{bmatrix} 0.573 & 0.567 & 0.591 \\ -0.761 & 0.631 & 0.136 \\ -0.296 & -0.530 & 0.795 \end{bmatrix}.$$

This is the rotation around the axis of orientation (–0.384,0.512,–0.768) by angle 60° screwwise.

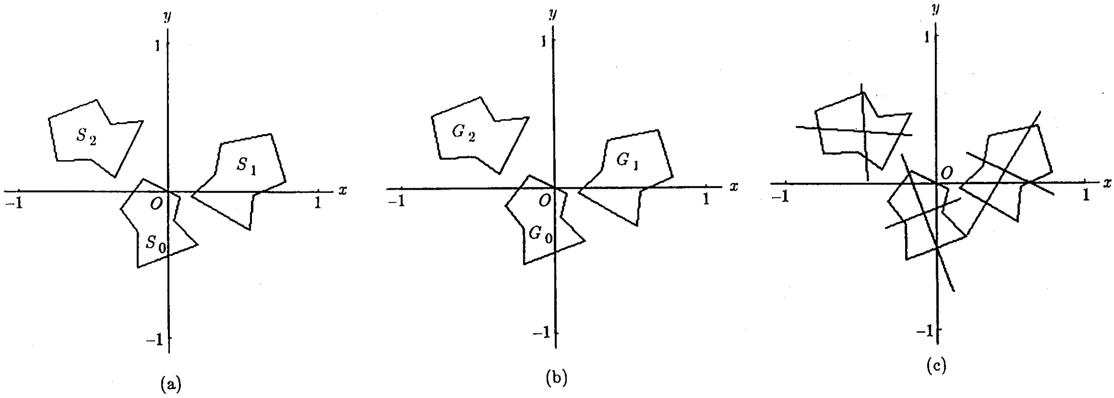

Fig. 2 (a) Three regions S_0, S_1, S_2 to be tested for equivalence. (b) Computed invariant centers of gravity G_0, G_1, G_2 of regions S_0, S_1, S_2. (c) Computed invariant principal axes of regions S_0, S_1, S_2.

Acknowledgement. The author thanks Prof. Azriel Rosenfeld and Prof. Larry Davis of the University of Maryland and Prof. Shun-ichi Amari of the University of Tokyo for helpful comments and discussions. He also thanks Tsai-Chia Chou of the University of Maryland, who double-checked all the results and prepared the numerical examples and figures. This work was supported in part by the Defense Advanced Research Projects Agency and the U.S. Army Night Vision and Electro-Optics Laboratory under Contract DAAK70-83-K-0018 (DARPA Order 3206).

REFERENCES

1. S. Amari, Invariant structures of signal and feature spaces in pattern recognition problems, *RAAG Memoirs*, **4** (1968), 553 - 566.
2. S. Amari, Feature spaces which admit and detect invariant signal transformations, *Proc. 4th Int. Joint Conf. Pattern Recognition*, Tokyo, 1978, pp. 452 - 456.
3. M. Hammermesh, *Group Theory and its Application to Physical Problems*, Addison-Wesley, Reading MA, 1964.
4. I. M. Gel'fand, R. A. Minlos and Z. Ya. Shapiro, *Representation of the Rotation and Lorentz Groups and their Applications*, Pergamon, Oxford, 1963.
5. A. J. M. Spencer, Theory of invariants, in *Continuum Physics*, Vol. 1 (ed. A. C. Eringen), Academic Press, New York, 1971, pp. 239 - 353.
6. C.-C. Wang, A new representation theorem for isotropic functions, Part 1, *Arch. Rational Mech. Anal.*, **36** (1970), 166 - 197.
7. C.-C. Wang, A new representation theorem for isotropic functions, Part 2, *Arch. Rational Mech. Anal.*, **36** (1970), 198 - 223.
8. G. F. Smith, On isotropic functions of symmetric tensors, skew-symmetric tensors and vectors, *Int. J. Engng Sci.*, **9** (1971), 899 - 916.
9. C.-C. Wang, Corrigendum to my papers on representation for isotropic functions, *Arch. Rational Mech. Anal.*, **43** (1971), 392 - 395.
10. K. Kanatani, Transformation of optical flow by camera rotation, *Proc. IEEE Workshop on Motion: Representation and Analysis*, Charleston, SC, 1986.