Pattern Recognition, Vol. 26, No. 5, pp. 667 681, 1993
Printed in Great Britain

0031-3203/93 $6.00 + .00
Pergamon Press Ltd

NORMALIZATION AND SHAPE RECOGNITION OF THREE-DIMENSIONAL OBJECTS BY 3D MOMENTS

J. M. GALVEZ † and M. CANTON ‡

† Departamento de Física Fundamental y Experimental, Centro Superior de Informática, Universidad de La Laguna, 38297 La Laguna, Tenerife, Canary Islands, Spain

‡ Departamento de Física, Facultad de Ciencias del Mar, Universidad de Las Palmas de Gran Canaria, 35017 Gran Canaria, Aptdo. 550, 35080, Canary Islands, Spain

(*Received* 8 *January* 1992; *in revised form* 12 *August* 1992; *received for publication* 18 *August* 1992)

Abstract—In this paper we are primarily concerned with the recognition of three-dimensional (3D) objects on the basis of the geometry of their physical surface. The 3D moments are defined on the object's surface with two main aims: (1) to derive a normalized version irrespective of position, size, and orientation for each imaged object; and (2) to establish global descriptors, extracted from normalized shapes, in order to make up a representative feature vector to be used in recognition tasks. The source data are multiple orthographic views of 3D objects with known viewpoint specifications. A volume intersection procedure is used in the recovery of 3D object surface from two-dimensional (2D) data. The problem of normalization to achieve recognition of 3D objects is dealt with and a heuristic solution is proposed to lessen the inherent ambiguity of the principal axes method for object orientation. Experimental results are described with ten object categories, showing excellent percentage classification success rates by using only a small number of normalized moments as elements of the feature vector.

Computer vision 3D shape recognition Surface representation Shape-from-contour
Shape normalization 3D moments

1. INTRODUCTION

The recognition of three-dimensional objects is a very important task which needs to be performed in many industrial applications of machine vision and in research areas, such as robotics and computer vision. A suitable way to tackle this problem is to carry it out in two stages. First, the three-dimensional (3D) surface of an object contained in a scene is recovered from a small number of two-dimensional (2D) images. To this end, several approaches were proposed. Stereopsis, motion, and analysis of shading stand in the forefront among the shape-from-x approaches. Shape-from-contour methods have been found to be effective in determining the shape of a visible surface. In the second stage, once the object's surface is properly reconstructed so that it can provide the relevant details, a pattern classification technique is accomplished. Here, the pattern features extracted from recovered surfaces are compared with a stored set of references until a match is found. The matching process is carried out by applying one of several well established similarity measure algorithms. Figure 1 shows a typical block diagram of the two-stage 3D object recognition procedure from multiple 2D views.

In the previous scheme there are two crucial points to take into account. The first point is related to the reconstruction method employed to obtain 3D surface models from 2D data. In this paper the source of shape information consists of the contours of multiple views of a 3D object, assuming the orthographic projection approximation in the sensing imaging device. A procedure based on volume intersection presented in papers by Martin and Aggarwal[1,2] has been developed[3] to derive a surface description, in contrast to the volumetric representation described in reference (2). The primary reason for surface representation is directly related to the application of shape recognition: that is, 3D object recognition based exclusively on the geometry of an object's surface.

The second point is that the application of pattern recognition techniques to shape recognition requires the definition of a suitable normalized feature vector, representative of both the general and the more subtle properties of a given shape. As the primary concern of this paper is shape recognition regardless of spatial position, orientation, and size, the associated feature vector for each 3D object must be normalized against shift, rotation, and scale change transformations.

Moment invariants were introduced by Hu[4] in 2D pattern recognition as features having the property of being invariant under the aforementioned transformations. Such moments were derived, in a limited and reduced number, upon the invariant algebra theory. Satisfactory applications, improvements, extensions and new kinds of normalizations, using 2D moments, have been proposed since then.[5-16]

The use of 3D moment invariants was first suggested by Sadjadi and Hall,[17] extending the concept of 2D moment invariants to three dimensions. The reduced number of invariants derived (two) may not be sufficient to identify objects having a complex shape.

In this paper the 3D moments are defined and evaluated on the object's surface for two specific

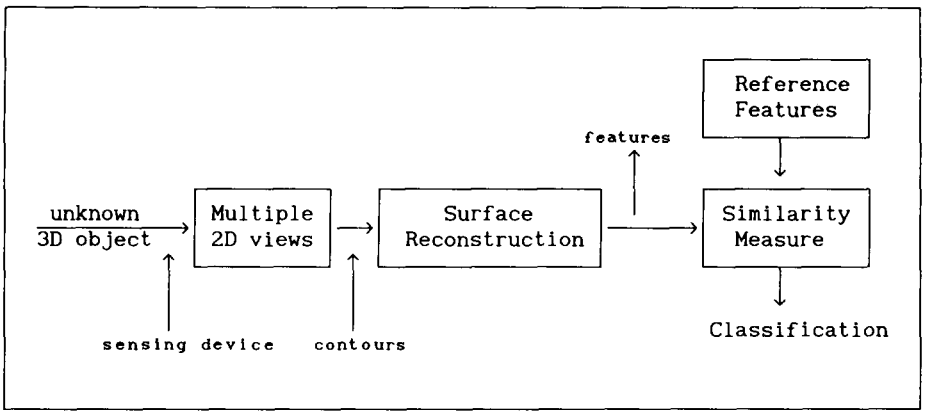


Fig. 1. Block diagram of the 3D object recognition process.

purposes: (1) to derive a normalized version independent of position, size and spatial orientation for each viewed 3D object; (2) as global descriptors extracted from normalized shapes, to make up the representative feature vector used in 3D object classification.

The organization of this paper is as follows. Section 2 describes the more relevent aspects of 3D surface reconstruction from contours of multiple orthographic projections. Three-dimensional moments evaluated on the object's surface are defined and presented in Section 3. The normalization problem, and a heuristic procedure to lessen the ambiguity problem related to the classical method of principal axes for 3D object orientation, are discussed in Section 4. Finally, in Section 5 experimental results are presented.

2. SURFACE RECONSTRUCTION FROM MULTIPLE VIEWS

A large class of 3D objects can be unambiguously identified on the basis of their enclosing surface regardless of colour, texture and other similar attributes. Being interested in 3D object recognition from 2D image data, we initially focused our attention on recovering the 3D structure of viewed objects.

We shall now sketch out a technique of volume intersection for visible surface reconstruction from contours of multiple views, and its more relevant characteristics as far as it relates to our particular concerns. A more detailed analysis can be found in references (1–3).

By the term "occluding contour", or more briefly "contour", we mean the silhouette boundaries of an orthogonally projected 3D object on an image plane.

Let us imagine a set of parallel planes with arbitrary orientation and uniform inter-plane separation. A 3D object characterized by a definite position, orientation and size is presented to the plane set. That object will be intersected by a set of planes defining a set of slices or cross-sections. Each slice will be composed of one or several disconnected sections. A volume intersection technique will be established in the pursuit of deriving a polygonal approximation for the boundary of each constituent 3D object real cross-section. This idea is illustrated in Fig. 2.

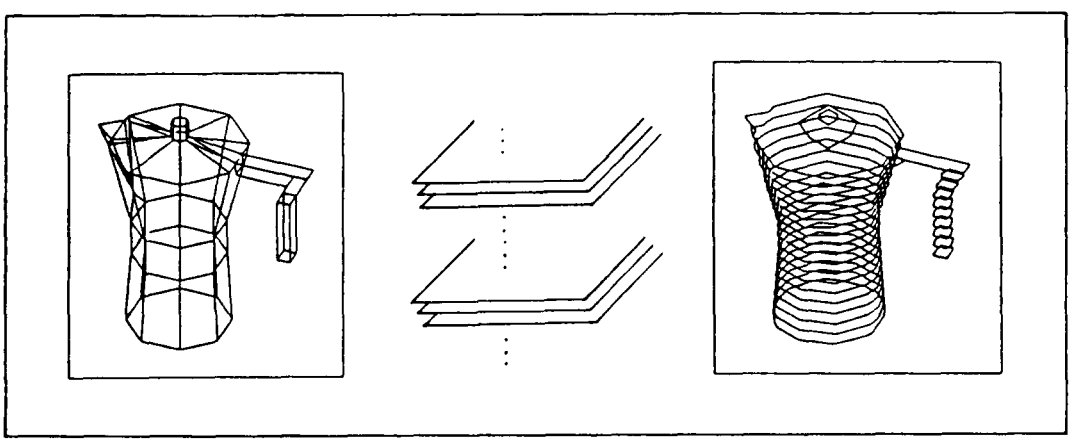


Fig. 2. Basic principle of contour line 3D object description.

It is clear that the inter-plane separation choice acts as a very powerful indicator of the kind of information made explicit on the final reconstructed 3D structure. Thus, the effect of reducing the number of slices makes the visible surface representation less descriptive of subtle details.

Having this in mind, let us now consider how the volume intersection technique can be implemented, in an elegant and easy way, from a small number of contours of 2D views. The process is developed in two stages. Initially the contours of two views, with no parallel plane images, are combined together to build a raw approach to the 3D shape of a viewed object. Then, the aforementioned reconstruction is refined as subsequent new contours are analysed.

The orientation of the plane set is derived from the first two views. The image planes of both 2D views share a common line. That common line is selected to specify the perpendicular direction to the set of planes. Then, the polygon that represents each occluding contour for the first two images is transformed in a new representation known as rasterized area (Fig. 3). The rasterized area is composed of an ordered set of rectilinear scan line segments, the endpoints of which specify the boundary points of the contour. Obviously, the rasterized area representation is obtained after the contour has been intersected with a family of parallel lines having uniform separation. The rasterization segments lie in a direction perpendicular to the common line, in such a way that the orthogonal projection of one segment meets the projection of the corresponding segment of the second contour on the common line.

Figure 3 shows the parallelogram structure that bounds the cross-sections of an object intersected by an imaginary set of planes. Each plane is associated with a common line point having features similar to the aforementioned ones. Every parallelogram is constructed by intersecting a pair of coplanar contour generating lines from the first view with another pair from the second view. It must be noted at this point that the contour generating lines, besides being perpendicular to each image plane, are chosen in such a way that a given scan line segment in one view has two contour generating lines.

It is worth mentioning here that unlike what is stated by Martin and Aggarwal[2] not all the parallelograms are bounding figures for the cross-sections of the 3D viewed object. Effectively, let us say that N_1 and N_2 are the numbers of scan line segments associated to the first and second view, respectively, for a given scan line (Fig. 4). Once the mutual constraints on contour generating lines of both views have been imposed, $N_1 \times N_2$ parallelograms are generated in the plane containing the scan lines. At this point, we can only state that the real cross-section of a 3D object must be circumscribed, at least, by N parallelograms, with $N = \max(N_1, N_2)$.[3]

In the second stage of the surface reconstruction method, the refinement process is achieved by updating the initial surface estimate in order to obtain increasingly finer approximations to the actual object's surface. This process is repeated by constraining each bounding surface description as new occluding contours are analysed. The underlying idea is to clip, in a slice-by-slice basis, every bounding polygon (parallel-

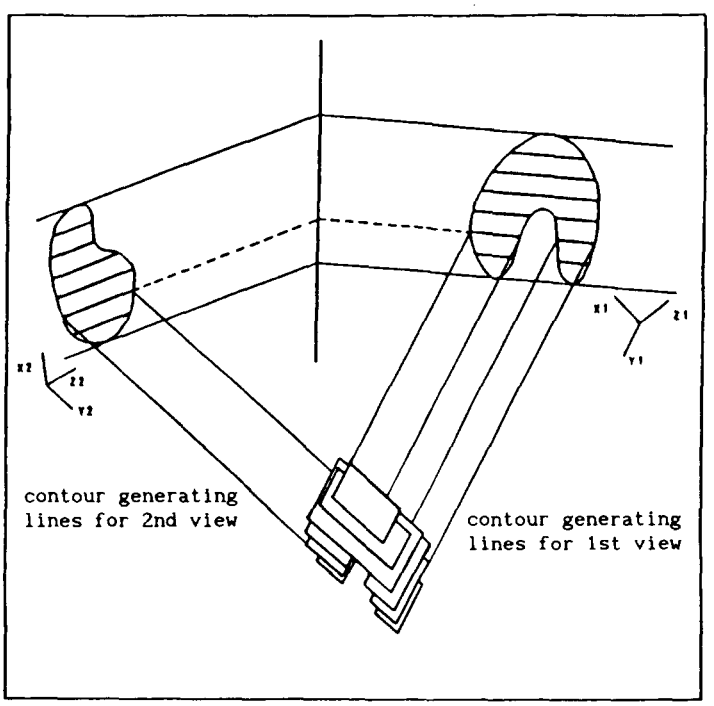


Fig. 3. Parallelogram structure of the initial reconstruction from two rasterized contours.

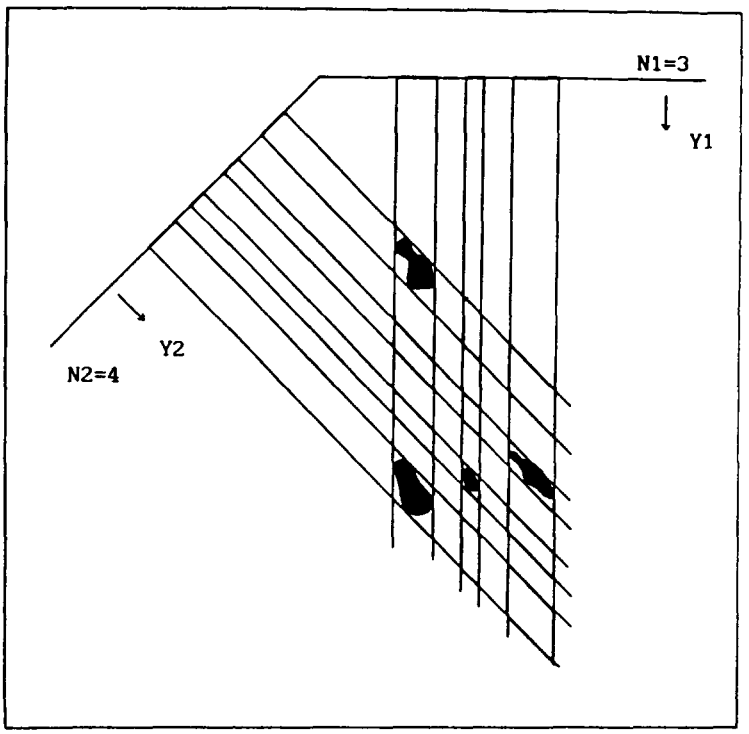


Fig. 4. Parallelogram structure associated with a slice. N_1 and N_2 are the number of scan line segments for the rasterization line shown in each contour. While the real slice comprises four disconnected sections of the actual object (shading), the total amount of generated parallelograms is 12.

ogram for the first new image) of the estimated reconstruction by the occluding contour in each new 2D image.

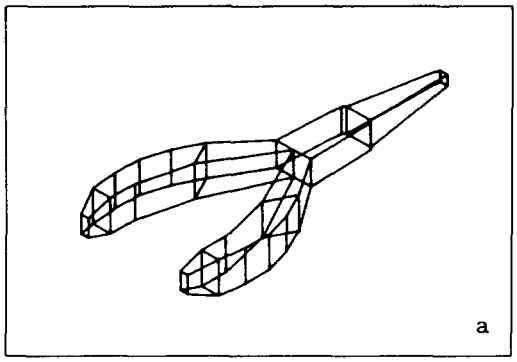
Thus, given the contour for a new view of a 3D object, any real cross-section of it should be orthogonally projected into the new contour. Obviously, the required direction of projection must be parallel to the line of sight for the new image plane.

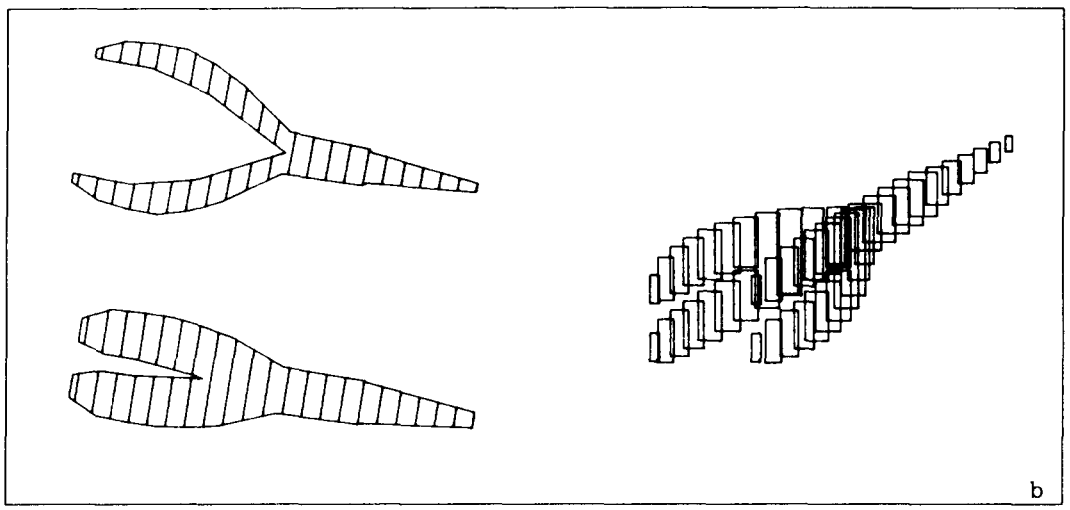
Whereas it is impossible to project an unknown real cross-section, its bounding polygons can be projected. Each projected polygon is in the new image plane and can be clipped by the contour of this view. Next, the resulting polygons of the clipping process are back-projected onto the original cross-section plane. Finally, the original bounding polygons are replaced by the aforementioned back-projected clipped polygons. It is worth mentioning that the result of clipping two polygons could be any number of new polygons, even zero. In the case of zero intersections, the polygon under refinement must be rejected as a bounding figure for the cross-section of the viewed object.

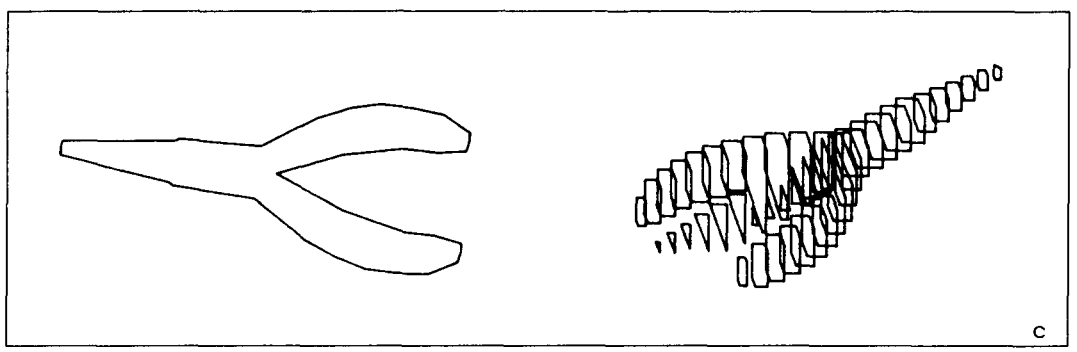
Figure 5 illustrates how the previous 3D reconstruction method works on estimating a contour line description for the surface of an object. Figure 5(a) presents a synthetic object to be imaged and subsequently reconstructed from a small number of 2D orthographic views. The initial reconstruction step applied to contours (with rasterization segments included) of the two views on the left-hand side of Fig. 5(b), yields the raw bounding surface (set of parallelograms) shown on the right of Fig. 5(b). The reconstructed surfaces after four successive applications of the refinement process are shown in Figs 5(c)–(f). The contour for each new orthographic view of Fig. 5(a) is included on the left-hand side. Each refined reconstruction is shown on the right.

The above contour line surface model, though it is easily interpreted by a human observer, may not be suitable for automatic 3D object recognition tasks. The primary reason lies in the fact that the orientation of the cross-section family is specified by the common line of the first two views. Thus, for a given set of 2D views of a 3D object, different sectional shape descriptions can be inferred depending upon the first two selected views at the initial step of the surface reconstruction process. Therefore, the numerical features extracted from the feasible different descriptions of the same 3D object, are strongly influenced both by the inter-slice separation and by their spatial orientation. We conclude, then, that the contour line surface descriptions do not observe the required invariance properties for automatic object recognition on the basis of numerical surface characteristics. So, the contour line representation should be converted to another surface model, in such a way that the numerical characteristics can be unambiguously evaluated independently of the processing order for the set of 2D views.

A suitable approach could be the construction of a polyhedral surface model, of triangular faces, from the







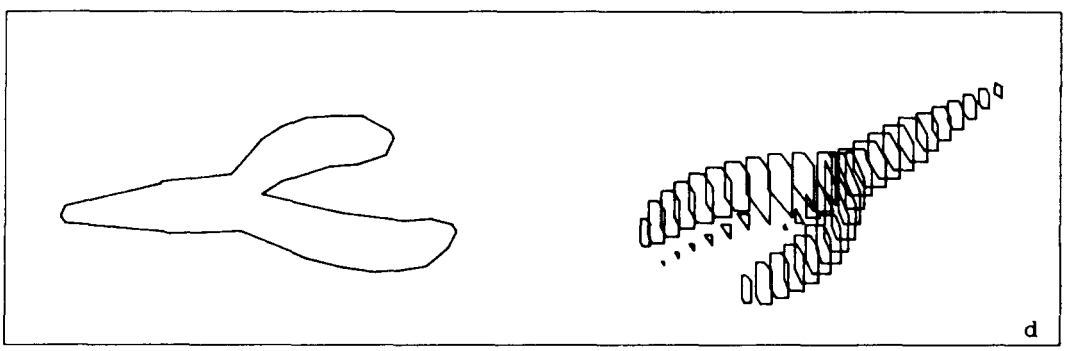


Fig. 5. Illustration of the surface reconstruction process. (a) Original 3D polyhedral model; (b) initial parallelogram structure recovered from two views; (c)–(f) subsequent refined versions after analysing four additional contours.

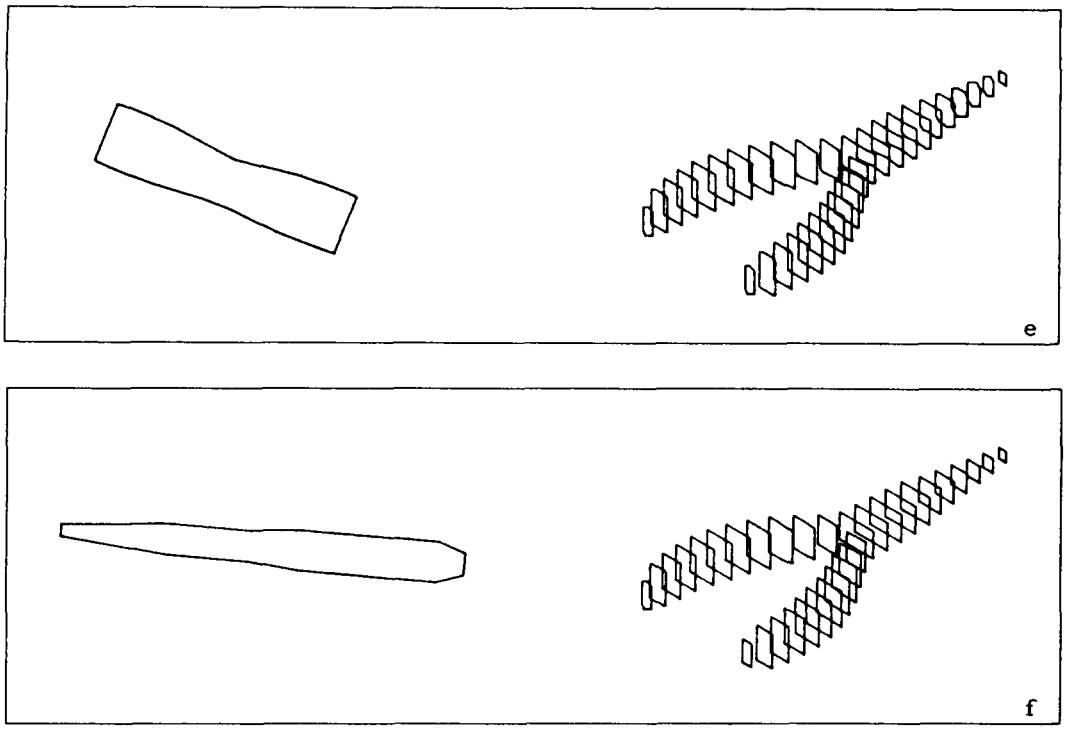


Fig. 5. (*Continued.*)

contour line representation. The resulting polyhedron has its vertices on the given planar contour lines, and each lateral surface band, delimited by two adjacent contour lines, is triangulated in accordance to the method developed by Christiansen and Sederberg[18] and Gálvez.[3] So, the 3D object surface is described unambiguously by a collection of planar triangles independently of the cross-sectional orientation.

The preceding observations are illustrated in Fig. 6. Figure 6(a) presents four contour line surface reconstructions for the object in Fig. 5(a). By considering four different processing orders for the six views of Fig. 5 we obtained the surfaces in Fig. 6(a). Their corresponding triangulated surfaces are shown in Fig. 6(b).

3. THREE-DIMENSIONAL MOMENTS

The 3D ordinary moments, m_{ijk}, of a 3D object characterized by a $f(x, y, z)$ density function are defined[17,19,20] as follows:

$$m_{ijk} = \int_{-\infty}^{+\infty} \int \int x^i y^j z^k f(x, y, z)\, dx\, dy\, dz \qquad i, j, k = 0, 1, 2, \ldots \quad (1)$$

where the integer $(i + j + k)$ is the order of m_{ijk}.

As pointed out in the previous section, the reconstructed surface of a 3D object can be effectively approximated by a tessellation of triangular faces, i.e. $S \equiv \{S_1, S_2, \ldots, S_N\}$. Here, the number of faces is represented by N. In this situation, expression (1) may be transformed so that m_{ijk} is evaluated by double integrals on the boundary surface S rather than on the whole volume extension. Furthermore, if the geometrical attributes of an object are the principal features to be used in shape recognition, the $f(x, y, z)$ function could be simplified by taking a constant value on surface points and zero elsewhere:

$$f(x, y, z) = \begin{cases} 1 & \forall (x, y, z) \in S \\ 0 & \text{elsewhere} \end{cases}. \quad (2)$$

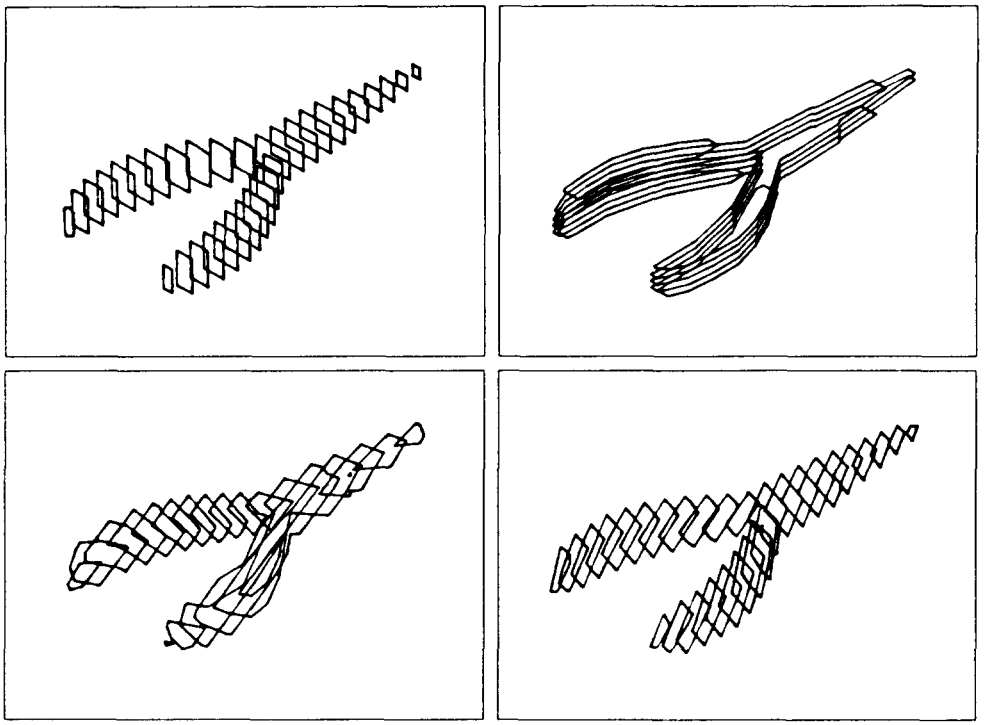
Given the previous considerations, the m_{ijk} ordinary moments may be obtained by evaluating double integrals on each constituent triangular face, S_t, of S:

$$m_{ijk} = \sum_{t=1}^{N} (m_{ijk})_t = \sum_{t=1}^{N} \int \int_{S_t} x^i y^j [z(x, y)]^k\, ds \quad (3)$$

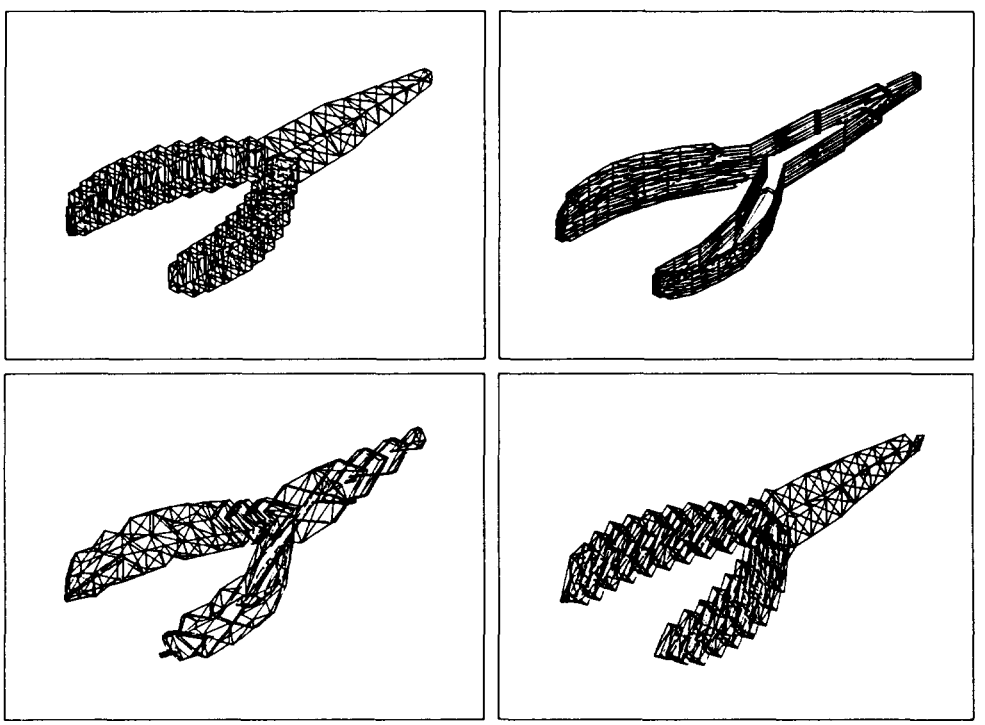
where $(m_{ijk})_t$ denotes the ordinary moment evaluated on S_t, ds is an element of the S_t surface, and $z(x, y)$ represents the equation for S_t.

4. THREE-DIMENSIONAL SHAPE NORMALIZATION

By the normalization term we mean the process of transforming the 3D object model into a standard version that retains all the relevant geometrical information of the original object, and also satisfies a set of conditions called the normalization criteria. These conditions will be imposed on the m_{ijk} moments evaluated upon the canonical or standard version. In geometrical terms, the canonical version will be characterized by a specific position, orientation and



(a)



(b)

Fig. 6. (a) Effect of recovering the surface of a 3D object from the same set of 2D views, but with a different processing order and a different intersection separation; (b) triangulated surfaces for objects in (a).

size. Thus, in categorizing 3D objects, only its canonical versions ought to be analysed and compared. Thereby, the position-, orientation-, and size-independence for the representative numerical feature vector of each normalized version will be successfully guaranteed.

The canonical or standard version is attained by the application of appropriate shift, rotation, and scaling transformations to the actual object. Such operations are defined by the expressions:

$$\boxed{\text{3D canonical version}} = A \cdot R \cdot T \boxed{\text{actual object}}$$

$$T: \begin{bmatrix} x \\ y \\ z \end{bmatrix} \longrightarrow \begin{bmatrix} x + T_x \\ y + T_y \\ z + T_z \end{bmatrix} \tag{4a}$$

$$R: \begin{bmatrix} R_{11} & R_{12} & R_{13} \\ R_{21} & R_{22} & R_{23} \\ R_{31} & R_{32} & R_{33} \end{bmatrix} \tag{4b}$$

$$A: \begin{bmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \alpha \end{bmatrix}. \tag{4c}$$

The unknown T_x, T_y, T_z, R_{ij}, and α coefficients in expressions (4a)–(4c), are derived according to the following normalization criteria set:

1. $\eta_{100} = \eta_{010} = \eta_{001} = 0$ (standardized position)
2. $\eta_{110} = \eta_{101} = \eta_{011} = 0$ (standardized orientation)
3. $\eta_{000} = C$ (standardized size) (5)

with η_{ijk} the ordinary moments extracted from the final canonical version.

Whereas the imposition of criterion 2 leads to aligning the principal axes frame of the normalized 3D shape with the x-, y-, z-directions, criteria 1 and 3 are respectively equivalent to locating the centroid at the origin of the coordinate system $OXYZ$ and to assigning a constant value, C, to the area of the boundary surface.

4.1. *Canonical position*

The centroid of a 3D shape can be defined as the point of coordinates $(\bar{x}, \bar{y}, \bar{z})$ in relation to the frame $OXYZ$ with associated unit and orthogonal vectors $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$

$$\bar{x} = \frac{m_{100}}{m_{000}}, \quad \bar{y} = \frac{m_{010}}{m_{000}}, \quad \bar{z} = \frac{m_{001}}{m_{000}}. \tag{6}$$

In expression (6), m_{ijk} are ordinary moments on the surface of the actual 3D object.

When the completion of criterion 1 is required, the centroid for any 3D object must be moved to the (0,0,0) position. The origin-centred condition is fulfilled after the application of the $T \equiv (T_x, T_y, T_z)$ shift transformation to the actual object. The appropriate values for T_x, T_y and T_z are specified by

$$T_x = -\bar{x}, \; T_y = -\bar{y}, \; T_z = -\bar{z}. \tag{7}$$

In this situation, the moments evaluated on the shifted version are known as central moments, μ_{ijk}, which are defined as

$$\mu_{ijk} = \iint_S (x - \bar{x})^i (y - \bar{y})^j [z(x, y) - \bar{z}]^k \, ds. \tag{8}$$

4.2. *Canonical orientation*

Given a 3D object located at the standard position, the canonical orientation is provided by rotating the position-normalized version to obtain a pre-established standard orientation. The canonical orientation has been selected so that the principal axes, associated to the normalized shape, lie in the x-, y-, z-directions, respectively.

When the bounding surface of a 3D object centred at the origin of the coordinate system is represented by moments up to second order, the object is equivalent to an ellipsoid of definite size and orientation, called the inertia ellipsoid. This ellipsoid carries with it three mutually perpendicular axes known as the principal axes, and denoted by $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ (Fig. 7). The principal axes of a 3D object centred at the origin of the $OXYZ$ coordinate system may be defined,[21,22] from a different point of view, as the eigenvectors of the inertia matrix

$$I = \begin{bmatrix} I_{xx} & -I_{xy} & -I_{xz} \\ -I_{yx} & I_{yy} & -I_{yz} \\ -I_{zx} & -I_{zy} & I_{zz} \end{bmatrix}, \quad \begin{aligned} I_{xx} &= \mu_{020} + \mu_{002} \\ I_{yy} &= \mu_{200} + \mu_{002}, \\ I_{zz} &= \mu_{200} + \mu_{020} \end{aligned}$$

$$\begin{aligned} I_{xy} &= I_{yx} = \mu_{110} \\ I_{xz} &= I_{zx} = \mu_{101}. \\ I_{yz} &= I_{zy} = \mu_{011} \end{aligned} \tag{9}$$

I is a real and symmetric matrix with real eigenvalues $\{\lambda_i, i = 1, 2, 3\}$ and orthogonal eigenvectors $\{\mathbf{u}_i, i = 1, 2, 3\}$ such as

$$I\mathbf{u}_i = \lambda_i \mathbf{u}_i. \tag{10}$$

The $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ eigenvectors define the rows of the rotation matrix S that aligns the principal axes with the $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ frame of reference. In other words, the $\{\mathbf{u}_i,$

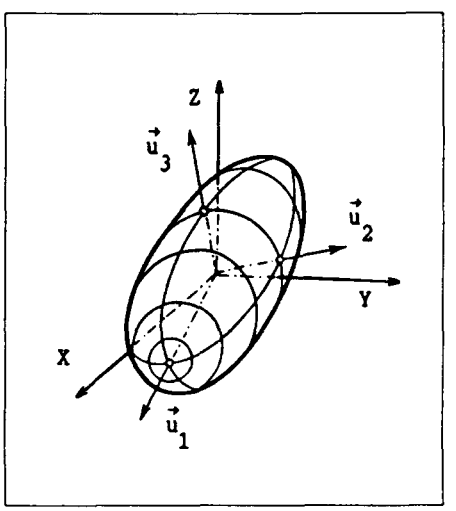


Fig. 7. Inertia ellipsoid with principal axes $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$.

$i = 1, 2, 3\}$ eigenvectors of the I matrix define a new orthogonal frame of reference so that the I matrix is diagonal when evaluated in relation to it. Thus

$$I'_{xy} = I'_{xz} = I'_{yz} = 0 \quad \text{or} \quad Q_{110} = Q_{101} = Q_{011} = 0 \tag{11}$$

where, I'_{xy}, I'_{xz} and I'_{yz} are the final elements of the I matrix once the object is reoriented so that its principal axes lie in x-, y-, z-directions. Q_{ijk} are the new moments evaluated on the object's surface after rotation S is applied to the actual position-normalized version.

Equation (11) represents the usual normalization criterion 2 with respect to orientation. As recently showed,[3,23] the imposition of equation (11), or the equivalent criterion 2, leads to a degeneracy problem in the selection of a unique rotation matrix from the $\{\mathbf{u}_i,\ i = 1, 2, 3\}$ eigenvectors of the I inertia matrix. Effectively, after sorting the eigenvalues $\{\lambda_i/\lambda_1 \leq \lambda_2 \leq \lambda_3\}$ in increasing order, the corresponding $\{\mathbf{u}_i,\ i = 1, 2, 3\}$ eigenvectors can be considered as three orthogonal and unit vectors so that $\mathbf{u}_i$ lies on the major ellipsoid axis direction, $\mathbf{u}_3$ lies on the minor ellipsoid axis direction and $\mathbf{u}_2$ is aligned with the intermediate ellipsoid axis. However, $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$ only define three perpendicular directions in 3D space, the same as their corresponding ellipsoid axes. Thus, $\mathbf{u}_i$ and $-\mathbf{u}_i$, $i = 1, 2, 3$, are two valid eigenvectors related to the same λ_i eigenvalue

$$I(\mathbf{u}_i) = \lambda_i(\mathbf{u}_i)$$
$$I(-\mathbf{u}_i) = \lambda_i(-\mathbf{u}_i). \tag{12}$$

Equation (12) reveals that there are exactly eight groups of unit and orthogonal vectors $\{\pm\mathbf{u}_1, \pm\mathbf{u}_2, \pm\mathbf{u}_3\}$ specifying the same principal axes directions of a 3D object. This states that any 3D object can be transformed by means of eight different rotation matrices in such a way that the eight rotated versions satisfy the normalization criterion 2 (equation (11)). In other words, whereas any 3D object has a unique associated ellipsoid, the opposite statement is not true. More specifically, an ellipsoid with certain size and orientation may be the same inertia ellipsoid of eight different versions (the same position and size but different orientation) of a 3D object.

Therefore, we conclude that classical normalization criteria (equation (11)) lead to eight different orientations of the same object.

4.2.1. *Elimination of the ambiguity in object orientation.* The problem we are now faced with is to find a strategy that allows to select a unique rotation matrix from the eight possible options provided by the fulfillment of criterion 2.

Although the $\{\pm\mathbf{u}_1, \pm\mathbf{u}_2, \pm\mathbf{u}_3\}$ vector system is made up of eight orthonormal frames, it is not possible to ensure that each of these frames is a right-handed coordinate system. The right-handed requirement reduces the ambiguity to four feasible combinations from $\{\pm\mathbf{u}_1, \pm\mathbf{u}_2, \pm\mathbf{u}_3\}$, but the degeneracy problem is not completely removed. Effectively, assuming $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ is one of the right-handed frames rigidly joined to a 3D object, any rotation around these directions by π radians yields a similar number of valid but different orientations for the 3D object, all in accordance with criterion 2. This fact is illustrated in Fig. 8,

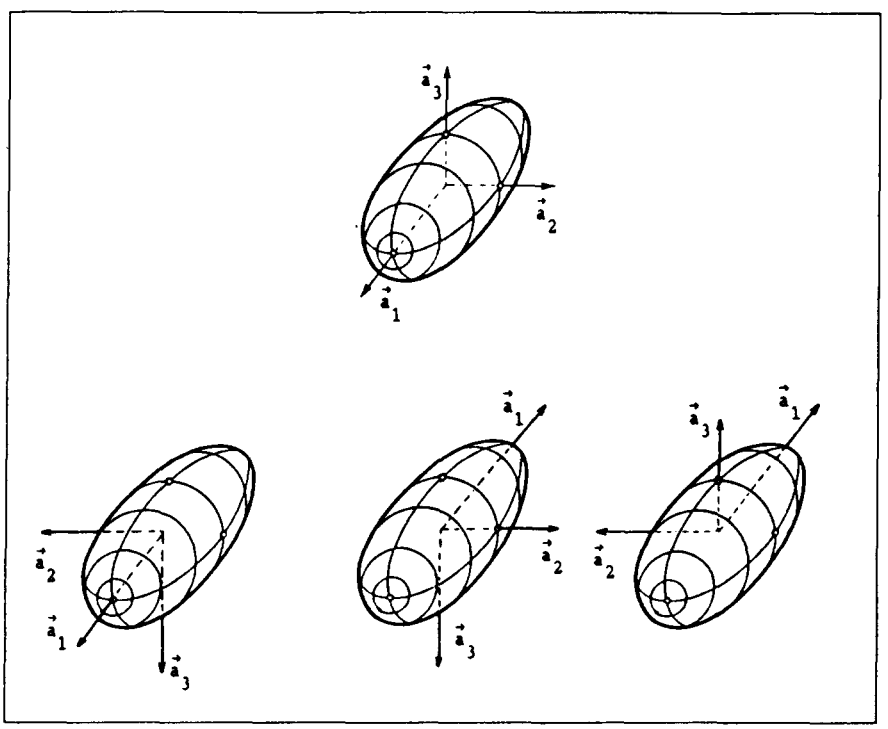


Fig. 8. The π radians rotations about $\mathbf{a}_1$, $\mathbf{a}_2$, and $\mathbf{a}_3$ lead to four different orientations of the same 3D object, keeping its corresponding inertia ellipsoid unchanged.

where the inertia ellipsoid associated with the four different instances of the same 3D object is kept unchanged after the above rotations are applied.

In order to lessen the orientation ambiguity, a unique object-centred and right-handed coordinate system $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ out of the $\{\pm\mathbf{u}_1, \pm\mathbf{u}_2, \pm\mathbf{u}_3\}$ set must be found. At this end, a heuristic procedure has been developed involving the following steps:

A.1. Let P_1 be the most distant point from the centroid, determined by the intersection of the $\mathbf{u}_1$ line with the bounding object surface. The $\mathbf{v}_1$ eigenvector, associated with the minor λ_1 eigenvalue, must be directed from the centroid to the P_1 point.

A.2. Let P_2 be the most distant point from the centroid, determined by the intersection of the $\mathbf{u}_2$ line with the bounding object surface. The $\mathbf{v}_2$ eigenvector, associated with the λ_2 intermediate eigenvalue, must be directed from the centroid to the P_2 point.

A.3. The third $\mathbf{v}_3$ eigenvector, aligned with the minor ellipsoid axis (major λ_3 eigenvalue), is selected so that the $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ frame defines a right-handed coordinate system, i.e.

$$\mathbf{v}_1 \times \mathbf{v}_2 = \mathbf{v}_3 \qquad \text{(cross-product)}.$$

Let R be a matrix formed from the three orthonormal row vectors $\mathbf{v}_1$, $\mathbf{v}_2$ and $\mathbf{v}_3$. Matrix R rotates the 3D object so that its principal axes lie in the $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ directions; furthermore, R is the only matrix satisfying the additional criteria A.1–A.3.

4.3. *Canonical size*

Given a position-, and orientation-normalized 3D object, the size normalization is carried out by determining the appropriate scaling transformation, A, that provides a fixed C value for the area of the final canonical version. The user-defined C value, in addition to the expression that relates the size-, orientation-, and position-normalized moments, n_{ijk}, to the Q_{ijk} moments extracted from position- and orientation-normalized surface shape, leads to the α scaling factor in equation (4c).

$$\left.\begin{aligned} \eta_{000} &= C \\ \eta_{ijk} &= \alpha^{(i+j+k+2)} Q_{ijk} \end{aligned}\right\} \rightarrow \alpha = \left(\frac{\eta_{000}}{Q_{000}}\right)^{1/2} = \left(\frac{C}{m_{000}}\right)^{1/2}. \tag{13}$$

The fact that the area of an object's surface may be expressed by the m_{000} ordinary moment ($m_{000} = Q_{000} = \eta_{000}$), irrespective of its position and orientation, has been implicitly assumed in deriving equation (13).

5. EXPERIMENTAL RESULTS

The recognition of an unknown 3D object as belonging to one of several pre-established categories can be sketched out in the following steps:

(1) Recovery of 3D shape for each imaged object.

(2) Normalization of a reconstructed shape independently of shift, rotation and scaling transformations.

(3) Generation of numerical pattern descriptors for recognition tasks.

To illustrate the appropriateness of 3D moments to normalize 3D shapes and to contribute to the process of identification of individual 3D objects, the following experiment was devised. Ten synthetic polyhedral 3D objects were created, showing different complexity levels in the shape of their visible surface (Fig. 9). Eleven transformed instances for each of the synthetic objects were derived after the application of appropriate shift, rotation and scaling operations. Ten different 2D views were generated for each resulting 3D object, assuming the approximation of orthographic projection. The shape-from-contour technique discussed in Section 2 was applied in rendering the boundary surface for each one of the 110 instances (with different position, size and orientation) associated to the ten fixed 3D object categories. Every reconstructed surface was approximated by a collection of planar and triangular faces from the boundary contour description.

The canonical version of each reconstructed surface was derived according to the criteria stated in Section 4, using only moments up to second order. The canonical versions for the actual 3D polyhedral models being imaged are shown in Fig. 10. Figure 11 presents the reconstructed and normalized shapes for the ten classes of 3D objects after the application of the first transformation. In this experiment, in order to improve the visual appearance, a canonical version was chosen having the following features:

(1) Object centroid located at the origin of the coordinate system (0, 0, 0).

(2) A value of $C = 100$ units for the area of the visible surface.

(3) Principal axes, or intrinsic and orthonormal right-handed $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ frame, aligned with the $\{\mathbf{k}, \mathbf{i}, \mathbf{j}\}$ frame of reference.

It is useful to note at this point that accurate reconstructed surfaces were obtained from a small number of 2D views. Likewise, from Fig. 11 it is clear the excellent concordance between the normalized version for actual objects and its corresponding canonical and reconstructed versions.

Once the normalized version is estimated for all instances of the ten 3D object classes, a feature vector with a maximum length of ten elements is used for shape description. Moments η_{ijk} ranged from zeroth order to fifth order, to the amount of 56, were employed to describe the boundary surface of 3D patterns. The seven moments ($\eta_{000} = 100$, $\eta_{100} = \eta_{010} = \eta_{001} = 0$, $\eta_{011} = \eta_{101} = \eta_{110} = 0$) involved in the normalization process were removed, since they take a constant value irrespective of the 3D normalized object we are working with. Thus, each element of the feature vector is selected from among the 49 remaining normalized moments.

A linear discrimination procedure was considered for 3D pattern recognition, with a minimum-distance classifier for similarity measurements.[3,24] In our work

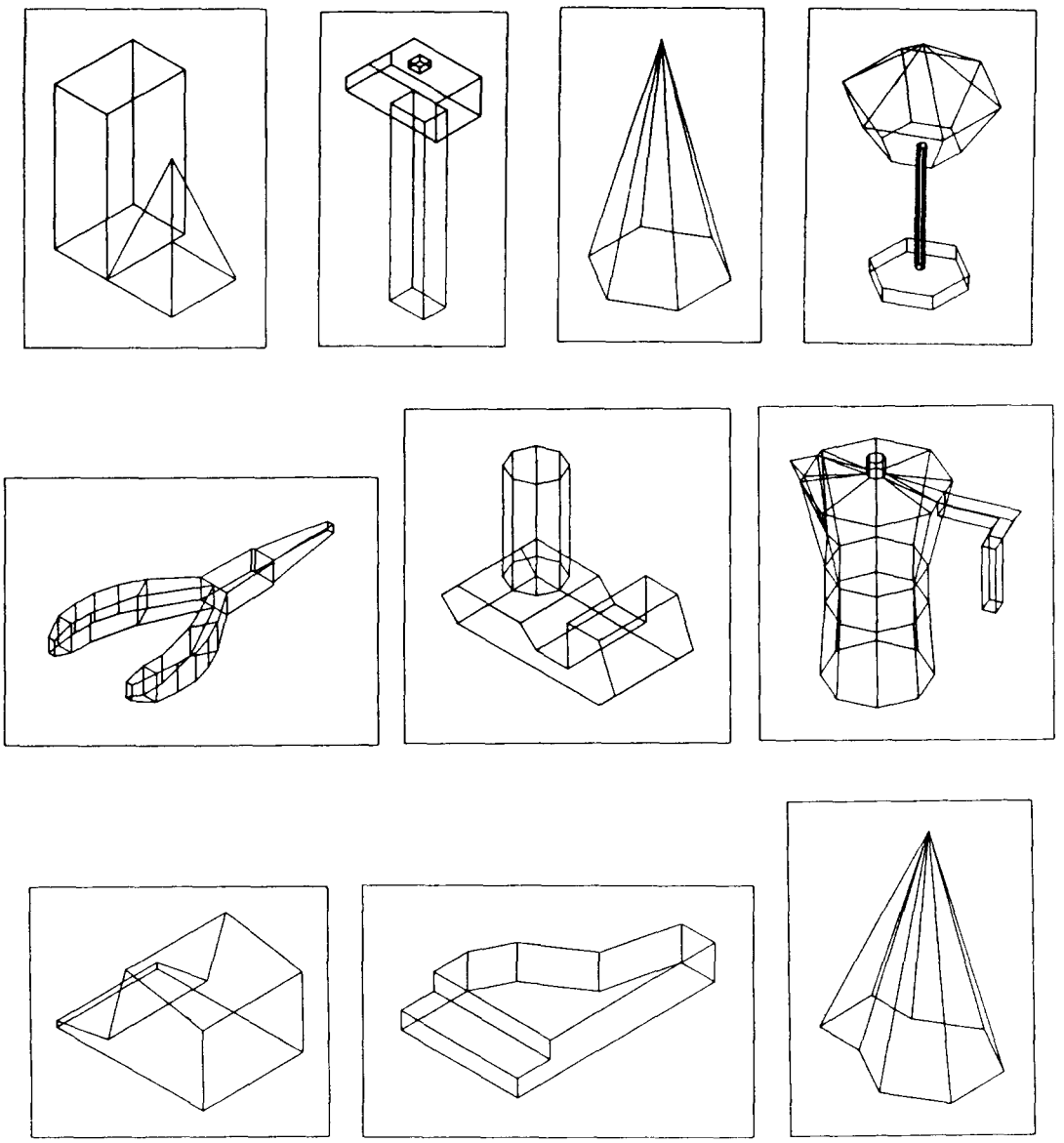


Fig. 9. Ten synthetic polyhedral 3D objects used in the reconstruction and recognition processes.

the metric defined by the total covariance matrix T was used for distance evaluation.

In order to establish a good partition of feature space for considered categories of 3D objects, a step-by-step technique for feature selection was employed to compound a relatively small feature vector representative of relevant properties of 3D patterns. For each step q, a η_{ijk} moment is incorporated as a new element of the feature vector. The additional η_{ijk} to be included is chosen from among the remaining set of moments, in such a way that the trace of the $(T_q^{-1} B_q^{-1})$ matrix is maximized. The process is sequentially repeated until a maximum of ten steps. T_q and B_q are the total and between-classes covariance matrices, respectively, derived from the corresponding T and B matrices by only taking into account the terms related to the q features.

The training set was constructed by randomly selecting 70 normalized 3D patterns from the 110 available 3D shapes (seven instances from each 3D object class). The remaining 40 instances of 3D normalized shapes were selected to build the test set (four patterns per class). It was assumed that the unknown patterns to be identified are those of the test set. That straightforward choice supplies us with the additional knowledge of possible misclassification for individual patterns, since their correct class membership is known in advance.

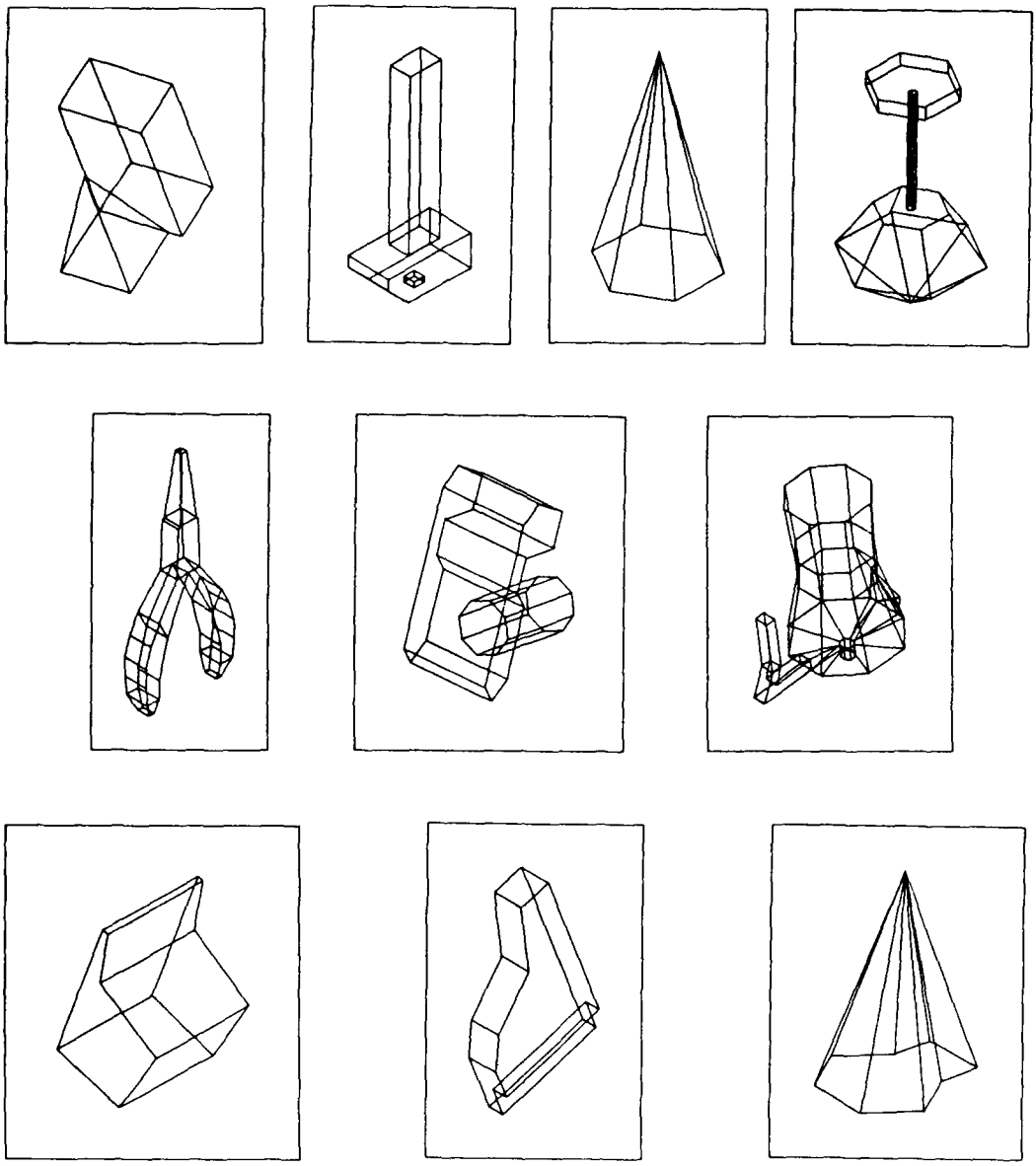


Fig. 10. Canonical versions of the polyhedral objects in Fig. 9.

Table 1 summarizes the identification results for 40 unknown 3D objects, assuming they belong to one of the ten existing categories. The classification procedure was repeated four times by considering η_{ijk} moments up through second, third, fourth, and fifth order as pattern descriptors.

The lowest classification success rate was estimated at about 86 and 93% for the training set and test set, respectively, by using the η_{002}, η_{020} and η_{200} moments as descriptors for 3D patterns. These results are in agreement with the discussion in Section 4.2, where the equivalence between a 3D pattern and a specific ellipsoid was justified by using only second order moments. Thus, subtle details are not considered in the recognition process. Furthermore, due to the fact that similarity measurements are performed over canonical versions, most of the unsuccessful classifications, even for the training set, are originated by the occasional resemblance between representative ellipsoids.

It is worth noting that identical moment descriptors have been selected as components of the feature vector for the seven first steps using moments up to fourth and fifth order. Furthermore, the success classification rates on working with fourth order moments are not improved in the last three steps when fifth order moments are considered. The selected moments —η_{021}, η_{002}, η_{201}—in the first three steps are the same as those chosen when working with third order moments, where it can be stated that the inclusion of additional moments into the feature vector does not substantially improve the success rate pointed out in the third step (98% for training set and 100% for test set).

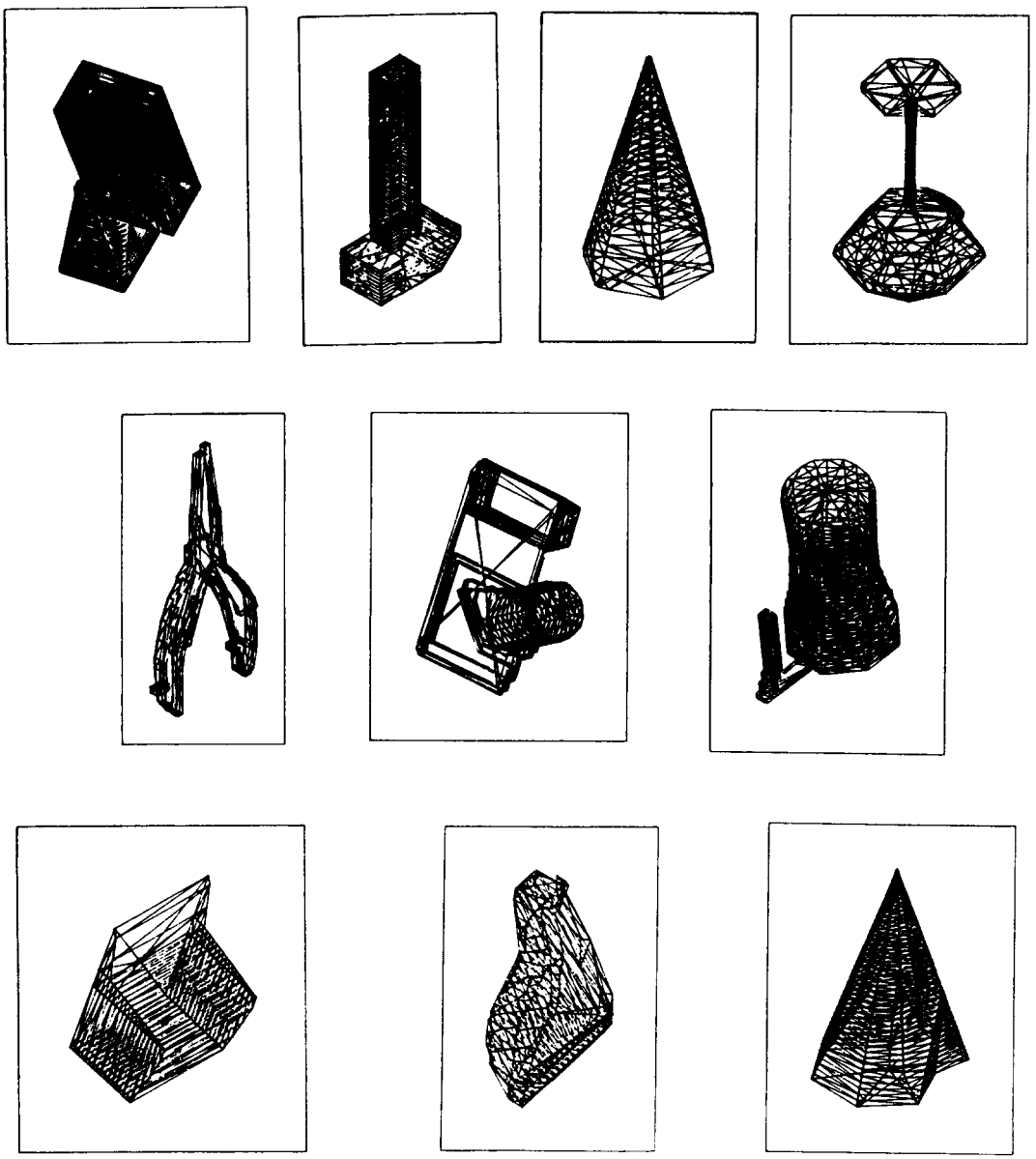


Fig. 11. Reconstructed and canonical versions of the objects in Fig. 9.

In the light of these considerations, and for the set of 3D objects shown in this paper, it seems reasonable to use only the discriminant functions provided after the execution of a few steps (three or four) of the linear discrimination algorithm. Finally, and for our particular 3D object working set, we conclude that the information carried by second and third order moments leads to a highly satisfactory percentage of classification success rates.

6. CONCLUDING REMARKS

This paper deals with the overall problem of shape reconstruction, normalization and recognition of 3D objects by means of 3D moments. Experiments have been performed on synthetic 3D data to infer the visible boundary surface from a small number of 2D views for each 3D object. The method proposed by Martin and Aggarwal, modified and made suitable for our purposes, has been tested on synthetic data. The experimental results are found to be highly satisfactory. Fine details on reconstructed objects can be inferred. Of course, the finesse in the recovered surface greatly depends on both the inter-slice separation and the number of 2D views employed, although very good results are often obtained with a few views (four or five).

Three-dimensional moments evaluated on the boundary surface of reconstructed objects have been used both to derive a position-, orientation-, and size-normalized 3D version and to make up a feature vector for 3D object recognition.

Table 1. Results of linear discrimination by a step-by-step procedure using moments up to: (a) second order; (b) third order; (c) fourth order; and (d) fifth order

Step	Moments	Success classification (%) Training set	Test set
	(a)		
1	η_{002}	55.7	72.5
2	η_{020}	74.3	95.0
3	η_{200}	85.7	92.5
	(b)		
1	η_{021}	62.9	67.5
2	η_{002}	94.3	100
3	η_{201}	98.6	100
4	η_{120}	95.7	87.5
5	η_{200}	94.3	90.0
6	η_{020}	95.7	92.5
7	η_{210}	98.6	92.5
8	η_{300}	98.6	92.5
9	η_{030}	100	92.5
	(c)		
1	η_{021}	62.9	67.5
2	η_{002}	94.3	100
3	η_{201}	98.6	100
4	η_{022}	94.3	100
5	η_{121}	94.3	95.0
6	η_{200}	95.7	95.0
7	η_{220}	98.6	95.0
8	η_{040}	98.6	95.0
9	η_{120}	98.6	95.0
10	η_{130}	97.1	95.0
	(d)		
1	η_{021}	62.9	67.5
2	η_{002}	94.3	100
3	η_{201}	98.6	100
4	η_{022}	94.3	100
5	η_{121}	94.3	95.0
6	η_{200}	95.7	95.0
7	η_{220}	98.6	95.0
8	η_{041}	98.6	95.0
9	η_{140}	98.6	95.0
10	η_{122}	98.6	95.0

The principal axes method used for object orientation does not lead to uniqueness in 3D orientation. Eight different orientations are derived from the diagonalization of the inertia matrix, which constitutes the classical normalization criterion with respect to orientation. That ambiguity situation is a very important problem in 3D object recognition. When additional restrictions are imposed to the eigenvectors of the inertia matrix, the heuristic strategy described in this paper can be used successfully in determining the standard orientation for any unknown object. However, this technique is not free of some drawbacks. Uniqueness in derivation of the rotation matrix that transforms any 3D object in its standard version cannot be guaranteed in the eigenvalue degenerate case ($\lambda_i = \lambda_j$, $i \neq j$). This situation is encountered when dealing with objects shaped like spheres, cylinders, etc. Moreover, if the centroid is outside the bounding surface, then the proposed heuristic solution may not always work.

The identification of unknown objects was achieved by a linear discrimination procedure. The similarity measure employed in this paper was constructed on a minimum-distance criterion, with the metric defined by the common covariance matrix. The elements of the feature vector—normalized 3D moments—were selected, by a step-by-step strategy, from normalized moments up to fifth order. The use of linear discriminant functions and normalized 3D moments as elements of the feature vector yielded a 3D object classification rate exceeding 95%, assuming the existence of ten different membership classes.

Finally it is worth noting that shape normalization by means of 3D moments does not limit the number of normalized moments to be extracted from the boundary surface of objects; this is not the situation with the moment invariants theory, where the invariants are restricted to a few moments.

REFERENCES

1. W. N. Martin and J. K. Aggarwal, Analyzing dynamic scenes, Laboratory for Image and Signal Analysis, University of Texas, Austin, Rep. TR-81-5 (1981).
2. W. N. Martin and J. K. Aggarwal, Volumetric descriptions of objects from multiple views, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-5**, 150–158 (1983).
3. J. M. Gálvez, Reconstrucción de Objetos a Partir de Vistas Bidimensionales y su Reconocimiento Mediante Momentos 3D, Ph.D. Thesis, Universidad de Las Palmas de Gran Canaria, Canary Islands, Spain (1990).
4. M. K. Hu, Visual pattern recognition by moment invariants, *IRE Trans. Inf. Theory* **IT-8**, 179–187 (1962).
5. F. W. Smith and M. H. Wright, Automatic ship photo interpretation by the method of moments, *IEEE Trans. Comput.* **C-20**, 1089–1095 (1971).
6. S. A. Dudani, K. J. Breeding and R. McGhee, Aircraft identification by moment invariants, *IEEE Trans. Comput.* **C-26**, 39–45 (1977).
7. S. Maitra, Moment invariants, *Proc. IEEE* **67**, 697–699 (1979).
8. S. S. Reddi, Radial and angular moment invariants for image identification, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-3**, 240–242 (1981).
9. M. R. Teague, Image analysis via the general theory of moments, *J. Opt. Soc. Am.* **70**, 920–930 (1980).
10. M. Cantón, La Teoría de Momentos en el Análisis de Formas: Formulación Tensorial, Invarianzas y Aplicaciones, Ph.D. Thesis, Universidad de La Laguna, Canary Islands, Spain (1982).
11. J. F. Boyce and W. J. Hossack, Moment invariants for pattern recognition, *Pattern Recognition Lett.* **1**, 451–456 (1983).
12. Y. Abu-Mostafa and D. Psaltis, Recognitive aspects of moment invariants, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-6**, 698–706 (1984).
13. Y. Abu-Mostafa and D. Psaltis, Image normalization by complex moments, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-7**, 46–55 (1985).
14. R. C. Luo and H. H. Loh, Tactile array sensor for object identification using complex moments, *J. Robotic Syst.* **5**, 1–12 (1988).
15. C. Teh and R. T. Chin, On image analysis by the methods of moments, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-10**, 496–513 (1988).
16. A. Abo-Zaid, O. R. Hinton and E. Horne, About moment normalization and complex moment descriptors, *Lecture Notes in Computer Science, Pattern Recognition*, J. Kitter, ed., pp. 399–409. Springer, New York (1988).

17. F. A. Sadjadi and E. L. Hall, Three-dimensional moment invariants, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-2**, 127–136 (1980).
18. H. N. Christiansen and T. Sederberg, Conversion of complex contour line definition into polygonal element mosaics, *Comput. Graphics* **12**, 187–192 (1978).
19. A. P. Reeves and B. S. Wittner, Shape analysis of three dimensional objects using the method of moments, *Proc. IEEE Computer Vision and Pattern Recognition*, pp. 20–26 (1983).
20. M. Cantón and R. Moreno Díaz, Normalized moments for 3D characterization of visual patterns, Int. Conf. on System Research Informatics, Baden-Baden, F.R.G. (1984).
21. G. Arfken, *Mathematical Methods for Physicists*. Academic Press, New York (1970).
22. T. L. Faber and E. M. Stokely, Orientation of 3-D structures in medical images, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-10**, 626–633 (1988).
23. J. M. Gálvez, M. Cantón, E. Rubio and F. Rubio, The use of 3D moments for three-dimensional object orientation, *Advances in Support Systems Research*, G. E. Lasker and R. R. Hough, eds, pp. 376–379. The International Institute for Advanced Studies in Systems Research and Cybernetics, Canada (1990).
24. J. Romeder, *Methodes et Programmes d'Analyse Discriminante*. Dunod, Paris (1973).

About the Author—JOSE M. GALVEZ LAMOLDA was born in Granada, Spain, in 1959. He received the B.S. and the M.S. degrees in physics from the University of Granada in 1981 and 1982, respectively, and the Ph.D. in computer science from Las Palmas de Gran Canaria University, Canary Islands, in 1990. After spending five years as Assistant Professor at La Laguna University, he joined the Computer Science Department at Las Palmas de Gran Canaria University, where he was Assistant Professor from 1988 to 1990. He is currently Associate Professor of Computer Science at La Laguna University. His research interests include computer vision and pattern recognition.

About the Author—MANUEL CANTON GARBIN was born in Almeria, Spain, on 26 September 1955. He received his B.S. degree in physics (electronics) from The University of Granada in 1979, and his M.S. and Ph.D. in 1980 and 1982, respectively, from La Laguna University. He is currently Professor in the Applied Physics Department at Las Palmas de Gran Canaria University, Canary Islands, and Director of the Physics Department. His research interests include computer vision and satellite image processing. He is a consultant of the ESA (European Space Agency) for the oceanographic applications of the ERS-1 satellite.