Pattern Recognition, Vol. 25, No. 10, pp. 1141–1156, 1992
Printed in Great Britain

0031-3203/92 $5.00 + .00
Pergamon Press Ltd
Pattern Recognition Society

SEGMENTATION OF RANGE IMAGES

JAYANTA MUKHERJEE,† P. P. DAS‡ and B. N. CHATTERJI†

† Department of Electronics and Electrical Communication Engineering, I.I.T., Kharagpur 721302, India
‡ Department of Computer Science and Engineering, I.I.T., Kharagpur 721302, India

(*Received* 30 *November* 1990; *in revised form* 11 *December* 1991; *received for publication* 9 *January* 1992)

Abstract—The analyses of three-dimensional (3D) scenes from range data need the segmentation of 3D surfaces into planar patches and quadratic surface regions. In this paper the concept of the Digital Neighbourhood Planes introduced earlier (*Pattern Recognition Lett.* **11**(3), 215–223 (1990)) for 3D binary data is suitably extended for the segmentation of range images. For range images the local neighbourhood of every point is virtually exploded (to $3 \times 3 \times 5$, $3 \times 3 \times 7$, etc.) and Neighbourhood Plane Set (NPS) values, indicative of the orientation of the surface normal at every range pixel, are computed. Subsequently a region growing technique is adopted to cluster points based on the NPS values which, with suitable post-processing, result in the final segments. The algorithm is simple and computationally efficient as it uses set-theoretic operations only. The algorithm is illustrated with the help of several examples. For most of them it produces good segmentation results. Finally, the algorithm has enough potential for parallelization which can be explored in the future.

Segmentation Range image Region growing technique Digital Neighbourhood Plane Neighbourhood Mapping Functions

1. INTRODUCTION

Segmentation can be defined as a process of fragmenting a given image into a set of meaningful components. Since range data is typically huge in volume and $2\frac{1}{2}$D in nature it is rather inconvenient to directly interpret the raw data for image understanding purposes. Hence segmentation in range images has gained an important dimension as it helps to partition the image array into low-level entities or surface patches which bear close similarity to higher-level entities commonly maintained in the knowledge base.[1] Every surface patch (or segment) is a set of pixels (points) which has some uniform property quite distinctive from the property of the other patches. It is often impossible to define this property from a strict mathematical context because the requirement of segmentation is frequently guided by the description/recognition stage or by the visual understanding of a human being. Various factors can contribute to the segmentation property like surface-curvature, physical proximity, orientation, moments, etc. Most of the attempts[2,3] for segmentation of depth maps are restricted to various classes of curve/surface fitting which exploit the curvature/orientation and position information extensively. These algorithms are often very computation intensive and not amenable to easy parallelization.

In this paper we present a new segmentation algorithm, called RIS, for range images. It is based on the concepts of 3D digital geometry. Here we exploit the notion of Digital Neighbourhood Planes (DNPs), introduced earlier[4] for segmenting 3D surfaces from 3D binary array form. With the use of Neighbourhood Mapping Functions (NMFs) and an extended version of DNP we have designed an algorithm for incorporating the 3D geometric constraints of segmentation in the setting of $2\frac{1}{2}$D information. The new algorithm has the advantage of being simple in design and efficient in execution. Its computation requirements are restricted to set-theoretic operations only. It has been coded in PASCAL as well as in C on the HP-9000/350 system. Experiments with different range images containing polyhedral as well as spherical, conical, cylindrical and other curved objects and having varying degrees of scene complexity show that RIS can produce good segmentation results in general. Finally RIS has enough inherent parallelism which can be easily exploited in an array processor.

In an attempt to compare and contrast, in Section 2 we present a brief survey of the algorithms for segmentation of range images. In Section 3, we present an overview of the concept of DNPs and the outline of a previous algorithm from reference (4). The segmentation algorithm RIS is discussed in Section 4. The extraction of segment information is given in Section 5. Section 6 highlights the experimental results and we conclude in Section 7.

2. RANGE IMAGE SEGMENTATION—A SURVEY

A commonly used definition of image segmentation[5] states that if I is the set of all image pixels and $P(\cdot)$ is a uniformity predicate defined on groups of connected pixels, a segmentation of I is a partitioning set of connected subsets or image regions $(R_1, R_2, \ldots, R_N)$ such that $\bigcup_{i=1}^{N} R_i = I$ where $R_p \cap R_m = \{\ \}$; $\forall\ p, m$;

$p \neq m$ the uniformity predicate $P(R_p) = \text{TRUE}$ for all regions .and $P(R_p \cap R_m) = \text{FALSE}$ whenever R_p is adjacent to R_m. Different segmentation algorithms may be viewed as implementation of different uniformity predicates.

The above definition is general enough to encompass most of the range image segmentation definition. The actual algorithms, however, vary owing to the differences in the definition of $P(\cdot)$ and the extraction methodology.

However, different approaches for surface segmentation can be broadly classified in the following categories:

(A) Region growing based on function approximation and local neighbourhood property.
(B) Edge based segmentation techniques.
(C) Clustering method.

We will now elaborate on various techniques based on this categorization.

(A) *Region growing based on function approximation and local neighbourhood property*

In this technique usually, a region is grown by fitting a surface function (planar or quadratic) over the region so long as the fit error is in the allowable range. In many cases, a preliminary segmentation of range data is obtained by computing the local neighbourhood properties at every pixel and subsequently clustering the pixels of similar properties. In the later stages, regions are grown by iteratively merging the adjacent clusters using additional surface fitting criteria or cluster attributes.

In reference (6) Milgrim and Bjorklund have fitted a plane for each range pixel in its surrounding 5×5 window by the least squares method and the information about the normal vectors and the planar fit error are used to form connected components of pixels satisfying planarity constraints.

Henderson[7] has developed a sequential region growing algorithm (via a planarity test) which creates convex planar faces from initial sets of three close non-collinear points.

Hebert and Ponce[8] suggested a method of segmenting depth maps into planar, cylindrical and conical primitives from the distribution of surface normals in the Gaussian sphere.

A similar segmentation technique has been described by Han *et al.*[9] where specified surfaces like planar, cylindrical and spherical regions are extracted in a sequence via the computation of surface normals. In this method planar regions are extracted by histogram analysis of normal values, the cylindrical regions are determined next by first computing the axis information and then by a projection to recognize the resultant circle data. Finally spherical regions are extracted by estimating possible centre points for each pair of surface points and deciding the majority value of the centre.

Sethi and Jayaramamurthy[10] handled spheres and ellipsoids in addition to planes, cylinders and cones by classifying surfaces using characteristic contours from its needle maps.

In the work of Oshima and Shirai[11] the points (range pixels) are grouped into small planar surface elements, which are merged to form planar or curved (spherical, cylindrical and conical) elementary regions. These are merged again into consistent global regions that are fitted with quadratic surfaces.

Faugeras and Hebert[12] isolated planar and general quadratic patches in range data by identifying internal boundaries and merging adjacent regions of small fit error within the internal boundaries.

Besl and Jain[1] have developed a segmentation algorithm based on variable order surface fitting. In their method any piecewise smooth digital surface is partitioned into smooth surface primitives of eight fundamental surface types. This lower level segmented description is then passed through a higher level accurate surface approximation stage by an iterative region growing method based on variable-order-surface-fitting.

Taylor *et al.*[13] have used a split-and-merge segmentation approach, where homogeneity criteria are based on a 3-parameter surface description technique. These three parameters are the two spherical angles indicating the directions of surface normals and range value at that point.

Abdelmalek[14] has carried out an algebraic error analysis prior to classifying each range pixel by sign of curvatures into three fundamental surface types. The error analysis automatically determines threshold levels for calculating the signs of curvatures.

(B) *Edge based segmentation techniques*

The other approach for segmentation of range data is to obtain the segments from the edge description of the range images. These approaches assume that the range and range gradient are continuous for all pixels in region interiors, where the region boundaries are computed explicitly.

Inokuchi *et al.*[15] presented an edge-region segmentation ring operator for depth maps which identifies different types of edges (step edge, convex roof edge, concave roof edge) and surface planar regions.

Mitiche and Aggarwal[16] have proposed an edge extraction procedure using a probabilistic model that attempts to account for range measurement errors. It uses a hypothesize and test strategy with planar fitting and Bayesian discriminator. This method can handle a large amount of noise.

Tomita and Kanade[17] segmented range image into edges and surfaces. The range image is first segmented into gap (step) edges and corner (roof) edges which are linked to obtain continuous edges. Surfaces are defined by their closed boundaries. A plane is fitted to the boundary points, and the surface is classified as either planar or curved depending on the error in the planar fit.

Bolles and Horaud[18] also analysed roof edges (convex and concave discontinuities) and jump edges in their 3DPO (3D Part Orientation) system. After the individual edge points have been detected, classified and filtered, they are linked together to form edge chains (edges). The edges are then partitioned into planar subedges which are further partitioned into circular arcs and straight lines.

Herman[19] used edges which may be occluding, convex or concave of the range images to obtain polyhedral object description.

Fan *et al.*[20] initially extracted edge-points using zero crossings and extreme of curvature along a given direction which form the boundaries of the surface patches. The boundaries are classified as jump boundaries, folds and ridge lines. The jump boundaries and folds are used to segment the surfaces into surface patches and a simple surface is fitted to each patch to reconstruct the original object.

Yokoya and Levine[21] used a combination of both region- and edge-based techniques for segmentation of range images. Segmentation is carried out in two stages. In the first stage, a region-based segmentation in the form of curvature sign map and two edge-based segmentations embodying jump and roof-edge maps are computed. In the next stage these initial segmentation maps are combined to produce the final segmentation.

(C) *Clustering method*

The clustering techniques for segmentation, commonly used for intensity images, are hardly used for range images. However, there have been some attempts in that direction too.

Hoffman and Jain[22] used clustering by converting each pixel into a 6D feature vector and thus obtained segmentation of range images.

Rimey and Cohen[23] used the maximum likelihood clustering technique for segmenting range images. The basic approach to segmentation is to divide the range image into windows, classify each window as a particular surface primitive, and group like windows into surface regions.

3. SEGMENTATION OF 3D SURFACES (BINARY FORM)

In this section we briefly review the concept of Digital Neighbourhood Planes (DNPs) introduced earlier[4] for segmenting 3D surfaces given in a 3D binary array form. Any reader acquainted with reference (4) may skip this section.

Throughout this paper single points are denoted by lower case letters and sets of points by upper case letters.

Let $Z^3 = \{x = (x_1, x_2, x_3): x_1, x_2, x_3 \text{ are integers}\}$ be the 3D digital space. A point $p \in Z^3$ can have 26

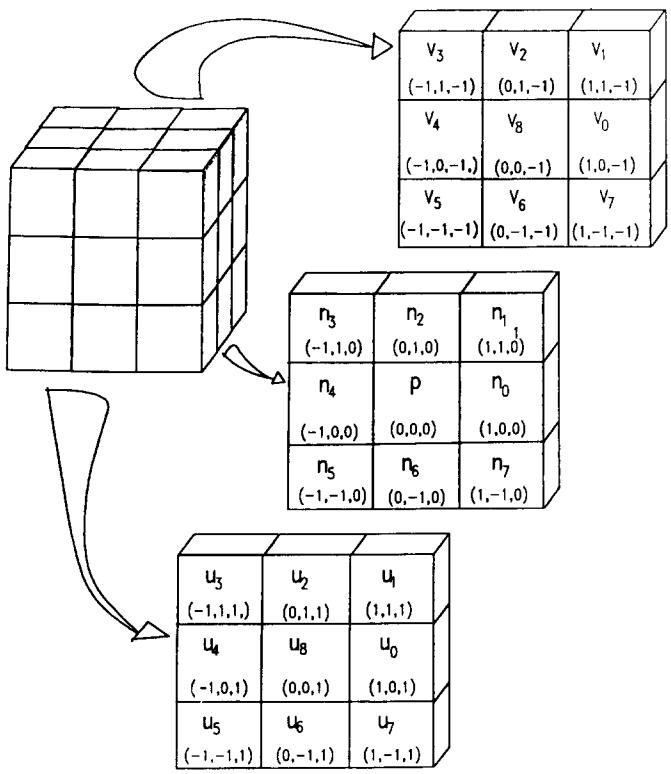


Fig. 1. Neighbourhood of "p" in 3D.

different neighbours, as shown in Fig. 1, in its $3 \times 3 \times 3$ neighbourhood $N(p) = \{q: |p_i - q_i| < 1, 1 \le i \le 3\}$. These points can be expressed by their logical names or the respective coordinate positions with respect to "p", as shown in Fig. 1, i.e.

$$N(p) = \bigcup_{i=0}^{8} (\{u_i\} \cup \{v_i\} \cup \{n_i\}), \quad n_8 = p.$$

The neighbours are, however, classified into three categories, 6, 18, 26 neighbours, according to their relative coordinates as follows.

A point $y \in Z^3$ is a 6 (18, 26) neighbour of x if x and y differ in at most one (two, three) coordinates by unity.

If $x, y \in Z^3$ and $y \in N(x)$ then we also write it as a binary relation xNy. Clearly relation N is symmetric but is not transitive. In an image $A \subset Z^3$, two points $x, y \in A$ are said to be h-connected,[4] $h = 6$, 18 or 26, if there is a sequence of points $u_0 = x, u_1, u_2, \ldots, u_t = y$, such that $u_i \in A$, $0 \le i \le t$ and (u_i, u_{i+1}) are h-neighbours, $0 \le i \le t-1$. The image as a whole is said to be h-connected[4] if for all $x, y \in A$, x is h-connected to y. Two images A and B are called adjacent if there is a point x in A and a point y in B such that xNy.

For xNy the unit vector along the straight line xy in the continuous space is called the gradient from x to y and denoted by $\overrightarrow{xy}$. There are thus, 26 different gradients in 3D. Using the continuity of identical gradient a (digital) straight line from x to y in Z^3 is defined as a sequence of points $u_0 = x, u_1, u_2, \ldots, u_t = y$, such that $u_i N u_{i+1}$, $0 \le i \le t-1$ and $\overrightarrow{u_{i-1}u_i} = \overrightarrow{u_i u_{i+1}} = \vec{m}$ (constant), $1 \le i \le t-1$.

Corresponding to 26 gradients, we count 13 distinct straight lines in 3D by assuming $\vec{m}$ and $-\vec{m}$ to refer to the same straight line. The gradient of a straight line L is $\vec{m}$ and that of straight line Q, perpendicular to L is $\vec{n}$ where $\vec{m} \cdot \vec{n} = 0$ and "$\cdot$" denotes the dot product of vector algebra.

A plane P is a connected set of points such that every straight line in P with gradient $\vec{m}$ is perpendicular to a given gradient $\vec{n}$ (i.e. $\vec{m} \cdot \vec{n} = 0$), the normal gradient of P. For $x \in P$, any straight line of gradient $\vec{n}$ containing x is called a normal to P at x.

For a $3 \times 3 \times 3$ digital cube at "p", there are nine planes containing P. They have different orientations and their normals through P are the 18-connected straight lines in the cube. These planes are termed as the DNPs of "p" and are illustrated in Fig. 2. The corresponding set of points are shown in Fig. 3. These can be easily expressed in the logical names of points in $N(p)$. The neighbouring condition of p will determine in which of these planes p will lie.

Let P_i be the set of points assigned to the ith neighbouring plane as described in Fig. 3. Now given an image A in Z^3, the Neighbourhood Plane Set (NPS) of point $p \in A$ is defined as: $[p]_k = \{i: |N(p) \cap A \cap P_i| > k\}$, $k \ge 3$, where $|\cdot|$ denotes the cardinality of a finite set. The value of k usually lies between 3 and 5. Here $[p]_k$ will be denoted by $[p]$ only since k will remain constant. So $[p]$ actually lists the serial numbers of those neighbourhood planes which have a sufficient (thresholded by k) number of points around p in the image, to be considered a candidate for a local neighbouring plane feature.

The neighbourhood plane set for a finite set of points $M \subset A$ is defined as

$$\text{If } M = \{m_1, m_2, \ldots, m_n\} \text{ then}$$
$$[M] = [m_1] \cap [m_2] \cap \cdots \cap [m_n].$$

In reference (4), $[x]$ is computed for all points x in the image. A seed point p is then selected from which a segment S is grown according to the following definition:

(i) S is 26-connected;

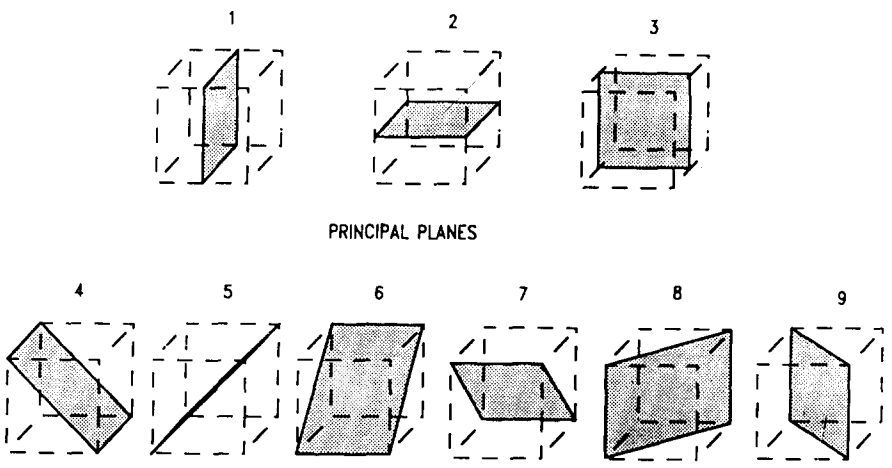


Fig. 2. Neighbourhood planes of "p".

(ii) $[S] \neq \{\ \}$;

(iii) if $x \in S$ and $z \in N(x)$ then $[z] \cap [S] = [S]$ implies $z \in S$.

As the segment definition does not imply disjoint segments it needs to be enforced directly and thus the choice of a seed is required. In the next section we present the necessary modifications in DNPs for the algorithm to work for range data. The segment definition is also refined to guarantee disjoint segments as well as better performance.

4. SEGMENTATION OF RANGE IMAGES

The segmentation algorithm developed for 3D binary images in reference (4) can be directly applied to a range image. Since a range image is represented in the form of a 2D array of depth values at the corresponding pixel locations (in raster form), range pixel $r(i,j)$ (where i denotes the row number and j the column number) can be represented by a voxel in a 3D binary array at $(i,j,\lfloor r(i,j) \rfloor)$. Then the segmentation can be invoked on the 3D. Such a direct application does not yield good segmentation results. This is due to the noise present in a practical range data, which is due to the limitation of sensors, quantization errors, etc. Hence in this paper we present suitable modifications to the above segmentation algorithm to handle range data in the proper form.

4.1. *Modification in the definition of DNP*

To be able to effectively handle range images we need to impart some tolerances to the previous definitions of DNPs. A modified DNP is defined in an extended neighbourhood, which includes a greater number of points. An easy way of realizing such tolerances is to define a new mapping function from the points in any reasonably sized neighbourhood (e.g. $a \times b \times c$) around a point "p", to the set of points in a $3 \times 3 \times 3$ neighbourhood of "p" (which are assigned to the corresponding set of logical points as shown in Fig. 1). It has the advantage of keeping the set of logical points in the definition of DNPs as given in Fig. 3 unchanged and hence the threshold value (k) in the definition of NPS need not be changed. The other advantage for this type of modification is that it reduces the storage space for representing the DNPs and thus restricts the search space.

We define the following terms for specifying the modification.

(1) For a point $q \in Z^3$ we denote $L(q) =$ logical name of q, $X(q) = X$-coordinate of q, $Y(q) = Y$-coordinate of q, $Z(q) = Z$-coordinate of q.

For the range pixel $r(i,j)$ which corresponds to the depth value at the jth column of the ith row in the range data, i, j, and $\lfloor r(i,j) \rfloor$ express the X-coordinate, Y-coordinate and Z-coordinate of that point, respectively.

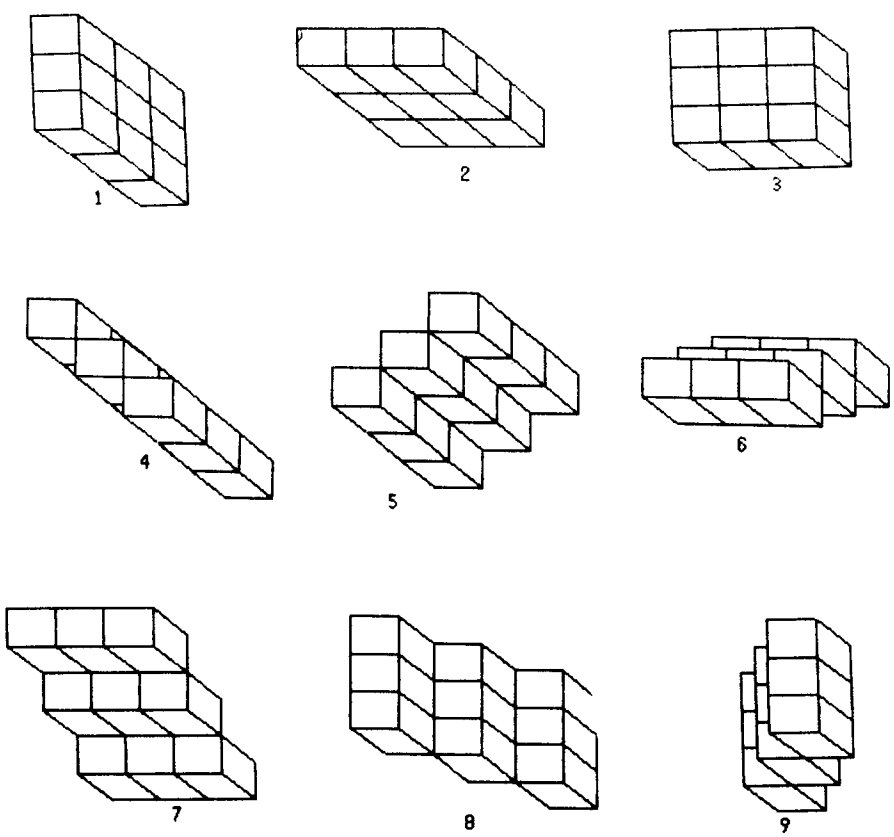


Fig. 3. Voxel diagrams of the DNPs (NMF = F_0).

(2) $N_{abc}(p)$ = extended neighbourhood of size $a \times b \times c$ around p, i.e. $\forall\ q \in N_{abc}(p)$, $|X(q) - X(p)| \leq \lfloor a/2 \rfloor$, $|Y(q) - Y(p)| \leq \lfloor b/2 \rfloor$, and $|Z(q) - Z(p)| \leq \lfloor c/2 \rfloor$ and $N(p) = N_{333}(p)$.

(3) Now, the mapping function can be defined as

$$F: N_{abc}(p) \longrightarrow N(p);$$

F is termed as a Neighbourhood Mapping Function (NMF), which should satisfy the following conditions:

(A) F is total and surjective (onto). That is $F(x)$ is defined $\forall\ x \in N_{abc}(p)$ and $\forall\ y \in N(p)$, $\exists\ x \in N_{abc}(p)$ such that $F(x) = y$.

(B) F should induce connected partitions of $N_{abc}(p)$ in the sense that $\pi_y = \{x: x \in N_{abc}(p) \text{ and } F(x) = y\}$ should be 26-connected $\forall\ y \in N(p)$. It follows from (A) that π_ys define a set of connected partitions of N_{abc}, i.e. $\bigcup_{y \in N(p)} \pi_y = N_{abc}(p)$ and $\pi_y \cap \pi_{y'} = \{\ \}$, $\forall\ y, y' \in N(p)$, $y \neq y'$.

(C) In addition to (B) induced DNPs should also be connected. That is for each P_i (ith Digital Neighbourhood Plane) $\forall\ x \in (P_i \cap N(p))$, $\bigcup_x \pi_x$ should be 26-connected where P_i is as defined in Section 3 (Fig. 3).

(D) There should be strong structural similarity (in a qualitative sense) between a modified DNP and the original one. Quantitatively it can be said that the normal to the plane through the set of points in a modified DNP should be within a tolerable limit of the normal to the original DNP. Some examples of F are:

F_0: Here $a = b = c = 3$ and $\forall\ x \in N_{333}(p)$, $F(x) = x$. This is the original neighbourhood definition (Fig. 1). The voxel diagrams of the DNPs are shown in Fig. 3.

F_1: Tolerance has been given in the mid-frame for a neighbourhood of $3 \times 3 \times 5$ around "p" as follows:

If $x \in N_{335}(p)$ and $y \in N(p)$
then if $(X(x) = X(y))$ and $(Y(x) = Y(y))$
  then if $(|Z(x) - Z(y)| \leq 1)$ and $(L(y) \in \{n_i\})$
    then $L(x) \longleftarrow L(y)$
    else if $((Z(x) - Z(y)) = -1)$ and $(L(y) \in \{v_i\})$
      then $L(x) \longleftarrow L(y)$
      else if $((Z(x) - Z(y)) = 1)$ and $(L(y) \in \{u_i\})$
        then $L(x) \longleftarrow L(y)$.

The voxel diagrams of the modified DNPs according to F_1 are shown in Fig. 4.

Similarly F_2 and F_3 are defined. In F_2 ($a = 3$, $b = 3$,

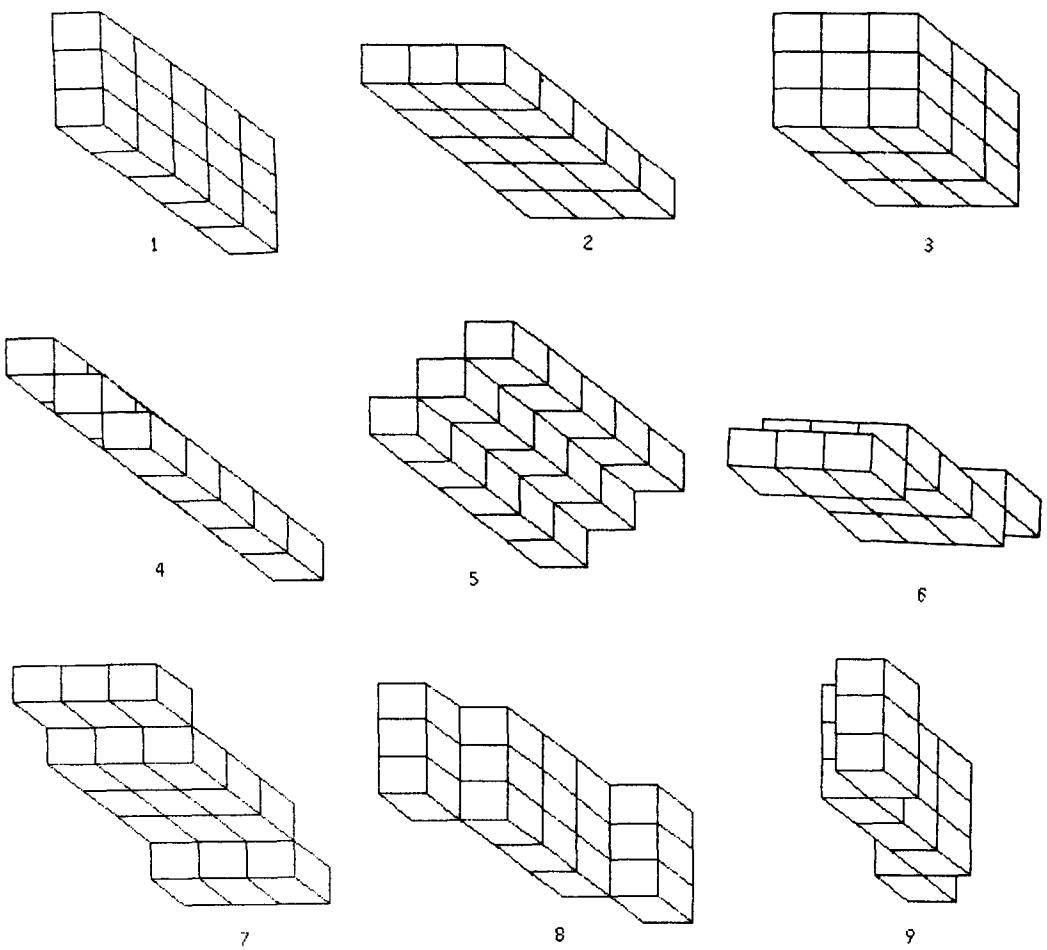


Fig. 4. Voxel diagrams of the DNPs (NMF = F_1).

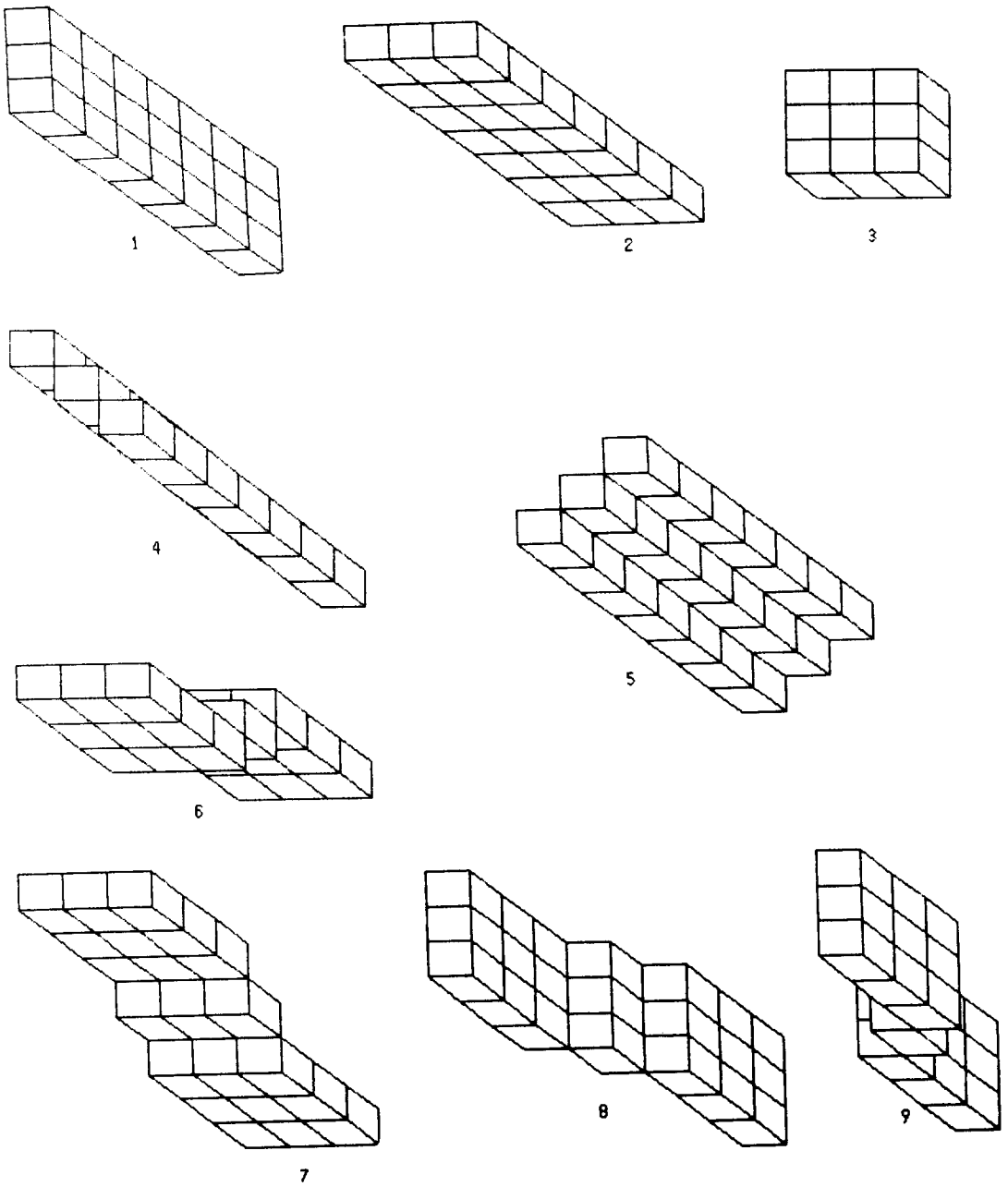


Fig. 5. Voxel diagrams of the DNPs (NMF = F_2).

$c = 7$) tolerances have been provided in front and back frames while in F_3 ($a = 3, b = 3, c = 9$) similar tolerances have been given in all the three frames, i.e. front, mid, and back frames (Fig. 1). These can be understood by the respective voxel diagrams of the DNPs for each mapping function F_2 and F_3 as given in Figs 5 and 6.

4.2. *Modification in the definition of segments*

Since the definition of DNPs are relaxed from reference (4), the definition of a segment needs to be more constrained. Otherwise, the ambiguity due to the given tolerance in the DNPs in the computation of NPS, may cause erroneous building up of large regions. Hence, the definition of the segment (S) is made more restrictive as follows:

(i) S is 26-connected;
(ii) $[S] \neq \{\ \}$;
(iii) $\forall\, x \in S$, $[x] = [S]$ (in reference (4) this condition reads as "if $x \in S$ and $z \in N(x)$ then $[z] \cap [S] = [S]$ implies $z \in S$").

It can be seen from the above definition that condition (iii) in this definition differs from the previous

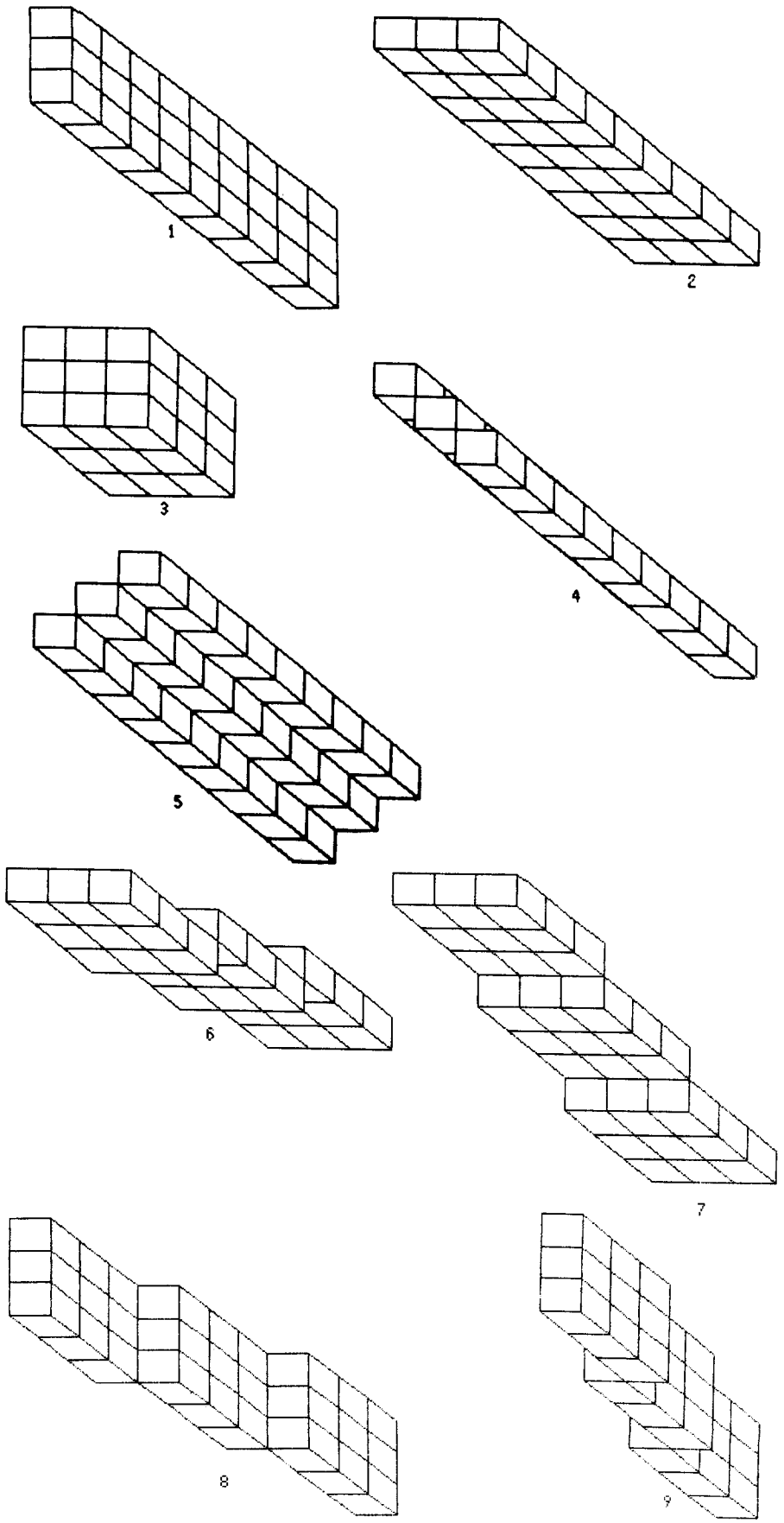


Fig. 6. Voxel diagrams of the DNPs (NMF = F_3).

one. This gives some other advantages over the previous segmentation algorithm. Under the modified definition, the set of segments of an image induces a unique partition in the image since the equality of NPS values is an equivalence relation.

As the definition of segment itself now yields a unique set of disjoint segments, there is no need of any criterion for choosing a seed point as in reference (4) for region growing. From any point of a segment, the full region (segment) can be extracted and it will not overlap with some other segments, due to the partition property. This reduces the time complexity by eliminating the search time for the proper seed point of the algorithm. Hence the algorithm can be stated as follows:

Algorithm: Range-Image-Segmentation (RIS)
Input: A range-image: R;
Output: A set of disjoint segments of R;

1. $\forall\, x \in R$, evaluate $[x]$;
2. Choose any point, $x \in R$;
3. Develop $S(x)$;
4. $R \longleftarrow R - S(x)$;
5. Repeat steps (2)–(4) until $R = \phi$;

End: Range-Image-Segmentation.

Theorem 1. The time complexity of the Range-Image-Segmentation (RIS) algorithm is $O(n)$, where $n = |R|$ = number of points in the range image, R.

Proof. The first step of the algorithm requires $O(n)$ operations for evaluating the NPS at every point of the image.

Let the set of segments produced at the end of RIS be $\{S_i\}$, $i = 1, 2, \ldots, m$. A point on the boundary (B_i) of a segment (S_i) may be checked for its inclusion to any segment, for the number of times, equal to the number of adjacent segments to that point. But, in general $|B_i| \ll |S_i|$ and also the number of adjacent segments to S_i (say a_i) is very very small with respect to the number of points in the range image, i.e. $a_i < m \ll n$. Hence, the approximate total search time for region expansion for all the segments is of the order of $\sum_{i=1}^{m} |S_i| = O(n)$. Hence the theorem. Q.E.D.

4.3. *Smoothing of the segments*

After partitioning the range image every resultant segment needs to be smoothed to obtain a continuous patch of points with definite boundary and interior. This is required for extracting the relevant information of the individual segments, such as area, perimeter, segment-adjacency lists, etc. In practice, due to the noise in the range data, the discreteness in the directions of DNP and the error occurring from tolerances to DNP, the segments are obtained as clusters of connected points. In between the cluster of points of a segment, the number of points having different NPS (including the empty set) may be found. That is why, the segments are smoothed so as to fill the gaps between successive points in a segment during a scan (if the gap is not too wide). This is done as follows:

Algorithm: Smooth-segment;
Input: A segment S_i obtained in ith iteration of step 3 of RIS.
Output: A smoothed segment (Sm_i)

0. $Sm_i \longleftarrow S_i$.
1. Scan row-wise/column-wise. Find out the gap (G) (i.e. a set of connected points not assigned to $S_j, j = 1, 2, \ldots, i$ between two successive points in that row/column belonging to S_i in the same row/column). If $(G \neq \{\ \})$ and $(|G| \leq$ Fill-thresh) then $Sm_i \leftarrow Sm_i \cup G$.
2. Repeat step 1 for every row and for every column of S_i.

End: Smooth-Segment.

In the above smoothing technique Fill-thresh is a threshold parameter (typically taken as 20% of the row size or column size of an image), which keeps a check over unnecessary filling of wide gaps.

It must be noted here that due to the incorporation of smoothing, the segmentation algorithm now depends upon the order of the selection of seed points for region growing and the smoothing of the grown region. For example, let a segment S_1 encompass a very small segment S_2 in the partitioned space. Now, if S_2 is grown before S_1, S_1 will loose a few of its points for filling the gap in S_2. On the other hand if S_1 is grown before S_2, it can fully subsume S_2 while filling its gap. The second alternative should be desirable, since it avoids unnecessary splitting in large segments. This can be enforced, by checking the validity of a grown segment S at step 3 of RIS. If the number of points in S is greater than a certain threshold parameter (called Seg-Thresh), S will be accepted and smoothed, otherwise S will be rejected and no smoothing is necessary. This makes the segmentation algorithm almost independent of the order of the choice of seed points for growing regions and makes it more tolerant to noise (or error in calculating DNPs at the points). The typical value for Seg-Thresh chosen for segmenting a range image of size 256×256 is 75 and of size 128×128 is 25.

4.4. *Effect of varying window size in region growing*

The present segmentation algorithm employs variable sized windows for region growing. So it is necessary to appreciate the effect of larger/smaller windows on the segmentation results. It is easy to see that larger windows help to avoid meaningless fragmentation of extended segments. With the use of smaller windows RIS has the potential of producing segments small enough to fail the segment validity test. On the other hand, the window size should not be too large, because different regions of the same NPS value may be merged into one (due to their relative proximity) and important segmentation information may be lost. In an attempt to achieve the proper trade-off several experiments have been conducted by varying the window size on

the same image. It is found that $5 \times 5 \times 5$ window size produces good segmentation results as judged from the correlation between the RIS output and visual understanding.

5. EXTRACTION OF THE RELEVANT INFORMATION FROM SEGMENTS

Let the boundary of smoothed segments Sm_i be denoted by Bm_i, i.e. $Bm_i = \{x: (x \in Sm_i)$ and $(\exists\ y \notin Sm_i$, such that $x N y)\}$. The following information can be obtained from Sm_i.

(1) Area: $|Sm_i|$.

(2) Centre (X_c, Y_c, Z_c): $X_c = \sum_{q \in Sm_i} (X(q)/|Sm_i|)$,

Y_c and Z_c are defined similarly.

(3) List of adjacent segments

The segments Sm_i and Sm_j are said to be adjacent iff

$$\min (|L_{ij}|/|Bm_i|,\ |L_{ji}|/|Bm_j|) \geq \text{Th-adj}$$

where $L_{ij} = \{x: (x \in Sm_i)$ and $(\exists y \in N(x)$, such that $y \in Sm_j)\}$ and $L_{ji} = \{x: (x \in Sm_j)$ and $(\exists\ y \in N(x)$, such that $y \in Sm_i)\}$ and Th-adj is the adjacency threshold with typical value 1/12.

(4) NPS value

NPS is important information for characterizing a segment. Hence its proper understanding is needed at this juncture. Essentially, the direction of the normal vector at any point of a surface has been quantized here by some predefined direction cosines and these quantized directions are coded by the planes 1–9. Hence a plane P having the normal $(0, 0, \pm 1)$ will have its NPS $= \{3\}$ and $(0, 1/\sqrt{2}, 1/\sqrt{2})$ {or $(0, -1/\sqrt{2}, -1/\sqrt{2})\}$ will have NPS $= \{6\}$. These cases are trivial (i.e. when $|[P]| = 1$) where a unique normal for a plane can be expressed by its $[P]$. But when $|[P]| > 1$, an ambiguity arises for finding out the normal to that plane. For example, for $[P] = \{3, 6\}$, two normals (by adding and subtracting normals for 3 and 6 vectorwise) can occur. But since the surface is assumed to be piecewise smooth for all practical purposes, no abrupt change in the orientation of the points is expected to occur. Hence, the set of points in a particular segment (P) having the same $[P]$ will in most cases have the same normal and by studying the variation of NPS in the vicinity of the segment, this particular direction out of the possible two normals (with respect to the example given here) can be found out. The algorithm essentially tries to exploit this contextual fact and does not attempt to calculate the exact direction cosines of the normal at any point. Rather it can effectively function in the NPS domain itself. That is the direction of a plane having NPS as $\{3, 6\}$, is somewhere in between directions $\{3\}$ and $\{6\}$. Hence one can precisely summarize the significance of NPS of a segment by assuming it as the average orientation of the pixels in that segment in its spatial context.

(5) Declaration of background segment

Let, nrow = number of rows in the image
ncol = number of columns in the image.

For a smooth segment Sm_i, BG is defined as

$$BG = \{q: (q \in Bm_i) \text{ and } ((X(q) = \text{nrow}) \text{ or } (X(q) = 1) \text{ or } (Y(q) = \text{ncol}) \text{ or } (Y(q) = 1))\}.$$

If $|BG| \geq$ Thresh-back then Sm_i is declared as the background segment. Thresh-back is a threshold parameter for declaring a segment as a background segment and it is typically taken as $0.25 * (\text{nrow} + \text{ncol})$.

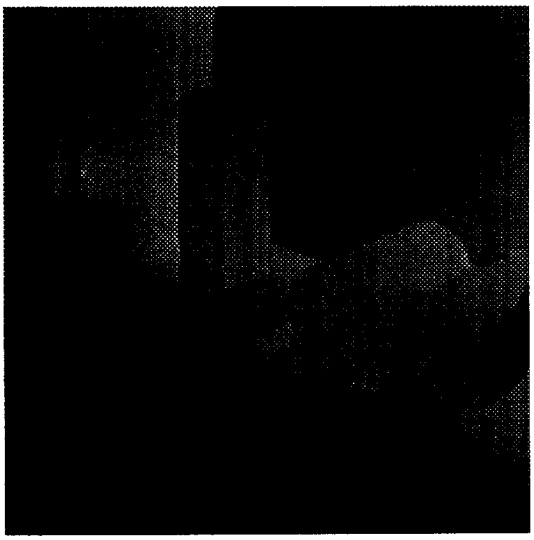
6. RESULTS

Experimental results are illustrated here for six range images. Four of them of size 256×256, have been obtained from the PRIMULA range database of Osaka University, Japan.[24] These are referred to as Quad (Fig. 7(a)), Plant (Fig. 7(b)), Poly (Fig. 7(c)) and Desktop (Fig. 7(d)), respectively. The other two are of size 128×128 and have been obtained from Environmental Research Institute of Michigan (ERIM), U.S.A.[25] They are referred to as Cup (Fig. 7(e)) and Ring (Fig. 7(f)). It may be noted that PRIMULA data have a dynamic range of -50 to 500 (approx.) while ERIM data have a range of 0 to 255. Both are presented here in 32 gray levels. In the segmentation output of an image each segment is shown by its boundary and marked with a unique integer to distinguish it from other segments.

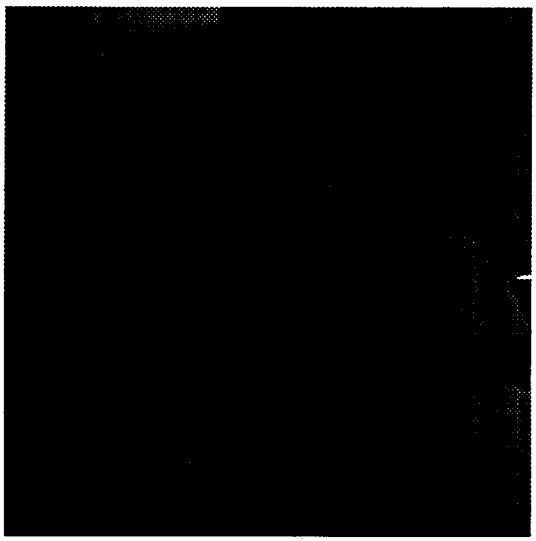
For the image "Poly" the segmentation outputs using F_0, F_1, F_2 and F_3 neighbourhood mapping functions are shown in Figs 8(a)–(d), where the window size is $5 \times 5 \times 5$. It can be seen that when NMF is F_1 and F_3 segmentation outputs are better. But since F_3 allows more tolerance in the DNPs than F_1, F_3 tends to form larger segments by merging those adjacent segments which do not differ much in the direction of the normals. This can be seen in Fig. 8(d) of "Poly". In a rule based object-description scheme reported elsewhere,[26] where this segmentation algorithm has been used, the boundary surfaces of an object should preferably be fragmented into a greater number of segments which can suitably capture the variations in the direction of the normals to that surface at different points. With this end use in mind, F_1 has been given preference over F_3 in the segmentation scheme.

The effect of variation of window size for region growing has been studied with various images. It is found that a window size of $5 \times 5 \times 5$ produces the best segmentation output. An example of such a comparative study is given in Figs 8(b), 9(a) and (b). The segmentation outputs for other images such as Quad, Plant, Desktop, Cup and Ring are illustrated in Figs 10(a)–(e), respectively (where NMF is F_1 and window size is $5 \times 5 \times 5$). An example of relevant information obtained after the early processing of range image is given in Table 1.

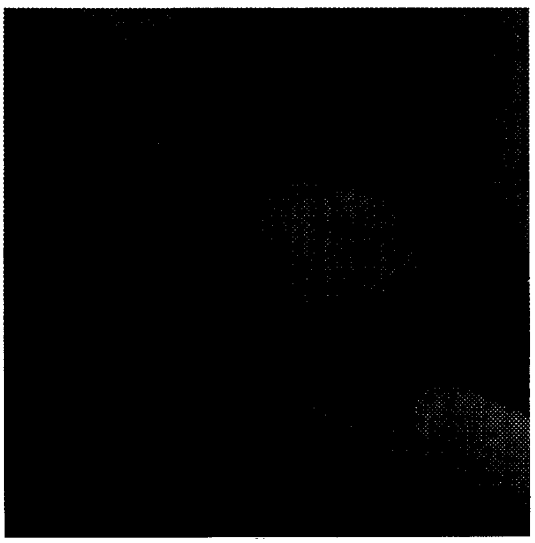
(a)



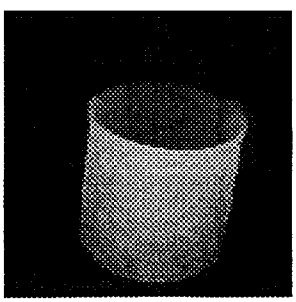
(b)



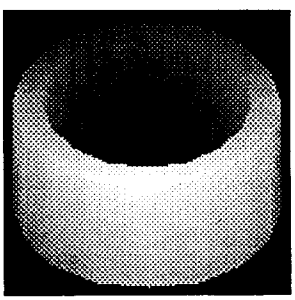
(c)



(d)

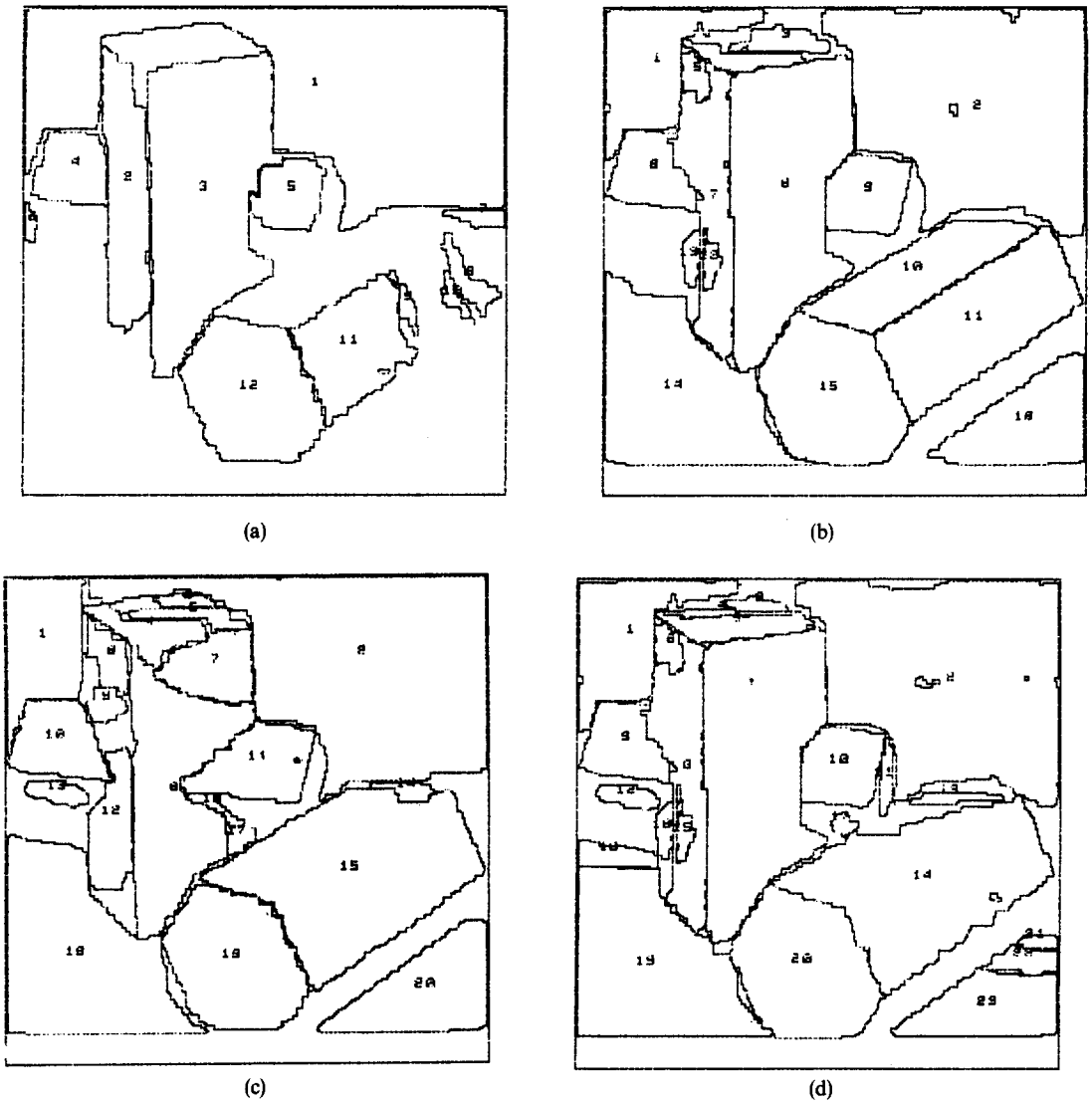


(e)



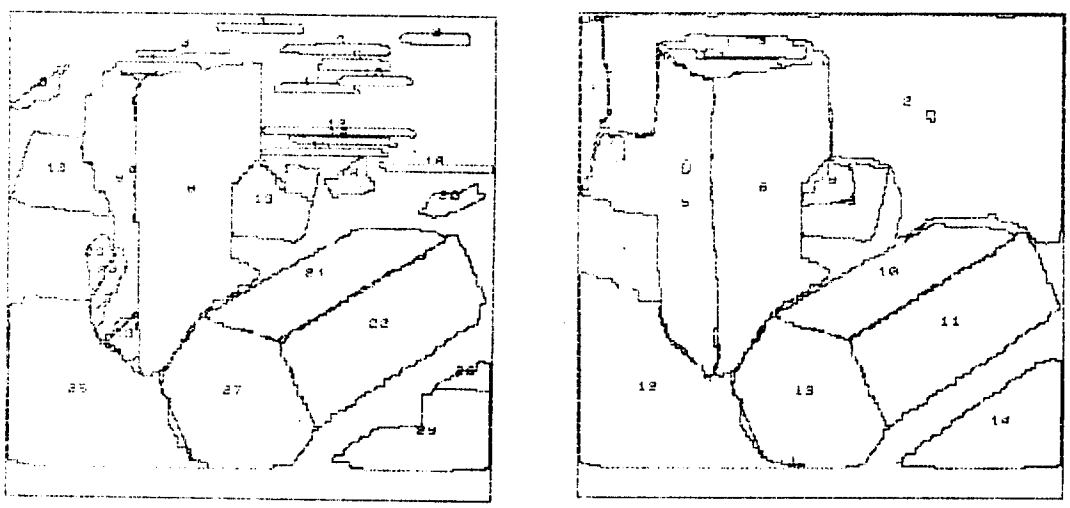
(f)

Fig. 7. Range images: (a) Quad; (b) Plant; (c) Poly; (d) Desktop; (e) Cup; (f) Ring.



(a) (b) (c) (d)

Fig. 8. Segmentation outputs of "Poly" (Fig. 7(c)) for NMFs: (a) F_0; (b) F_1; (c) F_2; (d) F_3 (window size = $5 \times 5 \times 5$).



(a) (b)

Fig. 9. Segmentation outputs of "Poly" (Fig. 7(c)) for window size: (a) $3 \times 3 \times 3$; (b) $7 \times 7 \times 7$ (NMF = F_1).

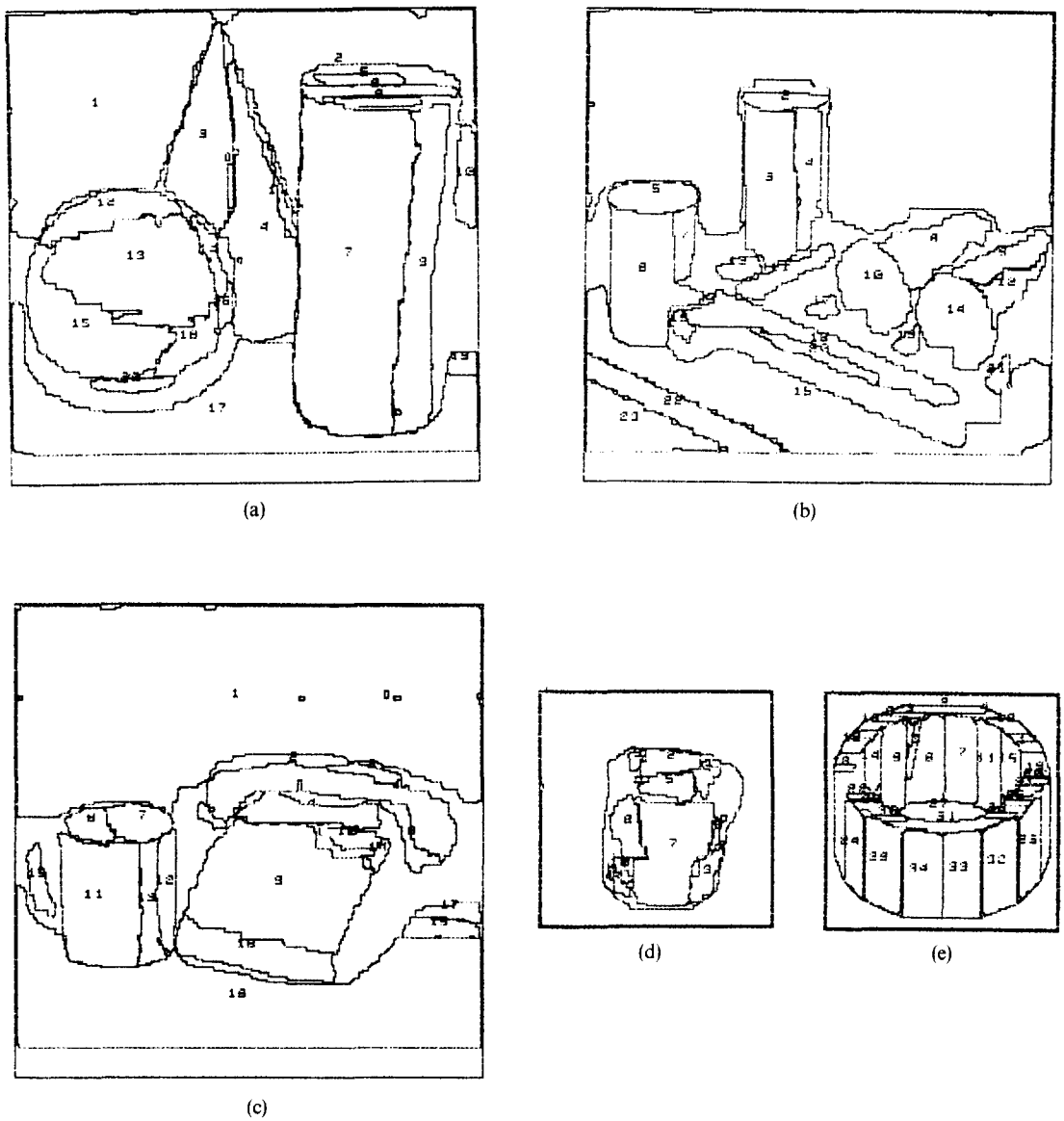


Fig. 10. Segmentation outputs of the range images: (a) Quad; (b) Plant; (c) Desktop; (d) Cup; (e) Ring (NMF = F_1; window size = 5 × 5 × 5).

7. DISCUSSIONS

In this paper we have presented a new algorithm for segmenting range images. The algorithm uses relevant results of 3D digital geometry and adapts a similar algorithm working with 3D binary data for depth maps. Analysing the segmentation output we make the following observations about RIS:

(1) RIS performs creditably with polyhedral objects (Figs 7(c) and 8(b)) and also with curved objects if the objects are not small (Figs 7(a) and 10(a)). In these cases the segmentation produced by RIS is better than those obtained by Hoffman and Jain[22] and Taylor *et al.*[13] and is comparable to the ones in Besl and Jain.[1] The point to note, however, is that RIS and other algorithms [1,11–13,22] work at different levels for segmentation. Most attempts to segment a range image involve two stages—first, primary level segments (typically tiny) are produced and in the second stage these primary segments are merged to form bigger segments using additional surface fitting or some such criterion based on segment attributes. It is important to note that RIS does not employ any such segment merging at all and thereby saves precious computation time. However, RIS has to compete in its quality of output with the likes of reference (1) which employs both stages. If the segments of RIS are compared with that of the primary level segments of other major schemes, RIS can outperform many of them.

(2) RIS fails to do well for small curved objects like the ERIM cup (Figs 7(e) and 10(d)) as compared to reference (1) because of the absence of the merging

Table 1. Relevant segment information obtained after the segmentation of Quad (Figs 7(a) and 10(a))

Segment number	List of adjacent segments	NPS	Area	Centre (x, y, z)			Back-ground? (Y/N)
1	3	38	10,203	53	46	142	Y
2	7, 10, 11	38	5938	30	180	295	Y
3	1, 4	3	2635	71	105	104	N
4	3, 11	38	3684	122	138	125	N
5	6	36	332	38	194	94	N
6	5, 7, 8	36	614	44	200	83	N
7	2, 6, 8, 9, 17	3	10,552	136	184	52	N
8	6, 7, 9	36	297	50	202	68	N
9	7, 8	38	3306	142	225	87	N
10	2	38	612	93	251	383	N
11	2, 4	8	274	102	148	155	N
12	13, 15	36	1227	108	54	−31	N
13	12, 14, 15, 18	368	4268	137	70	−39	N
14	13, 16	38	374	134	110	11	N
15	12, 13, 18, 20	3	2990	173	40	−73	N
16	14, 18	8	192	162	117	13	N
17	7, 19	36	8147	220	114	4	Y
18	13, 15, 16, 20	38	773	180	99	−18	N
19	17	36	229	193	248	182	N
20	15, 18	37	342	203	67	−110	N

stage mentioned above. This is due to the fast changing surface curvature of Cup (Fig. 7(e)) which can be captured effectively in meaningful segments only if a post-processing is performed. It, however, produced satisfactory curved surface segments in other cases (Figs 7(a) and 10(a)).

(3) Though we appreciate the significance of segment merging we still have preferred to omit the one in RIS primarily due to two reasons. First, merging often takes an excessive amount of time (e.g. if variable order surface fitting is used) to produce good results. Though in many cases RIS, can, even in the absence of merging, produce equivalent outputs. So we save in terms of time. Second, in many cases, the stages following segmentation in the recognition/description system have a bias of their own. Many of them will find it more convenient to use a low-level segmentation data with surface attributes than to use a higher-level segmentation data which has more rigidly formed regions. One such attempt in this regard can be found in reference (26).

Finally the distinctive features of this range image segmentation algorithm (RIS) can be summarized as follows:

(1) The RIS algorithm is computationally more efficient in comparison to the existing range image segmentation algorithms[2,3] which normally attempts either to fit the surface points to a given surface function or to calculate a local property (such as normals, curvatures) for growing the regions. The present approach is a set theoretic one in contrast to the analytical geometric approaches of the previous segmentation algorithms.

(2) RIS enjoys the flexibility of redefining a new NMF according to the requirements of a description/recognition scheme. For example, F_3 may be suitable for a description scheme, which may attempt to approximate the segments from the segmentation output of a range image with a global function in its later stages of understanding. For example, in this case the segmentation output of Cup in Fig. 11 is more suitable than the previous one (Fig. 10(d)).

(3) The segment characterization by NPS value is crude but simple and elegant in use. However, more stringent characterization of the segments may be needed in the later stages of description which has been suitably incorporated through a set of preprocessing rules used in reference (26).

(4) Last, but not least, RIS supports enough inherent parallelism in its structure. In RIS most operations are performed locally or through a local to global propagation. The similarity of the operations in RIS and the standard morphological operations is also

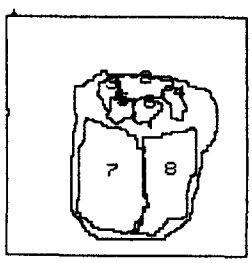


Fig. 11. Segmentation output of Cup (Fig. 7(e)) (NMF = F_3; window size = 5 × 5 × 5).

notable. For example, the NPS computation is necessarily a sort of 3D template matching which can be implemented by morphological operators with suitable modifications. Keeping this possible parallelization of the latter in view, RIS offers good parallelizability particularly in array processors or in special-purpose mesh-connected SIMD machines.

Acknowledgements—The authors are grateful to Prof. K. Sato of Osaka University, Japan, for providing the range data from the PRIMULA range database, and to the anonymous referee for pointing out the deficiencies of an earlier draft of the paper.

REFERENCES

1. P. J. Besl and R. C. Jain, Segmentation through variable-order surface fitting, *IEEE Trans. Pattern Analysis Mach. Intell.* **10**(2), 167–192 (1988).
2. P. J. Besl and R. C. Jain, Three-dimensional object recognition, *ACM Comput. Surv.* **17**(1), 75–145 (1985).
3. P. J. Besl, *Surface in Range Image Understanding*. Springer, New York (1988).
4. J. Mukherjee, P. P. Das and B. N. Chatterji, Segmentation of three-dimensional surfaces, *Pattern Recognition Lett.* **11**(3), 215–223 (1990).
5. S. L. Horowitz and T. Pavlidis, Picture segmentation by a directed split-and-merge procedure, *Proc. 2nd Int. Jt Conf. Pattern Recognition*, pp. 424–433 (1974).
6. D. L. Milgrim and C. M. Bjorklund, Range image processing planar surface extraction, *Proc. 5th Int. Conf. Pattern Recognition*, Miami, Florida, 1–4 December, pp. 912–919 (1980).
7. T. C. Henderson, Efficient 3-D object representation for industrial vision systems, *IEEE Trans. Pattern Mach. Intell.* **PAMI-5** (6), 609–617 (November 1983).
8. M. Hebert and J. Ponce, A new method for segmenting 3-D scenes into primitives, *Proc. Int. Conf. Pattern Recognition*, Munich, West Germany, 19–22 October, pp. 836–838 (1982).
9. J. H. Han, R. A. Volz and T. M. Mudge, Range image segmentation and surface parameter extraction for 3-D object recognition of industrial parts, *Proc. 1987 IEEE Int. Conf. on Robotics and Automation*, Vol. 1, pp. 380–386 (1987).
10. I. K. Sethi and S. N. Jayaramamurthy, Surface classification using characteristic contours, *Proc. 7th Int. Conf. Pattern Recognition*, Montreal, P.Q., Canada, 30 July–2 August, pp. 438–440 (1984).
11. M. Oshima and Y. Shirai, Object recognition using three-dimensional information, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-5**(4), 353–361 (July 1983).
12. O. D. Faugeras and M. Hebert, The representation, recognition and locating 3-D objects, *Int. J. Robotics Res.* **5**(3), 27–52 (Fall 1986).
13. R. W. Taylor, M. Savini and A. P. Reeves, Fast segmentation of range imagery into planar regions, *Comput. Vision Graphics Image Process.* **45**, 42–60 (1989).
14. N. N. Abdelmalek, Algebraic error analysis for surface curvature and segmentation of 3-D range images, *Pattern Recognition* **23**, 807–817 (1990).
15. S. Inokuchi, T. Nita, F. Matsuday and Y. Sakurai, A three-dimensional edge-region operator for range pictures, *Proc. 6th Int. Conf. Pattern Recognition*, Munich, West Germany, 19–22 October, pp. 918–920 (1982).
16. A. Mitiche and J. K. Aggarwal, Detection of edges using range information, *IEEE Pattern Analysis Mach. Intell.* **5**(2), 174–178 (1983).
17. F. Tomita and T. Kanade, A 3D vision system generating and matching shape descriptions in range images, *Proc. IEEE 1st CAIA*, pp. 186–191 (1984).
18. R. C. Bolles and P. Horaud, 3DPO: a three-dimensional part orientation system, *Int. J. Robotics Res.* **5**(3), 3–26 (Fall 1986)
19. M. Herman, Generating detailed scene descriptions from range images, *Proc. Int. Conf. Robotics and Automation*, St. Louis, Missouri, 25–28 March, pp. 426–431 (1985).
20. T. J. Fan, G. Medioni and R. Nevatia, Description of surfaces from range data using curvature properties, *Proc. Computer Vision and Pattern Recognition Conf.*, IEEE Comput. Soc., Miami, Florida, 22–26 June, pp. 86– 91 (1986).
21. N. Yokoya and M. D. Levine, Range image segmentation based on differential geometry: a hybrid approach, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-11** (6), 643–649 (1989).
22. R. Hoffman and A. K. Jain, Segmentation and classification of range images, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-9**(5), 608–620 (September 1987).
23. R. D. Rimey and F. S. Cohen, A maximum-likelihood approach to segmenting range data, *IEEE Trans. Robotics Automation* **RA-4**(3), 277–286 (1988).
24. K. Sato and S. Inokuchi, Range-imaging system utilizing nematic liquid crystal mask, *Proc. IEEE 1st Int. Conf. on Computer Vision*, London, England, 8–11 June, pp. 657–661 (1987).
25. D. J. Svetkoff, P. F. Leonard, R. E. Sampson and R. C. Jain, Techniques for real-time 3D feature extraction using range information, *Proc. Soc. for Photo-optical Instrumentation Engineers Conf. on Intelligent Robotics and Computer Vision*, Cambridge, Massachusetts, 5–8 November, Vol. 521. SPIE, Bellingham, Washington (1984).
26. J. Mukherjee, Some studies on the recognition of three-dimensional objects, Ph.D. Thesis, I.I.T., Kharagpur, India (1989).

About the Author—JAYANTA MUKHERJEE was born in Bongaon, India, on 8 January 1963. He received the B.Tech. (Hons.) in electronics and electrical communication engineering in 1985 and the M.Tech. in automation and control engineering in 1986 from the Department of Electronics and Electrical Communication Engineering, Indian Institute of Technology, Kharagpur, India. Subsequently he worked as a Senior Research Fellow of the Council of Scientific and Industrial Research of India at the same Department from 1987 to 1988 and obtained his Ph.D. degree in engineering in 1990. During 1989–1990 he served as a Junior Scientific Officer first in the Radar and Communication Centre and then in the Department of Computer Science and Engineering of I.I.T., Kharagpur. He is currently a lecturer in the Department of Electronics and Electrical Communication Engineering, I.I.T., Kharagpur. Dr Mukherjee has been selected for the INSA Medal for Young Scientists for 1992. He has been working in the areas of pattern recognition, image processing, computer vision and digital geometry.

About the Author—PARTHAPRATIM DAS was born in Calcutta on 30 July 1961. He received his B.Tech. degree in electronics and electrical communication engineering in 1984 and M.Tech. degree in computer

engineering in 1985 from the Department of Electronics and Electrical Communication Engineering, I.I.T., Kharagpur. Subsequently he worked as a Senior Research Fellow of the CSIR in the same department from 1986 to 1988 and obtained his Ph.D. degree in engineering in 1988. He served as a Junior Scientific Officer in the Radar and Communication Centre of I.I.T., Kharagpur, in 1988 before joining the Department of Computer Science and Engineering, I.I.T., Kharagpur, as a lecturer. He is currently an Assistant Professor in the same department. Dr Das received the INSA Medal for Young Scientists for 1990, the UNESCO/ ROSTSCA Young Scientist Award for 1989 and the Virendra Gupta Award of Computer Society of India for 1986. He is a member of IEEE (U.S.A.), IUIAPR, IETE and CSI. His research interests include digital geometry, computer vision and parallel algorithms. He has written more than 30 technical papers for international journals and conferences.

About the Author—BISWANATH CHATTERJI was born on 10 November 1942 at Varanasi. He obtained the B.Sc. degree from Allahabad University in 1961 and B.Tech. (Hons.) degree and Ph.D. degree in electronics and electrical communication engineering from Indian Institute of Technology, Kharagpur, in 1965 and 1970, respectively. He also carried out postdoctoral research work in pattern recognition at the University of Erlangen-Nurenberg in 1972–1973. He worked as an Assistant Development Engineer at Telerad Private Ltd, Bombay, in 1965 and as a Senior Research Fellow in the Electronic Instrumentation Division of Central Electronic Engineering Research Institute, Pilani, in 1965–1966. Since 1967 he has been a faculty member in the Department of Electronics and Electrical Communication Engineering of Indian Institute of Technology, Kharagpur, where he has been a Professor since 1980 and has served as the Head of the Department during 1987–1991. Professor Chatterji has published extensively and is the author of two textbooks. His area of research interests are pattern recognition, image processing, system identification, distributed and data flow processors, control systems and microprocessor based systems. He is the Fellow of Institution of Engineers, Institution of Electronics and Telecommunication Engineers, Indian Society of Engineers, Institution of Instrumentation Scientists and Technologists and members of several professional bodies. Professor Chatterji is the recipient of the prestigious Hari Om Ashram Prerit Dr Vikram Sarabhai Award for the year 1983 for his contributions in the area of electronics and telecommunication engineering.