JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 17, 347-369 (2001)

Combining Morphological Feature Extraction and Geometric Hashing for Three-Dimensional Object Recognition Using Range Images

CHU-SONG CHEN, YI-PING HUNG[+] AND JA-LING WU[*]
Institute of Information Science
Academia Sinica
Taipei, Taiwan 115, R.O.C.
[+]*E-mail: hung@iis.sinica.edu.tw*
[*]*Department of Computer Science and Information Engineering*
National Taiwan University
Taipei, Taiwan 116, R.O.C.

This paper presents a new approach for model-based object recognition with range images by combining morphological feature extraction and geometric hashing. In low-level processing, range images are segmented into 3D-connected surface patches. In middle-level processing, each connected component is processed by using morphological operations to extract the skeletons of high-variation regions. These skeleton points can be viewed as invariant salient feature primitives. In high-level processing, geometric hashing is used to recognize objects. We also use a basis-similarity constraint to reduce the number of spurious hypotheses. Experimental results have shown that the proposed method is effective and has great potential for model-based object recognition using range images.

Keywords: computer vision, object recognition, range image processing, feature extraction, geometric hashing

1. INTRODUCTION

Object recognition in a cluttered and partially-occluded 3D environment is an important but difficult problem. A reliable 3D object recognition system is useful for many applications in computer vision such as locating a 3D object from arbitrary camera positions, registering and integrating range images from different view points, and tracking a specific object in a dynamic environment. In this paper, a model-based 3D object recognition approach is proposed. In our experiments, range images of 3D objects were obtained using a stereo range finder [5]. Reliable invariant features based on the shape of 3D objects were then extracted from the range images and stored in a database which could be used for recognition.

Typically, there are three stages in a model-based recognition process. The first stage is a *low-level processing* or *pre-processing* stage. In this stage, the range images are filtered or smoothed to remove noise induced in range image acquisition. The second stage is a *middle-level* processing or *feature-extraction* stage. The feature primi-

Received August 29, 1998; revised September 30 & December 2, 1999; accepted January 18, 2000.
Communicated by Shing-Tsaan Huang.

tives to be extracted should be salient and invariant. With this feature extraction stage, the amount of data to be processed subsequently can be considerably reduced. Also, good feature selection can greatly improve matching efficiency. The third stage is a *high-level processing* or *recognition* stage which finds possible partial matches between the scene and the model-based database. To find the correct matches effectively, the recognition stage may include the following two tasks: *hypotheses generation* and *verification*. The hypotheses generation module selects matching hypotheses which have high scores, and then the verification module verifies these hypotheses by transforming the corresponding object model into the scene and computing the correlation between model and scene in the overlapping portions. These two tasks are performed iteratively until all the objects contained in the scene image find their matches, or all the hypotheses with high scores have been considered.

Along with the above framework, we develop an integrated approach for object recognition based on range images. In middle-level processing, feature primitives are extracted by using mathematical morphology (M.M.). The reason we adopt the M.M. for range image processing in this work is explained below. First, M.M. can deal with the shape of a function in an intuitive way; hence, it is suitable for range image processing since range images inherently contain plenty of shape information. Second, M.M. operations such as opening (or closing) used in this paper can inherently remove convex (or concave) bumping noises, and hence our feature extraction method is less sensitive to noise. On the other hand, most previous approaches used the differential-geometry based methods to extract invariant features from range images [1, 3, 23], which usually require the computation of the first-order or second-order derivatives of a range image, and are known to be noise-sensitive if without sophisticated preprocessing. Experimental results have shown that our feature extraction method is stable, and can extract salient features from noisy range images. In the past, not much work has been done to apply the M.M. to range image processing (a few examples are [15, 25]). In this work, we used the morphological technique to extract invariant salient features — skeletons of high-variation regions, from a range image. The characteristic of this method is that the residue computation is along the orthogonal direction, instead of along the parallel direction, as shown in Figs. 2(b) and 2(a). As a consequence, the extracted regions are invariant to 3D rotation and translation.

In high-level processing, geometric hashing (GH) is adopted for solving the recognition problem. GH is a general model-based scheme which can be used to solve many model-based recognition problems as long as the transformation class has been given. GH was first introduced by Lamdan and Wolfson [17]. Some error and sensitivity analysis was given in [13, 18]. In [24], the GH technique was generalized to use line features instead of point features. Gavrila and Groen [12] introduced a method to recognize polyhedral 3D objects using 2D images by increasing the voting complexity from all possible viewing directions under orthographic projection. Rigoutsos and R. Hummel [20] introduced a Bayesian approach for model matching and voting for the GH technique. In the previous literature [12, 13, 16, 18, 20, 24], GH was usually applied for the 2D case, especially for the 2D affine-transformation case.[1] In this paper, we use GH to solve the 3D object recognition problem, where the transformation class considered is 3D rotations and 3D translations.

[1] However, a recent research reveals that 2D affine transformation cannot account for human 3D object recognition [19].

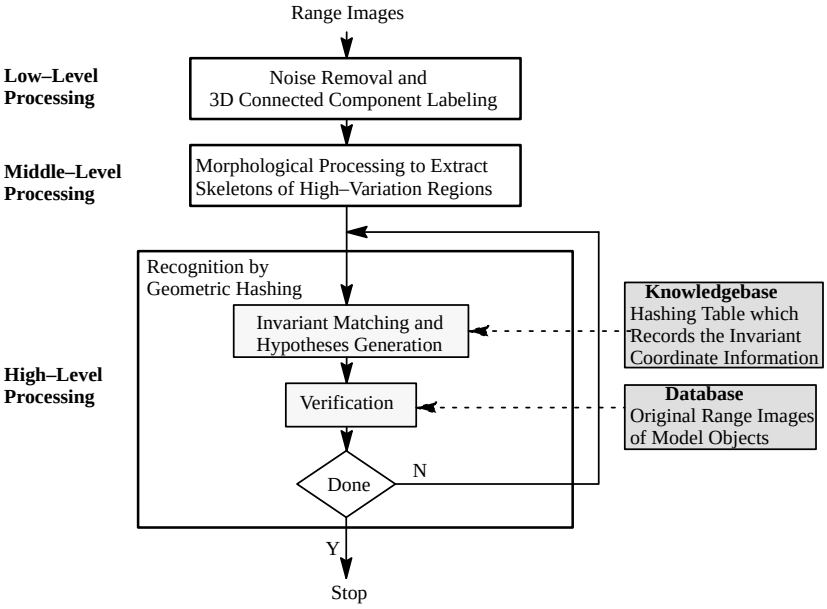


Fig. 1. Framework of our approach for 3D object recognition using range images.

Fig. 1 shows the framework of our approach for 3D object recognition using range images. In this figure, the input range images are first pre-processed to remove noise and then segmented as a set of 3D connected components. Each component is then processed using mathematical morphology (M.M.) to obtain the skeletons of high-variation regions. The knowledgebase used here is a hashing table which stores useful invariant coordinate information generated by the learning phase of GH. The original range images of the object models are also stored in the database for hypotheses verification. In this paper, point features are used. Edges or skeletons are viewed as a set of points, as in [20]. The reason for using point features is that this type of features can handle the partial-occlusion problem better than curves or surface patches can. As mentioned above, GH can exploit the similarities of feature attributes between the model and scene as a local constraint in the voting stage. To choose a local constraint, an intuitive method is to use the similarity of the neighboring range data of a feature point. However, the drawback of using such local primitives for range images is that it is easy for these primitives to be similar to each other. For example, the neighboring range data of all the feature points in an edge of a cube are similar to each other. An approach suggested in this paper is to use the similarity of the basis triangles (described in Section 3.2.1) as a local constraint to remove the incorrect hypotheses in voting, instead of using the similarities of the neighboring range data of feature points between the model and scene. In our experience, this useful local constraint (i.e., the similarity of the basis triangles) can significantly improve matching efficiency. Also, the relation between GH and a famous paradigm, the *random sample consensus* (RANSAC) [10], is pointed out and discussed in this paper. In fact, if the bases are selected randomly in the recognition phase of GH, then GH can be viewed as a generalization of the RANSAC approach,

where the model information is stored in a specified hashing table (see Section 3.3).

This paper is organized as follows. Section 2 describes our methods used for the low-level processing and the middle-level processing methods. The GH technique for 3D object recognition is proposed in Section 3. Some experimental results are given in Section 4. Section 5 gives the conclusions and some discussion.

2. LOW-LEVEL PROCESSING AND FEATURE EXTRACTION OF RANGE IMAGES

2.1 Low-Level Processing of Range Images

The purpose of the low-level processing stage is to delete jumping noise as well as to perform basic segmentation. These two tasks are unified into a single procedure called 3D connected component labeling. The procedure of 3D connected component labeling is similar to that of 2D connected component labeling introduced in [14], except that the depth differences between neighbor pixels are also considered. Those connected regions with areas smaller than a given threshold can then be deleted. Hence, this procedure can remove isolated jumping points. In our algorithm, a binary map of the processed range image is first generated. Two pixels in a range image are connected if (i) one of them belongs to the 8-connected neighbors of the other in the binary map, and (ii) the difference between the depths of these two pixels is smaller than a given threshold. The 2D connected component labeling algorithm introduced in [14] can be generalized to be a 3D algorithm by changing the definition of the NEIGHBOR() operator. In this paper, the algorithm in [21] is adopted to extend to the 3D case due to its time efficiency. According to our experience, this simple low-level processing strategy can successfully remove jumping noise and segment the range images into connected surface patches in almost all cases.

2.2 Feature Extractions of Range Images Using Morphological Operators

In middle-level processing, features were extracted by using the morphological opening or closing operation with spherical structuring elements. Our feature-extraction method can be divided into two stages: the first one is to compute the *orthogonal residue* described in Section 2.2.1, and the second one is to compute the *skeleton of high-variation region* (HV-skeleton) described in Section 2.2.2.

2.2.1 Orthogonal residue

In this work, the structuring elements (S.E.s) are chosen to be spherical since sphere is the only shape which can make the opening and closing results rotation invariant. For simplicity, we only describe in the following the case of using morphological opening, while this method can also be applied in the case of using morphological closing. Let $I(x, y)$ be a 2D range image, and k be a spherical S.E. with radius r. Traditionally, the opening residue [15] was defined as (where "o" denotes the operation of *opening*)

$\tau_p [I, k](x, y) = I(x, y) - [Iok](x, y)$

which is always non-negative and can be used for detecting the high-variation convex portion of a signal. Similarly, high-variation concave portion of a signal can be detected by using the dual operation – closing. However, such residue was measured in parallel along the direction perpendicular to the x-y plane, as shown in Fig. 2(a). This type of residue has the drawback of not being invariant to 3D rotations. Therefore, we introduced the orthogonal opening residue defined in the following (where "⊖" denotes the operation of erosion):

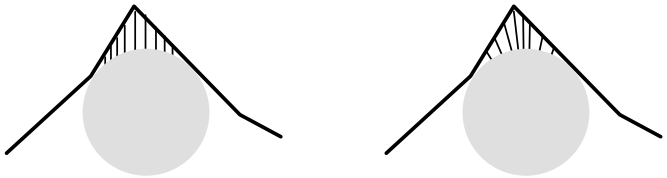


Fig. 2. (a) the parallel residue and (b) the orthogonal residue

Definition 1. Orthogonal Opening Residue:
Consider a range data point $(x_0, y_0, I(x_0, y_0)) \in \boldsymbol{R}^3$. Let $D_0 = \{(x', y', [I \ominus k](x', y') \mid (x', y') \in disk(r)$, where $disk(r)$ is a disk of radius r, centered at (x_0, y_0), in the x-y plane). The *closest eroded point* of $(x_0, y_0, I(x_0, y_0))$ is defined to be the point in $(x_0, y_0, I(x_0, y_0))$ that is closest to $(x_0, y_0, I(x_0, y_0))$. Let $d(x_0, y_0)$ be the distance from $(x_0, y_0, I(x_0, y_0))$ to its closest eroded point $(\underline{x}, \underline{y}, [I \ominus k]\ (\underline{x}, \underline{y}))$. Then, the orthogonal opening residue is defined to be $\tau_0[I, k](x_0, y_0) = d(x_0, y_0) - r$.

For simplicity of explanation, a 2D geometric illustration for the definition of the orthogonal opening residue is given in Fig. 3, although this paper actually deals with range data in 3D space. In the 2D illustration, $disk(r)$ degenerates to a line segment of length $2r$. The terms $d(x_0, y_0)$and $\tau_0[I, k](x_0, y_0)$ become $d(x_0)$ and $\tau_0[I, k](x_0)$, and the 3D points $(x_0, y_0, I(x_0, y_0))$ and $(\underline{x}, \underline{y}, [I \ominus k](\underline{x}, \underline{y}))$ become 2D points $(x_0, I(x_0))$ and $(\underline{x}, [I \ominus k](\underline{x}))$, respectively. For example, Fig. 4(b) shows the orthogonal opening residue of a range image of a cube-shaped object. An important property of the orthogonal opening residue is that, it is rotation-invariant and hence is suitable to be applied for 3D object recognition.

2.2.2 Skeleton of high-variation region

Definition 2. High-Variation Region:
A *high-variation region* H is defined as a set of 3D data points with large values of orthogonal opening residue, i.e., $H = \{(x_0, y_0, I(x_0, y_0)) \in \boldsymbol{R}^3 \mid$ to $\tau_0[I, k](x_0, y_0) > T\}$ (where T is a threshold).

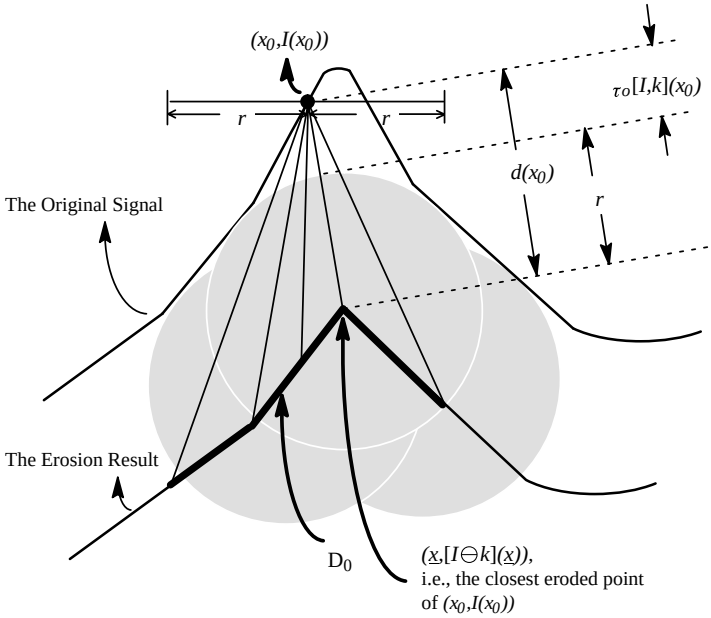


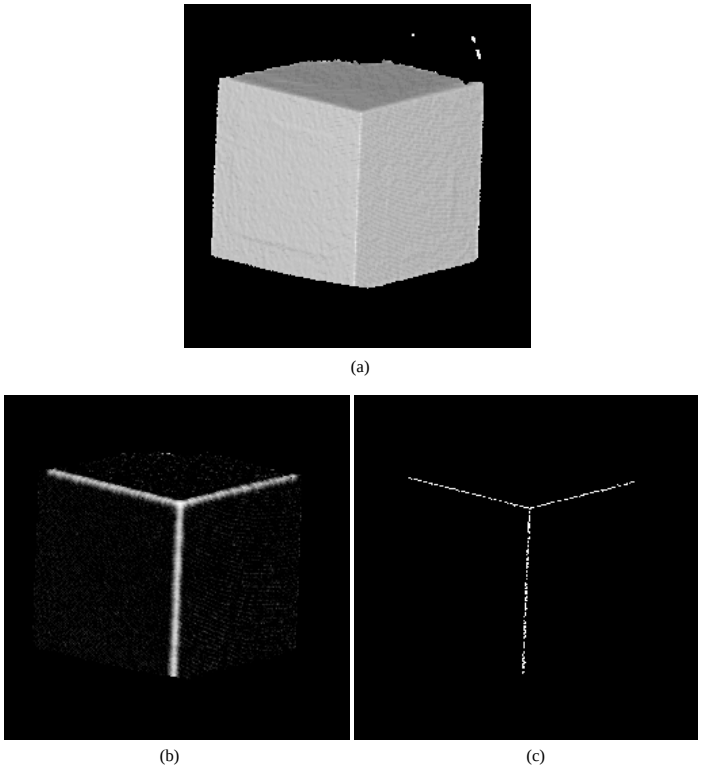
Fig. 3. 2D geometric interpretation of the orthogonal opening residue.

Definition 3. HV-Skeleton:
Let H be the set of all 3D data range data points contained in the high-variation region. The HV-skeleton of H is defined to be the set S = {($\underline{x}$, $\underline{y}$, [$I\ominus k$] ($\underline{x}$, $\underline{y}$)) $\in$ $\boldsymbol{R}^3$ | ($\underline{x}$, $\underline{y}$, [$I\ominus k$] ($\underline{x}$, $\underline{y}$)) is the closest eroded point of (x, y, $I(x, y)$), and (x, y, $I(x, y)$) $\in$ H}.

For simplicity, the HV-skeleton is directly applied for 3D object recognition without further classification in this paper. It is obvious that the HV-skeleton is also rotation-invariant. Intuitively, those regions which can not be filled by the spherical S.E. are treated as high-variation regions, and all the centers of the spherical S.E.s associated with a high-variation region constitute the HV-skeleton of that high-variation region. Fig. 4(c) shows the HV-skeleton of a range image of a cube-shaped object. Other examples of HV-skeletons can be found in Figs. 11-15. From these figures, it is easy to see that our feature extraction method is quite stable and can extract salient features for noisy range images in a simple and efficient way. The number of data points contained in the extracted features is considerably reduced compared to the original data set. In the following section, the extracted HV-skeletons will be used for object recognition.

3. GEOMETRIC HASHING FOR 3D OBJECT RECOGNITION

In this paper, the skeletons of high-variation regions are used as feature primitives. To deal with the partial-occlusion problem, the skeletons are viewed as a set of 3D points, and then GH is used to recognize 3D objects. GH is a general technique for model-based recognition which can be divided into two phases: the *preprocessing* (or



(a) (b) (c)

Fig. 4. (a) A range image of a cube displayed using a shading technique. (b) The orthogonal opening residue. The larger the intensity value is, the larger the intensity value is, the larger is the orthogonal residue. (c) The skeleton of the high-variation region (the HV-skeleton).

learning) phase and the *recognition* phase. In the preprocessing phase, a hashing table is established, which stores useful knowledge about the invariant coordinate information of the model objects. In the recognition phase, the established hashing table can then be used for hypotheses generation and invariant matching. In the following, the preprocessing phase will be described in Section 3.1, and the recognition phase will be described in Section 3.2.

3.1 The Preprocessing Phase

In our case, the transformation class is 3D rotation and 3D translation, which can be fully determined by three point-to-point correspondences. Hence, three points are needed to establish a single basis of a coordinate system. Given three points, p_0, p_1, p_2, which are not collinear, the coordinate system with respect to the basis (p_0, p_1, p_2) is established as follows: Let p_0 be the origin. Let the x axis be in the direction of $p_1 - p_0$ and the z axis be in the direction of $(p_1 - p_0) \times (p_2 - p_0)$, where "$\times$" is the outer product. The y axis is then defined as $-x \times z$. In order to obtain robust coordinate systems, the point set $\{p_0, p_1, p_2\}$ should not be selected as a basis if either the three points are nearly

collinear, or if any two of them are too close to each other. The configuration of the established coordinate system is shown in Fig. 5.

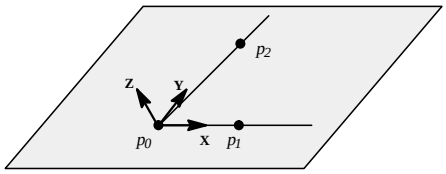


Fig. 5. The coordinate system established with respect to an ordered non-collinear triplet (p_0, p_1, p_2).

In the preprocessing phase, invariant information contained in the models is extracted and stored in a hashing table. Bases are formed for all the feasible ordered triplets. A coordinate system is established on each basis as described above, and the coordinates of all the other model points are computed in terms of this coordinate system. These coordinates can then be stored in a hashing table, in each of which we record a hash element (object, basis). The detailed procedure of the preprocessing phase is described as follows:

Procedure *Preprocessing*:

For each model object *M* and for each of its feasible basis *b*, do

Step 1. Compute the coordinates of all the other points of the model in terms of the basis *b*.

Step 2. Use the computed coordinates to index the hashing table entries.

Step 3. Add a *hash element* (*M*, *b*) in the hashing table entries indexed from Step 2.

After preprocessing, a hashing table is established, which can then be applied in the subsequent recognition phase. Notice that in the recognition phase, it is usually required that similar coordinates be put in the same or neighboring memory locations for possible retrieval since scene images may be corrupted by noise. To achieve this purpose, the hashing function used in this paper is the *identity / scaled* one which is similar to the one used in [12]. That is, if a point has a 3D coordinate (x, y, z) with respect to a given basis, then the corresponding pair (object, basis) will be linked in the memory location $(\frac{x}{w}, \frac{y}{w}, \frac{z}{w})$. The value w is referred to as the width of the hash cube, as shown in Fig. 6.

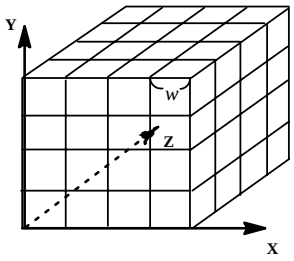
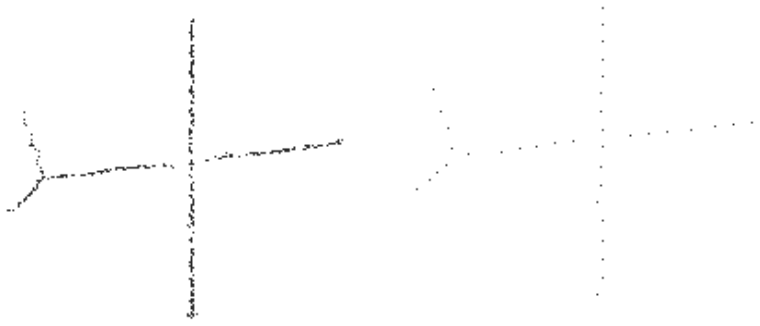


Fig. 6. Structure of the hashing table used in this paper.

This kind of hashing table can also be viewed as an associative or a content addressable memory, where three values of the basis triplets are directly used to retrieve the memory location.

To increase the processing speed and to reduce the required storage, a multiscale signal processing scheme is applied. In our method, two scales of data sets are used during the processing. One is the original HV-skeletons, which are referred to as *fine-scale* data. Another is a uniform subsampled version of the HV-skeletons, which is referred to as *coarse-scale* data. Fig. 7 shows an example of the fine-scale and the coarse-scale data in our processing. An advantage of using a two-scale approach is that it can increase the processing speed in the preprocessing stage since the number of data points involved in computation can be reduced significantly. Another advantage is that the number of the hash elements which need to be stored in the hashing table can also be decreased. The sampling is performed uniformly in the 3D space with a period t_p along each axis. A useful criterion for selecting the sampling period is to choose t_p to be equal to w, so as to avoid recording redundant information in the same hash cube. A major problem with this approach is that the induced 3D-transformation will be less accurate because the sampling positions of the model objects and of the scene may not be the same. In our approach, this problem can be solved in the verification stage of the recognition phase by using a refinement process to find the best matched basis triangle in the neighborhood.



(a) (b)

Fig. 7. (a) The original fine-scale HV-skeleton data obtained from the range image of a cross-shaped object. (b) The subsampled coarse-scale data of (a).

3.2 The Recognition Phase

The recognition phase of GH can be explained as a model-based recognition scheme as shown in Fig. 1. Two major tasks in Fig. 1 are hypotheses generation and verification, which will be introduced in Sections 3.2.1 and 3.2.2, respectively.

3.2.1 Hypotheses generation

In the recognition phase, the skeleton data obtained from the scene are also subsampled with a sampling period $t_r = r \cdot t_p$ $(0 \le r \le 1)$. That is, the skeleton data of the scene to be recognized are subsampled more densely than are those for building the hashing

table. In the beginning, an arbitrary ordered triplet is chosen from the scene image and forms a basis *b* as shown in Fig. 8(a). The coordinates of all the remaining points are then computed in terms of this basis *b*, as shown in Fig. 8(b). These coordinates are used as entries to retrieve information from the hashing table, and one vote is given to all the hash elements (object, basis) at the entry indexed by its coordinates (see Fig. 8(c)). Finally, pairs (object, basis) with high scores are then selected as possible hypotheses and are sent to the verification module (Fig. 8(d)). Hence, in the recognition phase, the hashing table serves as the knowledgebase for possible generation of hypotheses, where the invariant coordinate information of the model objects has been stored in the preprocessing phase.

A good local constraint is used to increase the recognition performance in this paper. From the above description, it can be observed that GH tends to utilize the information of spatial relationship among the model objects. Basically, good local constraints can increase both accuracy and the speed of matching by removing irrelevant hypotheses. Instead of using local characteristic contained in the neighborhood of the feature point, this paper exploits the basis-similarity constraint. Two bases (x, y, z) and (x', y', z') are similar if the length differences of the corresponding edges of their two basis triangles are all smaller than a given threshold. That is, $\overline{xy} - \overline{x'y'} < T$, $\overline{yz} - \overline{y'z'} < T$, and $\overline{xz} - \overline{x'z'} < T$, where T is a threshold, as shown in Fig. 9. This constraint is used in the voting stage. Pairs (object, basis) whose bases are not similar to the scene basis should not be counted in the voting stage. Using this constraint, the number of spurious hypotheses can be reduced.

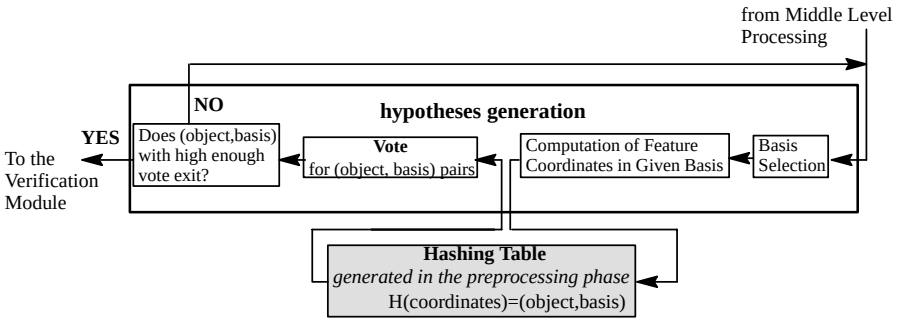


Fig. 8. The block diagram of the hypotheses generation in the recognition phase.

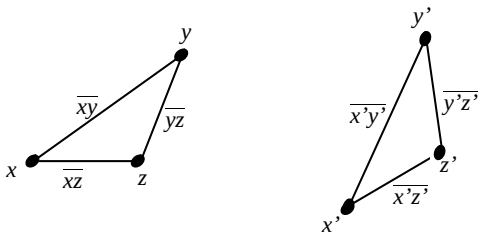


Fig. 9. The basis-similarity constraint.

3.2.2 Verification

In the verification stage, each of the generated hypotheses is examined through a more detailed procedure. Since a basis of a hypothesis is obtained from coarse-scale data, a refinement procedure is needed to find more accurate results. Given a hypothesized basis *b*, a search is performed in the fine-scale data within the neighborhood of each point of *b* to find the *most similar triangle* with respect to *b*. The most similar triangle is defined as the least error triangle, where the error is computed by summing up the absolute difference values of the three corresponding edge pairs. This triangle can then be viewed as a more accurate basis and is used in the subsequent processing. The block diagram of the verification module is described in Fig. 10. In Fig. 10(a), each possible pair (object, basis) generated from the hypotheses generation module is used to find a unique 3D rigid-body transformation (rotation and translation) which transforms the object to the basis. This transformation can be determined based solely on the correspondence of the object basis and the scene basis. However, to increase accuracy, the transformation is usually determined by exploiting all the corresponding skeleton point pairs in a least-square manner. In this paper, the least square error transformation is computed using the corresponding skeleton point pairs between the model and the scene by means of Arun's method. The range image of the hypothesized object is then transformed and compared with the range image of the scene to verify whether the hypothesis is correct or not, as shown in Fig. 10(b). If the overlapping region is larger than a given threshold, then the object can be treated as *matched*, and the data corresponding to it should be removed from the scene, as shown in Fig. 10(c). Otherwise, we shall go back to the hypotheses generation module. The verification procedure will continue until the remaining data in the scene image are few enough, or until no feasible basis exists.

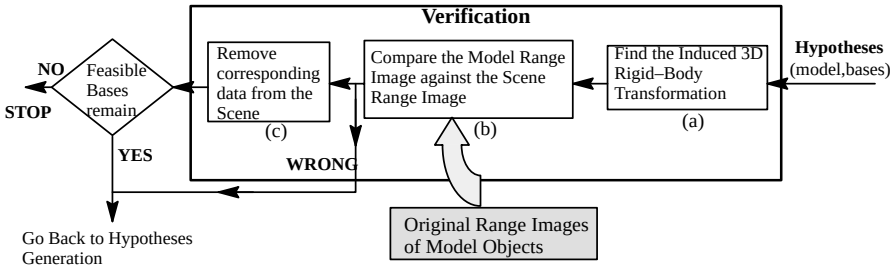


Fig. 10. The block diagram of the verification module in the recognition phase.

3.3 Relation Between Geometric Hashing and the RANSAC Paradigm

In the above, the major steps for applying GH to 3D object recognition have been described. In this section, the relation between the GH technique and the RANSAC paradigm [10] is discussed. RANSAC was proposed as a general robust estimation method for surface or model fitting. Robust estimation means that model fitting is not influenced by *outliers* (gross errors). For range image processing applications, the RANSAC technique has been adopted to segment range images into quadratic surface patches [26]. To avoid problems caused by outliers, the key concept of the RANSAC approach is to use the *random sampling* scheme. Assume a model that requires a minimum of *k* points to uniquely determine its free parameters. The algorithm begins

by randomly selecting a set of k points in the observed data set and then constructs a model M based on these k points. All other points in the data set which can be described by using the model M, within some error tolerance, are then collected as a consensus set. If, after a predetermined number of trials, no consensus set with enough members is found, we can either fit the model with the largest consensus set found, or terminate as a failure.

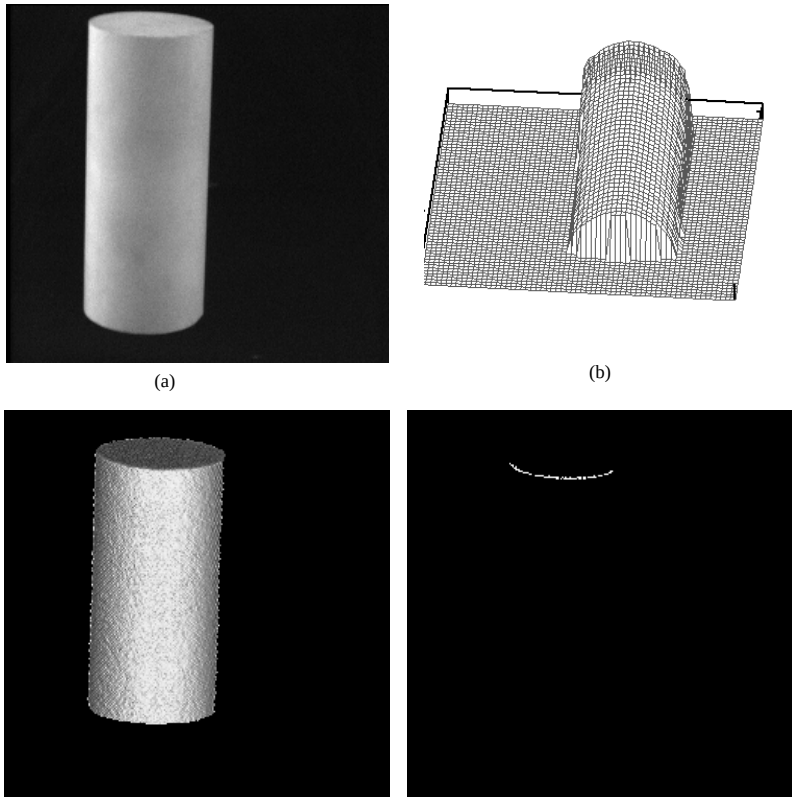
In the GH approach, if the random selection scheme is used for the basis choice step (Fig. 8(a)), GH can be viewed as a generalization of the RANSAC approach. In the RANSAC scheme, the model is usually represented as a parametric form such as a collection of quadratic surfaces or spline functions. On the other hand, the model information is stored in a well-organized structure of a hashing table in the GH approach. The GH approach tries to find the consensus set in a voting stage. If there are too many outliers in the data set, both RANSAC and GH will fail to find the correct solutions. In the GH approach, a necessary criterion for having a successful trial in the recognition phase is that the selected triplet should belong to a model object. Assume there are m feature points in a scene range image, in which totally qm (here $q < 1$ is referred to as the *object ratio*) points come from a model object M (i.e., the *outlier ratio* of M is $1-q$). Then, the probability of randomly selecting a basis (p_0, p_1, p_2) such that not all pi (i = 1, 2, 3) are contained in M is equal to $(1-q^3)$. Hence, the probability of not choosing a valid model triplet in n trials can then be computed as $(1-q^3)^n$. Although similar analysis has also be given in [16], our emphasis is on analysis from a RANSAC point of view. Table 1 lists the number of trials which need to be iterated so as to make the probability of *fail to match* smaller than 1.0% (i.e., $(1- q^3)^n < 0.01$). From this table, it can be observed that if the outlier ratio is equal to one half, then the correct model object has a 99% chance to being selected in 35 trials. However, if the outlier ratio is too large (0.9), then about 4600 trials are needed. As a consequence, it is important for the RANSAC approach to keep the object ratio larger than a given threshold. A reasonable number of maximal iteration steps should be given in the precedence. In our work, the object ratio is set to be 0.2, which leads to a maximal iteration number of about 600. That is, if the portion that the object data occupies in the scene is not larger enough (≤ 20%), then the system will probably give up on recognizing this object after 600 trials.

Table 1. Number of trials needed to make the probability of "fail-to-match" smaller than 1%.

Object ratio q	Outlier ratio $1 - q$	umber of trials which need to be iterated o as to make the probability of fail to atch smaller than 1.0%
0. 50	0.5	35
0.45	0.55	49
0.40	0. 60	70
0.35	0.65	106
0.30	0.70	169
0.25	0.75	293
0.20	0.80	574
0.15	0.85	1363
0.10	0.90	4603

4. EXPERIMENTAL RESULTS

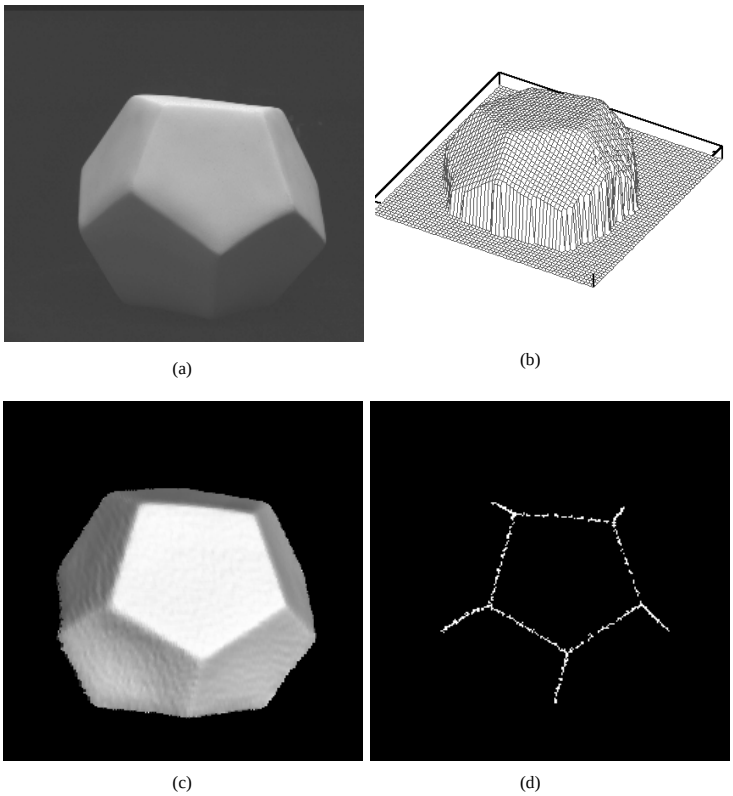
To verify the method developed in this paper, five objects are used to build a database for the following experiments. Figs. 11(a)-15(a) show the five objects (a cylinder, a dodecahedron, a model-head, a cross-shaped object, and a sculpture-head) used. The range images of the five objects are displayed with wire-frame models in Figs. 11(b)-15(b). The computer generated images of the range data are shown in Figs. 11(c)-15(c) using the standard phong shading technique [11]. Each of these range images contains about 10000 to 40000 3D data points, and all range images are stored in the database for the use of verification. After feature extraction using mathematical morphology, the number of the extracted skeleton points were considerably reduced. Figs. 11(d)-15(d) show the extracted skeletons of high-variation regions. The number of the skeleton points of the five objects (a cylinder, a dodecahedron, a model-head, a cross-shaped object, and a sculpture-head) are 77, 462, 119, 290, and 313, respectively. In the preprocessing phase, these subsampled skeleton points were sequentially used to construct a hashing table, as described in Section 3.1. In our work, the *identity / scale*



(a) (b)

(c) (d)

Fig. 11. A cylinder contained in the database: (a) An intensity image of this object. (b) A wire-frame display of the range image of this object. (c) A computer-generated image of (b) obtained by using the Phong shading technique. (d) The extracted HV-skeletons.



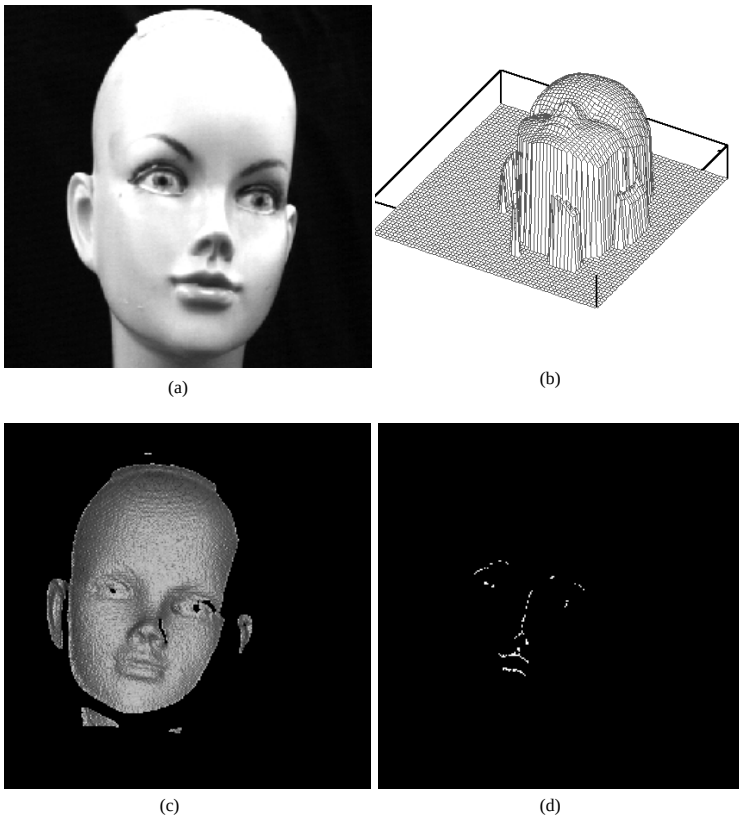
(a) (b)

(c) (d)

Fig. 12. A dodecahedron contained in the database: (a) An intensity image of this object. (b) A wire-frame display of the range image of this object. (c) A computer-generated image of (b) obtained by using the Phong shading technique. (d) The extracted HV-skeletons.

hashing function was used as explained in Section 3.1. To reduce the space complexity and also to avoid recording multiple invariant coordinates from the same local region, data points contained in the HV-skeleton were uniformly subsampled where the sampling period was equal to the width of the hash cube, w = 10 (mm). After subsampling, the number of the remained skeleton points were 10, 48, 17, 39, and 44, respectively. The value r introduced in Section 3.2.1 was selected to be 0.5 in our work; that is, the sub-sampled skeleton data for the scene was twice as dense as that for the model objects. The hash table built in our experiment contains $100 \times 100 \times 100$ (i.e., one Mega) entries. Hence, the range of the coordinate values which can be handled in the hash table was $10 \times 100 = 1000$ (mm).

In the recognition phase, several scene images were used to test the effectiveness of the proposed recognition scheme. Each scene range image contained a few objects such that these objects could be translated, rotated, and partially occluded in the 3D space. Scene 1 contained two geometric objects whose poses are different from those in Fig. 12(c) and 14(c) as shown in Fig. 16(a). The extracted HV-skeletons are shown in Fig. 16(b). After 7 iterations of feasible basis selection, the dodecahedron was first recog



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

Fig. 13. A model-head contained in the database: (a) An intensity image of this object. (b) A wire-frame display of the range image of this object. (c) A computer-generated image of (b) obtained by using the Phong shading technique. (d) The extracted HV-skeletons.

nized, and the matched object was then transformed for comparison with the scene. Fig. 16(c) shows the portion of the transformed model range image which was close enough to the scene data. Later, after 12 trials of feasible basis selection, the cross-shaped object was also recognized, as shown in Fig. 16(d). Fig. 17(a) shows another scene which contained three objects. The first object was recognized after 96 trials and is shown in Fig. 17(c). The second object and the third one were recognized after 221 and 229 trials, and are shown in Figs. 17(d) and 17(e), respectively. In Fig. 18(a), two human faces are contained in a single range image. Using the recognition scheme proposed in this paper, the object could be correctly recognized as shown in Figs. 18(c) and 18(d). The recognition process takes about 20 minutes in Sun SPARC 10 workstation. Although the objects contained in range images are all separated in the experiments, our method can be used for the case that the objects are touched because that the RANSAC principle can deal with the problem of partial matching.

Notice that in Fig. 17 (c) and Fig. 18(d), the matched objects are only portions of the existing object data in the scenes. This is because the estimated 3D rigid-body transformation obtained from GH was slightly different from the true one due to noise.

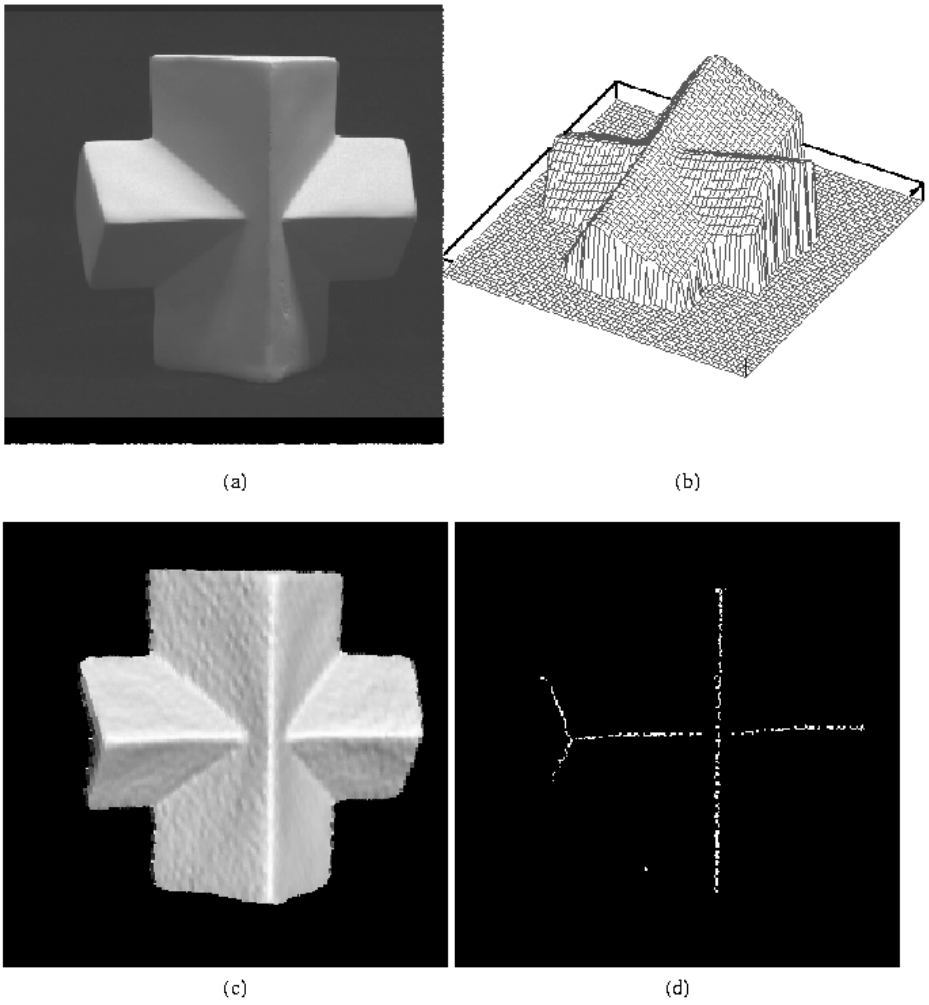


Fig. 14. A cross-shaped contained in the database: (a) An intensity image of this object. (b) A wire-frame display of the range image of this object. (c) A computer-generated image of (b) obtained by using the Phong shading technique. (d) The extracted HV-skeletons.

Since only the skeleton data were involved in this computation, the induced transformation was usually a *coarse* one. To improve the accuracy of the computed transformation, finer registration has to be performed, which uses not only the skeleton data, but also the original range images. In fact, the obtained coarse transformation can be used as an initial guess of the *iterative-refinement procedures* [3, 7] to find a better registration. In this study, only a coarse transformation was estimated in our experiment. In the future, we shall use iterative-refinement procedures to find a better transformation between the transformed model image and the scene image.

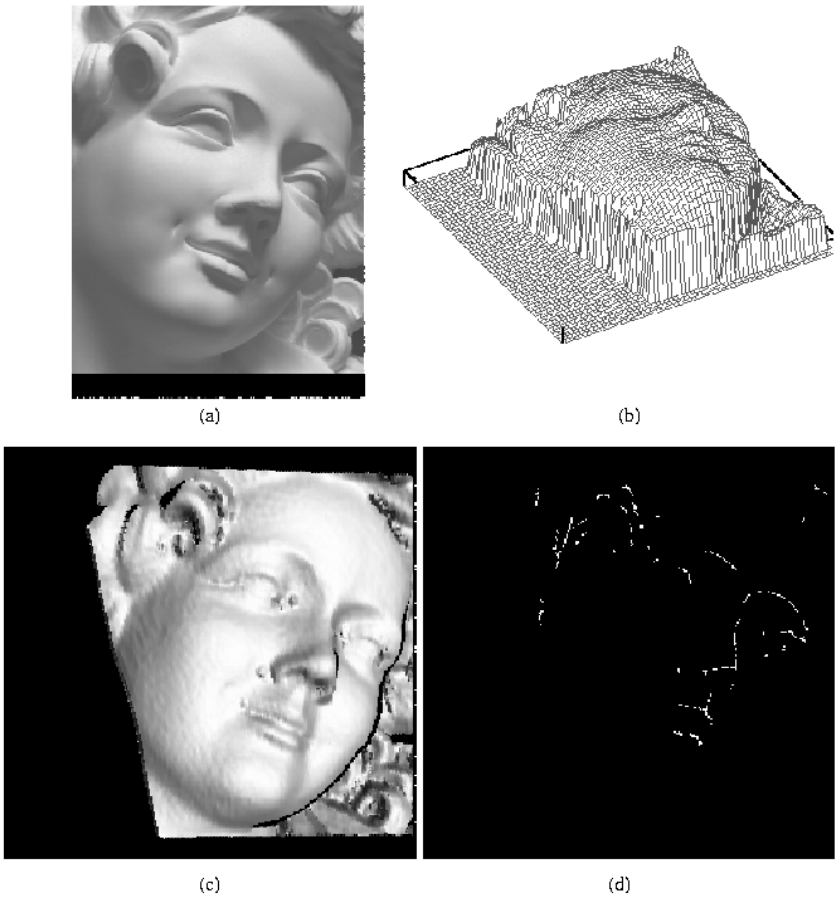


Fig. 15. A sculpture-head contained in the database: (a) An intensity image of this object. (b) A wire-frame display of the range image of this object. (c) A computer-generated image of (b) obtained by using the Phong shading technique. (d) The extracted HV-skeletons.

5. CONCLUSIONS AND DISCUSSION

In this paper, we have proposed a new integrated approach for model-based object recognition using range images. This approach combines the morphological feature extraction and the geometric hashing techniques to solve the 3D object recognition problem. The M.M. technique is inherently good for dealing with shapes of signals; hence, it is extremely useful for range images processing. However, less work has been done before on applying M.M. to range image. In this work, we have successfully applied the morphological technique in a simple and efficient way to extract invariant salient features, i.e., skeletons of high-variation regions, from a range image. Using the extracted invariant skeletons, we have also successfully applied the geometric hashing technique to recognize 3D occluded objects. The matching efficiency has been

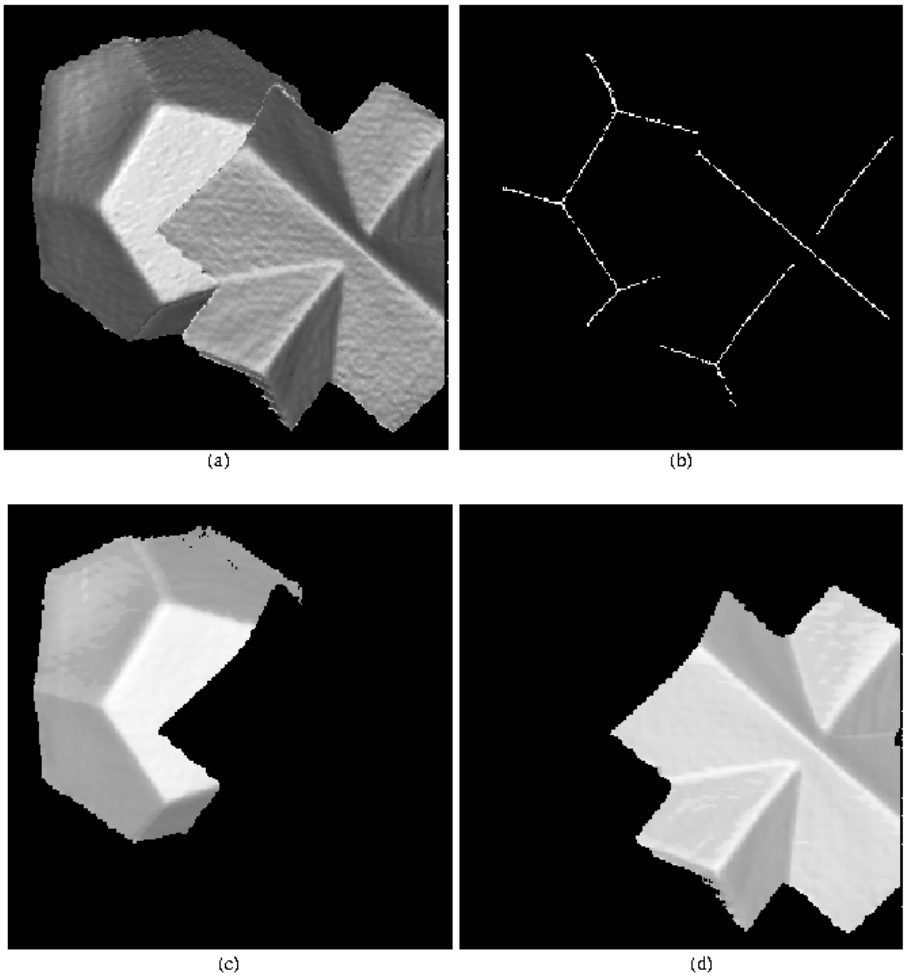


Fig. 16. (a) The shaded range image of scene 1 which contains a dodecahedron and a cross-shaped object. (b) The HV-skeletons extracted from the range image shown in (a). (c) The first recognized object (after 7 trials of feasible basis selection). (d) The second recognized object (after 12 trials).

significantly improved by using a basis-similarity constraint. The relationship between the GH technique and the RANSAC paradigm has been pointed out and discussed, which gives a useful guideline for determining the maximal iteration number in basis selection. Experimental results have demonstrated that the proposed method has great potential for model-based object recognition using range images. The 3D object recognition method developed in this paper can be used for many applications in computer vision such as locating a 3D object from arbitrary camera positions, registering and integrating range images from different view points [6], and tracking a specific object in a dynamic environment. In the future, we will integrate the proposed method into the active binocular vision system that we are developing.

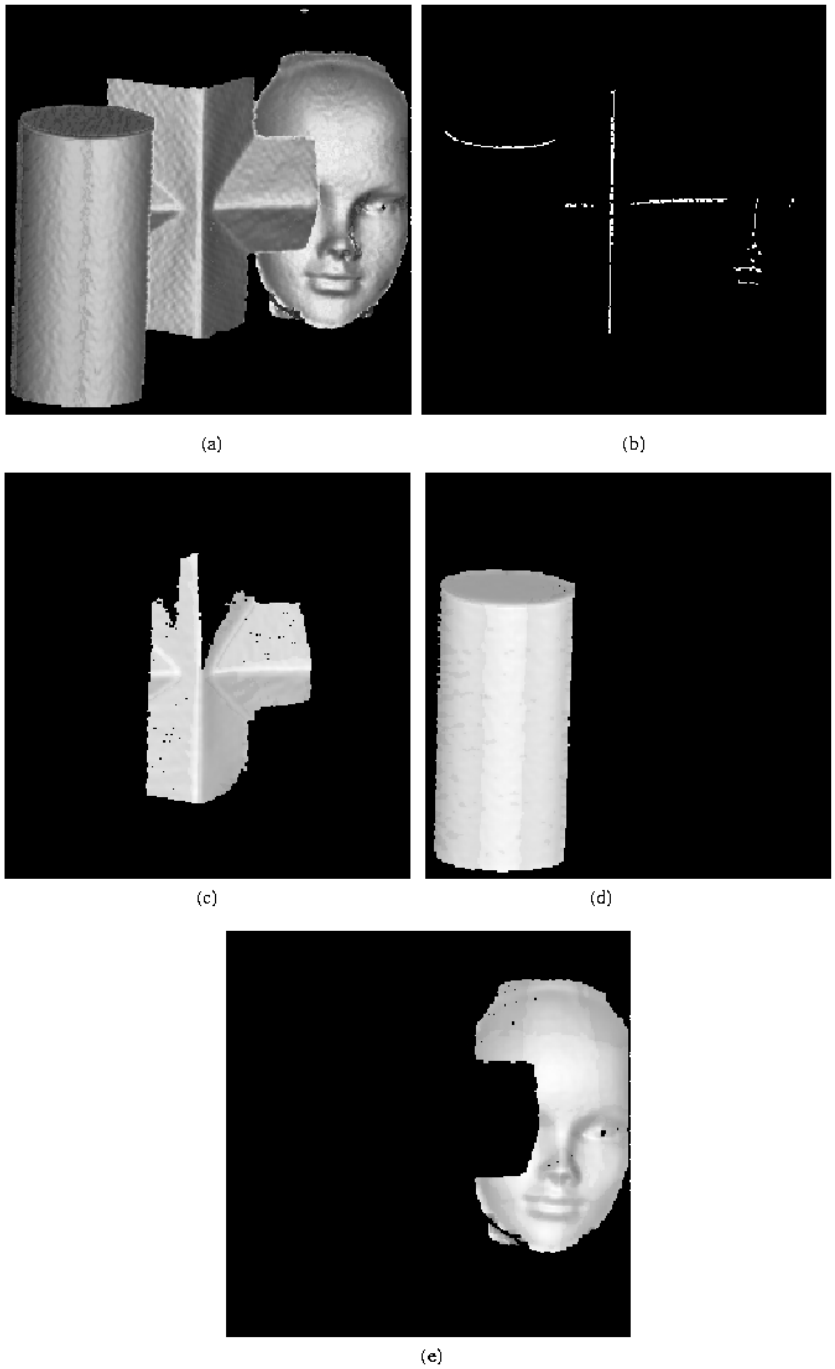


Fig. 17. (a) The shaded range image of scene 2. (b) The HV-skeletons extracted from the range image shown in (a). (c) The first recognized object (after 96 trials of feasible basis selection). (d) The second recognized object (after 221 trials). (e) The third recognized object (after 229 trials).

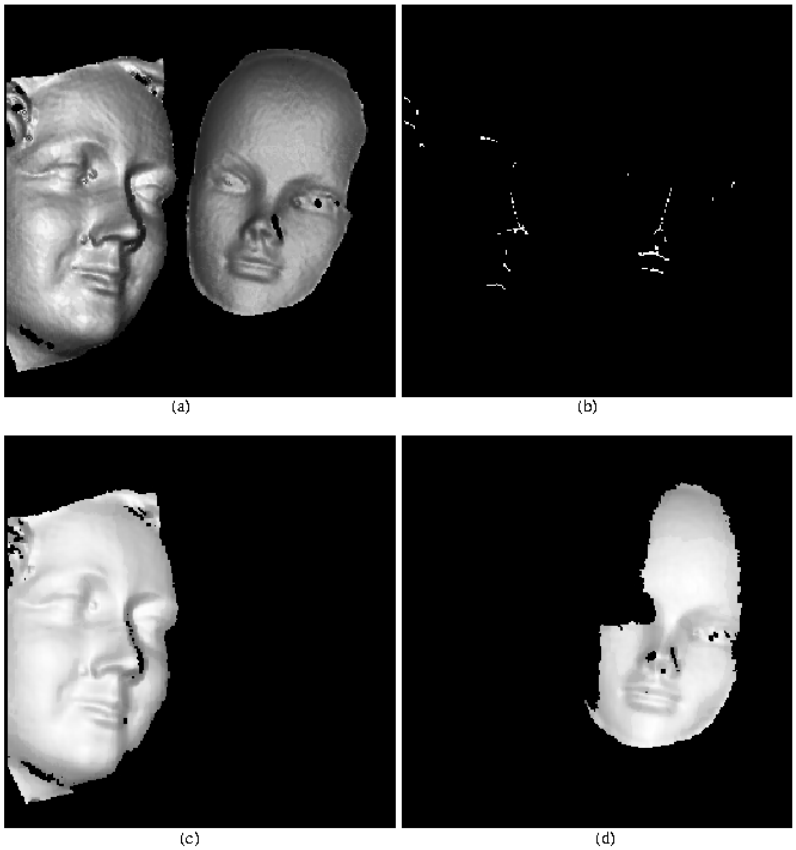


Fig. 18. (a) The shaded range image of scene 3. (b) The HV-skeletons extracted from the range image shown in (a). (c) The first recognized object (after 78 trials of feasible basis selection). (d) The second recognized object (after 95 trials).

ACKNOWLEDGEMENTS

This work was supported in part by the National Science Council, Republic of China, under Grant NSC 84-2213-E-001-007. We appreciate one of the anonymous reviewers very much for his careful correction of the grammatical errors of this paper.

REFERENCES

1. F. Arman and J. K. Aggarwal, "Model-based object recognition in dense-range images – A review," *ACM Computing Surveys*, Vol. 25, No. 1, 1993, pp. 125-145.
2. K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-square fitting of two 3-D point set," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 9, No. 5, 1987, pp. 698-700.
3. P. J. Besl and N. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, 1992, pp. 239-256.

4. C. H. Chen and A. C. Kak, "A robot vision system for recognizing 3-D objects in low-order polynomial time," *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 19, No. 6, 1989, pp. 1535-1563.
5. C. S. Chen, Y. P. Hung, C. C. Chiang, and J. L. Wu, "Range data acquisition using color structured lighting and stereo vision," *Image and Vision Computing*, Vol. 15, No. 6, 1997, pp. 445-456.
6. C. S. Chen, Y. P. Hung, and J. B. Cheng, "A fast automatic method for registration of partially-overlapping range images," in *Proceedings of Sixth International Conference on Computer Vision, ICCV '98*, 1998, pp. 242-248.
7. Y. Chen and G. Medioni, "Object modeling by registration of multiple range images," in *Proceedings of the 1991 IEEE International Conference on Robotics and Automation*, 1991, pp. 2724-2729
8. R. T. Chin and C. R. Dyer, "Model-based recognition in robot vision," *Computing Surveys*, Vol. 18, No. 1, 1986, pp. 67-107.
9. O. Faugeras, *Three-Dimensional Computer Vision*, Chap. 11. The MIT Press, 1993.
10. M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, Vol. 24, No. 6, 1981, pp. 381-395.
11. J. Foley, A. V. Dam, S. K. Feiner, and J. F. Hughes, *Computer Graphics : Principles and Practice*, Second Edition. Addison-Wesley, pp. 738-739, 1990.
12. D. M. Gavrila and F. C. A. Groen, "3D object recognition from 2D images using geometric hashing," *Pattern Recognition Letters*, 13, 1992, pp. 263-278.
13. W. E. L. Grimson and D. P. Huttenlocher, "On the sensitivity of geometric hashing," in *Proceedings of the Third International Conference on Computer Vision*, ICCV '92, 1992, pp. 334-338.
14. R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, Vol. 1, Addison-Wesley, 1992.
15. R. Krishnapuram and S. Gupta, "Morphological methods for detection and classification of edges in range images," *Journal of Mathematical Imaging and Vision*, Vol. 2, No. 4, 1992, pp. 351-375.
16. Y. Lamdan, J. T. Schwartz, and H. J. Wolfson, "Affine invariant model-based object recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, No. 5, 1990, pp. 578-589.
17. Y. Lamdan and H. J. Wolfson, "Geometric hashing: a general and efficient model-based recognition scheme," in *Proceedings of the Second International Conference on Computer Vision*, 1988, pp. 238-249.
18. Y. Lamdan and H. J. Wolfson, "On the error analysis of geometric hashing," in *Proceedings of the 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1992, pp. 22-27.
19. Z. Liu and D. Kersten, "2D affine transformations cannot account for human 3D Object recognition," in *Proceedings of Sixth International Conference on Computer Vision, ICCV '98*, 1998, pp. 549-554.
20. I. Rigoutsos and R. Hummel, "A Bayesian approach to model matching with geometric hashing," *Computer Vision and Image Understanding*, Vol. 62, No. 1, 1995, pp. 11-26.
21. A. Rosenfeld and J. L. Pfaltz, "Sequential operations in digital picture processing,"

Journal of the Association for Computing Machinery, Vol. 13, 1966, pp. 471-494.
22. F. Stein and G. Medioni, "Structural indexing: efficient 3-D object recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, 1992, pp. 125-145.
23. M. Suk and S. M. Bhandarkar, *Three-Dimensional Object Recognition from Range Images*, Chap 6, Springer-Verlag, 1992.
24. F. C. D. Tsai, "Geometric hashing with line features," *Pattern Recognition*, Vol. 27, No. 3, 1994, pp. 377-389.
25. J. G. Verly and R. L. Delanoy, "Adaptive mathematical morphology for range imagery," *IEEE Transactions on Image Processing*, Vol. 3, No. 5, 1993, pp. 272-275.
26. X. Yu, T. D. Bui, and A. Krzyzak, "Robust estimation for range image segmentation and reconstruction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 16, No. 5, 1994, pp. 530-538.

Chu-Song Chen (陳祝嵩) received a B.S. degree in Control Engineering from National Chiao Tung University, Hsinchu, Taiwan, in 1989. He received an M.S. degree in 1991 and Ph.D. degree in 1996, respectively, both from the Department of Computer Science and Information Engineering, National Taiwan University. He is now an assistant research fellow of the Institute of Information Science, Academia Sinica. Dr. Chen has received the good paper award and the outstanding paper award of the Image Processing and Pattern Recognition (IPPR) Society, Taiwan, in 1995 and 1997, respectively. He has received the best paper award of the Image Processing and Application Association (IPPA) of Taiwan, R.O.C. in 1997. His research interests include computer vision, augmented/virtual reality, and visual surveillance.

Yi-Ping Hung (洪一平) received his B.S. in electrical engineering from National Taiwan University in 1982. He received an M.S. from the Division of Engineering, and M.S. from the Division of Applied Mathematics, and a Ph.D. from the Division of Engineering at Brown University in 1987, 1988 and 1989, respectively. He then joined the Institute of Information Science, Academia Sinica, Taiwan, and has become a research fellow in 1997. He served as the deputy director of the Institute of Information Science from 1996 to 1997, and received the Outstanding Young Investigator Award given by Academia Sinica in 1997. He has been teaching in the Department of Computer Science and Information Engineering at National Taiwan University since 1990, and is now an adjunct professor. Dr. Hung has published more than 70 technical papers in the fields of computer vision, pattern recognition, image processing, and robotics. In addition to the above topics, his current research interests also include visual surveillance, virtual reality, human-computer interface and visual communication.

Ja-Ling Wu (吳家麟) received the B.S. degree in electronic engineering from Tamkang University, Tamshoei, Taiwan, R.O.C. IN 1979, and the M.S. and Ph.D. degree in electrical from Tatung Institute of Technology, Taipei, Taiwan, in 1981 and 1986, respectively.

From 1986 to 1987, he was an association professor of the Electrical Engineering Department at Tatung Institute of Technology, Taipei, Taiwan. Since 1987, he has been with the Department of Computer Science and Information Engineering, NTU, where he is presently a professor and the director of the communications and multimedia Lab. From 1996 to 1998, he was assigned as the head of the department of information engineering, National Chi Nan University, Puli, Taiwan.

He was the recipient of the Outstand Young Medal of the Republic of China and the outstanding research award sponsored by the National Science Council, from 1987 to 1994. He was the recipient of the Award for Distinguished Information People of the year (R.O.C. 1993), the special Long-Term Award for Collaboratory Research (1994), the Best Long-Term Paper Award (1995), and the Long-Term Medal for ten Distinguished Researcher (1996), all sponsored by the Acer Corp. Prof. Wu has published more than 200 technique and conference papers. His research interests include neural networks, VLSI signal processing, image coding, data compression, and multimedia systems.