Chapter 1

Introduction

1.1 Motivation

A computer model of shape data is a representation, possibly mathematical, that can be analyzed and used for a desired purpose. The ultimate objective of this research is to automatically construct a computer model of solid 3D (three-dimensional) polyhedral objects from a single range image. This model representation is based on extracting the boundary of the 3D object. The boundary is divided into a three-level hierarchy of entities: faces, edges, and vertices.

The traditional approaches to extracting 3D points from images are stereo reconstruction and structured-light analysis [29,48]. Stereo reconstruction requires establishing correspondences between points in two or more distinct images of the same scene. This approach is still a topic of research. Structured-light analysis methods attempt to infer 3D relationships from a single 2D intensity image of a scene. Although straightforward and generally reliable, structured light vision is applicable only in rather constrained situations, such as the inspection of industrial components of simple geometry. One of the situations in which this method fails is non-uniform illumination, which leads to shadow pixels that have invalid range measurements.

A range image is defined as a dense collection of range (distance) values. Range images directly provide geometrical information about the shape of visible 3D object surfaces. One goal of range image analysis is to extract and identify all objects of interest in a given image; the difficulty in this is range image segmentation. This typically refers to the subdivision of an image into non-overlapping regions, so that each region corresponds to a distinct 3D surface. Unfortunately, the problem of range image segmentation has proven to be surprisingly difficult, particularly in light of the fact that a range image directly provides geometrical information about a scene.

Automatic construction of computer models of 3D objects has important applications in the real world. The utilization of computer models is centered on tasks involving automation, measurement, analysis, entertainment, and rigorous processing power. For instance, in automated navigation—by land, air, or sea—the pertinent navigable space and obstacles must be modeled. In automated assembly, parts are modeled for purposes of interaction with robotic hardware such as arc welders, spray painters, and other tools. In manufacturing, products like automobiles are modeled for preliminary demonstration and testing before being built. In entertainment, real and imaginary objects are modeled for purposes of animation. In virtual reality, real and imaginary objects are modeled for purposes of simulated interaction, for both entertainment and pre-construction testing. In all these scenarios, computer models are needed.

1.2 Problem Statement

The overall goal of this research is to extract a computer model from an input range image so that it can be used in such post processing applications as 3D-object recognition. The constructed model is obtained as a high level description *D* from a given single range image *R*, as,

$$\begin{aligned}
D &= \{O_1, O_2, \ldots, O_n\} \qquad (1.1)\\
O_i &= \{S_i, V_i, G_i\}, \quad i = 1, \ldots, n\\
S_i &= \{s_{i1}, s_{i2}, \ldots, s_{im_i}\}\\
V_i &= \{v_{i1}, v_{i2}, \ldots, v_{ik_i}\}\\
G_i &= \{S_i, A_i\}\\
A_i &= \{\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{ih_i}\}\\
s_{ij} &= (v'_{i1}, v'_{i2}, \ldots, v'_{ip_i}), \text{ where } v'_{ij} \in V_i
\end{aligned}$$

The symbol O_i represents a description of object *i*, with *n* being the number of objects for the entire range image (*R*). For any object *i*, S_i represents the set of its visible faces that is composed of m_i planar faces, V_i represents the set of k_i visible vertices, and G_i represents a region adjacency graph (RAG) that consists of the face set S_i and an arc set A_i with h_i arcs. The first subscript for the upcoming symbols refers to the object number, while the second one refers to the entities within the object. The symbol v_{ij} represents a vertex *j* that has a 3D coordinate (*x*, *y*, *z*). The surface s_{ij} can be represented using a sequence of vertices, $v'_{ij} \in V_i$. The symbol α_{ij} represents an arc *j*.

The RAG represents the relative spatial position of the surfaces of the object in the image. A face is represented as a node in the RAG. An arc between any two nodes means that the two corresponding faces are adjacent in the image. The adjacency relation of the surfaces $\{s_{i1}, s_{i2}\}$ is represented as $s_{i1} \leftrightarrow s_{i2}$ which means, "s_{i1} is adjacent to s_{i2}."

For instance, Figure 1.1 shows an example of a range image and an extracted model of a 3D object. Figures 1.1(a) and (b) show the range and intensity images, Figure 1.1(c) shows the faces and vertices of the polyhedron, and Figure 1.1(d) shows the region adjacency graph of the polyhedron.

The model representation of the entire range image *R* of Figure 1.1 is illustrated as,

$$\begin{aligned}
D &= \{O_1\} \qquad (1.2)\\
O_1 &= \{S_1, V_1, G_1\}\\
S_1 &= \{s_{11}, s_{12}, s_{13}, s_{14}\}\\
V_1 &= \{v_{11}, v_{12}, v_{13}, v_{14}, v_{15}, v_{16}, v_{17}, v_{18}, v_{19}, v_{110}\}\\
G_1 &= \{S_1, A_1\}\\
A_1 &= \{\alpha_{11}, \alpha_{12}, \alpha_{13}, \alpha_{14}, \alpha_{15}\}
\end{aligned}$$

The image *R* has one polyhedron object ($n = 1$). This polyhedron has four faces ($m_1 = 4$), and ten vertices ($k_1 = 10$). Also, the number of arcs that represents the adjacency between faces is five ($h_1 = 5$). Table 1(a) illustrates each face as a sequence of object vertices and Table 1(b) presents the adjacency relation between the faces of the polyhedron.

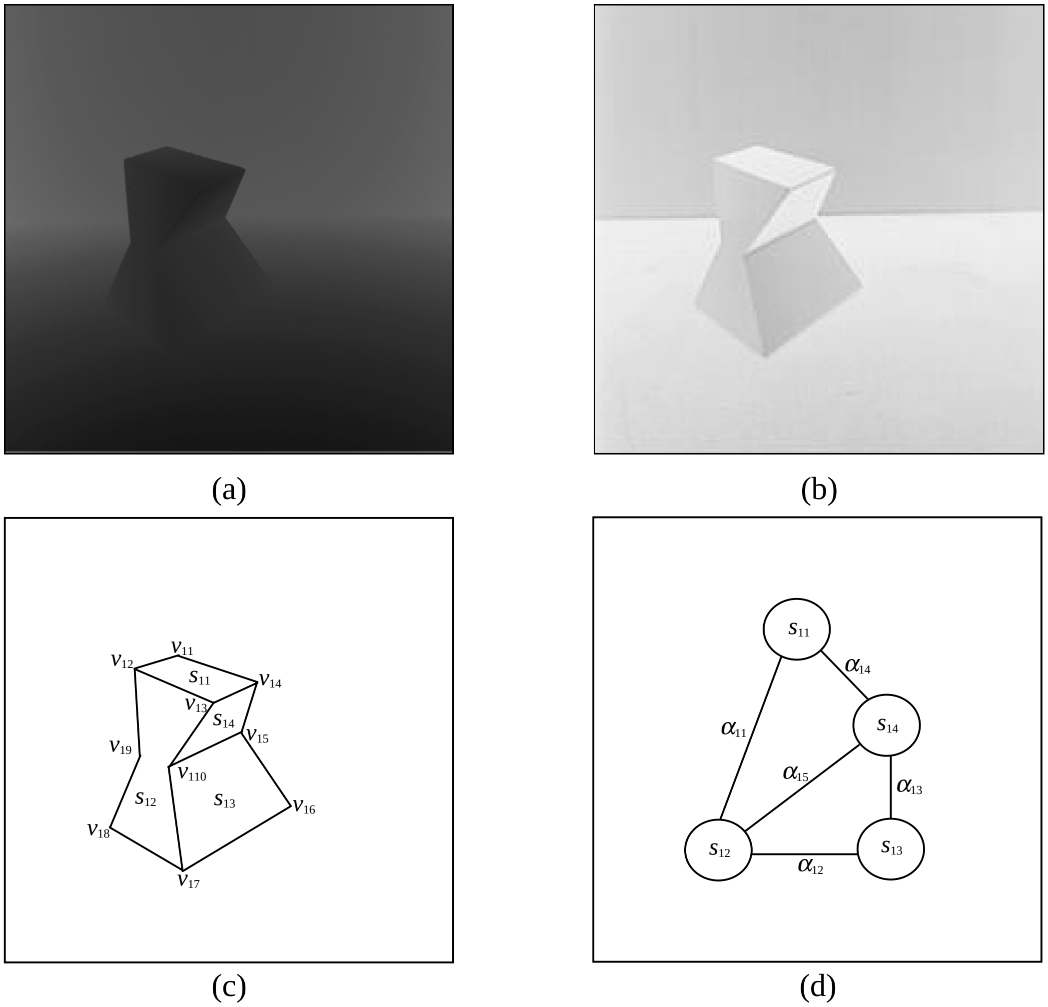


Figure 1.1 Example of model construction of single polyhedral object from a range image of size 512×512, obtained using a Perceptron range scanner. (a) Range image. (b) Intensity image. (c) Model representation. (d) Region adjacency graph (RAG) of the polyhedron.

Table 1.1 The face description in terms of vertices, and adjacency relation description in terms of faces. (a) The description of the faces as vertex sequences. (b) The description of the arcs as a mapping between each adjacent pair of faces.

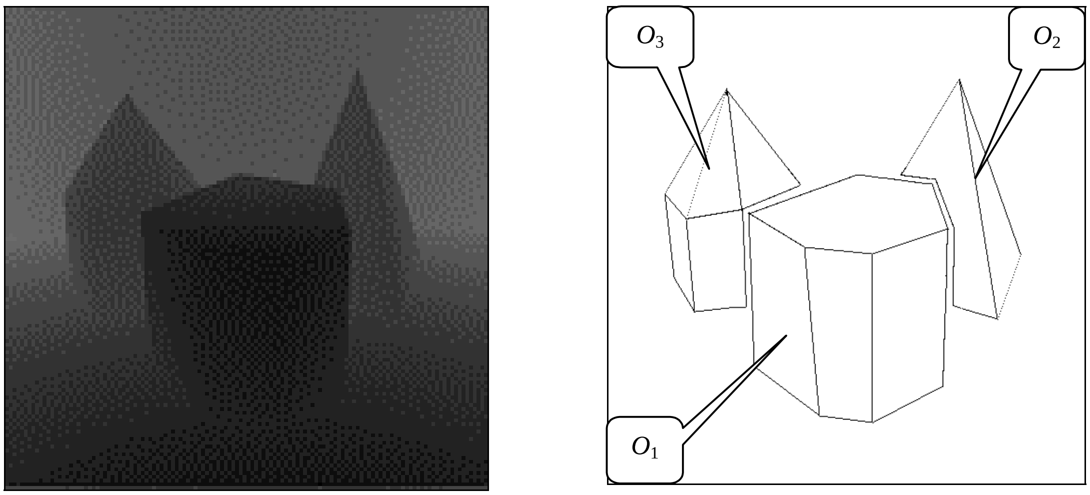
(a)

S_1	V_1 Sequence
s_{11}	$(v_{11}, v_{12}, v_{13}, v_{14})$
s_{12}	$(v_{12}, v_{19}, v_{18}, v_{17}, v_{110}, v_{13})$
s_{13}	$(v_{15}, v_{110}, v_{17}, v_{16})$
s_{14}	$(v_{14}, v_{13}, v_{110}, v_{15})$

(b)

A_1	S_1 Mapping
α_{11}	$s_{11} \leftrightarrow s_{12}$
α_{12}	$s_{12} \leftrightarrow s_{13}$
α_{13}	$s_{13} \leftrightarrow s_{14}$
α_{14}	$s_{14} \leftrightarrow s_{11}$
α_{15}	$s_{12} \leftrightarrow s_{14}$

Figure 1.2(a) illustrates an example of multiple overlapped polyhedral objects where one object (close to the camera) occludes two objects. In order to deal with such cases, a novel occlusion detection technique that isolates objects (O_1, O_2, and O_3) before representation has been developed, as shown in Figure 1.2(b). The occlusion detection technique is a part of the system of object extraction and representation and is discussed in Chapter 6 in detail.



(a) (b)

Figure 1.2 Example range image of multiple overlapped polyhedral objects, of size 512×512, obtained using a Perceptron range scanner. Inter-object occlusion example. (a) Range image. (b) Three extracted and isolated objects, with consideration of partially occluded faces.

1.3 Research Objectives and Contributions

The objective of this research is to develop an approach for the automatic construction of computer models of 3D solid objects from a single range image. This is obtained through bottom-up processing methods that will take a single range image as input and yield a computer model of the objects as output. The constructed model reduces the point set of the input image to a smaller data set in the form of a model.

To achieve this goal, the image must go through the following phases:

i. Range image noise reduction
ii. Edge detection
iii. Segmentation
iv. Initial vertex extraction
v. Occlusion detection
vi. Grouping planar faces into objects
vii. Global vertex optimization
viii. Model creation and object representation

i) *The range image noise reduction* phase is important due to the environment's effects on the image acquisition. Therefore, the first step is to reduce the noise level in the image. One of the most popular conventional techniques, referred to here as the mean-approach, is based on the idea of a neighborhood operator that estimates the mean value for the local neighborhood to compare it with the center pixel of the local neighborhood for outlier rejection [29,48]. The drawback of this operator is that it depends on the assumption that the data in the local neighborhood is normally distributed, which is not always true in practical applications. To address this drawback, the mean-approach is modified to use the least-median-square method. The main idea behind this adaptation is the robustness of the median to the outliers since the median has a 50% breakdown point, and this property carries over to least-median-squares method.

ii) In the *edge detection* phase, three methods of range image edge detection are considered. The first approach is based on extracting a set of points that encircle each point in the image and then applying the Fast Fourier Transform (FFT) to this set of points. Since the edge points are characterized by higher harmonics, when these harmonics exceed a certain threshold, the point is considered to be an edge point. This technique was used mainly in synthetic image [46] and is modified to be suitable for detecting the edges of real range image. The second approach is based on detecting the change of surface normal vector direction. A large change is associated with an edge. To compute these surface normals, traditional approaches use the biquadratic surface fitting locally around each point in the image [79,89]. This method did not yield good results for the real range images used in this work. However, after making some minor modifications to the method, much better results were obtained by increasing the order of fitting from biquadratic to the bicubic. Finally, a novel edge detection technique based on detecting the change in gradient direction using adaptive Gaussian filter is implemented. It uses a simple difference operator with adaptive window sizing to obtain estimates of 2-D gradient orientation. Edges are then detected by applying a difference operator to the gradient orientation and retaining those points for which the magnitude of the difference gradient exceeds a threshold.

iii) In the *segmentation phase,* a novel hybrid segmentation approach (region-edge based) is introduced to extract object faces from a single range image efficiently. The segmentation process depends on collaborating edge and region techniques where they give complementary information about the scene. Since this approach is dependent on the initial contours of the objects, its efficiency is affected by the accuracy of the edge detection technique. The novel edge detection technique described above provides high accuracy. Region boundaries are also improved using an iterative refinement method based on the geometry and topology of the objects. This step is very important in post-processing of vertex extraction and occlusion detection using the multiple evidence-based method.

iv) For the purpose of the *initial vertex extraction*, corner detection techniques have to be used. Corner detection techniques are commonly classified into two major categories; gradient-based and contour-based. The gradient-based approach is implemented by computing the second gradient of the image points and comparing these values of the second gradient to a certain threshold. If the gradient value exceeds the threshold, the point is considered to be a corner [23,68,71]. This approach is used mostly with intensity images. Applying this approach to the range images gave poor results, so; the contour-based approaches were evaluated. In this approach, the corners of the individual regions for the segmented image are identified in 2D. The Euclidean distances between the corners are calculated, adjacent corners are merged to form vertices, and a transformation to the 2D coordinate is applied to find the 3D coordinates. Determination of the 2D region corners is accomplished with two statistical approaches, both of which use chain encoded digital arcs for the region contour to isolate corners. The first approach is classified as an angle detection method [9,31,49,53]. In the second approach, the contour of a region is fit to a line model using the robust least-median-square method. A novel analysis of the residuals is performed due to fitting to find the corners. The performance of the second method is proven to be better than that of the first in terms of correct, missed, and false corner detection.

v) *Occlusion detection* is defined as the separation of 3D overlapped objects in a certain scene to construct the model of each individual object. The conventional methods of occlusion detection are divided into two categories: pattern recognition methods, and line and vertex labeling methods. Pattern recognition methods require the availability of a library of known shapes to guide the extraction process. The inefficiency of line and vertex labeling methods lies in assuming perfect segmentation and trihedral objects, which are impractical assumptions, and in the complexity of the approach. Therefore, two novel approaches were proposed for occlusion detection: layer-based and multiple evidence-based reasoning techniques. In the layer-based method, a histogram of the distance values from a given range image is clustered into separate modes. Ideally, each mode of the histogram will be associated with one or more surfaces having approximately the same distance from the sensor, which corresponds to a certain layer in the image. A procedural approach utilizes statistical reasoning for the topology and geometrical information of the 3D objects. The approach detects occlusion at the initially extracted vertices and checks all adjacent pairs of faces associated with these vertices to decide if there is occlusion between any adjacent pairs. A decision to disassociate a face is made based on the certainty factor value resultant from combining the measure of beliefs of the rules applied to the adjacent pair of faces. This approach is illustrated in detail in Chapter 6.

vi) The motivation behind *grouping planar faces into objects* is to reconstruct the polyhedrons in the range image after detecting occlusion, since occlusion detection is accomplished based on the local information of the pairwise face adjacency relationship. The approach depends on face adjacency and on the presence or absence of occlusion between adjacent faces. Face adjacency information is available in the region adjacency graph (RAG) that results from image segmentation. For any pair of regions that are adjacent according to the RAG, if one of the respective faces occludes the other, an arc is removed from the graph to separate them, indicating that they belong to different objects. This novel process of removing arcs and grouping faces based on occlusion has the effect of partitioning the RAG into subgraphs, each corresponding to a separate object.

vii) A g*lobal vertex optimization* procedure has been implemented to improve the 3D vertex locations for the polyhedral objects. The initial coordinates of the 3D vertices obtained via (2D) region shape and corner proximity are often quite poor in accuracy. This initial estimate of vertex coordinates is improved using an iterative search procedure that is formulated as a problem of energy minimization by minimizing the distances between the vertices and the faces to which the vertices belong. While traditional methods improve single vertex location, this method is a global improvement of all the vertices associated with one object. The improvement of the vertex locations procedure is novel from the point of global optimization.

viii) In the *model creation and object representations,* the polyhedral objects representations are described using the region adjacency graph (RAG) paradigm.

1.4 Outline of Dissertation

The rest of this dissertation is organized as follows. Chapter 2 presents a brief literature review and the background of some concepts of computer vision related to the research in this area. Chapter 3 presents an overview of laser camera types that are used to acquire range data for experimentation. Chapter 4 introduces the noise estimation, edge detection, segmentation, and initial vertex extraction that are used to extract 3D-object representation from a single range image. Chapter 5 presents an algorithm of occlusion detection that is based on analysis of the histograms for the range image. Chapter 6 introduces the algorithm of occlusion detection using a multiple evidence-based approach. Chapter 7 presents experimental, to demonstrate the robustness of the different algorithms used in the range images extraction, occlusion detection and vertex location optimization. Chapter 8 presents a conclusion and suggestions for future work.