Chapter 2

Background

2.1 Literature Review

Object extraction and surface description for 3D objects are active areas of research. In many cases, the extraction of surface representations is based on the image segmentation. Range image segmentation refers to the subdivision of an image into non-overlapping regions, so that each region corresponds to a distinct 3D (three-dimensional) surface. Common techniques for segmentation are classified as region-based, edge-based, and hybrid, which implies the consideration of regions and edges together [89]. In order to extract 3D objects, occlusion detection must be performed. Two common techniques for occlusion detection can be distinguished: pattern recognition methods, and line and vertex labeling methods. The following paragraphs present a brief review of work done in this area.

Besl and Jain [8] described a range image region-based segmentation technique. The mean and Gaussian curvature (H-K sign map) are computed by finding the first and second partial derivatives of the range image. The H-K sign map is used to classify each range pixel as one of eight qualitative surface types: peak, pit, ridge, saddle ridge, saddle valley, planar, minimal and valley. A region-growing algorithm is used to segment the range image surface into homogeneous surface patches based on the H-K sign map.

Taylor [81] introduced region-based range image segmentation by grouping homogenous regions based on surface orientation. Successive merging is applied to whole regions within threshold value.

Bose [12] proposed an algorithm that used differential invariants derived from classical differential geometry. An efficient algorithm for derivation of surface curvature at smooth surface patches is proposed such that the surface is approximated by Bezier and beta splines to qualitatively compare the proposed segmentation scheme. This scheme leads to the derivation of surface features, as seen in the integrated approach in terms of plane, quadric, and superquadric surfaces.

Hoover [42] developed an algorithm for range image segmentation based on a region growing technique. The main idea of this approach depends on the selection of initial seed points. These points are lying in the in the interior of each region. An iterative growing around each initial point is performed until all points of the image take labels. The growing criterion relies on the best plane fitted orientation parameters to each pixel. Flynn [43] developed an algorithm that uses clustering of pixels with similar 3D orientation. Both Hoover and Flynn used the region-based segmentation technique for range images.

Jia-Guu [35] proposed an edge-based approach that slices the range image at fixed increment range values to obtain a map of equidistant contours. The edges of the range image are determined by finding the places where a crossing occurs between the contours and object surface.

Jiang [50] developed an algorithm based on the fact that points on a scan line that belong to planar surfaces form a straight 3D line segment, and thus each scan line can be sub-divided into line segments lying in various planes.

Yokoya [89] presented a hybrid segmentation approach. The mean and Gaussian curvature are computed. A roof edge map is computed based on surface normal discontinuity. A step edge map is computed based on depth discontinuity. Finally, the three maps are combined to get the final segmentation.

The Zhang algorithm [90] consists of three steps: edge and critical point detection, triangulation, and region growing. Edge operators are applied to detect edges. Critical points are then extracted from the edge curve forming an initial set of segments. The constrained Delaunay triangulation is employed on the initial set to obtain a triangle-like connected graph by projecting the critical points and their connectivity relationship in parallel onto a 3D surface, and a 3D surface structure graph (SSG) is obtained. Hence, grouping these triangles like facets completes segmentation. This segmentation process could be classified as a hybrid approach.

Wu [87] proposed an approach modeled after electrical charges on the surfaces of a conductor. Charges tend to accumulate at areas of sharp convexity and to vanish at

locations of sharp concavity. The object boundaries are usually identified by a sharp concavity that can be detected by simulating the electrical charge density over the object surface. Surface points are located in a way to exhibit minimal local charge density.

Muraki [60] proposed a volumetric shape description using a “blobby model” for automatically generating a shape description from range data. This model can express a 3D surface as an isosurface of the scalar field, which is produced by a number of primitives generating fields. The fields from many primitives are blended with each other and can form a very complicated shape. To determine the number and distribution of primitives required to adequately represent a complex 3D surface, an energy function is minimized. This minimized function measures the shape difference between the range data and their model.

Chen [17] proposed the “balloon” model for multiple range images based on generating a triangulated mesh. The vertices in the mesh are linked to their neighboring vertices through springs to simulate the surface tension and to keep the shell smooth. This model is driven by an applied inflation force towards the object surface from inside the object. Next, this process is repeated until the mesh elements reach the object surface. Finally, the balloon model takes the shape of the object.

Allen [1] described a multi-sensor system for robotic recognition tasks that integrates passive stereo vision and active, exploratory tactile sensing. Also, Bay [5] presented a surface exploration approach via single and multi fingered hand.

Arman and Aggarwal [3] addressed the problem of recognizing an object in a given scene using a three-dimensional model of the object. The scene may contain several overlapping objects that are arbitrarily positioned and oriented. A laser scanner is used to collect three-dimensional (3D) data points from the scene. The collected data is segmented into surface patches, which enables the calculation of surface properties. The models are analyzed off-line to derive various geometric features, whose relationships and attributes are used for the task of object recognition.

Ding [24] addressed two important issues of shape from contour research. The first issue is the imperfection of input data, and the second is the completeness of output 3D object models. He presented a solution based on: (1) the analysis of imperfect line drawing configurations, especially the properties of L junctions, and (2) the implementation of a rule based system. Constraints imposed by line labeling, 2D and 3D geometry, and Gestalt principles of human visual perception are formulated into production rules. The system control mechanism is designed in a way that allows these rules to work interactively.

The above approaches provide a clear idea of the techniques that are commonly used in order to obtain 3D objects representations from range images. The hybrid method of range image segmentation will be introduced in this research because of its efficiency in collaborating information concerning both edges and regions of range images. Then, a multiple evidence-based occlusion detection technique will be implemented to detect occlusion, but in a different way than Ding [24]. The new method considers occlusion detection between each adjacent pair of regions for the segment image rather than the line and vertex labeling algorithm.

2.2 Segmentation of Range Images

2.2.1 Overview

The mathematical model for image segmentation is defined as follows:
Let R represent the entire range image. We may view the segmentation as a process that partitions R into n regions, R_1, R_2 … R_n such that [65]:

(a) $\bigcup_{i=1}^{n} R_i = R$,

(b) R_i is connected region, $i=1, 2, \ldots, n$,

(c) $R_i \bigcap R_j = \varnothing$ for all i and j, $i \neq j$

(d) $P(R_i) = 1$ for $i=1, \ldots, n$, and

(e) $P(R_i \bigcup R_j) = 0$ for $i \neq j$, (2.1)

where $P(R_i)$ is a logical predicate over the points in set R_i and $\varnothing$ is the null set.

Condition (a) indicates that the segmentation must be completed in a way that every pixel must be in a region. Condition (b) requires that points in a region must be connected. Condition (c) implies that the regions must be nonoverlapping. Condition (d) deals with the properties that must be satisfied by the pixels in a segmented region; e.g., $P(R_i)=1$ if all pixels in R_i have the same range value. Finally, condition (e) indicates that regions R_i and R_j are different in the sense of predicate *P*. Another important condition for the segmentation process is that the smallest allowable region size $A_{\min}$ should be specified to avoid small false regions:

$$A_{\min} \leq \min_{1 \leq i \leq n} |R_i| \tag{2.2}$$

The most common techniques for segmentation of range images can be classified as region-based [42,44], edge-based [46,79], and hybrid, which combine both regions and edges together [89]. The following sections present a brief discussion of these segmentation techniques.

2.2.2 Edge-Based Segmentation Techniques

Edge-based segmentation techniques depend on edge detection for images. The two edge types of particular interest in range images are step (jump) edges and roof (crease) edges. A step edge is a discontinuity in depth, whereas a roof edge refers to a discontinuity in surface normal direction. Step edges are often between visible surfaces and occluded (non-visible) surfaces, whereas roof edges are usually between two visible surfaces. Roof edges are more difficult to detect reliably, and traditional approaches of edge detection often do not perform well for roof edges in real images.

Common edge detection approaches for range images are known for being mathematical morphological operations and directional derivative methods. These methods have some drawbacks and advantages. A major disadvantage is that the edges may not be connected or their contours may not be closed. Also, these methods do not provide information about different surfaces. A key advantage with these methods is that they provide information about the boundaries of each region.

2.2.3 Region-Based Segmentation Techniques

The main idea of the region-based approaches is to estimate the surface curvature at each range pixel and to cluster range pixels with homogeneous surface curvature properties. These methods often yield poor performance in delimiting the region boundaries. One of the most commonly used methods of region-based segmentation depends on the calculation of Gaussian and mean curvature values, which are used to characterize the point of interest on the surface as one of eight descriptive types.

Calculating derivatives of digital surfaces is essential in Gaussian and mean curvature computation. Direct numerical differentiation is generally discouraged on all but the cleanest signal data because it tends to amplify noise and obscure signal content. Better derivative estimates have been obtained by combining data smoothing techniques with the difference operations needed to estimate a derivative. The basic approach is to determine a continuous differentiable function that best fits a discrete sample data with respect to some criterion. The next step consists of computing the derivatives of the continuous function analytically and evaluating them at the corresponding discrete data points. A local bicubic surface fitting method is used to estimate the function $z = f(x, y)$ for each pixel within a specific window such that the pixel of interest is the center of that window.

The partial derivatives $z_x, z_y, z_{xx}, z_{yy}, z_{xy}$, and z_{yx} of the fitted functions are computed [38,89]. Then, the mean curvature (H) and Gaussian curvature (K) are computed at each pixel denoted by r and c, the row and column indices in the image, as follows [6]

$$H(r,c) = \frac{\left(1+z_y^2(r,c)\right)z_{xx}(r,c) + \left(1+z_x^2(r,c)\right)z_{yy}(r,c) - 2z_x(r,c)z_y(r,c)z_{xy}(r,c)}{2\left(\sqrt{1+z_x^2(r,c)+z_y^2(r,c)}\right)^3} \tag{2.3}$$

$$K(r,c) = \frac{z_{xx}(r,c)z_{yy}(r,c) - z_{xy}^2(r,c)}{\left(1+z_x^2(r,c)+z_y^2(r,c)\right)^2} \tag{2.4}$$

Due to noise variation and quantization effects, the thresholds about zero for the Gaussian and mean curvatures are selected symmetrically. They are referred to as K_0 and H_0, respectively; that is, they are assigned $K = 0$ for $-K_0 \le K \le K_0$ and $H = 0$ for

$-H_0 \le H \le H_0$. From the definition of *K* and *H*, the constraint on the selection of K_0 and H_0 are $K_0 \approx H_0^2$ and $K_0 \ge H_0^2$.

Table 2.1 represents the eight surface types that each pixel can have, and this depends on the Gaussian and mean curvature signs [37,38]. As an example, if the value of *H* and *K* are equal to zero, then this pixel represents a flat surface point, etc.

Table 2.1 Surface types determined by surface curvature signs.

	$K<0$	$K=0$	$K>0$
$H<0$	Peak	Ridge	Saddle
$H=0$	(None)	Flat	Minimal
$H>0$	Pit	Valley	Saddle Valley

Let $\text{sgn}_\varepsilon(x)$ be defined as follows

$$\text{sgn}_\varepsilon(x) = \begin{cases} +1 & x > \varepsilon \\ 0 & x = \varepsilon \\ -1 & x < \varepsilon \end{cases} \tag{2.5}$$

The *H-K* sign map is then constructed by means of labeling each pixel in the range image with a value from 1 to 9, excluding 4 corresponding to the surface type:

$$\text{sgn}_\varepsilon(r,c) = 1 + 3(1 + \text{sgn}_\varepsilon(H(r,c))) + (1 + \text{sgn}_\varepsilon(K(r,c))) \tag{2.6}$$

Three non-negative images $|H|$, $|K|$, and $\sqrt{H^2 - K}$ describe the magnitude of the mean curvature, Gaussian curvature, and scaled magnitude of the difference of the principle curvature, respectively. These images are also used in finding the step and roof edges. Surface area $\sqrt{g}$ and smoothness/flatness *Q* are defined as:

$$\sqrt{g} = \sqrt{1 + z_x^2 + z_y^2} \tag{2.7}$$

$$Q = z_{xx}^2 + 2z_{xy}^2 + z_{yy}^2 \tag{2.8}$$

The fitting error (ε) defined below is a measure of the reliability of the partial derivative estimates, and therefore indicates the reliability of the H-K sign map and all characteristics:

$$\varepsilon^2 = \frac{1}{n}\sum_{i=1}^{n}\left(z_i - z(x_i, y_i)\right)^2 \tag{2.9}$$

Figure 2.1 demonstrates a common approach of region-based range image segmentation after Besl [6,8]. This approach begins by finding the partial derivatives for the range image, computing the mean and Gaussian curvature (H, K) at each point for the image, and labeling each range pixel with the surface type. The regions need to be shrunken to eliminate false labels near region boundaries.

A sequential connected component algorithm is applied to identify core regions. The core regions that are too small have to be removed. For the initial core regions, a least square method is applied to fit a bivariate patch. For each region, the set of neighboring pixels with values close to that of the surface patch evaluated at the location of the neighbor is determined (region growing). These pixels are considered candidates for inclusion in the region. The union of the set of pixels in the region and the set of candidates are refitted to construct a new surface patch. If the fit is acceptable, the set of candidate pixels is added to the region; otherwise, the set of candidate pixels is discarded. The steps of region growing and refitting are repeated until no region is changed.

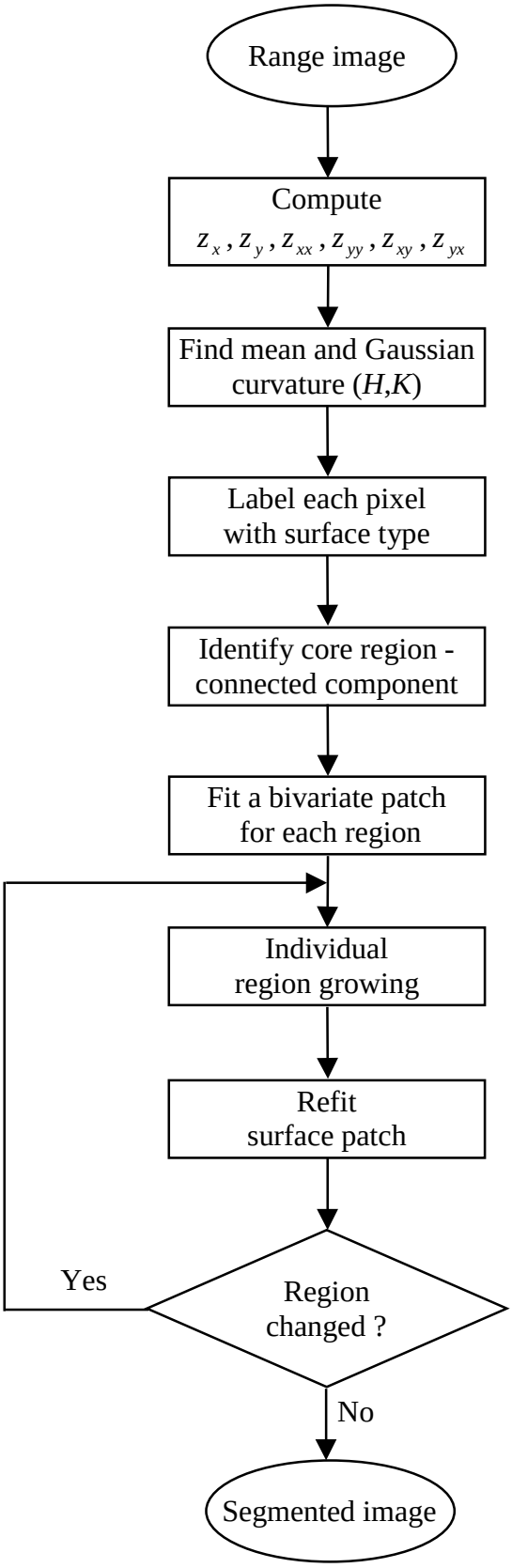

Figure 2.1 Algorithm for region-based segmentation of range images [6,8].

2.2.4 Hybrid Segmentation Techniques

Edge-based and region-based methods for range image segmentation provide complementary information about the scene; hybrid approaches attempt to combine the best aspects of these approaches. These approaches utilize the information inferred about the edges from different techniques, two of which are common. The first technique is the construction of closed contours. These contours may not be correct due to the problems of edge detection (e.g. missing and false edges). Missing and false edges can be solved using split and merge processes, respectively. In the second technique, the surface region is allowed to grow but not pass over initial edge points considered to be the confident surface boundaries. This process converges after a small number of iterations and produces a better correspondence between the shapes of the regions and the surfaces of the objects.

As an application for the second technique, Yokoya [89] used the *K*-*H* sign map with point values ranging from 1-8 according to the surface types. Meanwhile, the edge map has taken –2 for jump edges and –1 for the roof edges. Combining both maps, a region-edge map with values (-2, -1, 1-8) is obtained. Based on this combination, regions with the same label are clustered to constitute an initial region, ignoring regions with an area less than the minimum area size according to equation (2.2). Finally, the initial regions are expanded to erase the negative labels of the edges.

2.3 Segmentation Errors

The segmentation process can lead to several different types of low-level error [65]. The boundary error occurs when most of a region is correctly segmented, but the boundary of the region is not correctly positioned in the image. Therefore, the region boundary must be substituted. The second error is the region merge error, which occurs when two regions are merged even though a step or orientation discontinuity exists between the regions.

A new region boundary must be inserted to correct the error. The failure-to-merge error occurs when two regions that should have been merged are not merged and are separated by a meaningless boundary. A region boundary must be deleted to correct the error. In the same-region-label error, two non-adjacent regions are given the same label even though they should not be grouped together. This requires a new region label. The

final error is a different-region-label error, which when two non-adjacent regions are given different labels even though they should be grouped together. One label must be substituted for another to correct the error [6,65].

These errors require higher-level knowledge for detection and correction. If such errors could be detected using low-level mechanisms, those mechanisms could theoretically be used to correct the segmentation errors. The concept of distance between two regions is considered a reasonable metric for quantitative evaluation and comparison of segmentation, although it does not detect all errors.

Let $S = \{R_i\}$ be the "correct" segmentation of the digital images R with n_R regions and let $\hat{S} = \{\hat{R}_i\}$ be the automatically computed segmentation derived from the data, with $\hat{n}_R$ regions. Let n = min (n_R , $\hat{n}_R$) be the number of correspondence between regions in the two segmentations. A simple image segmentation error metric $e(S, \hat{S})$ for the image R is shown as follows

$$e(S, \hat{S}) = \frac{1}{2|R|} \sum_{i=1}^{n} \left| R_i \Delta \hat{R}_i \right| \tag{2.10}$$

where the symmetric difference operation between two sets is defined as $R_i \Delta \hat{R}_i = \left| (R_i - \hat{R}_i) \cup (\hat{R}_i - R_i) \right|$. The error metric ranges from zero (if both segmentations are identical) to one (if a completely wrong correspondence is established where all corresponding regions are disjoint).