RANGE DATA RECOGNITION: SEGMENTATION, MATCHING, AND SIMILARITY
RETRIEVAL

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

NESLİHAN YALÇIN BAYRAMOĞLU

IN PARTIAL FULLFILLTMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILISOPHY
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

SEPTEMBER, 2011

Approval of the thesis

RANGE DATA RECOGNITION: SEGMENTATION, MATCHING, AND SIMILARITY RETRIEVAL

Submitted by **NESLİHAN YALÇIN BAYRAMOĞLU** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy** in **Electrical and Electronics Engineering Department, Middle East technical University** by,

Prof. Dr. Canan ÖZGEN
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. İsmet ERKMEN
Head of Department, **Electrical and Electronics Engineering** _____

Prof. Dr. A. Aydın ALATAN
Supervisor, **Electrical and Electronics Engineering Dept.** _____

Examining Committee Members:

Prof. Dr. Uğur HALICI
Electrical and Electronics Engineering Dept., METU _____

Prof. Dr. A. Aydın ALATAN
Electrical and Electronics Engineering Dept., METU _____

Assist. Prof. Dr. Pınar DUYGULU ŞAHİN
Computer Engineering Dept., Bilkent Univertsity _____

Prof. Dr. Gözde BOZDAĞI AKAR
Electrical and Electronics Engineering Dept., METU _____

Assoc. Prof. Dr. Uğur GÜDÜKBAY
Computer Engineering Dept., Bilkent Univertsity _____

Date: _____

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name :   Neslihan, Yalçın Bayramoğlu

Signature   :

ABSTRACT

RANGE DATA RECOGNITION: SEGMENTATION, MATCHING, AND SIMILARITY RETRIEVAL

Yalçın Bayramoğlu, Neslihan

Ph.D., Department of Electrical and Electronics Engineering

Supervisor: Prof. Dr. A. Aydın Alatan

September 2011, 132 pages

The improvements in 3D scanning technologies have led the necessity for managing range image databases. Hence, the requirement of describing and indexing this type of data arises. Up to now, rather much work is achieved on capturing, transmission and visualization; however, there is still a gap in the 3D semantic analysis between the requirements of the applications and the obtained results. In this thesis we studied 3D semantic analysis of range data. Under this broad title we address segmentation of range scenes, correspondence matching of range images and the similarity retrieval of range models. Inputs are considered as single view depth images. First, possible research topics related to 3D semantic analysis are introduced. Planar structure detection in range scenes are analyzed and some modifications on available methods are proposed. Also, a novel algorithm to segment 3D point cloud (obtained via TOF camera) into objects by using the spatial information is presented. We proposed a novel local range image matching method that combines 3D surface properties with the 2D scale invariant feature transform. Next, our proposal for retrieving similar models where the query and the database both consist of only range models is presented. Finally, analysis of heat diffusion process on range data is presented. Challenges and some experimental results are presented

Keywords: Range image, segmentation, similarity retrieval, correspondence matching, 3D data, semantic scene analysis.

ÖZ

DERİNLİK VERİLERİNİN TANINMASI: BÖLÜTLEME, EŞLEME VE BENZERLİK ÇIKARIMI

Yalçın Bayramoğlu, Neslihan

Doktora, Elektrik ve Elektronik Mühendisliği Bölümü

Tez Yöneticisi : Prof. Dr. A. Aydın Alatan

Eylül 2011,   132 sayfa

3B tarama teknolojilerinin gelişmesi, derinlik görüntüsü veri tabanlarının yönetilmesi ihtiyacını doğurmuştur. Bu nedenle, bu tip verilerin tanımlanması ve indekslenmesi gereksinimleri ortaya çıkmıştır. Günümüze kadar çoğunlukla 3B verilerinin elde edilmesi, iletimi ve görüntülenmesi üzerinde çalışmış, ancak hala 3B verilerin anlamsal analizi açısından uygulamaların ihtiyaçları ve eldeki sonuçlar arasında bir boşluk bulunmaktadır.. Bu tezde 3B derinlik görüntülerinin anlamsal analizi çalışılmıştır. Bu geniş başlık altında derinlik görüntülerinin bölütlenmesi, karşılık eşleşmesi ve derinlik modellerinin benzerlik çıkarımı konuları incelenmiştir. Tezde ele alınan görüntüler tek bakışlı derinlik haritalarından oluşmaktadır. Öncelikle 3B verilerin anlamsal analizi ile ilgili olası araştırma konuları tanıtılmıştır. Düzlemsel yapıların belirlenmesi araştırılmış, varolan yöntemler için değişiklikler önerilmiştir. Uçuş zamanı kameralarından elde edilen nokta kümesi verilerinin uzamsal bilgilerinin kullanılarak bölütlenmesi için özgün bir yöntem de tezde önerilmektedir. Ayrıca derinlik görüntülerinin eşleşmesi için 3B yüzey özelliklerinin 2B değişimsiz özellik dönüşümü ile birleştirilmesi önerilmiştir. Sonra, derinlik görüntülerinin benzerlik çıkarımı için önerilen yöntem sunulmuştur. Son olarak derinlik görüntülerinde ısı transferi incelenmiş, problemler ve deneysel sonuçlar sunulmuştur.

Anahtar Kelimeler: Derinlik görüntüsü, bölütleme, benzerlik çıkarımı, karşılık eşleme, nokta kümesi, anlamsal sahne analizi.

This thesis is dedicated to
Dr. Muhammet Fatih Bayramoğlu
Hatice Yalçın
Ali Yalçın
Fatma Yalçın Öz
Oğuzhan Yalçın
Enes Öz

ACKNOWLEDGMENTS

everything in the world. I thank her son, Enes, my little nephew, for being the sweetest child, for being so cute, clever, and loving me.

My parents Ali Yalçın and Hatice Yalçın deserve to have all the best things on Earth. I wish I would be able to provide some. But rather than proving, I brought with me sorrow, trouble and distress. I want them to forgive me. I thank them for being altruistic, supportive, and compassionate. My mother is the best mother in the world. I remember her constant support, her love, and her goodness. I remember many sleepless nights with her accompanying me when I was suffering from tonsil. I love you mother.

And the Oscar goes to my husband Dr. Muhammet Fatih Bayramoğlu. I want to thank my life companion for sharing my burden. Without his support and love, I would not be able to complete this work. Almost everything that made the life bearable can be traced to his support without hesitation. I am indepted for his helps to overcome the problems that would otherwise insuperable. I thank him for his endless love and the support in every part of my life

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

The motivation of semantically describing and representing objects arise from the fact that there is a significant need for recognizing, organizing, classifying and searching the *content* of visual data. Since capturing, displaying and storing the 3D media increased rapidly and 3D data takes place in our daily life, such as TV (3DTV), laptops, chemistry (e.g. protein modeling), archeology (museum data), medicine, geography, military, industry (CAD), computer games, architecture, medical surgery, virtual reality programs, education and entertainment. Up to now, rather much work is achieved on capturing, transmission and visualization; however, there is still a gap in the 3D semantic analysis between the requirements of the applications and the obtained results.

Semantic information retrieval from 3D data is a broad title containing many different problems within. Segmentation, correspondence matching, verification (self matching), similarity retrieval (indexing), and classification can be considered in the first place (Figure 1-1). Problem variety also increases as the 3D input type (polygon mesh, point cloud, range image) vary. According to this difference, some possible research areas are presented in Table 1. This dissertation addresses segmentation, correspondence matching and similarity retrieval problems, where the input data are range images. Methods proposed to solve these problems are evaluated based on their performances on the single view depth images. However, we focus exclusively on similarity retrieval of range models with an extensive survey on the corresponding literature.

If the input is a scene containing numerous 3D objects with occlusions and clutter then the segmentation is inevitable for further processing. It is an important step that needs to be

performed prior to the object recognition, model fitting, tracking, indexing, etc.. For range data analysis, segmentation can be defined as the process of labeling range measurements belonging to the same surface or region with the same label. Segmentation is an ill-posed, open ended and a quite complex research within computer vision community regardless of the input type. One of the main reasons is the ambiguity in defining the ground truth.

Correspondence matching can be considered as an intermediate step during verification. Pose normalization and tracking type of application would also benefit from correspondence matching. Simply, locally similar points are matched among the two inputs. Typically, scaled, rotated and/or translated version of a particular object is presented to both of the inputs. Local descriptors are utilized for finding correspondences.
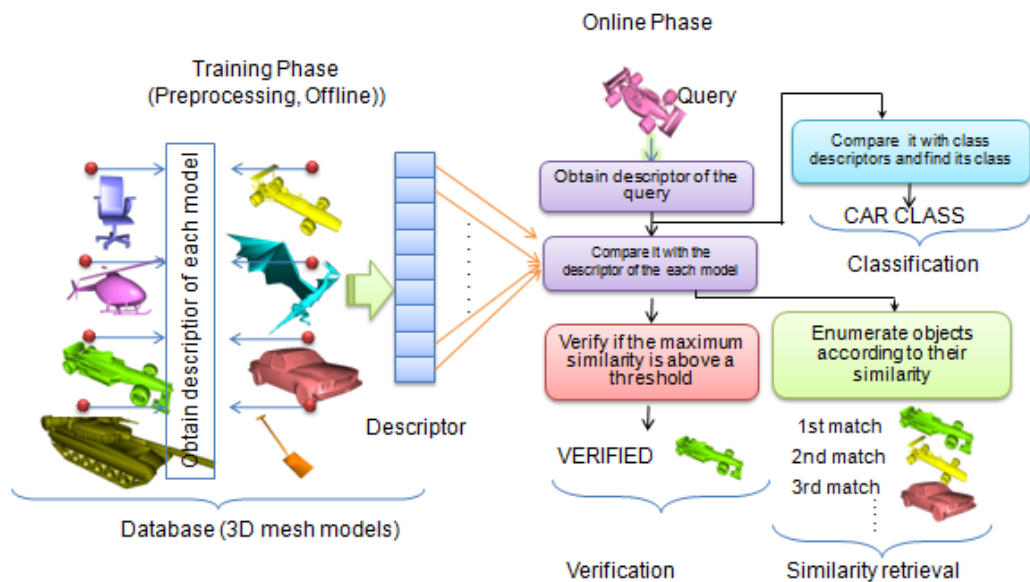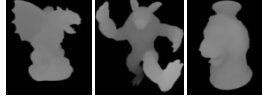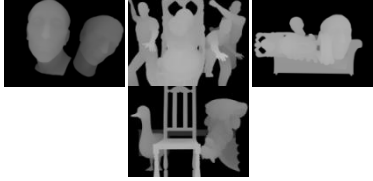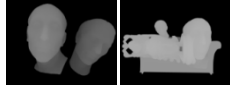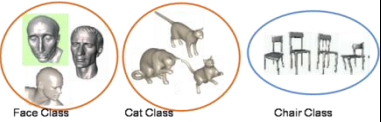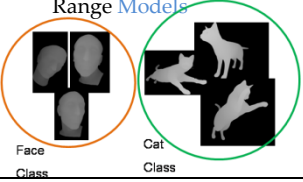
Figure 1-1. Steps and differences in 3D shape verification, classification and similarity retrieval methods

Table 1. Possible research topics related to 3D recognition

	Input	Compared Data Type	Output Type	Comments
	SELF MATCHING (VERIFICATION)			
	Input	**Compared Data Type**	**Output Type**	**Comments**
1A	A 3D Mesh Model	Database Consisting of Mesh Models	Verified (+Matched Model) or Not Verified VERIFIED	Global methods are certainly superior to local methods
1B	Single Range Model	Database Consisting of Mesh Models	Verified (+Matched Model) or Not Verified VERIFIED	Registration Type Methods (i.e. ICP) can be adopted. Slow but accurate. **Major Problems:** Scale Self Occlusion
1C	Single Range Model	Database Consisting of Range Models *Nuance: Range image of a rotated query will be verified or not? Self matching or partial matching?*	Verified (+ Matched Model + % overlapping) or Not Verified VERIFIED, %20 Overlap	Registration Type Methods **Major problems:** -Partial information on both side -Self Occlusion -Rotation, **-Scale**
1D	Single Range Model	Database Consisting of Range Scenes	List of Scenes Containing Query	Problems: Self Occlusion Rotation Scale Segmentation Boundary Occlusion Clutter
	SIMILARITY RETRIEVAL			
	Input	**Compared Data Type**	**Output Type**	**Comments**
2A	A 3D Mesh Model	Database Consisting of Mesh Models	Ranked List	Global methods should be considered.
2B	Single Range Model	Database Consisting of Mesh Models	Ranked List	Scale and Self Occlusion are major problems
2C	Single Range Model	Database Consisting of Range Models	Ranked List	A very challenging problem. Local matching could a solution with many existing difficulties
2D	Single Range Model	Database Consisting of Range Scenes	Ranked List	Problems: Self Occlusion, Rotation, Scale, Segmentation, Boundary, Occlusion, Clutter
	CLASSIFICATION			
	Input	**Compared Data Type**	**Output Type**	**Comments**
3A	A 3D Mesh Model	Database Consisting of **Classes** of Mesh Models	Class ID Chair Class	Similar to problem-2A, but utilization of class properties is an advantage.
3B	Single Range Model	Database Consisting of **Classes** of Mesh Models	Class ID	Similar to problem-2B but utilization of class properties can be an advantage.
3C	Single Range Model	Database Consisting of **Classes** of Range Models	Class ID Face Class	A very challenging problem. Local matching seems to be a solution with many difficulties

Today information retrieval in web and other databases, such as library, personal computers, medical records are mainly achieved on the basis of text comparison. A word is usually given as the input to the algorithm (query) and the records containing the query text are voted according to some similarity measure and enumerated, accordingly, which is called indexing. Research on similarity retrieval of shapes has some common properties with text retrieval research and some different properties from it. A 3D model possess properties including shape, color and texture. Among these properties, shape is the most characteristic one. Color and texture can also be used for identifying objects, but they have limited use. Thus, in order to discriminate 3D models, a comparison is conducted between their shapes. Existing 3D model retrieval methods primarily use geometric properties; however, depending on the application, more information can be gathered, if color and texture are also used.

Available feature based 3D retrieval systems can be implemented as in Figure 1-2. They usually have an offline (training) and an online phase. During the training phase, the descriptors of each model are calculated; the system will go into this phase once and works on the resulting feature vector at the end of this stage by the following steps: When a query is placed, the online phase starts by calculating the descriptor of the query object. Next, the descriptor of the query and descriptors of the database models are compared. Comparison is simply the calculation of distance between the feature vectors. Several distance functions, which will be explained in detail in subsequent chapters, are available and must be selected according to the descriptor properties and considering the whole system.

Existing systems depend on synthetic models that are usually polygon meshes consisting of vertex and face information. Besides, these models should not appear in a scene or they must be segmented before the processing. In fact, in real life scenarios, objects appear in scenes, they do not lie alone in space and most of the time their shape is partially available to the viewer. However, it should be noted that 3D retrieval would be a necessity for applications using or manipulating real life data. As an example, researchers are studying in the field of 3D television for years and it gives promising results such that in the very near future, most of the consumers are expected to watch 3D television and record programs to their own storage disks. Three dimensional televisions currently operate on depth (range) data and objects are not fully described with their vertices and corresponding faces. Another example could be the systems getting data from laser scanning devices. This type of data is

4

considered as 2.5D data (range data, range image, depth data, 2.5D data are used interchangeably throughout the dissertation).
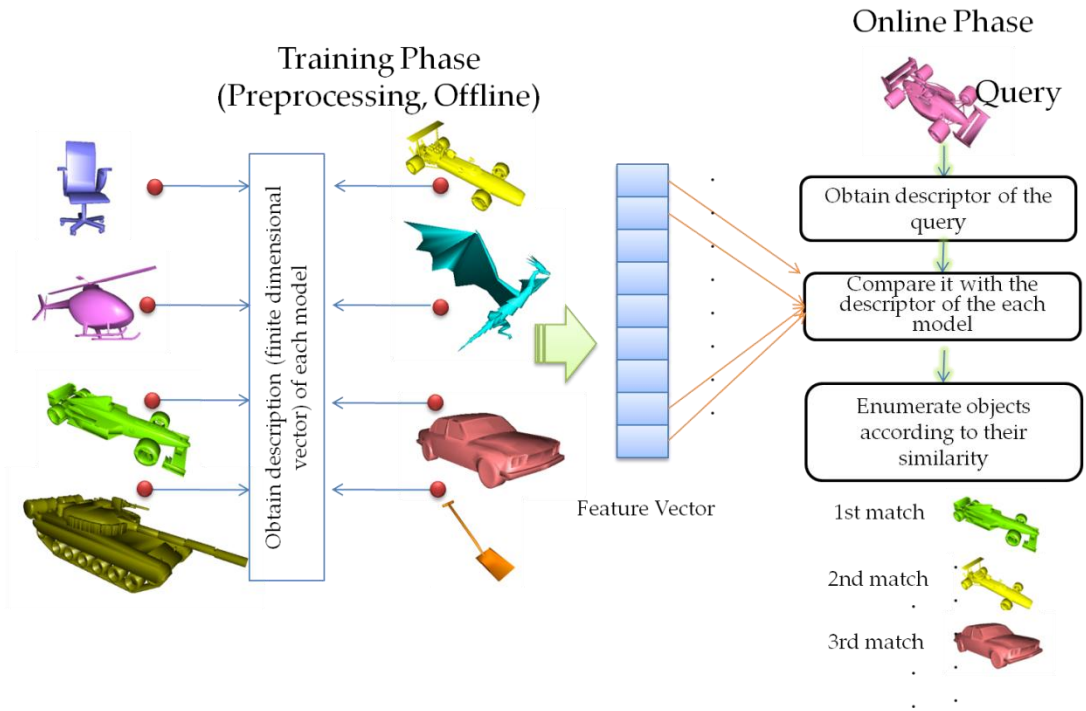
Figure 1-2. Basic steps in descriptor based 3D shape retrieval systems

Similarity retrieval of range data has unique problems. Occlusions, partial information and affine transformations are some of the major challenges. Although range data contains 3D information, similarity retrieval of range data differs significantly from 3D mesh similarity retrieval.

1.1   Contributions of the Thesis

The following contributions are achieved within this thesis:

- Efficient planar structure detection is proposed. The performance of the proposed Recursive Hough Transform in finding planar structures in range images is shown to be superior to the classical Hough Transform. While the proposed modification in Hough Transform introduces additional complexity in space and time, it is quite successful in detecting planar structures in range data. This is mainly due to eliminating the fine parameter discretization ambiguity that emerges in the classical Hough Transform.

- A novel algorithm to segment 3D point cloud (obtained via TOF camera) into objects by using the spatial information is presented. The proposed algorithm exploits the fact that many objects stand orthogonal to the ground plane due to the gravity and the projection of 3D points onto this plane could be equivalent to Fisher Linear Discriminant methodology. The projection of 3D points is followed by a kernel density estimation process in which a saliency map is generated. Salient regions that represent the most probable object locations have high values in the generated map. Then the points with a probability value larger than some certain threshold are assigned to the closest local maxima segmented into objects. The experimental results show that the locations of the objects are determined quite accurately. Compared to the tested methods the proposed algorithm is less sensitive to the noise, less parameter dependent and leads more accurate segmentation results.

- We also developed a technique for correspondence matching in depth data. Shape index is utilized with scale invariant transform for matching. The proposed local surface description method does not require any initial segmentation step; it can also handle affine transformations up to a scale. The experimental results indicate that the proposed approach outperforms the recent two state-of-the-art methods. Moreover, clutter and occlusion do not significantly affect the efficiency of the proposed method.

- Lossless global description of range data by Spherical Harmonic for similarity retrieval is proposed. We utilize Spherical Harmonic Transform for 2.5D range images by representing the models in a reciprocal world observed from the camera. The difference, as well as the advantage of our algorithm, is being information lossless. In other words, the available shape information is completely exploited for obtaining the descriptor, whereas other mesh retrieval applications utilizing SHT "approximate" the shape that yields information loss. The descriptor is invariant to rotations about z-axis. The proposed technique is tested on a large database having high diversity and its performance is superior to the performance of popular *D2* distribution and the classical SHT.

- Analysis of heat diffusion process on range data is presented. Challenges and some experimental results are presented. Also some future directions are proposed.

1.2 Outline of the Thesis

In Chapter 2, possible descriptor based 3D similarity retrieval basics are introduced. Representation formats (3D data types), challenges, distance measures and evaluation metrics are defined. Classification of similarity retrieval methods is made and a comprehensive literature survey is presented.

In Chapter 3, segmentation of range scenes is evaluated. A sub problem of the segmentation is planarity detection. The chapter begins by planar structure detection. Problem definition, proposals, modifications on available methods and results are presented. Later, state-of-the-art methods in segmentation is adapted for real range scenes, where the inputs are the Time of Flight Camera outputs. Color information is integrated with the spatial information and its effects on the performance of the segmentation methods are evaluated. Several local surface properties, such as curvatures, surface variances, normals are utilized with region growing. Finally, a novel range scene segmentation algorithm depending on the kernel density estimation method is proposed and results are presented.

In Chapter 4, correspondence matching of range images is examined. A method for keypoint matching is proposed. Range image is represented with its corresponding shape index image. This representation magnifies salient regions by the nonlinearity structure in the shape index function. Keypoint selection and description is performed using scale invariant feature transform.

In Chapter 5, a global description that relies on Spherical Harmonics Transformation (SHT) is proposed. Although *SHT* is a relatively mature concept in shape retrieval research, we propose to utilize it for range images by representing the models in a reciprocal world observed from the camera. The difference and advantage of our algorithm is being information lossless in this manner. The descriptor is invariant to scale and rotations about z-axis. Proposed method is tested on a large database having high diversity.

In Chapter 6, heat diffusion theory and the usage in shape analysis is presented. The challenges of this approach in case of range image description are presented.

In Chapter 7, the conclusions of this thesis are presented, as well as some future research directions.

CHAPTER 2

3D SIMILARITY RETRIEVAL

This chapter introduces representation formats of 3D data and reviews challenges in 3D shape retrieval research. Classification of similarity retrieval methods is proposed and a comprehensive literature survey is presented. Common resources of 3D data are also addressed at the end of this chapter.

2.1   3D Representations

A taxonomy of 3D object (or scene) representations is given in Figure 2-1 [1-3]:
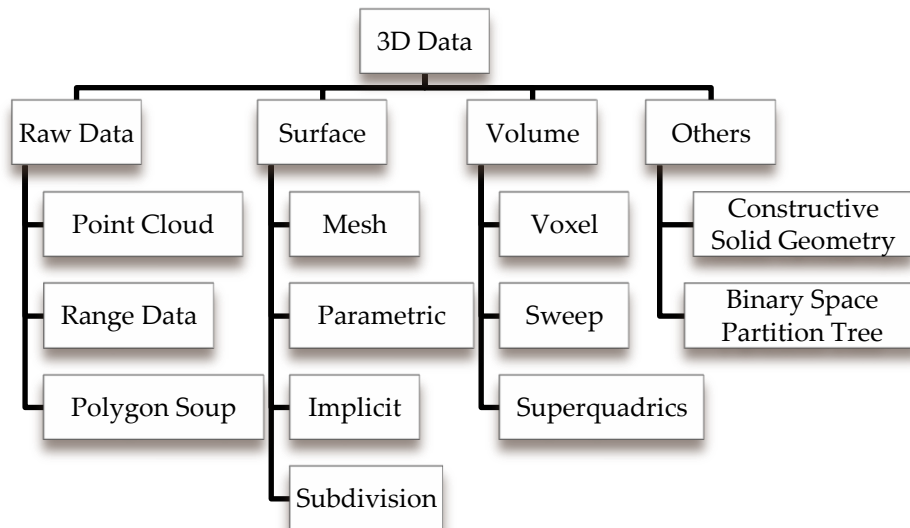
Figure 2-1. Classification of 3D data formats

Figure 2-1 covers nearly all state-of-the-art representations and the examples of point cloud, triangular mesh, and range data representation are given in Figure 2-2. Point cloud representation is the simplest format; it can be considered as 3D point samples hooked in 3D

space with no other information. It is important to develop retrieval algorithms, which are capable of supporting point clouds, since other formats can easily be converted into this representation. Range sensors and computer vision algorithms, such as 3D reconstruction from multiple views, are the sources of point cloud models. Range data can also be considered as raw data, since it is a 2D image, where each pixel value represents the *distance* of that scene point to the sensor. Outputs of some scanning devices, such as Time of Flight (TOF) cameras, are typical range data. Haphazard polygon models are named as *polygon soups*, which are one step ahead from being raw.
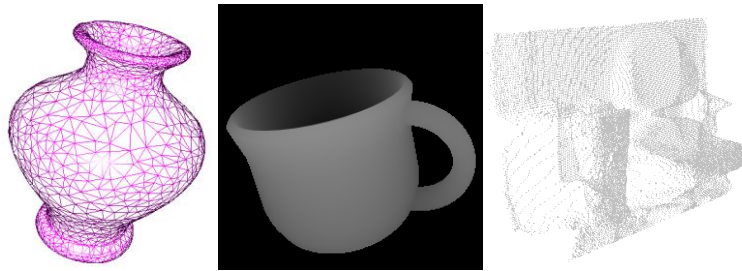
Figure 2-2 Samples of 3D formats of polygon mesh, range image, and point cloud from left to right

In Computer Aided Design (CAD) systems, Constructive Solid Geometry (CSG) and parametric surfaces are preferred. Parametric surfaces can be formulized by mathematical equations; they are easily sampled and used for generating realistic models [1]. The most popular parametric representation is non-uniform rational B-Spline format (NURBS) [4, 5].

Voxel-based applications are usually used in medical imaging. Voxel is simply a volumetric element; its counterpart in a 2D image is pixel. Voxel does not carry position information, however, it might contain multiple values. Medical resonance imaging, ultrasound imaging, and computed tomography devices works with voxel data [6].

The most popular representation is the polygon mesh representation especially the triangular mesh. There are various formats for polygon mesh representation. Virtual Reality Modeling (VRML) is used commonly, object file format (OFF), object file format (OBJ) developed by WaveFront Technologies company, polygon model format (PLY), 3D studio format (3DS) developed by Autodesk Media and Entertainment company can be taken as other popular formats. The idea behind every model is simple; each model consists of polygon surfaces (usually called as *faces*) which are connected to each other and each

polygon consists of some number of 3D points (vertices). Triangle mesh format is the easiest one to handle and usually polygon mesh models are converted to triangle mesh format before processing. Conversion is done without an effort. More detailed information on other 3D representation formats can be found in [1, 2].

Every format includes 3D information, while it differs in some aspects. Some of the differences between point cloud format, mesh representation and range imaging are:

- ***Density.*** *The number of points describing the object.* This criterion usually depends on the application. Laser scanned data (unorganized point cloud) are typically the most discriminative among the others in terms of density. High resolution polygon mesh models can also be generated in artificial environments. Practical 3D acquisition systems (i.e. TOF cameras) for range imaging are able to capture quite low frame-rates at present. On the other hand, dense depth estimation algorithms can produce high resolution range images from multi-views.

- ***Completeness and View Dependency.*** *The amount of the information about the geometry of the object contained in the data.* The most descriptive type is the mesh format. They contain the full geometry of the objects. Besides, normal estimation can be achieved easily, which is another important clue for shape identification. Certainly, parts of a full model can be represented by using mesh format; although information is partial it is view independent. On the other hand, laser scanned data or range images that are obtained in various ways could only contain partial information about the geometry of the object. This view dependent information yields loss in the geometry information mainly due to projection and self occlusions.

- ***Signal to Noise Ratio (SNR)***. As mentioned previously, the mesh models are usually synthetic objects and can be made noise free easily; however, in real life, noise is unavoidable for 3D applications. High SNR prohibits accurate local feature (i.e. normal, curvature) estimation.

- ***Neighborhood Relations.*** *Information of the connectivity of a point with the other object points.* Mesh models consist of polygon structures and provide vertex coordinates plus edge information. Neighborhood of a point could be extracted precisely. Unorganized point clouds and range images lack such information. In these situations, neighboring relations are determined by spatial closeness. However, occlusion, view dependency

and other objects might lead grouping points belonging to different parts as neighbors.

2.2   Background and Related Work in 3D Object Retrieval

There are various precious surveys [1, 7, 8], and theses [9-12] on 3D object retrieval and will be referred through this dissertation. There are many 3D retrieval methods, while precise classification of these approaches does not exist. In [9], a five-class taxonomy is adopted, which are histogram-based, transform-based, graph-based, and 2D image-based. Moreover, there are unclassified methods which are considered under the others label. Iyer et al [13] classified 3D shape searching methods in five categories: global feature based, manufacturing feature recognition based, graph based, histogram based, and 3D object recognition based. We adopt a grouping similar to [9].

For comparing and matching 3D shapes, the adopted method is describing models mathematically by a compact formulation, which is called the description of the model, and matching the query by comparing its descriptor with the models' descriptor in the database by some distance metrics. Representing the shape geometry by a mathematical formulation such as a function or a vector is desired. If it could be achieved, it would be practical, fast, easy to manipulate, less space consuming, easy to compare and easily adapted to new object classes.

The task is hard and complicated and there are many technical challenges. As mentioned previously, there is plenty of 3D representation formats; hence, constructing a representation independent retrieval system is a challenge all by itself; thus, usually a preprocessing step is required. Next difficulty is related to the orientation and scale variations. Different pose and geometric ratios between models constitutes huge problems during comparison steps. Describing an object in an effective and efficient manner is also another challenge. The correct comparison metric is another unanswered question for the 3D object retrieval methods. The last problem is segmentation which emerges from the cluttered scenes. Registration, in the case of different views to be combined for obtaining a single 3D model, is another challenge to work upon.

2.3 Challenges

2.3.1 Representation Formats

Since there are many 3D representation formats, it is not possible to obtain a representation independent method. Some of them work with triangular meshes that requires triangulation of other formats; some of them work with voxel data, especially medical applications; thus, requiring voxelization. Absolutely a format must be adopted in a retrieval system; however, an obligatory condition is converting the system easily in such a way that it can work with other representation modalities. Thus, a point cloud is the fundamental structure in all of these formats and it is important to develop a point cloud based retrieval method.

2.3.2 Alignment

A well-known dictionary [6] defines to the word *shape* as

> "part of space occupied by the object as determined by its external boundary — abstracting from other aspects the object may have such as its color, content, or the substance of which it is composed, as well as from the object's position and orientation in space, and its size".

It is highly important having an orientation and scale invariant descriptor, which is called invariance. However, it is claimed in [9] that forcing to make an invariant descriptor leads sacrificing from being discriminative. We argue that scale information can only be discriminative if the true dimensions are known. For example, geometrically similar shapes belonging to different classes such as a book and a door can be distinguished by their scales. Therefore, inputs should be acquired from a device such as TOF (Time of Flight) camera, LIDAR ((Light Detection and Ranging; or Laser Imaging Detection and Ranging, also LADAR) or Microsoft Kinect sensor [14] that measures the true dimensions. The methods, which are not invariant to the pose and size, normalize their 3D object model before processing. This normalization is usually achieved by using principal component analysis (PCA) method (Karhunen Loeve Transform, KLT, or Hotelling Transform). PCA is used to find the major axis (principal axis) of the model. It can be summarized as follows [15, 16]:

1. Center of mass is translated to the origin
2. Calculate the covariance matrix

$$C = \begin{pmatrix} cov(x,x) & cov(y,x) & cov(z,x) \\ cov(x,y) & cov(y,y) & cov(z,y) \\ cov(x,z) & cov(y,z) & cov(z,z) \end{pmatrix} \qquad 2.1$$

3. Calculate the eigenvectors and eigenvalues of the covariance matrix

4. Eigenvector corresponding to the largest eigenvalue is the major axis; perform rotation according to this direction.

Drawbacks of the PCA algorithm are its time and space complexity and being unstable for symmetric objects.

2.3.3 Descriptor Requirements

Following properties are desirable for a good descriptor:

- Discriminative (effective): it should capture details fine enough to differentiate objects belonging to different classes

- Stable (Robust): small shape changes would result small alterations in the descriptor. This constraint is necessary to classify similar objects. Within class variations should also be small.

- Low Dimensional (efficient): in order to retrieve objects faster, it should be low dimensional.

These properties sometimes contradict with each other. There is a tradeoff between being discriminative and robust. Similarly, effectiveness and low dimensionality are mostly opposing conditions.

2.3.4 Similarity Measure

Decision on the similarity measure is a critical question. A distance function, or a norm, is selected as a dissimilarity measure. Usually trial and error approach is used for deciding the proper distance function. To be on the safe side, distance between descriptors should be metric. In other words, the following properties should hold:

Let M be the metric space and d is the distance function defined on M, then

- Positivity

$$d(x) \geq 0 \quad for \ \forall \ x \ \epsilon \ M$$

- Identity

$$d(x,y) = 0 \quad iff \ x = y, \ for \ \forall x, y \ \epsilon \ M$$

- Symmetry

$$d(x,y) = d(y,x) \quad for \ \forall x, y \ \epsilon M$$

- Triangle Inequality

$$d(x,z) < d(x,y) + d(y,z) \quad for \ \forall x, y, z \ \epsilon \ M$$

14

Some measures (some of them are not metric) that are used in content based retrieval algorithms are as follows:

- Minkowski Norms, [17-23]
- Histogram Intersection, [24, 25]
- Kullback-Leibler Divergence, [20, 25-27]
- Jensen-Shannon Divergence, [22, 28, 29]
- Log-Probability Distance,
- Manhattan Distance (L_1),
- Chebyshev Distance (L_∞),
- Quadratic Form distance, [30]
- Earth Mover's Distance, [20, 31-33]
- Chi Squared, [22, 23, 34]
- Bhattacharyya Distance, [20, 23]
- Mahalanobis Distance, [35, 36]
- Hausdorff Distance, [37]

Minkowski Norms

The most commonly used distance metric is the Minkowsky norms defined as:

$$d(x,y) = \left(\sum_{i=1}^{N} |x_i - y_i| \right)^{\frac{1}{L}}$$
$$2.2$$

where N is the dimension of the vectors x and y, and L is the degree of the norm. L_1 and L_2 norms have a special name, called the city block distance (Manhattan) and Euclidean distance, respectively. These metrics are preferred due to their simplicity and successive performance.

Kullback-Leibler Divergence & Jensen–Shannon Divergence

Kullback-Leibler divergence (KL) is a non-symmetric measure of similarity between two probability distributions P and $Q.$ It is defined as follows:

$$KL(P,Q) = \sum_i P_i log \frac{P_i}{Q_i}$$
$$2.3$$

KL divergence is not symmetric so it is not a metric, that is *KL(P,Q) ≠ KL(Q,P).* Based on this idea, symmetrized and smoothed version of KL divergence, namely Jensen-Shannon Diverge (JS) is proposed. JS is defined as follows:

$$JS(P,Q) = 0.5\ KL(P,M) + 0.5\ KL(Q,M)$$

2.4

where *M = 0.5(P+Q)* as in [23], shape retrieval algorithms that utilize shape distributions generally prefer this measure.

Log-Probability Distance

A new metric, Log-Probability (LP) Distance metric [38] is proposed for comparing two distributions. It is defined as follows:

$$LP(P,Q) = \sqrt{\sum_{i=1}^{N}\left(log\frac{P_i}{Q_i} - \left(\frac{1}{N}\sum_{j} log\frac{P_j}{Q_j}\right)\right)^2}$$

2.5

Unlike to the KL divergence and JS divergence, LP distance is a metric satisfying positivity, identity, symmetry and triangle inequality properties. This recent distance has not been utilized in computer vision algorithms due to its originating discipline (information theory).

Chi Squared:

The Chi-Squared distance (denoted by χ^2) is used to compare histograms and defined by the following equation:

$$d(x,y) = \sum_{i=1}^{N} \frac{(x_i - y_i)^2}{x_i + y_i}$$

2.6

Bhattacharyya Distance:

Another measure which is also used to compare probability distributions is Bhattacharyya distance defined by the following equation:

$$d(x,y) = \sum_{i} \sqrt{x_i y_i}$$

2.7

2.3.5 Segmentation

Independent of the application, segmentation is a great challenge all by itself, since it requires an intelligence. Geometry and color information, which are usually available, are not sufficient for obtaining any desired segmentation result. Clustering techniques are applied prior to processing to identify objects, since infinite number of object combinations could be present in a single scene. Region growing, K-means [39], Gaussian Mixture Models [40], Mean-shift [41], and graph based method Normalized Cuts [42] are the powerful image segmentation methods. Some of them can be modified to n-dimensional space; thus, applicable to 3D segmentation. However, in this case, the sole spatial information (color and texture usually are not available for 3D inputs) introduces some extra challenges.

2.3.6 Registration

The most popular algorithm used in registration of point clouds is iterative closest point (ICP) method [43]. There are many variations on this algorithm [44, 45]. Main steps of the algorithm are given as follows: *i) selecting random points on models, ii) matching these points to the closest samples on the other model (correspondences), iii) calculate error with a weighting by some metric iv) minimize the error.*

2.3.7 Evaluation Metrics

In order to compare and evaluate the retrieval algorithms there are various performance evaluation tools. Each performance measure favors a property of the algorithms which will be explained in detail later. SHREC shape retrieval contest [46] evaluates competitor retrieval systems comparing each evaluation tools. These are Precision-Recall, Nearest Neighbor, First-tier and Second-tier, E-Measure, Discounted Cumulative Gain (DCG), Normalized Discounted Cumulative Gain (DCG):

- **Precision-Recall:**

  The most commonly used statistics for measuring the performance of retrieval algorithms is the Precision-Recall values. Precision is defined as the ratio of the number of relevant items that are retrieved (namely true positives) to the number of items that is retrieved as a result of a search. Ideally it should be equal to one; that is

all retrieved items should be (or preferably are) similar to the query that is being searched. Recall is the ratio of the number of the retrieved items that are similar to the query to the total number of similar items (true positives plus false negatives) be in the database. Ideally recall should be equal to one, in other words, all items that are similar to the query should be retrieved from the database as a result of a search.

$$Precision = \frac{number\ of\ relevant\ items\ retrieved}{number\ of\ retrieved\ items}$$

$$Recall = \frac{number\ of\ relevant\ items\ retrieved}{number\ of\ total\ relevant\ items\ in\ the\ database}$$

- **Nearest Neighbor:**

  It is the percentage of the first closest retrieved models that belongs to the same class of the query. Ideally it should be 100% that is each item belonging to the same class with the query appears as an outcome of the search at first top places.

- **First-tier and Second-tier:**

  First-tier is the percentage of the matches belonging to same class of the query which appears at the top *N* matches where *N* is the number of items in the query's class. Second-tier is the percentage of the top relevant matches (items that are in the same class with the query) appear in the top *2*N* matches. Ideal value for both First-tier and Second-tier is 100%. In this context query is assumed to lie in the test data not in the training set.

$$First\ Tier = \frac{\#relevant\ items\ appear\ at\ the\ top\ N\ matches}{number\ of\ items\ in\ the\ query\ class = N} x100 \qquad 2.8$$

$$Second\ Tier = \frac{\#relevant\ items\ appear\ at\ the\ top\ 2N\ matches}{number\ of\ items\ in\ the\ query\ class = N} x100 \qquad 2.9$$

- **E-Measure:**

  It is defined as follows:

$$E - Measure = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \qquad 2.10$$

In the ideal case E-Measure must be equal to one, *(precision=1, recall=1)*. While calculating precision and recall for computing E-Measure, number of retrieved items is considered to be 32. This is due to the fact that people making a search will be interested in the limited number of top retrieved results [47].

- **Discounted Cumulative Gain (DCG):**

It is introduced by Järvelin & Kekäläinen [48]. This is a weighted measure, taking into consideration the ranking of the results. Consider two matches where the first one is a true negative (irrelevant match) appearing at the first place and the second one is also a true negative appearing at the end of the list. Seriousness of appearing an irrelevant match at the top of the list is larger than appearing at the end. Similar arguments hold for true positives (relevant matches). A relevancy array is used to calculate DCG where the value of the array element is equal to one if the top corresponding index match is relevant with the query. The relevancy array (RA) is defined as:

$$RA_i = \begin{cases} 1 & \text{if the match at the } i^{th} \text{ position is in the same class with the query} \\ 0 & \text{if the match at the } i^{th} \text{ position is irrelevant} \end{cases}$$

Then DCG at the ith rank is given by the following equation:

$$DCG_i = \begin{cases} RA_i & i = 1 \\ DCG_{i-1} + \dfrac{RA_i}{\log_2 i} & otherwise \end{cases} \qquad 2.11$$

Final DCG value for a query is:

$$DCG = \frac{DCG_{i_{max}}}{1 + \sum_{i=2}^{N} \frac{1}{\log_2 i}}, \qquad 2.12$$

where i_{max} is the maximum index number (i.e. total number of objects in the database) and first N matches is relevant with the query. Higher values indicate better performance.

Example:

Search Result:

Rank	1	2	3	4	5	6	7	8	9	10
*Match Relevance, **R**:Relevant, **IR**: Irrelevant*	R	R	R	IR	IR	R	IR	R	R	IR

$$RA = [1,1,1,0,0,1,0,1,1,0]$$

$$DCG_{i_{max}} = 1 + \frac{1}{\log_2 2} + \frac{1}{\log_2 3} + \frac{0}{\log_2 4} + \frac{0}{\log_2 5} + \cdots\cdots + \frac{0}{\log_2 10} \cong 3.2797$$

$$DCG = \frac{3.2797}{1 + \frac{1}{\log_2 2} + \frac{1}{\log_2 3} + \cdots + \frac{1}{\log_2 10} = 5.2544} \cong 0.624$$

- **Normalized Discounted Cumulative Gain (NDCG):**

  Normalized Discounted cumulative gain is usually used in comparing retrieval algorithms. It scores both the relevancy and the position of matches as does DCG but also when used by a standard database it will yield a quick idea about the performance of a retrieval algorithm among the competitors. In fact it indicates the statue of the algorithm among others. To achieve this aim, DCG of an algorithm i is divided by the average of DCG values of all algorithms and one is subtracted from this fraction. If an algorithm's NDCG value is positive it can be concluded that its retrieval performance is above the average and vice a versa if it is negative. It is defined by the following formula:

$$NDCG = \frac{DCG_i}{DCG_{avg}} - 1 \qquad\qquad 2.13$$

2.3.8 Additional Challenges in Range Image Description

In range image similarity indexing research, challenges are not limited with the aforementioned ones. Occlusions (self-occlusions) are another problem in surface description. Moreover, due to view-dependent structure of range imaging, affine transformations introduce some other difficulties which are not present in complete 3D mesh retrieval research. Detailed discussions about range image descriptor challenges are presented in Chapter 5.

2.4 Literature Survey on 3D Retrieval

After a brief introduction about the retrieval tools, in this part some significant works on 3D object retrieval will be explained in detail. Our taxonomy for 3D shape matching methods is shown in Figure 2-3 and Figure 2-4. Two survey papers [7, 8] are mainly utilized in this section. We use multiple classifications for algorithms, instead of a single taxonomy, according to : i) *descriptor properties,* ii) *mathematical methods they exploit.*

The vast majority of algorithms are statistically described global geometric based techniques. Other classifications can be achieved according to their complexity, partial similarity support, and dependency on the 3D model format (mesh, range, or point cloud).
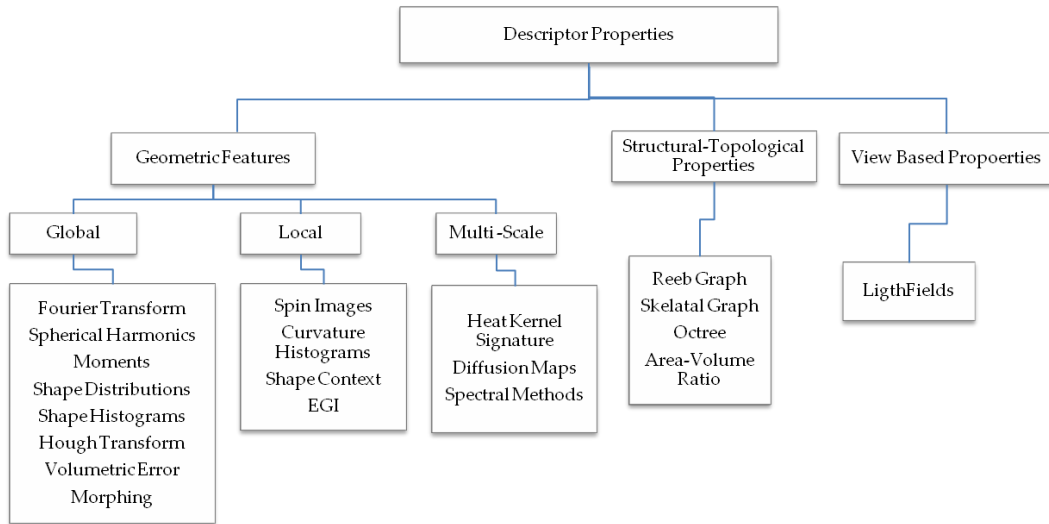
Figure 2-3. Classification of 3D shape retrieval methods depending on the descriptor properties
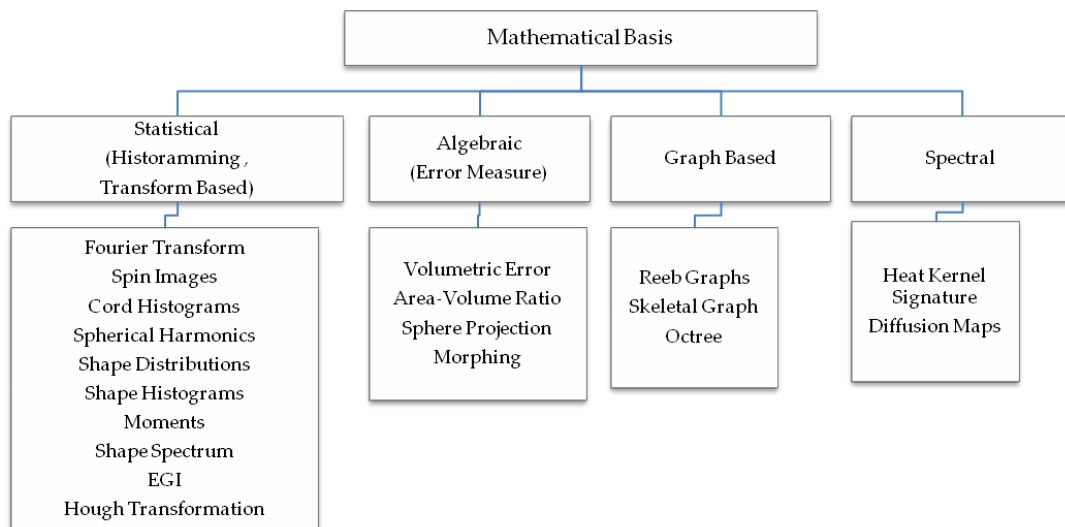
Figure 2-4 Classification of 3D shape retrieval methods depending on the mathematical methodology

In the following paragraphs some significant works of 3D shape description literature will be explained in detail. At the end, some of the publications that utilize local surface properties are listed (Table 2). Some of the significant works such as Spherical Harmonics Transform and Heat Kernel Signature is not presented in this section intentionally, since they will be explored in detail in the upcoming chapters.

21

Extended Gaussian Imaging (EGI):

EGI is proposed by Horn [49] in 1984. It is a spherical histogram based surface descriptor method. Histogram is obtained by cumulating surface normals direction for each surface (usually used with mesh models) where normal directions are regularly spaced on sphere (each bin is represented by (θ, φ), no radial information is used). Figure 2-5 is an example of Gauss sphere mapping of surface normals for a rectangular pyramid. It is weighted histogram with a weight equal to the area of the corresponding surface. It is invariant to translation and can easily be made scale invariant; however, it is not invariant to rotation. On the other hand, EGI approach owns a nice property about rotation of the model inducing the same rotation in the histogram. However, EGI is not a suitable descriptor for partial shape matching and point cloud representations.
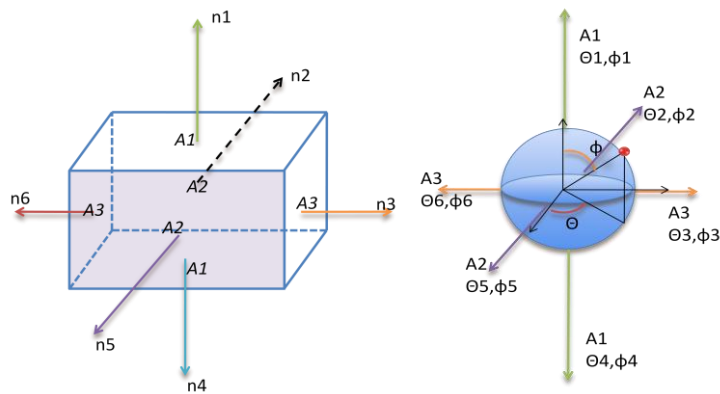
Figure 2-5. Extended Gaussian Imaging of a Prism

Complex Extended Gaussian Imaging (CEGI):

Kang and Ikeuchi [50] improved the EGI by including radial information. CEGI histogram is composed of complex numbers $\alpha + j\beta$, where α denotes the area of the surface and β denotes the distance of the surface to the origin. Figure 2-5 show a mapping of surface normals of a rectangular prism onto a Gaussian sphere with EGI principle, if CEGI is used then weights (A1,A2,A3 which are real numbers, areas of surfaces of prism in the EGI representation) will be complex containing surface position information. Similar to EGI, CEGI is not a discriminative descriptor for partial, noisy and occluded object indexing.

Cord and Angle Histograms:

This method is proposed by Paquet et al. [51, 52]. Cord is defined as a vector joining the center of mass of the model with the surface point. Surface point is selected as the center of mass the triangle for mesh models; a weight measure is used as the area of the triangle. Normalized histograms of three properties of cords are generated: i) length of cords, ii) the angle between the cord and principal axis, iii) the angle between the cord and the second principal axis; and used as the descriptor of the model. Method is noise sensitive, invariant to rotation and translation, but not scaling. Non-star shaped objects will be represented by its convex-hull resulting loss in details.

Moments:

Moments, especially Hu moments are popular tools in 2D image recognition. Also 3D counterparts are proposed and utilized for 3D shape retrieval: geometric moments [53], Zernike moments [54], spherical moments [55]. Moments are scalar quantities describing the distribution of the points belonging to models. They are simple descriptors, whereas not invariant to rotation scaling and translation. Low order moments capture coarser details hence, they are not powerful in discriminating; however, increasing the order improves the discriminative ability of any such descriptor. There is a tradeoff between the order of moments and the efficiency in terms of space and time and noise sensitivity. Geometric moments formulation is given with the following general formula [56]:

$$\mu_{ijk} = \sum \sum \sum_{p \in M} w_p x_p^i y_p^j z_p^k \quad , \qquad\qquad 2.14$$

where p is the point in model M.

Shape Histograms:

Histogram based classification method shape histograms are introduced by Ankerst et al. [57]. Method divides space into shells (concentric spheres), sectors or combination of the two so called *spiderweb* (Figure 2-6). In the first one, namely *shell model*, boundary sphere of the model is divided into concentric spheres, while each sphere is an accumulator for the points, whose distance to the center of mass lies within the sector's radii range. It can be easily observed that shell model is invariant to rotation and translation, but not invariant to scale changes. This approach can be accepted as a simple and a weak descriptor in terms of distinguishability as it is shown in [47] it finds a place at the bottom of the Precision-Recall

graph of 12 algorithms. *Sector model* gives the distribution of the area of the models as a function of spherical angle. Sector model is better than the shell model according to the experiments represented in [47]. The best effort comes from the combined model.
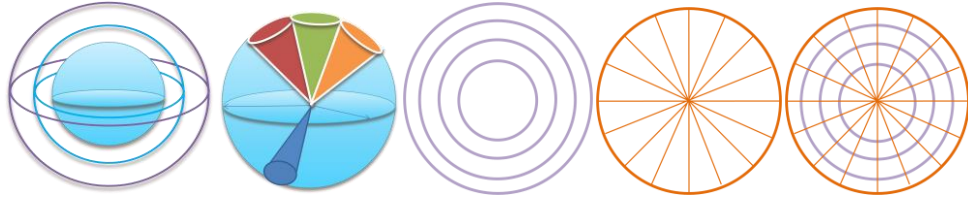
Figure 2-6 Shell, sector model and 2D projections of shell, sector and compound model respectively

Spin Images:

A local shape descriptor method, namely *spin images* is introduced by Johnson and Hebert [58]. About an interest point on the surface of the model 2D histograming is obtained. Normal vector for interest point p is required; thus, the method cannot be directly applied to point cloud representations. Algorithm can be summarized as follows: given set of surface points and corresponding normal vectors $S = \{(p_i, n_i)\ i = 1, 2, \dots, k\ \}$, accumulate surface points which are (α_i, β_i) distant from the interest point. A cylindrical coordinate frame is assigned whose origin is the interest point p, and z axis points along the normal vector. Radius and elevation components are discretized to obtain histograms where β_i the distance is measured along z axis and α_i is the radial distance (Figure 2-7). For a given 3D model, the descriptor is the set of 2D histograms that is computationally costly and space requiring. The authors [58] suggest compression techniques to reduce the number of spin images for an object to make an efficient comparison. However, this technique has the following disadvantages: it is not scale invariant, it requires high storage and results multiple ambiguous correspondences, since spin images of near points are similar; i.e. spin images are not unique [59].
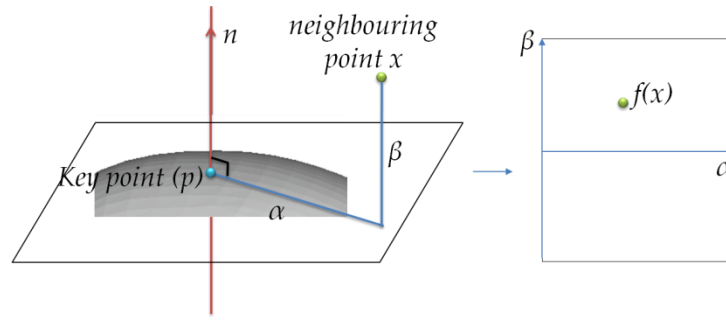
Figure 2-7. Illustration of radial (α) and axial (β) distance in spin image generation

Shape Distributions:

Osada et al. [23] propose to use shape features' distributions as a descriptor. For various features of an object algorithm calculates histograms and comparison is done using several metrics (χ^2, *Bhattacharya, Minkowsky L_N norms*). Distributions of the following features are selected:

- *A3: Angle between three random points selected on the surface of the model.*
- *D1: Distance between a fixed point and one random point on the surface. They use center of mass of the boundary of the model as the fixed point.*
- *D2: The distance between two random points on the surface.*
- *D3: The square root of the area of the triangle between three random points on the surface.*
- *D4: The cube root of the volume of the tetrahedron between four random points on the surface*

Sampling of the surface points is achieved randomly. *D2* distribution is similar to the cord histograms [51, 52] but as reported in Akgül et al.'s study [9, 60] random sampling over the surface improves the performance.

Similar to Osada's work [23] , Akgül [9] extracts density based 3D descriptors. However they use probability density estimation by using kernels instead of using directly histograms. They use three local geometric features, as shown in Figure 2-8:

i)   Radial feature: is a four-tuple including distance of surface point to the origin and unit vector pointing that point,

ii)   Tangent plane-based feature: is a four-tuple including the normal vector at the surface point and the component along this normal of the distance between the point and the origin (center of mass).

25

iii)   Cross-product feature: aims to give a relation between previously mentioned features, simply the cross product of the radial and tangential features.
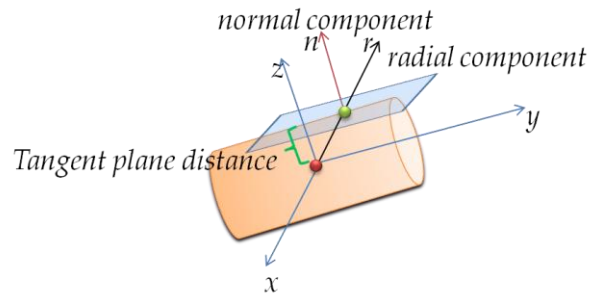
Figure 2-8. Shape features used in [9, 60]

According to their experiments, their DCG values for their combined feature descriptor is the best among 13 other algorithms. However, in that comparison the combined feature descriptor of Osada [23] is not available, whereas only *D2* feature based descriptor performance is included.

Light Fields:

This method is a multi-view approach for indexing 3D models proposed by Cheng et al. [61]. All models in the database are viewed from the predetermined point of views and a number of silhouettes (2D images) are obtained. Then, for each silhouette, a feature vector is calculated, so that set of features constitutes the shape descriptor. They use a large number of views exactly one hundred orthogonal projections of an object. These silhouettes are then described by their Zernike moments and Fourier transformations. According to the experiments conducted by the authors, LigthField descriptor provides considerable improvements over existing methods, such as spherical harmonics [61]. Similar to Light Fields idea, the approaches described in [62-64] and [28] are view based 3D retrieval methods.

Reeb Graphs:

Graph based method Reeb graphs introduced by Hilaga et al. [65] is a topological descriptor. Skeletal structure of an object is represented in a graphical way and used for matching and retrieval. Basically, an object is divided into partitions according to some criteria, called μ function (e.g., curvature value of the point or geodesic distance at that point), each partition

corresponds to a node in the graph structure. Links between the nodes represent a μ function relations of adjacent partitions.

Curvature histogram (Surface Curvature, shape Spectrum):

In this method, descriptor depends on the distribution of the curvatures. The method is proposed by Zaharia and Preteux [66] within the MPEG-7 framework for multimedia retrieval. Shape spectrum description is obtained by histograming the shape index values for the surface points. The shape index is introduced by Koenderink [67] is defined as a function of the two principal curvatures. The maximum and minimum values of the normal curvature at a point on a regular surface are called as *principal curvatures* κ_1 and κ_2. Then, *shape index (*SI*)* is defined as follows:

$$SI = \frac{1}{2} - \left(\frac{2}{\pi}\right) arctan \left(\frac{\kappa_1 + \kappa_2}{\kappa_1 - \kappa_2}\right) \qquad 2.15$$

The descriptor is invariant to translation and rotation. Scale invarince is not satisfied, if the underlying geometry is estimated from the local point samples. Since estimation is dependent on the size of local patch. However, for parametric surfaces, SI is scale invariant. Figure 2-9 represent shape index values for some basic shapes.
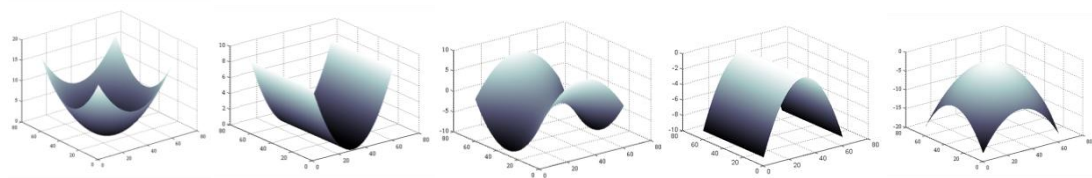
a) Spherical Cup (0) b) Rut (0.25) c) Minimal Saddle (0.5) d) Ridge (0.75) e)Spherical Cap(1.0)
Figure 2-9. Shape index values of basic shapes

Fourier Transform Descriptor:

Fourier Transformation is a popular descriptor in 2D content based image retrieval. Vranic and Saupe [68] is adapted this technique to three dimensional case. Simply, a model is represented in frequency domain, a vector descriptor is obtained from the Fourier transformation coefficients while ignoring the high frequency components. The lowest frequency components capture the models' major topology, whereas the high frequency components contain details. Vranic and Saupe use 172 components at maximum.

A voxelization step is required before processing and since descriptor is not invariant to rotation and scaling, pose normalization (using a modified version of PCA, called continuous PCA in [68]) is performed afterwards. In spite of preprocessing steps, which are disadvantages of the algorithm, it is efficient in terms of space, and as reported in [10], it is also efficient in terms of discrimination power. Fourier transformation on the voxelized and normalized data g (g_{ijk} represents a single voxel at position ijk of a 3D array) is given by the following equation [68]:

$$f_{uvw} = \frac{1}{\sqrt[2]{N^3}} \sum_{i=-\frac{N}{2}}^{\frac{N}{2}-1} \sum_{j=-\frac{N}{2}}^{\frac{N}{2}-1} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} g_{ijk} \; e^{-j\frac{2\pi}{N}(iu+jv+kw)} \qquad 2.16$$

Fourier Mellin Transformation:

Another frequency domain representation is Fourier Mellin transformation (FMT) which combines the spherical harmonics with the Mellin transformation. Resulting descriptor is rotation, scale and translation invariant. The underlying theory takes place in [69], but no experimental study is presented. Mellin transform of a function is given by the following equation:

$$F(m) = \int_{0}^{\infty} f(x)x^{-jm}\frac{dx}{x} \qquad 2.17$$

The Mellin transform of a scaled function $f(\alpha x)$ is equal to $F_\alpha(m) = \alpha^{jm}F(m)$ however the influence of the scaling is reflected only the phase; thus, scale invariance is obtained after taking the modulus and rotation invariance is achieved by applying spherical harmonics transformation. FMT of a continuous function f, in polar coordinates is given by:

$$F_{FMT}(k,l,m) = \frac{1}{2\pi^2}\int_{0}^{\infty}\int_{0}^{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} f(r,\theta,\phi)r^{-jm}\,e^{jk\theta}\,e^{-jr\phi}\,\frac{dr}{r}d\theta d\phi \qquad 2.18$$

Due to numerical difficulties in computing the Fourier-Mellin transform, the Analytical Fourier-Mellin Transform [70], which adopts polar coordinate instead of the Log-polar coordinate, is proposed by the following formulation:

$$F_{AFMT}(k,l,m) = \frac{1}{2\pi^2}\int_{0}^{\infty}\int_{0}^{2\pi}\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} f(r,\theta,\phi)r^{\sigma-jm}\,e^{-jk\theta}\,e^{-jr\phi}\,\frac{dr}{r}d\theta d\phi \qquad 2.19$$

Coefficients' amplitudes of of low order frequency components are used as to obtain rotation, scale and translation invariant shape descriptors.

Structural Indexing: Efficient 3-D Object Recognition

In this early work [71] Stein and Medioni detects interest points at high curvature values. In the vicinity of these interest points, they form the "splash" structure. For a given interest point, *P*, on a surface, a contour, which is ρ distant from the point, is formed. The distribution of the normals on the contour is called as *"splash"*. Definition of the surface normals can be understood from the Figure 2-10. For every sample point of a splash, they obtain a pair $v(\theta)=(\phi(\theta),\psi(\theta))$ and mapping this parameterization to a 3D space (θ,ϕ,ψ) they obtain a 3D curve . Then straight line segments are fitted to this 3D curve. Finally 3D curve is encoded using curvature and the torsion angles.
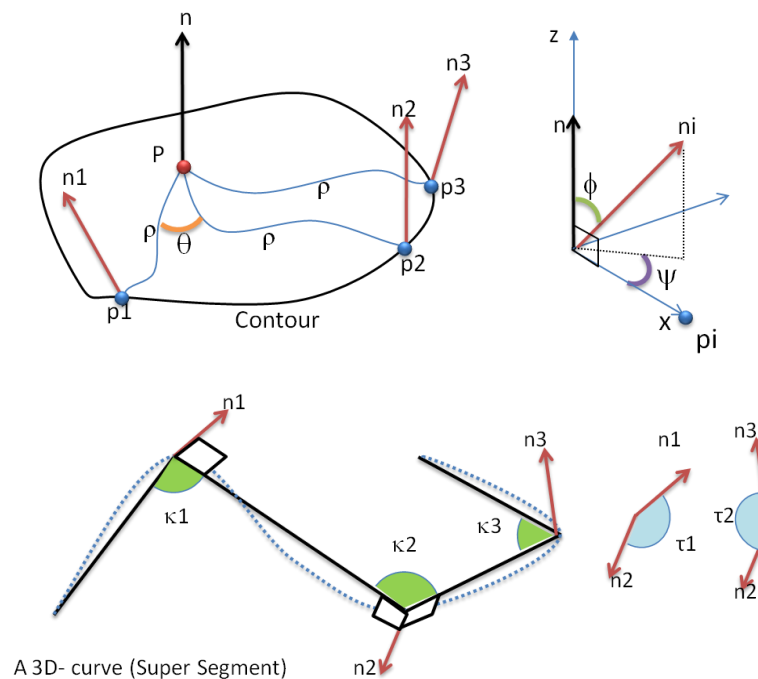
Figure 2-10 Splash and 3D curve

Point Signatures: A New Representation for 3D Object Recognition

Chua and Jarvis [72] inspired their work from that of Stein and Medioni [71]. The authors place a sphere centered at the interest point. Intersection of the sphere and the surface forms a contour. By Principal Component Analysis of this contour, a plane is fitted. This is the

29

plane, where the distance from the points on the contour to the plane is minimum. Normal vector at the interest point is assumed as the normal vector of that plane. The plane is translated to the interest point. Then, a 1D parametric curve $d(\mu)$, descriptor of the region, is formed from the signed distances from the points on the contour to the plane.
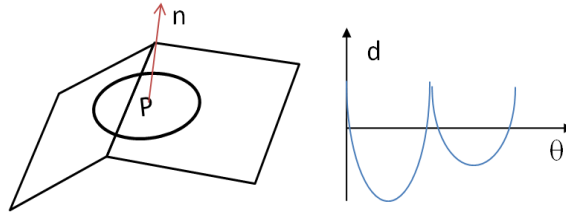
Figure 2-11 Surface and the corresponding "Point Signature" function

Recognizing Objects in Range Data Using Regional Point Descriptors

Frome et al. [73] extends the shape context idea from 2D to 3D. It is similar to "Shape Histograms" [30]. It is simply a 3D histogram accumulating the shape points by using their spherical coordinates (θ, ϕ, R). For scale invariance, the volume is partitioned into 15 equal divisions along the radial direction. Since the composed histogram is not rotation invariant, shape is rotated about its azimuth direction in L different positions and L different histograms are stored as the descriptor of the object. As an obvious solution to this rotation invariance problem, they also propose harmonic shape context description, which is a Spherical Harmonic Transform (SHT) of the 3D shape context matrix. The descriptor is a vector of the amplitudes of this transformation. As a similarity measure, L_2 distance for 3D shape contexts and the inverse of the normalized correlation for harmonic shape contexts are used. In the presence of clutter and noise, 3D shape context and harmonic shape context outperforms spin image method.

Salient Geometric Features for Partial Shape Matching and Similarity

Ran and Daniel [74] fit implicit quadric surface locally such that f(x,y,z)=0 by using least squares minimization. Then, for each vertex in the mesh, they find the projections onto this surface. After that for each vertex they compute the Gaussian curvature. A region growing considering the vertex Gaussian curvature values is performed and local pathes are formed. Then, a "saliency grade" function over the regions is assigned. Later, again with region

30

growing clusters are formed that maximizes the saliency grade of the cluster. Clusters with the highest grades gives the salient geometric features of the given shape. Hence, they extract the top 10% features.

Distinctive Regions of 3D surfaces

The main emphasis of the paper [75] is to find the distinctive regions of object classes. This aim is achieved by analyzing the classified database. The algorithm can be summarized as follows: Random points on the surfaces are selected, spheres at different scales are placed from regions, spheres are centered at those points, then for every region spherical harmonic transform is applied to obtain a local shape description. After these steps, the authors compare all pairs of descriptors and rank them from the best result to the worst. Then, this ranking is examined for finding the most discriminating parts of the classes. These parts are the regions from the same class of objects whose descriptors appear at the top of their ranked list. Next, they represent the mesh in the database with its most distinctive shape descriptors. In this approach, as an important constraint, the database should be classified and the intra-class similarity should be low.

3D free-form object recognition in range images using local surface patches

Chen and Bhanu [76] build up a range image recognition sytem. Authors initially fit a quadric surface, $f(x, y) = ax^2 + by^2 + cxy + dx + ey + f$, to a local window by using least squares method. Then, for each surface point *Gaussian, Mean*, and *principal curvature* values are calculated analytically. After obtaining those values, *shape index* values at every point is computed and local maxima and local minima of this mapping are selected as the salient points of the object. Finally, local descriptor is formed as a 2D histogram around these salient points. 2D histogram accumulates the shape index values and the angle between the normal of the interest point and that of its neighbors.

Thrift: Local 3D Structure Recognition

Thrift approach [77] makes a strong emphasis on the repeatability of the interest point (keypoint, salient point and interest point is used interchangeably throughout the dissertation). The authors first estimate the density function, say f(x,y,z) (surface points) by regular sampling method and obtain a voxel representation of the surface . Then, construct a

scale space by convolving the voxel F (3D matrix) by down sampling at each time with a 3D Gaussian Kernel.

$$S(x, y, z, \sigma) = \left(F \otimes g(\sigma)\right)(x, y, z) \ \text{ where } g(x, y, z\sigma) = \exp\left(\frac{-x^2 - y^2 - z^2}{2\sigma^2}\right) \qquad 2.20$$

In order to locate an interest point, the authors use determinant of the 3D Hessian matrix. Gradient vector for three principal directions should be large. In fact, they use the term curvature, but the proposed term corresponds to the gradient vector magnitude.

$$H(p, \sigma) = \begin{pmatrix} S_{xx}(p,\sigma) & S_{xy}(p,\sigma) & S_{xz}(p,\sigma) \\ S_{xy}(p,\sigma) & S_{yy}(p,\sigma) & S_{yz}(p,\sigma) \\ S_{xz}(p,\sigma) & S_{yz}(p,\sigma) & S_{zz}(p,\sigma) \end{pmatrix}, \ \ \text{where } S_{xx}(p,\sigma) = \frac{\partial^2}{\partial x^2} S(p,\sigma) \qquad 2.21$$

Finally, the interest points become the local maxima of $|\det(H)|$. They use the normal vector as the descriptor of the salient point. Their work presents the repeatability of this interest point detector about %50 in the presence of a 10% percent noise. However, matching results are relatively poor and conducted on few examples, which are not obvious.

Part Analogies in Sets of Object:

In this study [78], 3D models are partitioned into parts and a signature based on the geometric characteristics and its relation to the complete model is associated to them. First of all, the algorithm starts by evaluating the Shape Diameter Function (SDF) for each point on the model. Let p be a point on the surface of the mesh and let n be the normal vector. A cone is centered around, inward normal direction is placed and rays from the point p are sent to the other side of the mesh within the cone. Mean value of all lengths of the rays are calculated and the ones within one standard deviation of the mean are selected. SDF is calculated as the weighted average of the lengths of the selected rays. Weights are the inverse of the angles between the ray and the center of the cone. The cone angle is specified in their earlier work [79] as 120°.

After evaluating SDF values for each point, mesh is partitioned. They utilize the Gaussian Mixture Model by fitting k different Gaussians to the histogram of SDF values. They select k as 5 during their experiments. Signature of a part consists of the normalized SDF histogram

of the corresponding part and the ratio of the part to the whole model. The local description of the method requires full model; hence, it cannot be applied to the range imaginary. User specified segment numbers can be argued and the method works fine with articulated objects.

Table 2 gives a summary of the literature which utilizes local surface properties for keypoint selection, description, and matching. Some of them are already explained in detail in the previous paragraphs; however, included in the list for the sake of completeness.

Table 2. Summary of a some of the local surface description studies

	Name	Authors	Year	Methodology & Keywords	Data Format
1.	Structural Indexing: Efficient 3-D Object Recognition [71]	Fridtjof Stein, Gerard Medioni	1992	3D curve and Splashes	Range Scenes
2.	Point Signatures: A New Representation for 3D Object Recognition [72]	Ching Seng Chua, Ray Jarvis	1997	Signed Distance	Range Scenes
3.	Using spin images for efficient object recognition in cluttered 3D scenes [58]	Andrew E. Johnson, Martial Hebert	1999	Spin images	Laser Scanned Scenes
4.	3D Object Recognition From Range Images Using Local Feature Histograms [29]	Günter Hetzel et al.	2001	Global histograms are computed using local features: Normal, depth, Shape Index	Range Models
5.	A New Paradigm for Recognizing 3-D Object Shapes from Range Data [80]	Salvador Ruiz Correa et al.	2001	Manual keypoint selection, Spin image Description, SVM training	Range Scenes
6.	Multi-scale Feature Extraction on Point-Sampled Surface [81]	Mark Pauly et al.	2003	Surface variation calculation at different neighborhood sizes (scale), not for matching	NA
7.	Recognizing Objects in Range Data Using Regional Point Descriptors [73]	Andrea Frome et al.	2004	3D shape context and harmonic shape context, pick random local points	Laser Scanned Scenes
8.	Salient Geometric Features for Partial Shape Matching and Similarity [82]	Ran Gal, Daniel Cohen-Or	2006	Saliency Grade, Curvature	Full Models

33

Table 2. Continued. Summary of a some of the local surface description studies

9.	Rapid Object Indexing Using Locality Sensitive Hashing and Joint 3D-Signature Space Estimation [83]	Bogdan Matei et al.	2006	KeyPoint Selection: High surface variation, description: Spin images	Laser Scanned Scenes
10.	Distinctive Regions of 3D surfaces [75]	Philip Shilane, Thomas Funkhouser	2007	Random sampling, Scale Space, Spherical Harmonic Transform	Full Mesh Models
11.	3D free-form object recognition in range images using local surface patches [76]	Hui Chen, Bir Bhanu	2007	Surface fitting, shape index, histogram	Range Models
12.	Thrift: Local 3D Structure Recognition [77]	Alex Flint et al.	2007	3D Hessian, Surface Normal vector. Keypoint repeatability.	Laser Scanned Scenes
13.	3D object recognition from range images using pyramid matching [84]	Xinju Li, Igor Guskov	2007	Scale Space in 3D Spatial movement along normal direction, Spin image	Range Models
14.	Multi-scale Feature Extraction for 3D Surface Registration Using Local Shape Variation [85]	Huy Tho Ho, Danny Gibbins	2008	Scale Space using surface variation, Spin image	Range Models
15.	Sparse points matching by combining 3D mesh saliency with statistical descriptors [86]	U. Castellani et al.	2008	Scale Space, Spatial movement along normal direction, principal curvatures, HMM	Range Models
16.	Salient Region Detection and Feature Extraction in 3D Visual Data [87]	Min Dong, Yanhua Chen	2008	Scale Space, Gradient location orientation histogram in 3D	Volumetric MRI images
17.	Part Analogies in Sets of Object [78]	S. Shalom et al.	2008	Shape diameter function, Normalized Histogram of SDF	*Requires* Full Models
18.	Local Feature Extraction and Matching on Range Images: 2.5D SIFT [88]	Tsz-Wai Rachel Lo, J. Paul Siebert	2009	Range-> Sift -> keypoint localization->shape index	Range Models
19.	On the Repeatability and Quality of Keypoints for Local Feature-based 3D Object Retrieval from Cluttered Scenes [89]	A. Mian et al.	2009	Keypoint selection based on symmetric deviation (k1/k2), depth histogram	Range Models

2.5   Discussion on 3D Shape Retrieval Techniques

Histogram based methods, such as EGI, CEGI, shape histograms, shape distributions, and shape spectrum, are sensitive topology changes and are not discriminative for partial similarity indexing. View based approaches, such as Light-Fields, requires multi-view images of the query, which is not available in single depth searching. Graph based methods, such as Reeb graphs, are constructed on the connectivity of the parts. Range images usually contain occlusion and self-occlusion resulting in numerous disconnected regions. This property makes graph based approached non-practical in range image description. Transform based approaches, e.g., DFT and Spherical Harmonics, are also sensitive to global characteristics of the shape, but represents all available information in the frequency domain. It is a way of changing the representation domain of the signal. With this property, transform based methods differ from the previous ones in terms of discrimination.

Based on these discussions, Spherical Harmonics Transform is adapted for range model similarity indexing (Chapter 5). Other local surface properties studied in the aforementioned works, such as normals, curvatures, and shape index, are also utilized during segmentation, keypoint selection and correspondence matching stages.

CHAPTER 3

SEGMENTATION OF RANGE DATA

With the increase in the performance of the range detector technologies, and the decrease in costs of off-the-shelf hardware for 3D sensing and computing, 3D systems are applied to a number of different applications, such as 3D television, robotics, virtual reality. Most of these applications are utilized in the real-world environments in which the object and semantic models are far from their synthetic counterparts. The differences between the objects residing in the scene and their synthetic models make the segmentation process a crucial and prerequisite step for further tasks such as recognition, tracking, shape retrieval.

Many methods have attempted to solve the segmentation problem in range images [90]. Most of these methods exploit the local surface properties and/or edge information, and performed on laser scanner outputs or synthetic images generated via computer graphics. The local surface properties and edge information can be extracted more precisely from laser scanners and synthetic images in compared to the data generated by Time-of-Flight (TOF) cameras due to the high Signal to Noise Ratio and high resolution space discretization of the laser scanner devices. However, the real-time performance and the small dimensions of the TOF camera make it also a strong candidate for sensor selection.

This chapter begins with planar structure detection in range scenes. Problem definition, proposals, modifications on available methods and results are presented. Later, well-known segmentation methods are adapted for segmenting real range scenes, where the inputs are the Time of Flight Camera outputs. Color information is integrated with the spatial information and its effects on the performance of the segmentation methods are evaluated. Several local surface properties such as curvatures, surface variances, normals are utilized

with region growing. Finally, dissertation presents a novel algorithm to segment 3D point cloud (obtained via TOF camera) into objects by using the spatial information.

In the proposed approach, initially, regions with similar surface normal and curvature values are segmented and planar surfaces models are estimated for these regions. Estimated plane with smallest residual and largest member point number is chosen as major plane which represents the surface that objects can be placed. Afterwards, 3D points are projected onto the major plane and a saliency map is generated by using direct density estimation method which gives likelihood of an object existing on a particular position on the major plane. In other words, the local peaks in the saliency map represent the most probable object locations. Finally, points which are closer to the local peaks over a certain threshold are segmented into objects.

3.1   2.5D Data

As mentioned previously scenes of 2.5D data are more probable to be encountered within the real 3D applications than the watertight 3D mesh models (e.g., robots, TV, tele-immerse systems, machine vision, Human-Computer interaction). With this motivation, we tested segmentation methods on 2.5D point clouds acquired from a range camera.

We employ a time of flight (TOF) camera Swiss Ranger 3000 (SR3000) produced by Swiss Center for Electronics and Microtechnology (CSEM) (Figure 3-1).

Figure 3-1. The CSEM SwissRanger SR3000 TOF Camera and its illumination. Second photograph is taken with a standard camera in night vision mode.

SR3000 acquires both gray level image and depth of the scene. Depending on the modulation frequency setting of the camera, distance range varies between 5.00-7.9 meters. Distance calculation of TOF cameras depend on the propagation time of the light.

$$Distance = \frac{Speed\ of\ Ligth\ X\ \Delta t}{2},$$ 3.1

where the speed of light is 299792458 m.s^{-1} and Δt is the elapsed time of the specific light. SR3000 emits light (870nm) which is invisible with a naked eye (Figure 3-1) with 55 LEDs. A CMOS/CCD sensor with a spatial resolution of 176x144 active pixels senses the reflected light and calculates the distance. Detailed information on this calculation can be found in [91, 92].

There are distance resolution limitations such as thermal noise, quantization noise, reset noise, and electronic shot noise and practical limitations, such as noise, blur, overflow, reflectance, sensitivity, and multiple targets [92]. Figure 3-2 is an illustration of the multiple reflection problem.
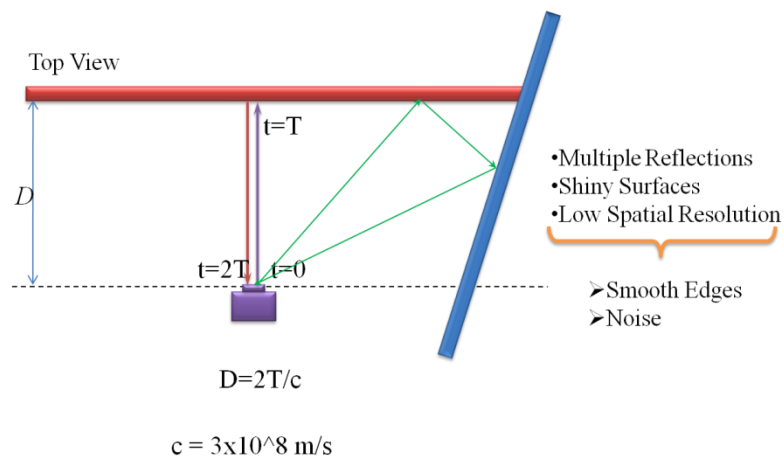
Figure 3-2. Multiple target case illustration and measurement errors

Figure 3-3 shows a sample depth data in point cloud representation and its corresponding intensity image. As shown in Figure 3-1, which is taken with a standard camera in night vision mode, camera's illumination is not ambient. Thus, intensity image of SR3000 is not handy.

Figure 3-3. Point cloud and corresponding intensity image of the scene acquired by SR3000 TOF camera

3.1.1 Color Information

Laser scanning devices or TOF cameras do not capture color image of the scene, except the intensity like image (Figure 3-3), which is practically useless. However, color information can be handy for analyzing the data. To obtain the color information of each point, we use TOF cameras' intensity images and a standard camera with the calibration tool obtained from [93]. Due to internal and external calibration errors exact match between two views cannot be achieved.

Figure 3-4. Point cloud obtained by TOF camera. Color image of the same scene acquired by a standard camera. Calibration of both images results colored point cloud data

3.2 3D Plane Extraction

When we examine the surroundings around us, it could be observed that most of the (man-made) objects have planar, nearly planar or piecewise planar geometric structure, such as walls, floors, doors, roads, etc. Thus, detecting and describing planar structures on a scene from set of 3D points could be a meaningful effort before any high level global or local 3D representation. We test and compare Hough Transform, RANSAC and HK segmentation

methods. Moreover, we propose modifications on Hough Transform and RANSAC that improves the performance of the planarity detection.
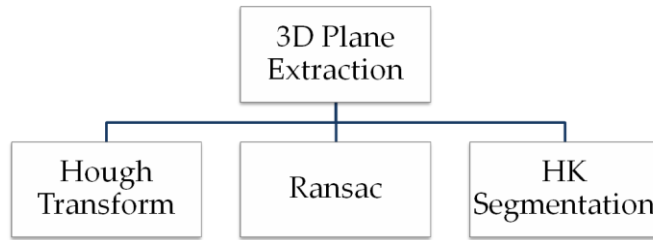
Figure 3-5. Methods used in planar structure detection in range images

3.2.1 Hough Transform

Hough transform is a popular algorithm in computer vision, proposed by Hough [94] for line detection in 2D images, later extended for detecting circles, ellipses, and planes. Hough transformation for 3D object recognition is applied by Zaharia and Preteux [95, 96] and used to define planar structures in depth map data, LIDAR data and X-Ray data [97-99]. The fundamental assumption is piecewise planarity of objects. Objects, buildings, even people can be represented by combinations of planar surfaces. Thus detecting planar patches in 3D data is a fundamental step in recognizing objects. Applications working with LIDAR data (airborne laser scanning system) widely use Hough transform to detect planes (parks, roofs, etc).

Introducing Hough Transformation for line detection is a good starting point for explaining 3D Hough Transformation (3DHT). It is a voting scheme, each point in a set of points $S = \{P_1, P_2, \dots, P_n\}$, votes for the lines passing through it. For a continuous space, there would be infinite number of lines (Figure 3-6), yet, we usually discretize the space.
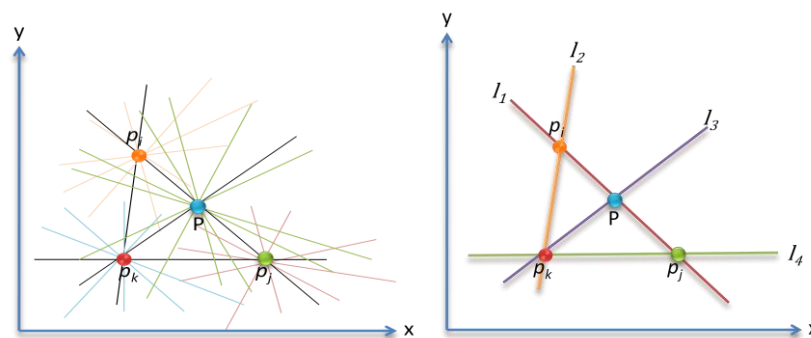
Figure 3-6. Line Detection using Hough Transformation

A line can be represented with the following equation:

$$y = mx + n$$

<div align="right">3.2</div>

Our aim is to find all (m,n) pairs for each (x,y) pair in our point set. However, the equation $m = \frac{y-n}{x}$ or equivalently n= $y - mx$ has a problem for a constant x. In order to overcome this problem, parametric line equation is suggested, instead of the previous one having the following form:

$$x \cos \theta + y \sin \theta = \rho$$

<div align="right">3.3</div>
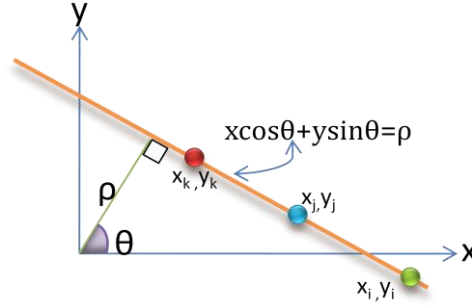
Figure 3-7. Parametric line equation

Thus, histograming (counting number of points voting for this line) is achieved in the parameter space (θ, ρ) where

$$-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$$

$$-(\sqrt{|x|_{max}^2 + |y|_{max}^2}) \leq \rho \leq (\sqrt{|x|_{max}^2 + |y|_{max}^2}$$

<div align="right">3.4</div>

Similarly, plane equation can be given in both Euclidean space and in parametric space:

$$ax + by + cz = d$$
$$x \cos \theta \cos \varphi + y \sin \theta \cos \varphi + z \sin \varphi = \rho$$

<div align="right">3.5</div>

Each point P=(x,y,z) votes for several (θ, φ, ρ) trio depending on the space discretization. Coarser parameter space division will result erroneous planes, whereas finer discretizing costs time and space. Thus, decision is achieved in a trial and error fashion. Parameters for the best plane are the ones taking maximum votes, i.e. the most probable plane contains the maximum number of points (in Figure 3-6 best line is ℓ_1 with a maximum vote of three, while ℓ_2, ℓ_3, and ℓ_4 takes two votes each which are the next most probable lines). Applications working with watertight models compare histograms, accumulators, to obtain a similarity measure.

Hough Transform with a fine parameter discretization works well for finding the major plane in our data; in this context, a major plane is the flat surface owing plenty portion of the scene points. However, Hough Transformation is not successful for finding "*distinct*" remaining planar structures in a scene, even with a finer parameterization. The reason is illustrated in Figure 3-8; human perception for planar structures in a point cloud (first image) results with the two planes shown in second image of Figure 3-8, but Hough transformation will come up with a different decision, a probable one is shown in third image of the figure.
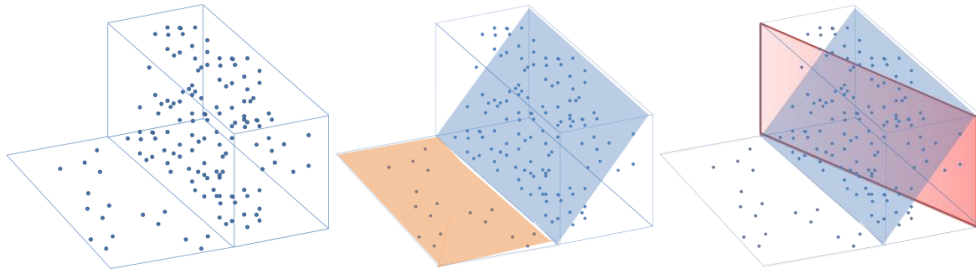
Figure 3-8.*(Left to Right):* point cloud, expected plane extraction, probable plane detection with Hough transformation
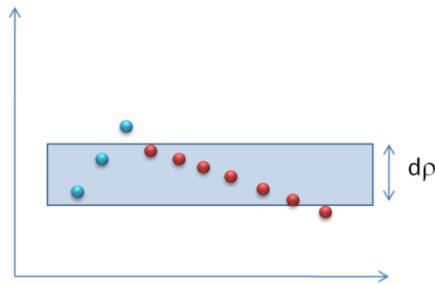
Figure 3-9. Hough Transformation parameter sensitivity

In fact, Hough algorithm exactly fulfills what is expected from itself, since it is demanded for the planes which have higher votes. As shown in Figure 3-9 maximum vote does not always correspond to our perception.

Based on our experience, we could argue that primary plane can be detected by Hough, whereas remaining ones might fail. Thus, we modify the algorithm to obtain a recursive one as follows:

- Apply 3D Hough Transformation (3DHT) on the 3D point cloud
- Detect the plane that has the maximum vote
- Remove points belonging to this plane

- Run 3DHT with the remaining points until the number of points is smaller than a threshold.

Recursive Hough transformation results are shown in Figure 3-11. Scenes used in this study are shown in Figure 3-10 (First scene is used only in segmentation section).

Data1   Data2   Data3   Data4

Data5
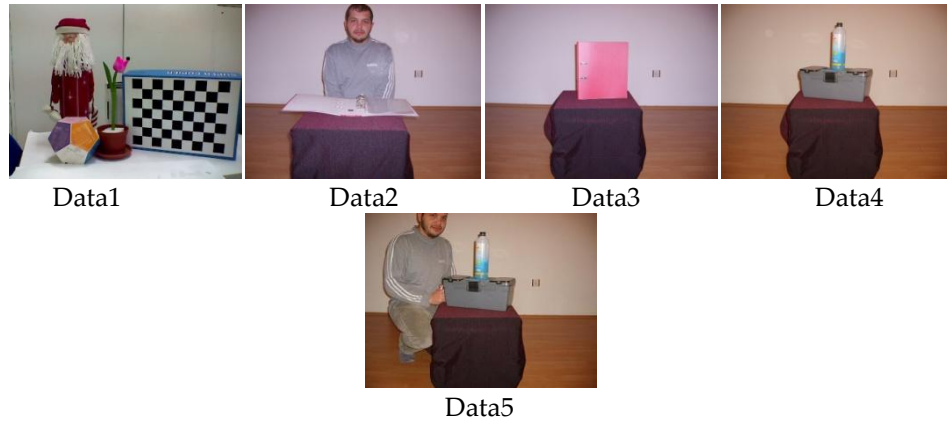
Figure 3-10. Scenes used in plane extraction and segmentation tests
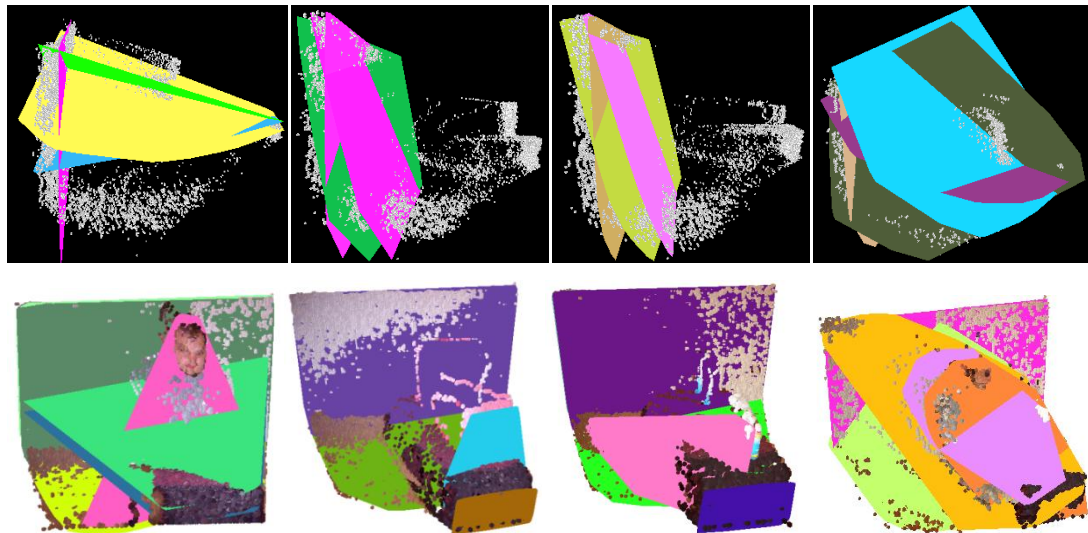
Figure 3-11. *(Top Row)* Planes obtained by standard Hough Transformation, side views are used for easy comprehension, *(Bottom Row)* Planes obtained by Recursive Hough transformation
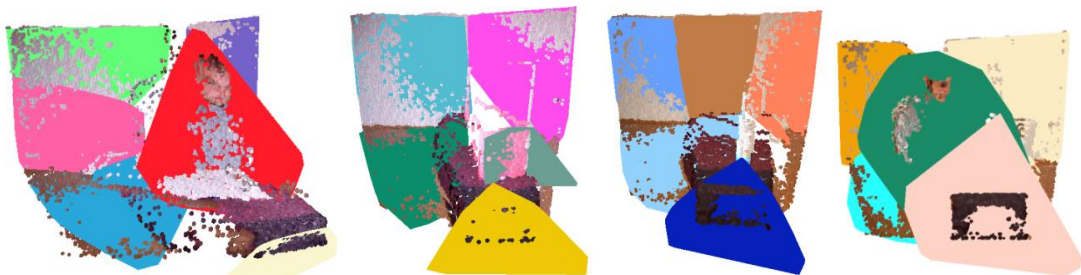
Figure 3-12. Planes obtained by Hough Transformation after colored K-Means Segmentation

Another technique to improve Hough Transformation is to import color information. There can be two alternatives for intensity assisted Hough Transformation for plane detection. The first approach can be summarized as follows:

- Segment regions according to their intensities, and
- Run 3DHT on each segment.

For segmentation, we prefer to use K-Means. The results for colored K-Means segmented Hough Transformation are shown in Figure 3-12. Details of segmentation algorithms and their performance measures will be given in the next section and they are shown in Figure 3-18 and in Figure 3-19.

The second approach has also two main steps:

- Obtain a 4D Hough Matrix consisting of (ρ,θ,ϕ, rgb), i.e. a point can vote for a plane with the ones having similar intensity values, and
- Find the maximum of 4D Hough Matrix.

At first glance, the second approach seems to be more effective, while it is more complex; whereas, the first approach might suffer from the segmentation errors in spite of its time and space efficiency. The second approach is left as a future study.

3.2.2 Random Sample Consensus (RANSAC)

Random Sample Consensus (RANSAC) is a general outlier removal approach that could also be exploited for finding planar structures in 3D [100-103]. The algorithm can be summarized as follows:

- Select three non-collinear points (p_i, p_j, p_k) randomly from the input set $S=\{p_1, p_2, …, p_n\}$;
- These three non-collinear points forms a triangle. From these points plane parameters are extracted and the number of points is calculated which belong to this plane (i.e., if the distance of a point to the plane is smaller than a user defined threshold t, than the point belongs to the model);
- If the number of points is above a threshold, again a user defined parameter, then the model is accepted.

Algorithm finds numerous noisy planes, since our data is noisy and there are many combinations of triple non-collinear points. Selection of initial triangle is critical and affects

results significantly. Hence, we insert some constraints for initial plane (triangle) selection besides the non-collinearity:

- Select points (p_i, p_j, p_k) such that $distance(p_i, p_j) > t$, $distance(p_i, p_k) > t$, and $distance(p_j, p_k) > t$, i.e. edges are not too small compared to the scene's maximum dimensions (Figure 3-13) and nearly equilateral; $distance(p_i, p_j) \approx distance(p_i, p_k) \approx distance(p_j, p_k)$.

- Select points if their colors are similar ($r_i \approx r_j \approx r_k$, $g_i \approx g_j \approx g_k$, $b_i \approx b_j \approx b_k$)
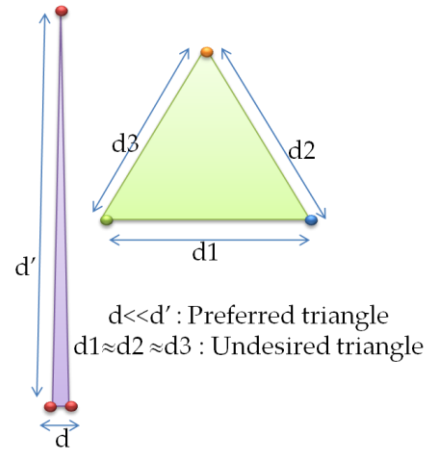
Figure 3-13. Triangle selection for RANSAC

The results are shown in Figure 3-14. Performance of the algorithm is similar to the recursive Hough. In spite of the constraints, RANSAC is not stable for our input. At each run, the method detects different planes for the same scene. This result is due to numerous combinations of the initial triangle selection. For reducing the number of triangle candidates, more elegant constraints should be introduced.
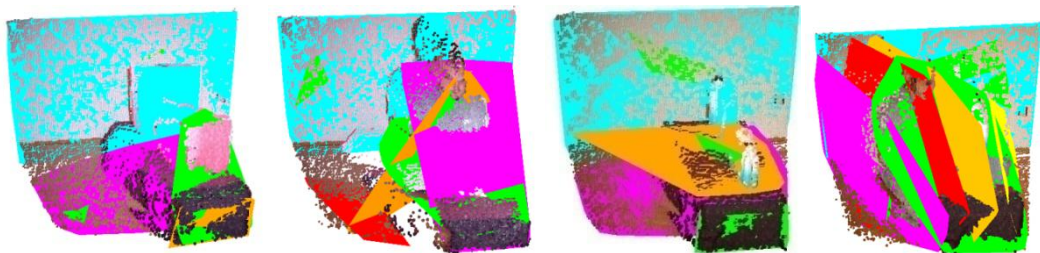
Figure 3-14. Ransac plane fitting results

3.2.1 HK Segmentation

Some local surface properties can be identified by utilizing differential geometry. Signs of the Mean Curvature (H) and Gaussian Curvature (K) provides us classifying each scene point into geometric classes shown by Figure 3-15, [104].

H\K	negative	0	positive
negative			
0			
positive			

K	H	Shape Class
0	0	Plane
0	+	Concave cylindrical
0	-	Convex cylindrical
+	+	Concave elliptic
+	-	Convex elliptic
-	Any	hyperbolic

Figure 3-15 Shape classes according to the HK curvatures [104]

Mean curvature (H) and Gaussian Curvature (K) is defined as follows:

$$K = \frac{h_{xx}h_{yy} - h_{xy}^2}{(1 + h_x^2 + h_y^2)^2}$$

$$2H = \frac{(1 + h_x^2)h_{yy} - 2h_x h_y h_{xy} + (1 + h_y^2)h_{xx}}{(1 + h_x^2 + h_y^2)^{3/2}}$$

3.6

where h is the range data, h_x is the first and h_{xx} is the second derivative along the "x" *direction*, similar notation holds for h_y and h_{yy}. These derivatives can be evaluated by surface fitting and direct differentiation. However, noisy measurements should be avoided, since it affects the HK segmentation. Besides a threshold should be utilized for zero values, since H and K is hardly equal to zero. Due to low SNR in our inputs, surface fitting is preferred to estimate derivatives.

3.2.1.1 *Surface Fitting*

Like all other sensors, laser scanners and Time of Flight cameras introduce noise over the input data. For reducing the effect of this poor quality surface representation, one could apply quadric surface fitting. Many efforts are put forward for surface fitting to point clouds [74, 105-107]. In these studies, the B-Spline and the quadric surface fitting are usually adapted after utilizing the Least Squares Estimation (LSE) method.

Considering the piecewise continuous nature of the surfaces appearing in the typical scenes, we fit quadratic surfaces to local patches. Mathematical details of the surface fitting can be found in the appendix. When we find the polynomial coefficients we can directly calculate

the local surface properties like curvature and the normal by differentiation. Experimental results for this estimation will be presented in the following sections.
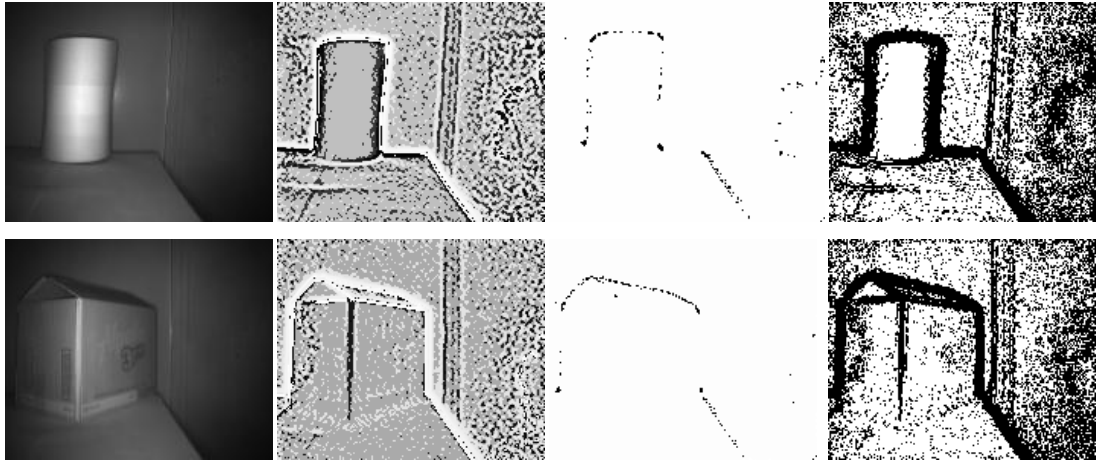
Figure 3-16. (Left to Right) TOF intensity image, H curvatures values mapped to gray level, K curvature values mapped to gray level, Planar regions shown in white (binary images)

Figure 3-16 illustrates the mapping H and K curvatures values to gray level for two test images. Planar regions obtained according to the HK values are shown in white in the last column. Even the HK values are calculated by surface fitting, which smoothes the surface, results are noisy. Due to noisy structure of the input and the noise sensitivity of the H and K computations, planarity structures are not detected directly by examining the HK values. A post processing such as region growing should be performed.

3.3   3D Object Segmentation

3.3.1 K-Means

K-means is one of the classical iterative clustering methods. It can be summarized as follows:

1. Assign initial random mean vector with K values $(m_1, m_2, m_3, …, m_K)$;
2. Cluster data according to the initial mean vector (assign data to a class whose mean is closest to it);
3. Calculate mean value for each cluster mean and update mean vector;
4. Assign data according to the new mean vector;
5. Repeat third and fourth steps until there is no change in the cluster mean values.

Essentially, K-means try to minimize total cluster variance and can be formulated by:

$$D = \sum_{i=1}^{K} \sum_{j=1}^{n} \left\| x_j^i - m_i \right\|^2, \qquad\qquad 3.7$$

where x_j^i represents the jth data in the ith cluster and m_i denotes the ith cluster's mean. Initial mean assignment is the critical step in the algorithm and effects results considerably. The accustomed strategy is picking up random K samples from the data set. The desired property of the algorithm is ability to handle different types of data.

We supplied only the spatial information of the point clouds to the implemented K-means algorithm. Figure 3-10 shows the tested scenes that we use throughout the experiments and Figure 3-17 presents the point cloud representation of the scenes acquired by TOF camera and their ground truth for segmentation. The results for K-means are shown in Figure 3-18. With the calibration of TOF and standard camera, we can access the color information of the points. Thus, modified the algorithm such that sample vectors are in the form of $p_i = \{x_i, y_i, z_i, r_i, g_i, b_i\}$. Since spatial and color space is different, they are normalized to have a maximum value of one. The simulation results of Colored K-means algorithm is also shown in Figure 3-18.

3.3.2 Mean-Shift

In Mean-Shift paradigm, the local peaks of a probability density function is tried to be estimated from the observed samples out of this density. For this probability density function, Mixtures of Gaussian distributed points are usually assumed and their related parameters (means) are estimated. Dense regions in the input space correspond to local maxima in the probability density functions (called *modes*). Modes of input points are obtained and clusters around these modes are generated [108]. A formal definition of the algorithm can be summarized as follows:

Given k input points $x = \{x_1, x_2, \dots, x_k\}$, in d dimensional space R^d, then the density estimation with the window size w is :

$$f(x) = \frac{1}{kw^n} \sum_{i=1}^{i=k} K\left(\frac{x - x_i}{w}\right), \qquad\qquad 3.8$$

where *K(x)* is the kernel function (usually Gaussian). The modes of the density function are located at the zeros of the gradient function $\nabla f(x) = 0$. It can be shown that [41] the displacement between ensemble average of the samples within an

arbitrary window and the center of this window always points in the direction of the mode; hence, a window moves toward the mode of the density, if the center of the window is shifted towards the ensemble average position . For segmentation purposes, this paradigm is exploited by grouping the starting points (pixels) whose modes turn out to be the same [41].

Mean-shift segmentation is implemented with and without the color information of the points. Mean shift segmentation results are shown in Figure 3-19.

In order to evaluate segmentation performance of the algorithms, we segment the regions manually to obtain the ground truth and compare them against the result of the algorithms. Let $c = \{c_1, c_2, \dots, c_k\}$ be the regions that are obtained by the segmentation algorithm where $c_i = \{p_1, p_2, \dots, p_w\}$ is a list of points constituting the region and let $s = \{s_1, s_2, \dots, s_n\}$ be the ground truth regions where $s_i = \{p_1, p_2, \dots, p_m\}$. Then, for each segment in c, one should decide the corresponding region in the ground truth by looking at the which region majority of the points in this region fall in and obtain correspondence vector $cs = \{cs_1, cs_2, \dots, cs_k\}$. Then, the error is computed as follows:

$$error_{(in\ number\ of\ points)} = \sum_{i=1}^{i=k} \sum_{j=1}^{j=w} d(i, p_j), \quad where\ d(i, p_j) \begin{cases} 1 & if\ p_j \notin s_{cs_i} \\ 0 & o.w. \end{cases} \quad 3.9$$
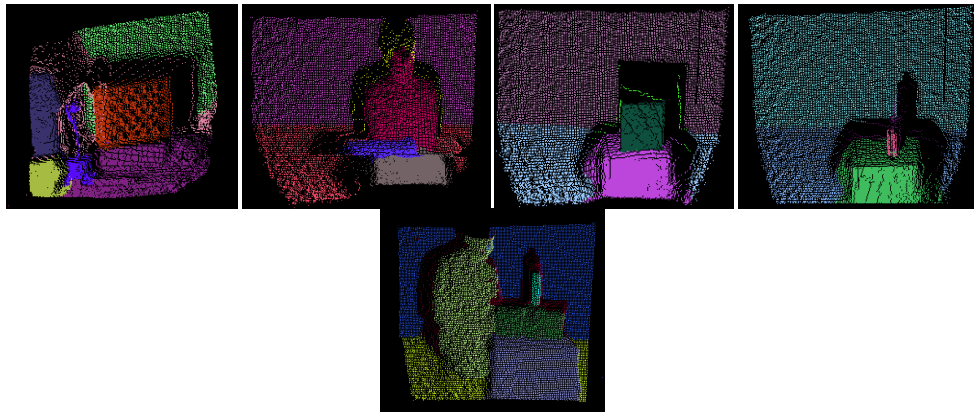
Figure 3-17. Point cloud captured by TOF camera and ground truth (each segment colored with different color)

K-Means (First Row) & Colored K-Means (2nd Row)

Error	Data1	Data2	Data3	Data4	Data5
K-Means	2754/18371≈0.15	2923/18491≈0.15	1740/18512≈0.09	1326/18514≈0.07	3778/18487≈0.2
Colored K-Means	7869/18371≈0.42	2369/18491≈0.13	1139/18512≈0.06	982/18514≈0.05	2960/18487≈0.16
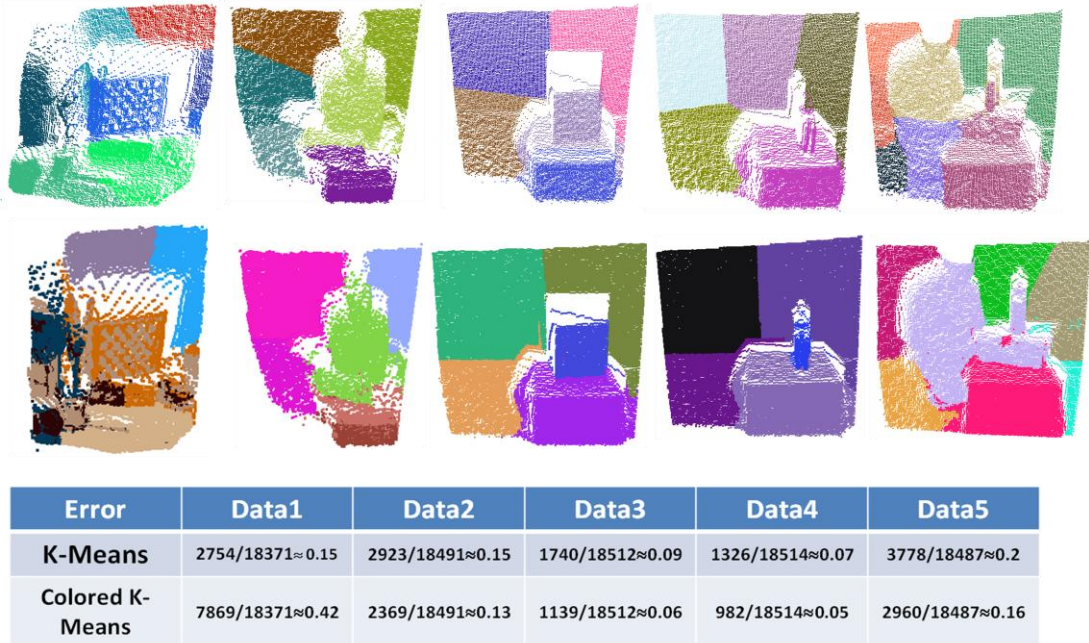
Figure 3-18. K-means and Colored K-Means Results. Error is given as the ratio of the number of points segmented incorrectly to the total number of scene points.

Mean Shift (1st Row) & Colored Mean Shift (2nd Row)

Error & (Cluster Difference)	Data1	Data2	Data3	Data4	Data5
Mean Shift	2754/18371≈ 0.15 (23)	3633/18491≈0.19 (-1)	3407/18512≈0.18 (5)	1697/18514≈0.09 (2)	4084/18487≈0.22 (0)
Colored Mean Shift	10531/18371≈0.57 (3-7=-4)	2453/18491≈0.13 (5-6=-1)	1071/18512≈0.06 (6-5=1)	739/18514≈0.04 (6-5=1)	3276/18487≈0.18 (7-7=0)

Figure 3-19. Mean Shift and Colored Mean Shift Segmentation Results. Error is given as the ratio of the number of points segmented incorrectly to the total number of scene points. Difference of the number of regions between the ground truth segmentation and the Mean Shift segmentation is given in parentheses.

First data set is quite noisy and occlusion is distinct; therefore error values for this data set is high. For evaluation of average error ratio, other data and corresponding error values are used and shown in Table 3. According to this table, color information improves the segmentation quality and both colored K-means and colored Mean Shift segmentation algorithms achieve around 10% error.

Table 3. Performance of the segmentation methods

Error (Number of misclassified points / Total num. Points x 100)	K-Means	Colored K-Means	Mean Shift	Colored Mean Shift
Data2	15 %	13%	19%	13%
Data3	9%	6%	18%	6%
Data4	7%	5%	9%	4%
Data5	20%	16%	22%	18%
Average	**12.75%**	**10%**	**17%**	**10.25%**

3.3.3 Region Growing

Rabbani [109], proposed a segmentation algorithm based on the surface normals. The author preferred surface normals over curvature, because of the high rates of over-segmentation due to the unreliable estimation of the curvature from noisy point clouds. However, both of these surface characteristics are considered in this work for geometric information. The proposed segmentation method consists of two main stages; normal and residual estimation, and region growing.

Point clouds can be used to find geometric information in the scene, since they explicitly represent the surfaces. However, surface properties, surface normals and curvatures, of the points in the point cloud are defined by their local neighbors rather than a single point. Therefore, local surface properties must be estimated from the local neighborhood of a query point. Eigen analysis of the covariance matrix of a local neighborhood can be used to estimate local surface properties [110] [111].

3.3.3.1   Local Surface Properties

- **Normal Estimation**

In their survey paper [112] Dey et al. compare different normal estimation methods for point clouds. Weighted plane fitting (WPF), adaptive plane fitting and big Delaunay balls (BDB)

methods are considered and tested for their performances. Their conclusions of the authors about these three methods are as follows: the observed performances are almost equal (though WPF gives the best results) to each other, when the noise level is low and the point cloud samples the surface more or less evenly. By following their conclusion, their experimental results and the resolution of MESA range camera (176x144) we decided to focus on WPF (weighted plane fitting) method to estimate the surface normals.

The estimated normal n of the point p in the point cloud data can be approximated with the normal of the local neighborhood of p. The normal of local neighborhood of p can be found by estimating the best plane fitting to the local patch.
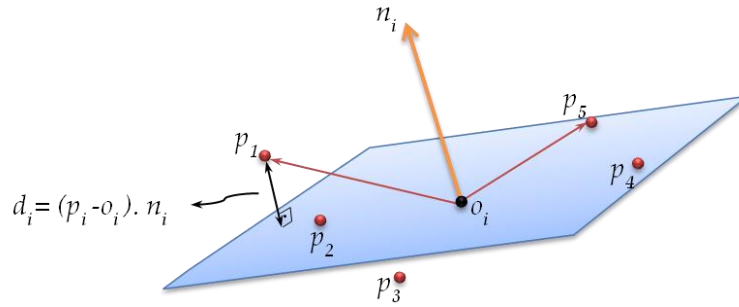
Figure 3-20. Tangent Plane Estimation

Let o_i is the center point of the point cloud $(p_1, p_2, p_3, p_4, p_5)$

The distance of a point to the plane (error):

$$d_i = (\bar{p}_i - \bar{o}_i).\bar{n}_i$$
$$d_i = \bar{n}_i^T (\bar{p}_i - \bar{o}_i)$$
$$let\ \bar{v}_i = \bar{p}_i - \bar{o}_i$$
$$d_i = \bar{n}_i^T \bar{v}_i$$

3.10

Sum of Squared distances:

$$D^2 = \sum_{i=1}^{n} d_i^2$$

$$D^2 = \sum_{i=1}^{n} (\bar{n}_i^T \bar{v}_i)^2$$

3.11

$$D^2 = \sum_{i=1}^{n} \bar{n}_i^T \bar{v}_i \bar{v}_i^T \bar{n}_i$$

$$C \triangleq \sum_{i=1}^{n} \bar{v}_i \bar{v}_i^T > 0$$

3.12

$$D^2 = \bar{n}_i^T \left(\sum_{i=1}^{n} \bar{v}_i \bar{v}_i^T \right) \bar{n}_i = \bar{n}_i^T C \, \bar{n}_i \quad such \ that \ \|n_i\| = 1$$

<div align="right">3.13</div>

$$\min(D^2) = \min(\bar{n}_i^T C \bar{n}_i) \ subject \ to \ \bar{n}_i^T \bar{n}_i = 1$$

Introducing the Lagrangian multiplier is equivalent to minimize Lagrangian

$$L(x) = \bar{n}_i^T C \, \bar{n}_i + \lambda(\bar{n}_i^T \bar{n}_i - 1)$$

<div align="right">3.14</div>

Differentiation and equating to zero gives:

$$C \, \bar{n}_i = \lambda \, \bar{n}_i$$

<div align="right">3.15</div>

Meaning that $\bar{n}_i$ is an eigenvector of C and the solution is:

$$\therefore \bar{n}_i = c_0$$

<div align="right">3.16</div>

where c_0 is the eigenvector corresponding to the minimum eigenvalue (λ_0) of the covariance matrix C.

The weighted covariance matrix from the points p_i of the local neighborhood where $i=1...k$ is:

$$C = \sum_{i=1}^{k} w_i (p_i - \bar{p})^T (p_i - \bar{p})$$

<div align="right">3.17</div>

The weight w_i for point p_i is defined as:

$$w_i = \begin{cases} \exp\left(-\dfrac{d_i^2}{\mu^2}\right) & if \ p_i \ is \ outlier \\ 1 & if \ p_i \ is \ inlier \end{cases}$$

<div align="right">3.18</div>

where μ is the mean distance from the query point p to all its neighbors p_i, and d_i is the distance from point p to a neighbor p_i. This weighting method reduces the effect of the outliers in the surface normal calculation process.

- **Curvature Estimation**

Surface curvature (surface variance) estimated at point p is computed using the following equation, called *surface variation*,

$$c = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2}, \quad \lambda_0 \leq \lambda_1 \leq \lambda_2,$$

<div align="right">3.19</div>

by comparing the three eigenvalues λ_i that are obtained from Principle Component Analysis (PCA) of the local neighborhood of the query point p. Since C is symmetric and positive semi-definite covariance matrix, its corresponding eigenvalues are real and nonnegative [113]. Eigenvalues λ_i measure the variation of the point set p_i along the associated eigenvector. Thus, λ_0 estimates how much the points deviate from the tangent plane (Figure 3-21). As a result surface variation is closely related with curvature [111]. If λ_0=0 then surface variation results in *0* which can be interpreted as a zero curvature (all points lie on a planar surface).

$$L(x) : (\overline{x} - p).v0 = 0$$
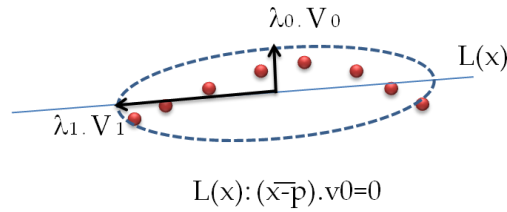
Figure 3-21. Curvature Estimation

Segmentation via region growing algorithm explained in [114] returns a set of labeled pixels, meeting the adjacency and similarity criteria. The iterative algorithm is as follows:

1) Begin with the center pixel at *(i; j)*.

2) Consider the pixels at 4 or 8 neigbourhood of the center pixel. Is that pixel 'similar' to the center pixel? If yes, push it onto the stack.

3) Mark the center pixel so it will not be considered again

4) Pop the stack to choose a new center pixel, thus redefining *i* and *j*.

5) If the stack is empty, stop. The set of marked pixels constitutes the region.

6) If the stack is not empty, go to the step two.

Previously defined local surface properties in the scene are considered as 'similarity measures', which are mentioned in step two. Therefore, region growing algorithm uses the point curvatures, normals and their residuals, in accordance with user specified parameters to group points belonging to the similar surfaces. The points in a segment should make a locally smooth surface, in which the normal vectors do not vary "significantly" from each other. This constraint would be enforced by having a threshold θ_{th} on the angles between the current seed point and the points added to the region. Additionally, a threshold on residual values r_{th} makes sure that the smooth areas are broken on the edges. A smoothness threshold in terms of the angle between the normals of the current seed and its neighbors is selected. If the smoothness angle threshold is expressed in radians, it can be enforced through dot product, as follows

$$\|n_p \cdot n_s\| \geq \cos(\theta_{th})$$

As the direction of normal vector has a 180° ambiguity the absolute value of the dot product is considered.

54

Segmentation results of the region growing method are shown in Figure 3-22, Figure 3-23, Figure 3-24, and Figure 3-25. Since planar regions are clustered separately even they belongs to the same object, object segmentation is not achieved, whereas partial segmentation is successful with region growing via surface normals and spatial closeness.
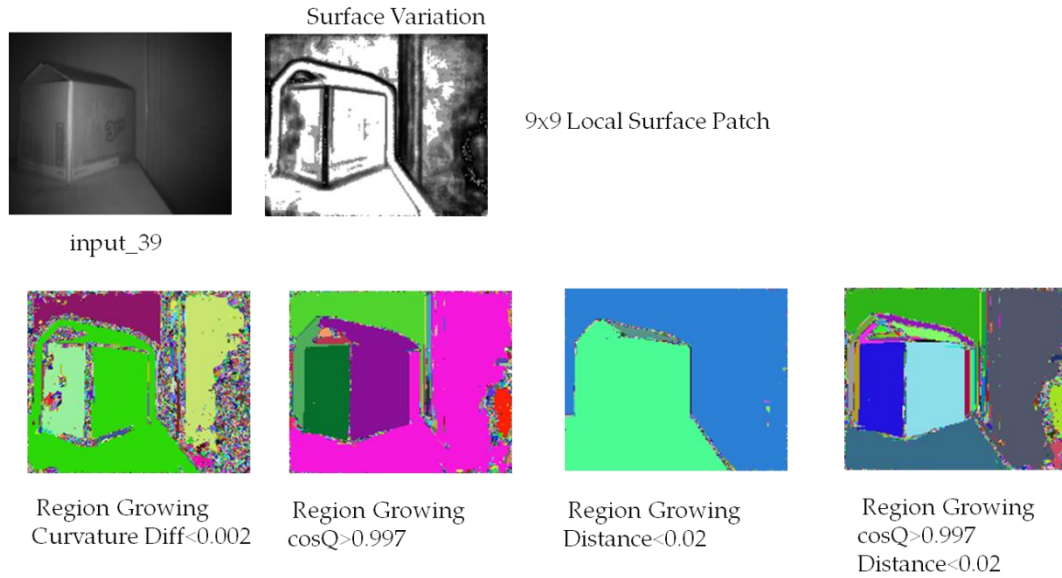
Figure 3-22. Region growing results based on distance and/or curvature similarity measure
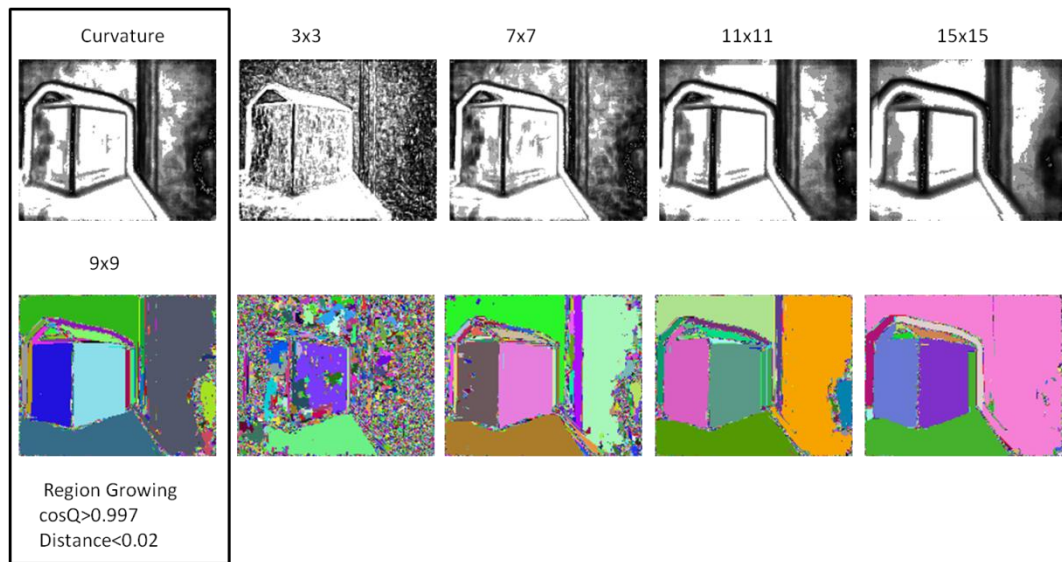
Figure 3-23. Region growing results with varying curvature estimation. *(Top row)* Surface variation estimated using different neighboring size mapped to gray level, *(Bottom row)* Corresponding region growing results

Surface Variation

9x9 Local Surface Patch

input_36

Region Growing
Curvature Diff<0.002

Region Growing
cosQ>0.997

Region Growing
Distance<0.02

Region Growing
cosQ>0.997
Ditance<0.02
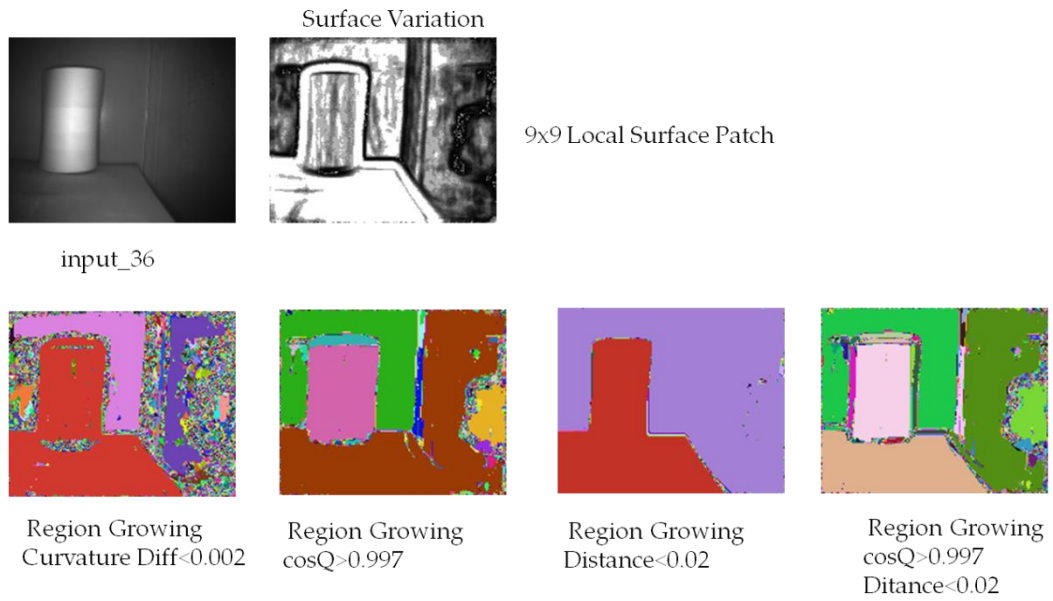
Figure 3-24. Region growing results based on distance and/or curvature similarity measure

input_61

9x9 Local Surface Patch

Region Growing
cosQ>0.997
Distance<0.02

Region Growing
Curvature Diff<0.002

Region Growing
cosQ>0.997

Region Growing
Distance<0.02

cosQ>0.990

Distance<0.05

cosQ>0.990
Distance<0.05
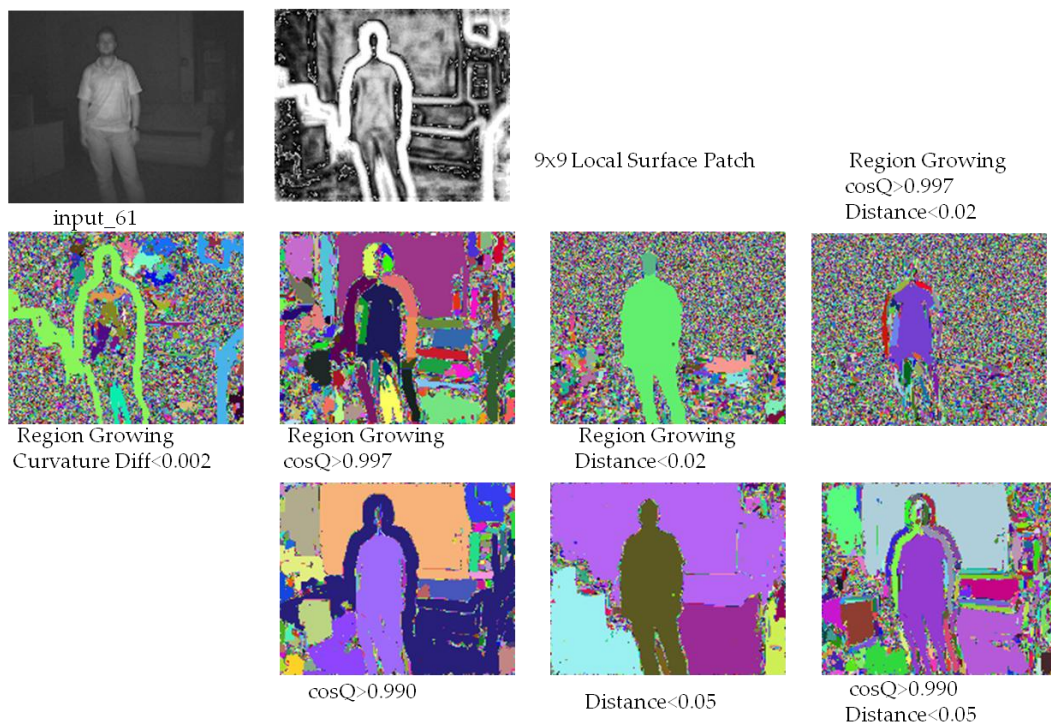
Figure 3-25. *(Top Row)* Corresponding intensity image of the input range image, curvature values mapped to gray level, *(Second and Third Rows)* Region growing results with different parameter selection

56

3.3.4 Mixture of Gaussians (Expectation Maximization)

We also consider mixture of Gaussians (MOG) based density estimation method ([115, 116]). However, similar problems arise with the K-means segmentation algorithm due to the necessity of giving number of Gaussians as the input to the algorithm. By the changing number of Gaussians, possible object regions are oversegmented or undersegmented resulting in false objects segmentation. Figure 3-26 shows MOG based segmentation with k=10.
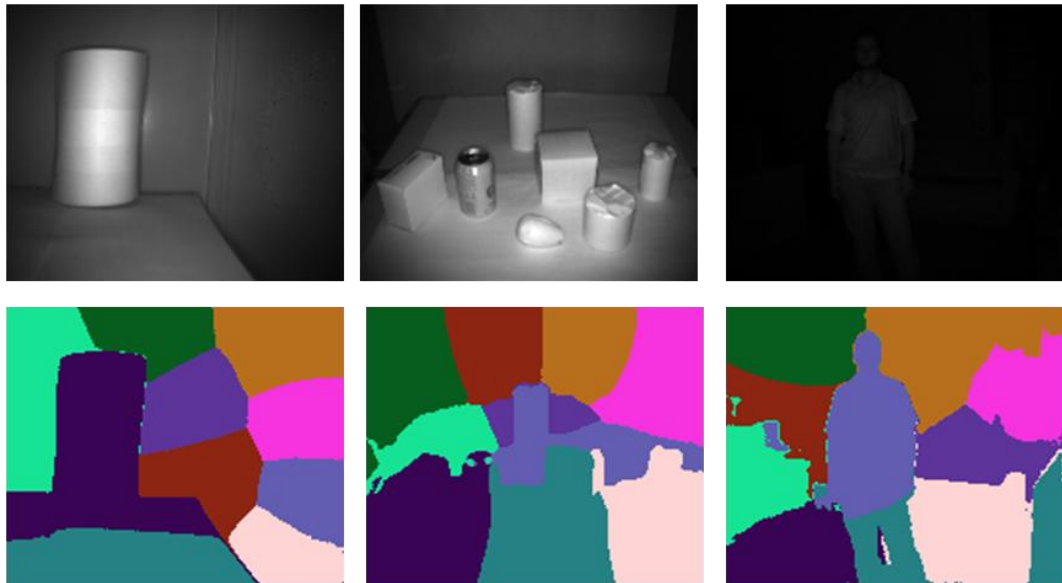
Figure 3-26. MOG based segmentation, number of clusters = 10

3.4 Segmentation based on Kernel Density Estimation

The low spatial resolution of the TOF camera decreases the accuracy of the local surface property estimation process and makes it more sensitive against noise. Moreover, false measurements due to the multiple reflections might result in significant variance in the estimated normal values within a small surface patch. This variance leads to oversegmented or undersegmented regions. Furthermore, color-based segmentation algorithms are ineffective, when the objects in the scene have similar color values with the background or consist of different colored patches. 3D information or edge based segmentation methods also fail due to the continuous and smooth transitions, originated from the multiple reflections, between object surfaces and background surfaces on which the objects are placed. However, utilization of the information about the background (surfaces) and the fact that the objects are placed on the background surfaces can be exploited to overcome most of

the aforementioned shortcomings. Therefore, we propose a segmentation method which exploits the available spatial and geometric information in the scene instead of using only local surface properties or only color information for segmenting surfaces.

Flowchart of the proposed method is shown in Figure 3-27. First, sparse (noisy) points (outliers) are eliminated from the point cloud. Then, the remaining points are projected onto a major plane which maximizes the separation of regular shaped items from each other. Major plane is extracted by using the plane fitting algorithm explained in the previous section. Afterwards, a saliency map is generated by kernel density estimation method for finding the most densely populated regions on the projection plane. Afterwards, local maximum points in the saliency map represent the most probable object locations. Finally, the points with a probability value larger than some certain threshold are assigned to the closest local maximums and labeled as objects.
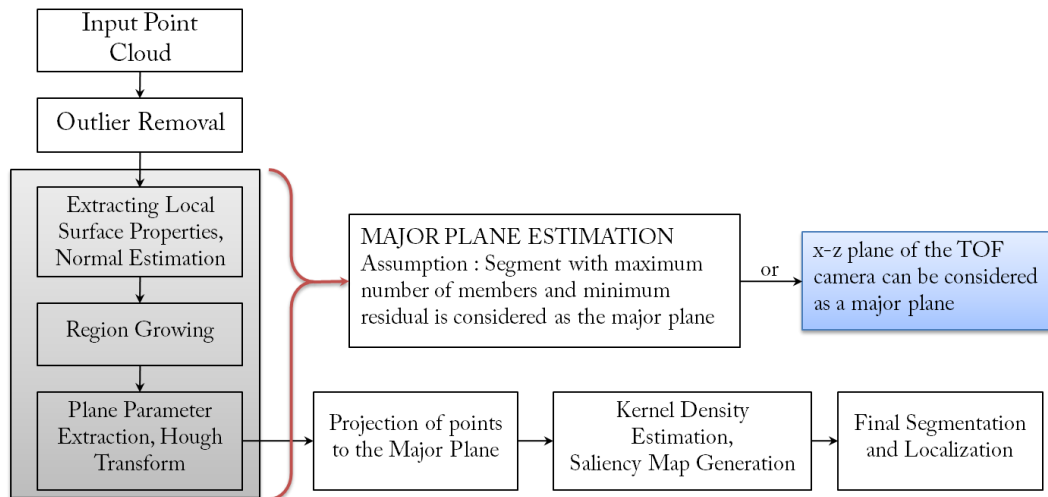
Figure 3-27. The flowchart of the proposed segmentation algorithm

3.4.1 Outlier Removal

Noisy measurements, especially at the object boundaries and on the relatively less visible surfaces, result in outliers in the point cloud. However, these outliers can be eliminated, since they generally situated far from their neighboring points. A small sphere with radius r is fitted to each point and the number of neighboring points inside the sphere is counted. If the number of points is greater than some certain threshold then the point is labeled as inlier. Otherwise, it is labeled as outlier and eliminated from the point cloud.

3.4.2 Kernel Density Estimation

Due to gravity, many objects stand orthogonal to the ground plane; hence, projecting 3D points onto this plane can be intuitively accepted as the "optimal" projection towards the classification of objects in the scene. In other words, such a projection could be equivalent to Fishers Linear Discriminant methodology [117] in Pattern Recognition area. Almost all regular shaped items occupying a space in an environment generate densely populated regions, when their surface points are projected onto the surface that they are standing on which we consider as major plane (Figure 3-28). Furthermore, projection of the background points which belongs to the surfaces, such as tables, ground etc. is sparse. Therefore, evaluating the 3D point cloud of the scene after projecting the points onto a major plane would be a convenient way to handle separation/segmentation problem. This novel plane minimizes the occlusions between items (as long as the objects are placed on the major plane, not on top of each other).

We assume that the plane with maximum number of inliers and minimum residual is the major plane in the scene. This plane is obtained by using the plane fitting method explained in Section 3.3.3.1. 3D points with similar local surface properties are grouped together and a plane model is estimated for each group. Then, the group with maximum number of members and minimum residual is considered as the major plane. When the major plane is not available due to some constraints (e.g when the camera's principal axis is parallel to the surface or the major plane cannot be found), x-z plane of the TOF camera can be considered as a major plane.

After projecting 3D points onto the major plane, densely populated regions represent the salient regions, where objects should exist. The idea of defining densely populated regions as salient/important regions resembles to the main idea of kernel density estimation in which the estimated density function has high values, where the samples are highly populated. In kernel density estimation method, instead of grouping observations together in bins, like in histogram density estimators, a small kernel is placed on each observation and summed. Therefore, high local peaks occur in high density regions.
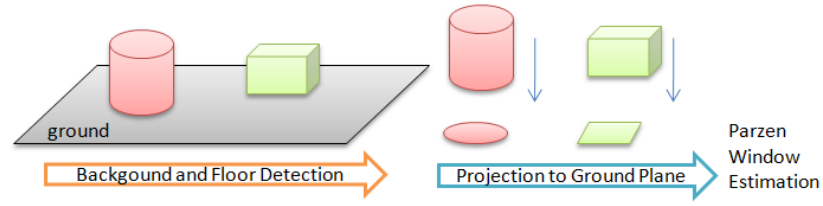
Figure 3-28. Illustration of the proposed method

Parzen window based saliency map generation [118] depends on estimating the probability density of observing objects in a non-parametrical manner, considering the density of 3D points projected on the major plane. Each 3D point contributes to the saliency map that has local (and sometimes global) maxima inside the highly dense regions.

The value of the estimated function at the point x, $\rho_n(x)$ is given as:

$$\rho_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_n}\varphi(\frac{x-x_i}{h_n})$$

3.20
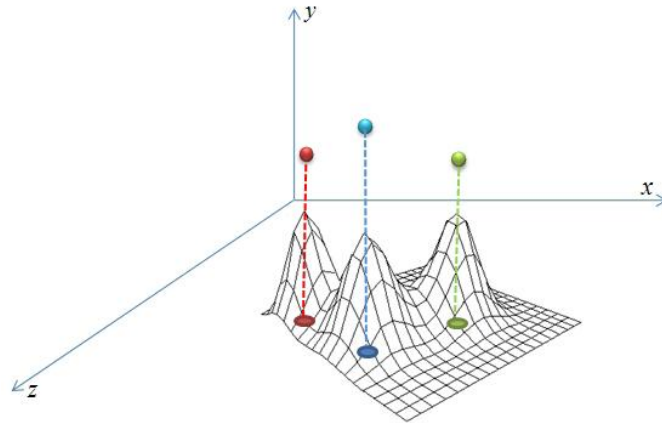
Figure 3-29. Illustration of the Parzen window approach

where $\varphi(u)$ is a window function and h_n is the smoothing parameter called the bandwidth. Gaussian kernel would be a proper choice since the contribution of each 3D point to the density estimation is equal to each other and decreases as the distance to the point increases. Then, the overall contribution of the 3D points to the saliency map and the estimate of the probability of observing an object at a particular location x on the saliency map is given as:

$$\rho_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma_i}e^{-\frac{(x-x_i)^2}{2\sigma_i^2}}$$

3.21

60

where n is the number of points, x_i is the projection of point i and σ_i is the kernel bandwidth. Hence, in the estimated saliency map regions where the projected points x_i are densely populated have high values (Figure 3-29). After generating the saliency map, modes (local maximums) in the map represent the possible object locations (salient regions where the objects can be situated).

As a next step, the points with a probability value greater than some certain threshold are assigned to the closest local maximums and labeled. Later, labeled points are segmented into objects. Moreover, the 3D points belonging to the background surfaces are eliminated, since they have low probability values.

The suitable kernel bandwidth σ_i must be carefully selected, since quite small bandwidth might result in very low resolution, while large bandwidth might lead to an over-smoothed density estimate. Thus, the size of the image and the average height of the arm/camera with respect to the ground should also be considered, while choosing the kernel bandwidth. In our experimental setup, we used kernel bandwidth $\sigma_i = 0.005$ which corresponds to 5 mm on the major plane.

3.5 Experimental Results and Discussions

Various scenes are tested by using the examined segmentation algorithms. K-means and mean-shift segmentation algorithms, which use only 3D point coordinates fail, since the points do not generate separable clusters in the 3D space due to the continuous nature of the 3D information in the scene. Integration of the color information with the spatial information results in minor improvements in the results which can be explained by the variation of color values on the object surfaces due to the shadows and direction of the light source. Moreover, color information would be not functional, if the objects have similar color values with the background or consist of different colored patches. Mean-shift algorithm has an advantage over K-means which is the number of the clusters are determined automatically. On the other hand, mean-shift algorithm is also parameter dependent and the selection of the hypersphere radius is critical.

Region growing algorithm tends to oversegment the objects, since the low spatial resolution and the noise of the TOF camera decreases the accuracy of the local surface property estimation process. Furthermore, the objects with multiple surfaces are oversegmented since the local surface properties significantly change at the edges, which stress the dissimilarity.

Results for the proposed method are given in Figure 3-30, Figure 3-31, and Figure 3-32. Figure 3-30 shows the generated saliency map via kernel density estimation from different angles. In the last row, image with red spots represent the local maximum points where the objects can be placed, and the last image shows the segmentation results. Proposed algorithm achieved better segmentation results on tested data in compared to the other methods. Considering the overall performance, the advantages of this algorithm are twofold. Segmentation results are more accurate and less sensitive to the noise, and the algorithm is less parameter dependent.

Proposed algorithm exploits the fact that many objects stand orthogonal to the ground plane due to gravity and the projection of 3D points onto this plane could be equivalent to Fisher Linear Discriminant methodology. The projection of the 3D points is followed by a kernel density estimation process in which a saliency map is generated. Salient regions which represents the most probable object locations have high values in the generated map. Then the points with a probability value larger than some certain threshold are assigned to the closest local maximums segmented into objects. The experimental results show that the locations of the objects are determined quite accurately. Compared to the tested methods proposed algorithm is less sensitive to the noise, less parameter dependent and leads more accurate segmentation results.
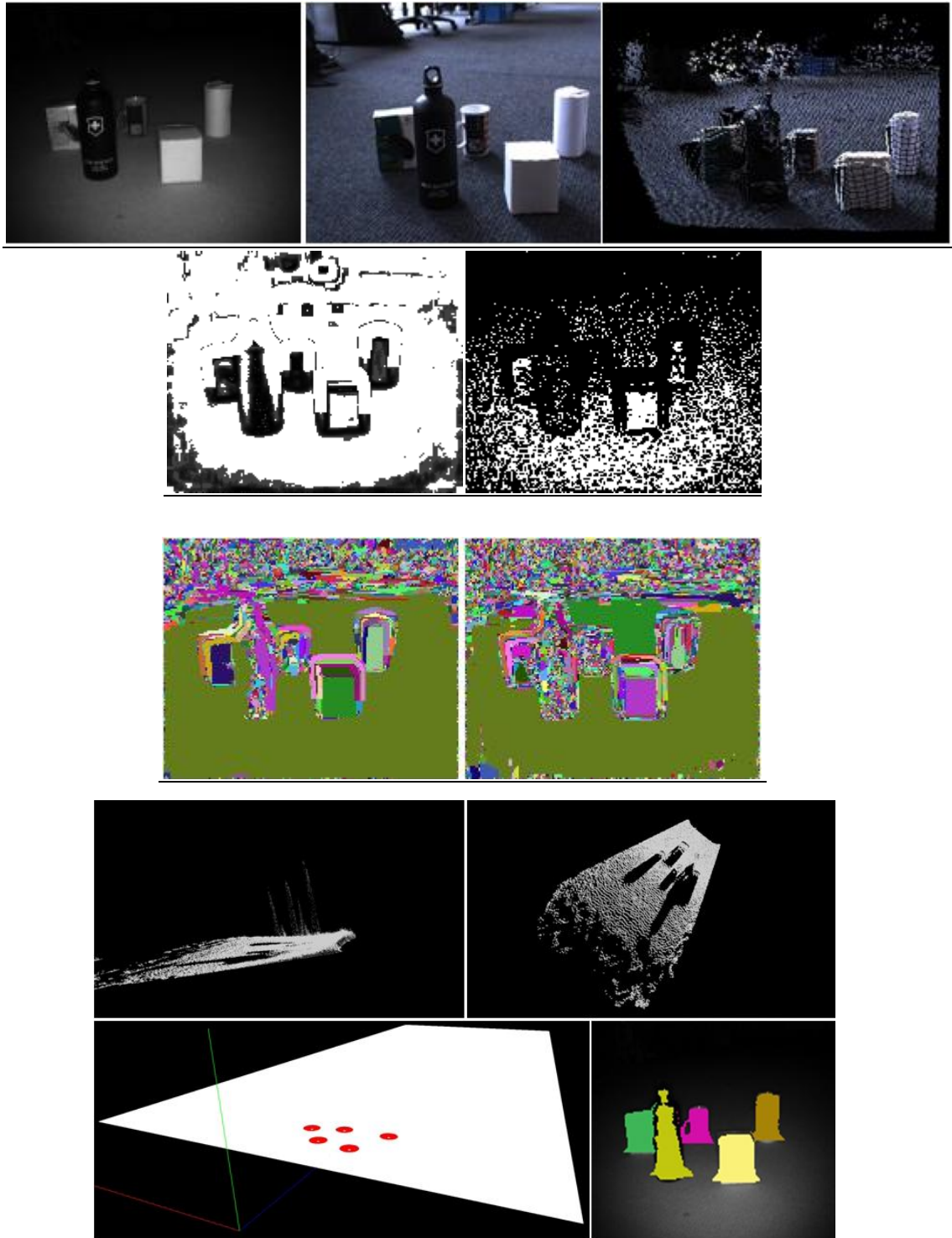
Figure 3-30. *First Row*: TOF camera intensity image, color image, calibrated point cloud. *Second Row:* PCA based curvature values (darker regions indicate high curvature), HK curvature (white regions indicate planar regions), *Third Row*: region growing with surface normal, region growing with surface normal and PCA based curvature. *Fourth Row:* Parzen window based segmentation results two different views of generated saliency map via projection and KDE. *Fifth Row*: Segmented object locations and segmentation results based on segmented object locations

K-Means

Mean- Shift

Region Growing
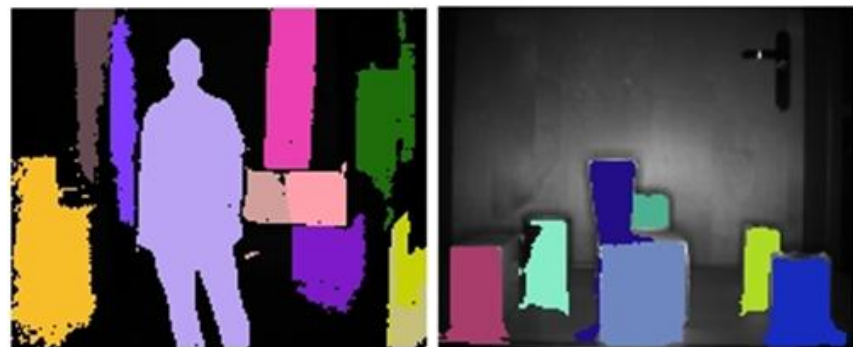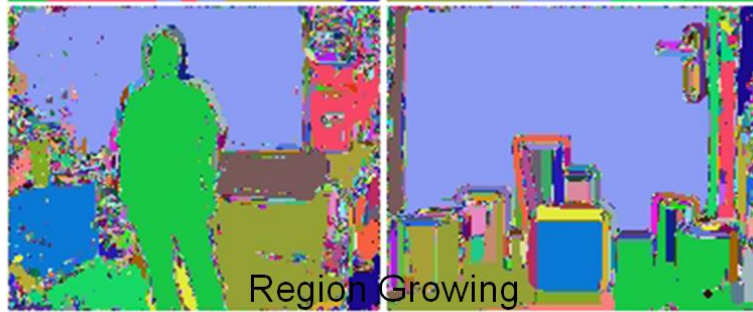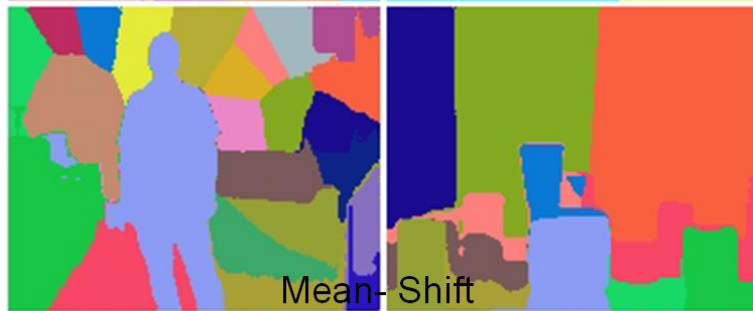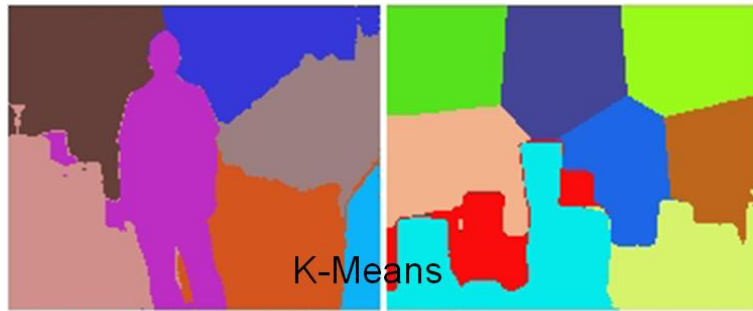
Figure 3-31. Segmentation results of the exemplar ranges images; top images are the corresponding intensity images, last row shows the segmentation result of the proposed algorithm.
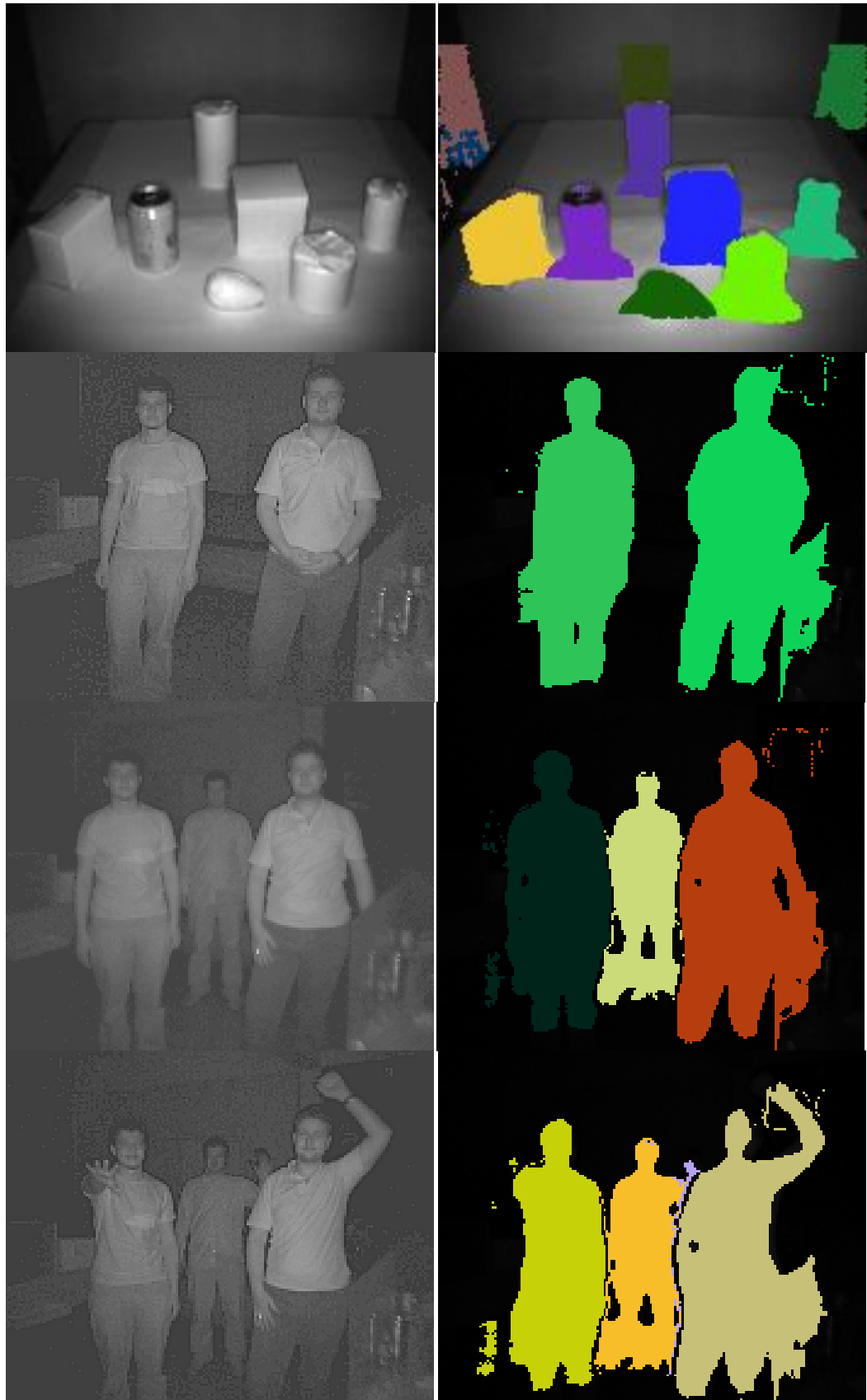
Figure 3-32. Segmentation results of the proposed method, intensity images are given in the right.

CHAPTER 4

CORRESPONDENCE MATCHING IN RANGE DATA

Correspondence matching can be defined as detecting similar structures in different views of a surface. Correspondence matching is the first stage in any registration attempt and can also be utilized in verification type of problems. This chapter presents a local range image matching method that combines 3D surface properties with the 2D scale invariant feature transform. First, proposed local surface properties are explained. Next, feature extraction methodology is stated, and finally matching results and comparison to the previous studies are presented.

4.1 Introduction

With the increase in the performance of the active range data generators, such as time-of-flight cameras that are assisted with multi-view dense depth estimation algorithms, as well as the decrease in the cost of off-the-shelf hardware for 3D sensing, computing, and displaying, 3D commercial systems are becoming commercially available. This kind of data could be applied to a number of different applications, such as 3D television (3DTV), virtual reality, etc. In case of widespread usage of these systems, the capability of searching any content in such range data becomes more critical. Hence, description, indexing and search techniques should be utilized in archive systems for 3D video, which is expected to be consisting of video and range data (or dense depth), considering upcoming ISO MPEG standards.

The problem of 3D shape matching and retrieval is studied extensively in the past [8, 119]. There are two main approaches for shape identification: *global description* and *local description*

[8]. First type of approach [68, 120-122] is utilized in retrieving geometrically similar shapes and requires complete models. In the second type, descriptors are obtained and evaluated on some local keypoints and utilized for complete or partial matching. However, there are also some "partial" matching methods that also require global models [75, 78]. For local descriptions, *spin images* [58], *splashes and 3D curves* [71], *point signatures* [72], *local feature histograms* [29], and *regional point descriptors* [73] can be listed as leading efforts for local surface description.

Unfortunately, the majority of the aforementioned methods consider objects to be isolated and pre-segmented. However, in many applications (3DTV archives, robotics, laser scanning), the objects appear as a part of the scene and only some portions are visible. These scenes might also contain clutter, occlusion or noise. Thus, the problem of matching objects in range images is more challenging compared to the conventional 3D shape matching problem.

4.2 Global vs. Local Descriptions

Firstly, it should be noted that global description of range object is only possible after segmentation from the scene, but segmentation is a very challenging task; besides, object structure is still partially available even after segmentation. Thus, global methods are likely to fail for the applications based on 3D video.

Local description could be more promising compared global description for such problems in many aspects: segmentation is avoided, complete models are not necessary, occlusion and self occlusion can be handled more easily. Moreover, in 2D object detection and matching, local feature extraction methods [123] are shown to perform notably better.

There is also an important difference between these two approaches; while the global methods ignore shape details and gather the main properties of objects, the local descriptors put more emphasis on the details. As a result, the global methods perform better in similarity search, while local methods are more suited for matching, verification and identification. The drawbacks of local descriptors in similarity retrieval of range images are presented in Chapter 5.

4.3 Local Descriptors for Range Data

In range image recognition, surface properties, such as *normal* or *curvatures*, are widely used [29, 71, 76]. The early work splashes and 3D curves [71] makes use of distribution of normal vectors in the neighborhood of a keypoint. Although splash technique is a promising approach, the representation is complex. Similarly, *point signatures* [72] obtain a contour by intersecting the surface with a sphere centered at a keypoint. Then, a plane is fitted to these contour points and translated to this keypoint. After this fitting, a 1D signature, which is the signed distances of contour points to the fitted plane, is utilized as the descriptor. However, both of these methods do not address the scale and rotation invariance problem.

Another well-known technique, namely *spin images* [58], are widely used as local shape description method and applicable for scenes containing occlusions and clutter. It is a 2D histograming around a keypoint in which spatial configuration is represented by a cylindrical coordinate system. However, it has the following disadvantages: it is not scale invariant, it requires high storage and results multiple ambiguous correspondences, since spin images of near points are similar; i.e., spin images are not unique [59]. Yet another similar histograming technique is *3D shape contexts* [73]. In that approach, spatial distributions of the surface points are accumulated utilizing their spherical coordinates. The same arguments for the spin images also hold for this description as well.

4.3.1 Shape Index as a Local Descriptor

There are two key features of the 3D surfaces: *normal vector* and *curvature*. A 3D point can be described by its minimum and maximum curvatures (*principal curvatures*) or some functions of these principal curvatures (κ_1, κ_2) at the point of interest. One of these measures is the *Shape Index* (SI) which is introduced by Koenderink [67]. Firstly, it is used to obtain global description, named as *shape spectrum* [95]. Later, as mentioned previously, it is used both for local description and keypoint selection [76, 88].

The curvature values on 3D surfaces can be obtained robustly after fitting a quadric surface to a local patch as explained in Chapter 3. Then the shape index is calculated using principal curvatures as follows:

$$SI = \frac{1}{2} - \left(\frac{1}{\pi}\right) arctan \left(\frac{\kappa_1 + \kappa_2}{\kappa_1 - \kappa_2}\right) \qquad 4.1$$

There are two recent studies that use *SI* within the context of local description. In the first study [76], the authors identify keypoints, where the shape index values are extremum. Around these keypoints, 2D histograming is formed, in which one of the dimensions belongs to the SI values of the neighboring pixels, whereas the other one denotes the angles between the normal of the keypoint and the normal of the neighboring pixels (SI-Normal-Hist). In their work, scale invariance problem is not addressed. In a different recent effort, namely *2.5D Scale Invariant Feature Transform* (2.5D SIFT) [88], keypoint detection is achieved on 2D range images similar to the Lowe's work [123]. However, an extra preprocessing step is required. After this step, using the histogram of the shape index values and the range gradient orientations around these keypoints, a description is obtained. Both of the methods use isolated objects in which problematic boundary regions can be eliminated by simple thresholding; however, it is not applicable in range scenes due to occlusion. Besides, the matching performance in complicated scene with the presence of other objects, clutter and occlusion are not evaluated in these studies.

4.3.2 Proposed Local Surface Descriptor

We propose a local range image matching method which combines 3D surface properties with the 2D scale invariant feature transform. Figure 4-1 shows the basic steps in our proposal.

The main difference of the proposed approach against related methods is conversion of range data into SI representation before exploitation of SIFT during local description. Range data does not have sufficient blob-like properties for SIFT keypoints.
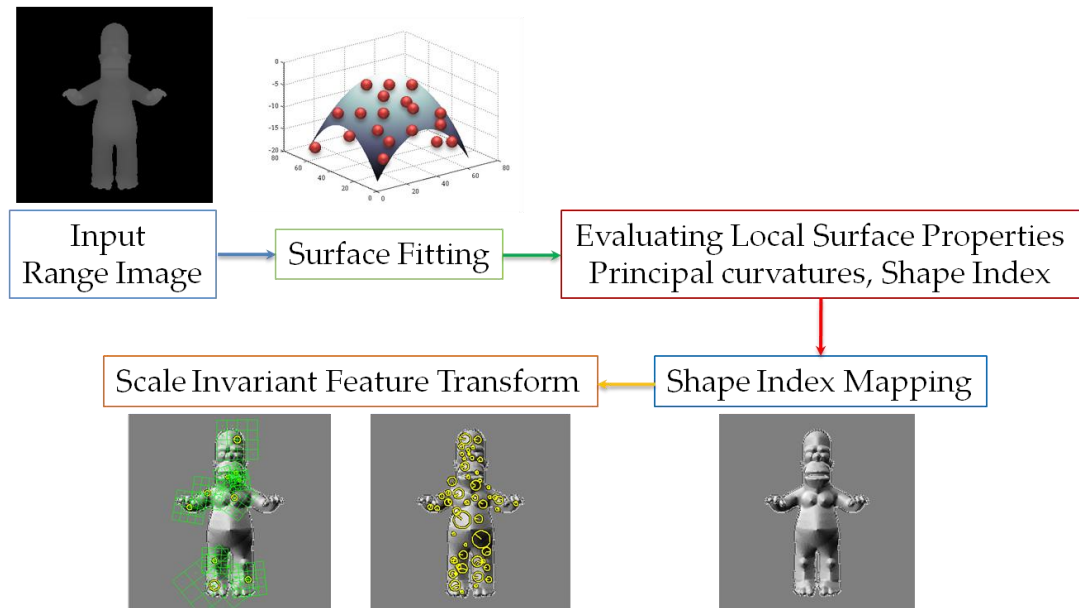
Figure 4-1. Basic steps in the proposed correspondence matching method

The most appealing property of SI is being scale, translation, and rotation invariant. However, in range imaging, the scale invariant property holds if and only if the ratio of the scale to the resolution is constant. Moreover, SI makes strong emphasis on points, where surfaces deviate from being smooth, even for small changes. Thus, representing the range image with its SI values which are mapped to 0-255 strengthen the features. Shape details become clearly visible. Therefore, utilization of SIFT-type features becomes more feasible.

Figure 4-2 shows exemplar range images and their corresponding SI images. In Figure 4-5 and Figure 4-6, SIFT matching results on range images without SI mapping are shown. There is a 15 ° rotation among one of the axis between the two "Face" models in those figures. SIFT parameters in [124] are utilized during matching shown in Figure 4-5 . In Figure 4-6, local extremum threshold of the DOG scale space is reduced, number of octaves and number of levels are increased, and local extrema localization threshold is decreased. This is utilized for detecting more keypoints, that are localized on the surface, for increasing the matching performance. Keypoints are mainly localized at the boundaries when SIFT is applied directly on the range images. Moreover, a few keypoints, which are on the surface, match with false keypoints on the target images. This observation is due to sharpness of the boundaries. If the edge threshold is further decreased, either no keypoints are detected or a few keypoints are detected that match incorrectly. To reduce the effect of the sharp edges present on the

boundaries, the intensity of the background is modified such that background has a similar intensity with the intensity of the object. Previous SIFT tests are performed based on this modification. The results are shown in Figure 4-4 and in Figure 4-7. This modification further decreases the number of detected keypoints. Only a few keypoints are detected on the object surface.

Figure 4-5 and Figure 4-8 shows SIFT matching results based on shape index mapping. The number of the keypoints and the number of true positives are much higher than the previous attempts. Moreover, keypoints are localized across the surface.

4.3.3 Keypoint Selection and Feature Extraction

SIFT is an invariant 2D local descriptor which is robust to some specific transformations [123]. SIFT has three main steps: scale space construction, keypoint detection, and feature extraction. Scale space is constructed by taking the difference of the Gaussian (DoG) blurred images at different scales [123]. Next, keypoints are detected as the local extremum of the DoG images across scales. A gradient orientation histogram is computed in the neighborhood of the keypoint as the feature vector [123]. In this study, SIFT keypoints and the features are computed by using the publicly available software[1].
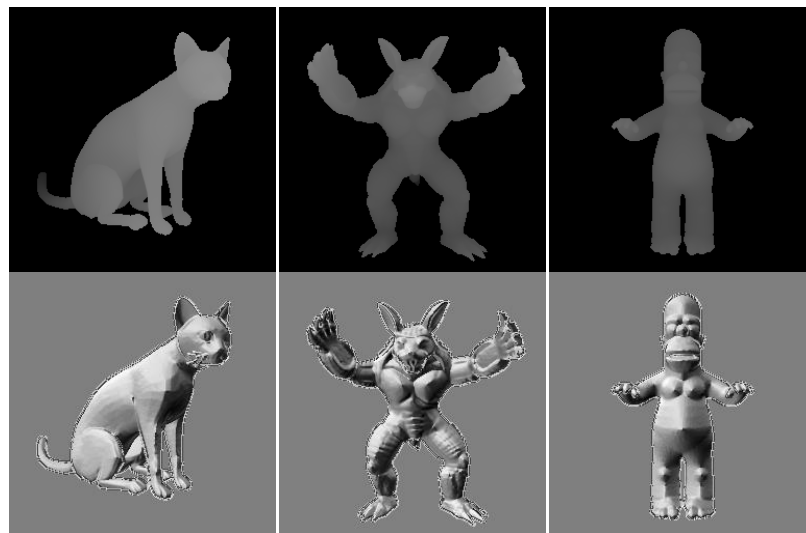
Figure 4-2. (Top Row) Exemplar range images, (Bottom Row) Shape index mapping

---

[1] http://www.vlfeat.org/~vedaldi/
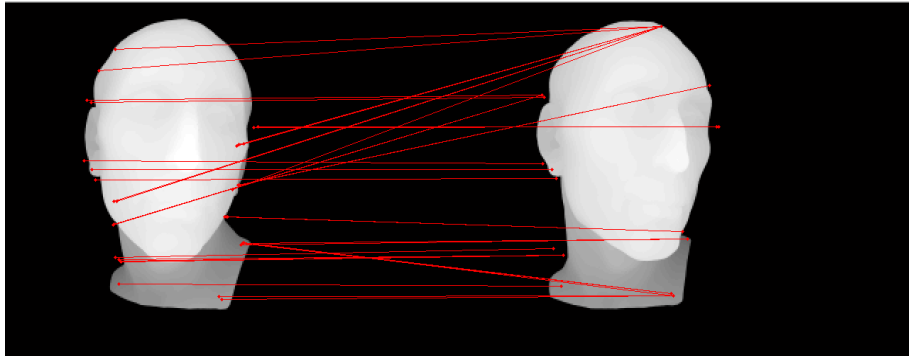
Figure 4-3. SIFT matching result with Lowe's parameters
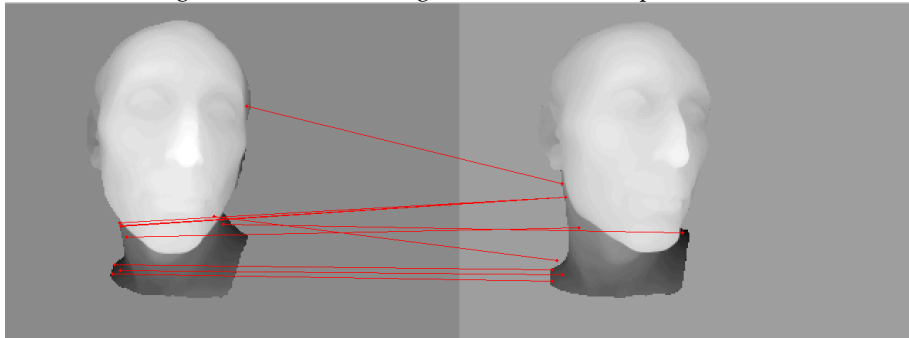
Figure 4-4. SIFT matching result with Lowe's parameters with reduced boundary effect
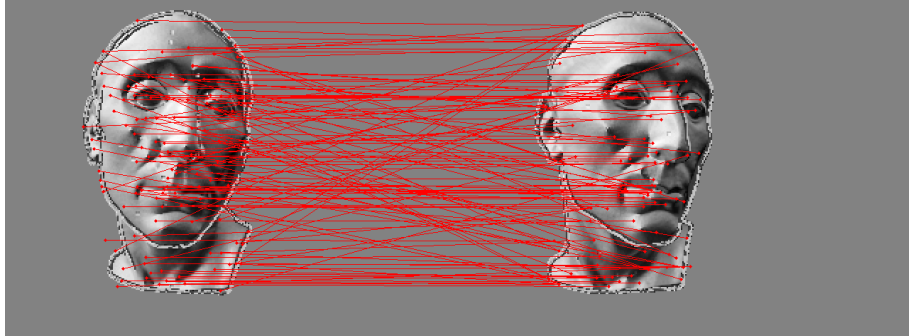
Figure 4-5. SIFT matching result with Lowe's parameters with proposed shape index mapping
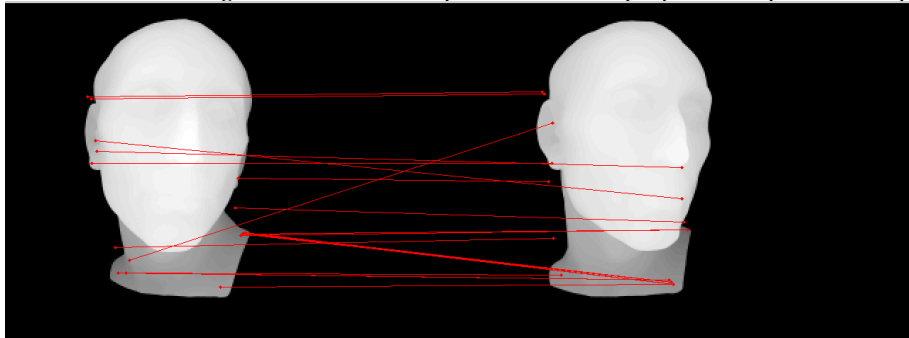
Figure 4-6. SIFT matching result with reduced local extremum threshold of the DOG scale space, increased number of octaves and increased number of levels, and decreased local extrema localization threshold compared to the original Lowe's parameters
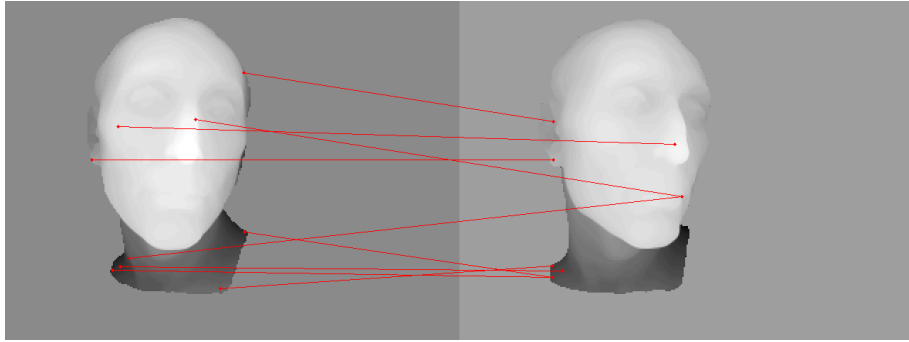
Figure 4-7. SIFT matching result with reduced boundary effect . Local extremum threshold of the DOG scale space is reduced, number of octaves and number of levels are increased, and local extrema localization threshold is decreased compared to the original Lowe's parameters
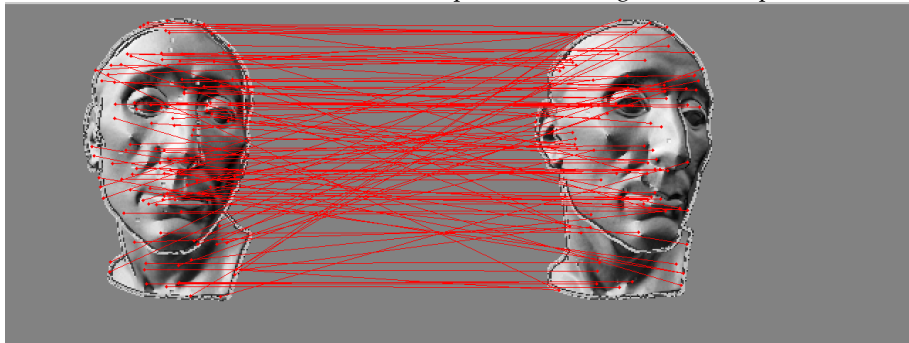
Figure 4-8. SIFT matching result with proposed shape index mapping . Local extremum threshold of the DOG scale space is reduced, number of octaves and number of levels are increased, and local extrema localization threshold is decreased compared to the original Lowe's parameters

4.4   Experimental Results

Experiments are performed on range images from the Ohio State University range database[2] and range images formed by combining models collected from AIM@SHAPE repository[3] and ISDB database.

---

[2] http://sampl.ece.ohiostate.edu/data/3DDB/RID/minolta/

[3] http://shapes.aim-at-shape.net/
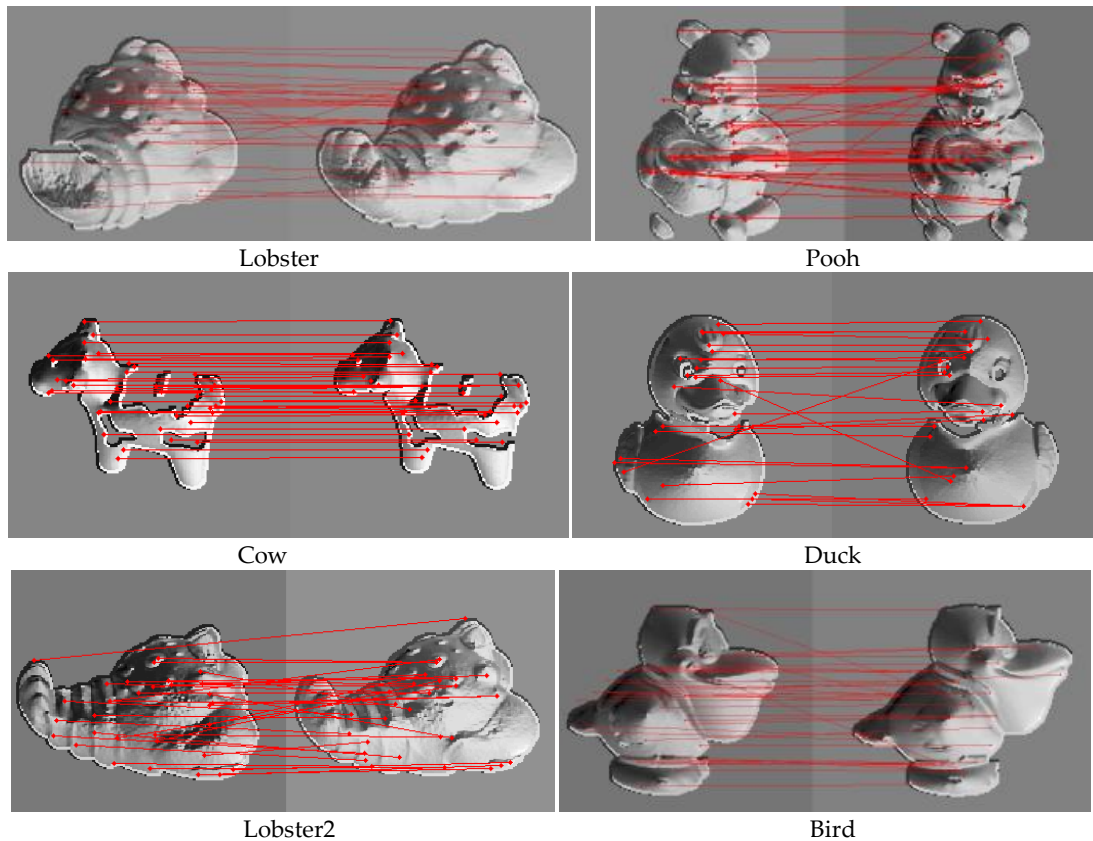
Figure 4-9. Matched pairs obtained by the proposed method are shown on the SI images.

Table 4 shows the comparison of the proposed method against 2.5D SIFT [88] and SI-Normal-Hist. [76]. The experiments are performed on the same data (Figure 4-9). Based on the experimental results, the number of matches and the number of correct matches are obtained more than three times of the number of matches that are obtained by [88]. Although limited comparison against LSP is achieved due to the data availability, the proposed method still shows a better performance. The reason for this superior performance could be due the fact that SIFT keypoint selection and feature extraction on SI images is more effective than selecting the keypoints as SI extremum and describing features as SI histogram around the keypoints. The former method is also more effective than applying SIFT directly to the range image. This result can also be explained as follows: SI comprises the understanding of the neighbor geometry whereas range image pixel only indicates its depth. Besides informative regions are visually strengthen by the nonlinearity behavior of the SI function. Additionally, SIFT features are considerably discriminative [123].

Table 4. Local descriptor matching results for proposed method, SI-Normal-Hist., and 2.5D SIFT

	Total #	Correct	False
Lobster			
Proposed	33	19	14
SI-Normal-	30	13	17
Pooh (20°)			
Proposed	37	27	10
2.5DSIFT	9	8	1
Cow (15°)			
Proposed	37	37	0
2.5DSIFT	10	10	0
Duck (20°)			
Proposed	25	19	6
2.5DSIFT	6	4	2
Lobster2			
Proposed	32	20	12
2.5DSIFT	8	8	0
Bird(20°)			
Proposed	37	30	7
2.5DSIFT	9	8	1

In order to demonstrate the versatility of the proposed method in the cluttered scenes, some tests are also conducted (Figure 4-10). There are two other faces synthetically included into the first scene for obtaining more complicated scenarios and the rotated query object takes part in the target image; the mismatch is only 3 out of 20 matches, whereas the correct match ratio is higher in the second example. Although scaled and occluded query objects take part in the last two examples, matching results are still considerably high.

4.5   Conclusions

The proposed local surface description method does not require any initial segmentation step; it can also handle affine transformations up to a scale. The experimental results indicate that the proposed approach improve the performance of two recent methods from the literature. Moreover, clutter and occlusion do not affect the efficiency of the proposed method, significantly.
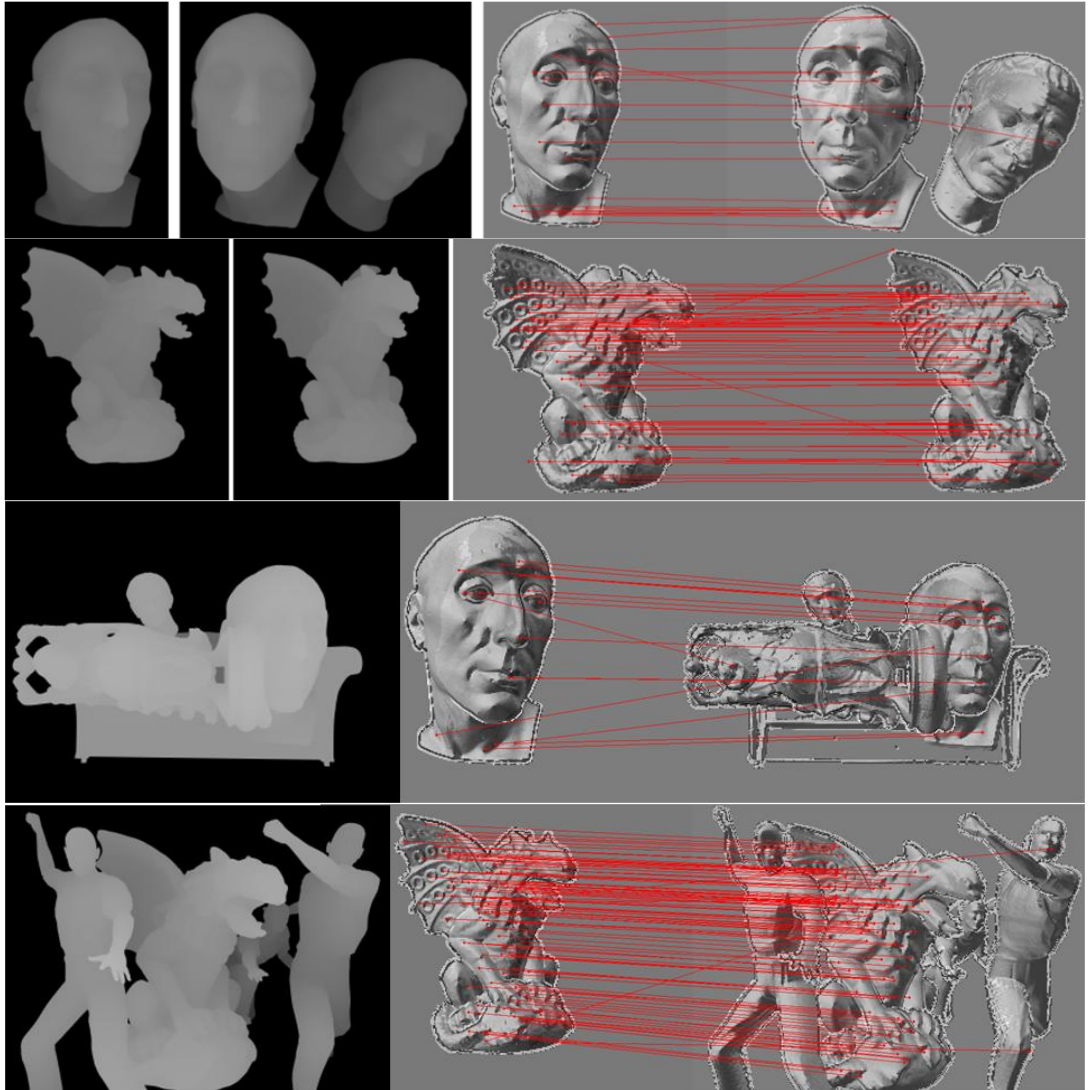
Figure 4-10 First two rows: (left to right) Query range images, target range images, combined shape index images with matched pairs. Rotated query objects are included in the target images. Last two rows: (left to right) Target range images, combined shape index images with matched pairs. Scaled and occluded query objects are included in the target images.

CHAPTER 5

LOSSLESS DESCRIPTION OF RANGE MODELS

The improvements in 3D scanning technologies have led the necessity for managing range image databases. Hence, the requirement of describing and indexing this type of data arises. Since a range model has different properties compared to complete 3D models, we propose a novel method that relies on Spherical Harmonics Transform (SHT) for retrieving similar models where the query and the database both consist of only range models. Although SHT, is not a novel concept among shape retrieval research for 3D complete models, we utilize SHT for 2.5D range images by representing the models in a reciprocal world observed from the camera. The difference, as well as the advantage of our algorithm, is being information lossless. In other words, the available shape information is completely exploited while obtaining the descriptor. On the other hand, some other mesh retrieval applications utilizing SHT [121] "approximates" the shape that yields some information loss. The proposed descriptor is also invariant to rotations about z-axis. The proposed technique is tested on a large database having high diversity and its performance of the proposed method is superior to the performance of popular D2 distribution.

5.1   Introduction

3D object description and retrieval have become popular research topics during the last decade. The field is attracting more and more people every day due to increasing availability of 3D models (with the use of highly developed scanning technologies and computer graphics software), increased processing power, increased storage capabilities, and the progress in visualization technologies, as well as the consumer penetration of 3DTV. These

improvements facilitate obtaining and managing large 3D model databases which arises the need of describing and indexing these models and similarity retrieval systems.

3D object description is treated in computer vision research and also in sole shape analysis discipline with some differences. In shape analysis research, similarity retrieval and part-based matching studies among watertight mesh models are popular with its contests (i.e. SHREC). Computer vision side is mainly interested in verification, self-matching, registration, and point correspondences; moreover, their inputs are usually obtained from scanners. Considering both modalities, the methods used in 3D description can be classified into *feature based*, *structural-topological based*, and *view based* approaches. Feature based methods can further be classified into *local description* and *global description*. Global methods usually preferred in similarity retrieval whereas local ones are popular in partial matching and point correspondences. Among the global description studies, *cord and angle histograms* [52], 3D *Zernike moments* [125], *shape histograms* [30], *spherical harmonics* [120, 121, 126, 127], *shape distributions* [23], and *diffusion distances* [22] can be listed. On the other hand, *shape spectrum* [95], *splashes and 3D curves* [71], *point signatures* [72], *spin images* [58], *local feature histograms* [29], *multi-scale features* [81, 85], *auto diffusion function/ heat kernel signatures* (*HKS*) [128, 129] are some notable local descriptors. Most of these local shape descriptors can be extended to contain more global information by adjusting the size of the local region that is being described. *Reeb graphs* [65, 130], *skeletons* [131], *curve-skeletons* [132] are *structural-topological* based approaches which are efficient in articulated shape description. In view based methods, 3D objects are represented by several 2D images (depth buffers or silhouettes) obtained from various viewing angles. *Lightfield descriptor* [61], *compact multi-view descriptor (CMVD)* [133], *bag-of-features SIFT (BF-SIFT)* [134], and *panoramic views* [62] are view based studies. The literature certainly contains many other studies and we refer the reader to Tangelder and Veltkamp [8] and Bustos *et al.* [135] for detailed surveys and Bronstein *et al* [136] for a more recent synopsis.

The aforementioned approaches usually assume that complete geometric information of the object is available. This information contains explicit neighboring relations between points, surface patches or polygons, as in the case of mesh representations. Considering the growing and improving scanning technologies, databases consisting of range images (scanned data),

which are different from 3D mesh models in many ways, will arise. Our proposed technique address this problem by a global descriptor that relies on (SHT).

5.2 Related Work

To the best of our knowledge, there is no reported range image retrieval system, in which the query and the database consisting of *only* range images. The motivation for obtaining a range image retrieval system could be due to the some new paradigms, such as 3DTV archive systems or LIDAR databases. There are some similarities between range image similarity retrieval and partial matching or query by range image research; however, the differences are quite crucial. In partial matching research [74, 75], query is a part of a 3D model where the part is usually identified with topologically valid mesh, as well as the database consisting of 3D complete models. In this case, local descriptors are evaluated and a matching score is used to obtain a similarity degree between the query and the database models. Latter type of studies query range images [133, 134, 137] among a database consisting of complete 3D models. In this case, database models are viewed from several viewing angles to get a similar viewing with the query. Then, they search the best match among the views for indexing. The only difference between these types of study from *view based* approaches is that single view of a query is used instead of multiple views. The descriptors are again obtained from 2D images.

If the database images and the query are both range images, then partial matching and view based approaches become deficient. Although range images contain 3D information, they are different from the actual 3D model representation. These differences are due to *i) self-occlusion* (Figure 5-1, Figure 5-2), *ii) transformations* and *iii) view dependent partial geometry*. Self-occluded regions which are formed around the salient regions violate neighboring relations. This effect can be observed in when a hand model is rendered with a scanner type of camera (see Figure 5-1); the informative details, such as fingers, due to self-occlusion are lost and the convex-hull of the shape is left. Local descriptors that should be selected to capture highly discriminative regions contain distant region features, if they are formed around these salient regions. This descriptor eventually resembles to a representation of another local geometric structure (consider the side view of a tip of a nose). If the salient regions are avoided, then the local descriptors lose their discriminative power and represent

only smooth regions which will be ineffective for describing shapes. Occlusion awareness [138] can overcome these shortcomings up to a level, since occlusion detection is not available in many cases and cannot be obtained easily from a single range image. Translation, rotation and scale changes introduce self occlusion and information loss. Fine details disappear as the objects become distant from the camera. Independent of the previous discussion local descriptors in 3D shape analysis are considered to be less discriminative and far from being robust. Since 3D shapes have insufficient features and keypoint repeatability is not satisfied [136].
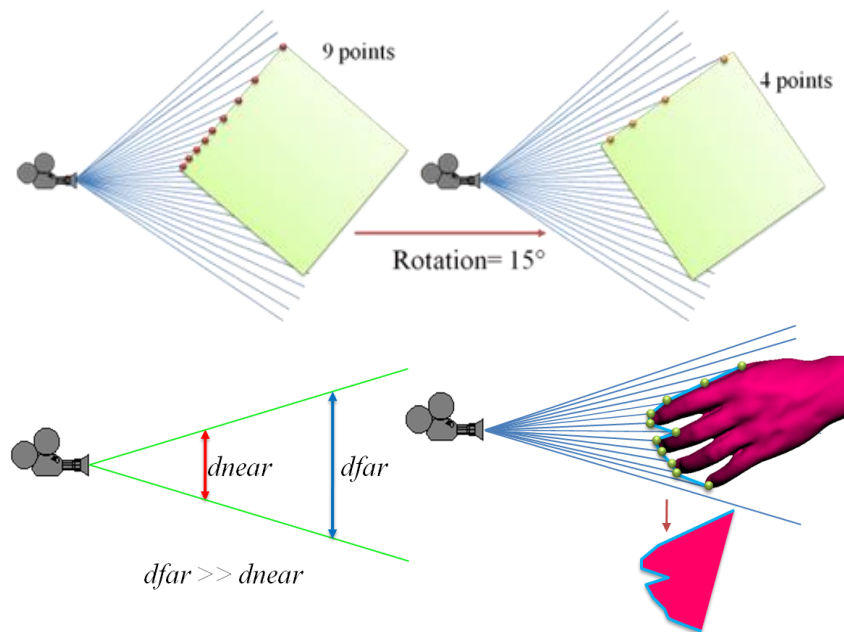
Figure 5-1 Effects of transformations; rotation *(top)*, translation *(bottom-left)*; in range image acquisition. *(Bottom-right)* Information on local structures is highly influenced by position.
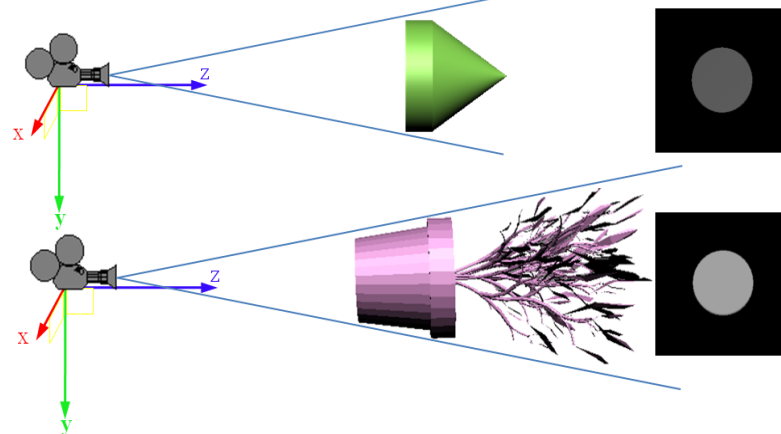
Figure 5-2 Self-occlusion example, *(left)* orientation of objects w.r.t. camera coordinate system, *(right)* range images of two objects in different classes have rather similar views.

We propose a method for similarity retrieval of range images that relies on a global description method *Spherical Harmonics Transform* (SHT). We use spherical harmonics with the view-dependent structure of the range images.

5.3 Spherical Harmonics Transform

The square integrable complex functions defined on two-sphere S^2 form a Hilbert-space where the inner product of two functions $f(\theta, \phi)$ and $g(\theta, \phi)$ in this space is defined as follows:

$$\langle f, g \rangle = \int_0^\pi \int_0^{2\pi} f(\theta, \phi)\overline{g(\theta, \phi)} sin\theta d\phi d\theta \tag{5.1}$$

The Spherical Harmonics Y_l^m of degree l and order m ($|m| \leq l$) form an orthonormal basis in this space. In Figure 5-3, visual representation of spherical harmonics $Real\{Y_l^m\}^2$ is shown up to degree 3. They are related with the associated Legendre polynomials P_l^m as follows:

$$Y_l^m(\theta, \phi) = \underbrace{\sqrt{\frac{(2l+1)(l-m)!}{4\pi\,(l+m)!}}}_{K_{l,m}} P_l^m(cos\theta)e^{im\phi} \tag{5.2}$$

$$P_l^m(x) = \frac{-1^m}{2^l\,l!}(1-x^2)^{\frac{m}{2}}\frac{d^{(l+m)}}{dx^{(l+m)}}(x^2-1)^l \tag{5.3}$$

Consequently any function $f(\theta, \phi)$ defined in this space can be written as a combination of these basis functions as follows:

$$f(\theta, \phi) = \sum_{l=0}^{\infty}\sum_{|m|\leq l} \hat{f}_l^m\, Y_l^m(\theta, \phi) \tag{5.4}$$

where expansion coefficients $\hat{f}_l^m$ are projections of the function $f(\theta, \phi)$ on the basis functions. They can be obtained utilizing the inner product (Equation 5.1) defined in this space as follows:

$$\hat{f}_l^m = \langle f, Y_l^m \rangle = \int_0^\pi \int_0^{2\pi} f(\theta, \phi)K_{lm}P_l^m(cos\theta)e^{-im\phi}\, sin\theta d\phi d\theta \tag{5.5}$$

If the function $f(\theta, \phi)$ is bandlimited with B then it can be written as a finite weighted summation of the basis functions (Discrete Spherical Harmonics Transform, *DSHT*). For a

function $f(\theta, \phi)$ sampled in an equiangular grid ($2B$ x $2B$) with a sum of $4B^2$ points, expansion coefficients $\hat{f}_l^m$ are obtained follows [139]:

$$\hat{f}_l^m = \frac{\sqrt{2\pi}}{2B} \sum_{j=0}^{2B-1} \sum_{k=0}^{2B-1} w_j f(\theta_j, \phi_k) P_l^m(cos\theta) e^{-im\phi} \qquad 5.6$$

The coefficients $\hat{f}_l^m$ is equal to zero for $l \geq B$ for functions bandlimited with B. Consequently, the number of non-zero coefficients is B^2. The original function can be recovered from these coefficients, when the inverse Spherical Harmonics Transform is applied. If the function is not bandlimited, then the recovered function using the expansion coefficients obtained from *DSHT* is an approximation of the original function. As B increases, the error between the approximate function and the original one decreases.
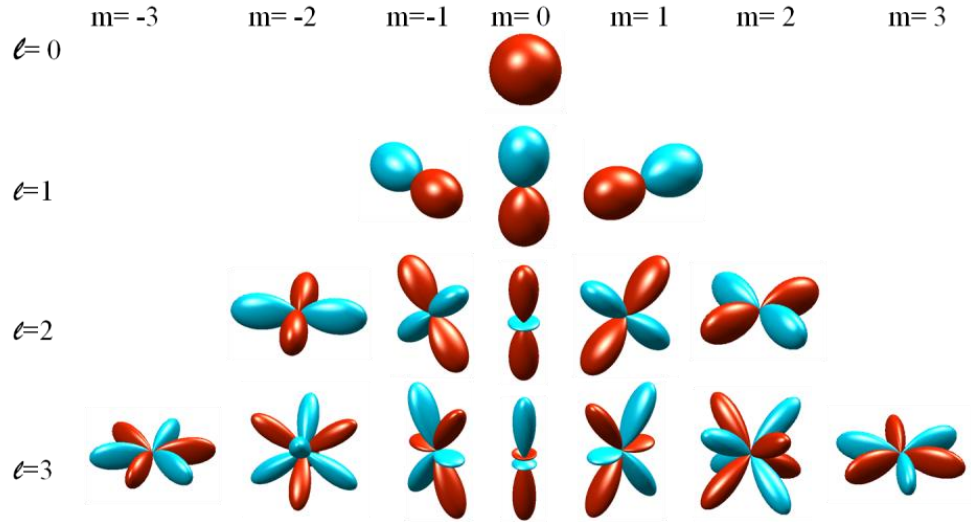
Figure 5-3. Visual representation of spherical harmonics up to degree 3. $Real\{Y_l^m\}^2$ is plotted, positive and negative portions are colored with red and blue respectively.

5.4 Spherical Harmonics in Shape Analysis

Vranic *et al.* proposed to use Spherical Harmonics Transform in 3D model retrieval [126]. They describe the shape as a spherical function, $f(\theta, \phi)$ where the origin is selected as center of the mass of the model. The value of the function $f(\theta, \phi)$ is the length of the ray that is emanating from the origin and ending at the outermost intersection of the 3D model. They perform *DSHT* on this functional representation. Magnitude of the expansion coefficients are utilized as a feature vector. These descriptors are compared by L_1 norm. Their method has two disadvantages: firstly, pose normalization, for this Vranic propose a modified Principal

Component Analysis (PCA) method, should be performed; secondly, this functional representation ignores interior structure of shapes. Later, Funkhouser *et al*. propose to decompose a 3D model into a collection of functions defined on concentric spheres to use spherical harmonics [121]. This representation preserves interior structure of shapes up to a level. Initially, they obtain the binary voxel grid of a model. Then, by restricting to the different radii, they obtain a collection of binary spherical functions. Their approach does not require pose normalization, since their descriptor is rotation invariant. This is achieved by a property of Spherical Harmonics Transformation. The amount of the energies contained at different frequencies does not change when the function is rotated. That is:

$$\sqrt{\sum_{m=-l}^{m=l} |\hat{f}_l^m|^2} = \sqrt{\sum_{m=-l}^{m=l} |\hat{f}_{l,ROTATED}^m|^2} \qquad 5.7$$

Their feature vector for each spherical function is formed by collecting these scalars for each frequency (l) and the overall shape descriptor is obtained by concatenating these feature vectors. L_2 norm is used to compare two descriptors. Vranic [127] argue that many fine details are lost in the binary voxel grid representation and propose a ray-casting method that finds all points of intersection. Therefore, Vranic uses concentric spheres with ray based descriptor with normalization step. It is argued [127] that this method outperforms rotation invariant spherical harmonics descriptor based on binary voxel grid [121]. Moreover, Kazhdan et al. [120] used spherical harmonics as a general tool to transform rotation dependent shape descriptors into rotation independent ones. Apart from the aforementioned approaches, spherical harmonics transform is used in many other shape analysis studies.

5.5   Proposed Lossless Description Technique

Obviously, SHT can describe functions defined on two-sphere. Since many 3D shapes are not star shaped, i.e. spherical representations are not single valued, in literature concentric spheres are proposed to define shapes on spheres. In that case, information loss depending on the radius discretization is inevitable. However, a range image can be represented with a spherical function. Besides, all available information is preserved with this representation. Instead of describing the shape, we describe the world captured from the camera. The main

83

steps for computing our spherical harmonics descriptor for range models is shown in Figure 5-4:

 i. First, background is removed,

 ii. The origin of the coordinate frame is assigned as the center of the camera (Figure 5-4),

 iii. Spherical coordinates is used and the 3D space is discretized according to the resolution of the input image and $f(\theta, \phi)$ is initialized with zeros,

 iv. Each point coordinate on range model is expressed with respect to the camera frame; associated θ, φ and the length of the ray connecting the point with the origin is calculated. The ray length is assigned as the value of $f(\theta, \phi)$,

 v. The maximum ray length is determined and the function is normalized such that the function has the maximum value of one,

 vi. For non-object parts of $f(\theta, \phi)$ the value one is assigned (Figure 5-5),

 vii. With the use of spherical harmonics transform, the function is expressed as a finite weighted summation of the basis functions,

 viii. Utilizing the rotation invariant property of the spherical harmonics, the amplitude of the coefficients within each frequency band (l) is computed:

$$\|f_l\| = \sqrt{\sum_{m=-l}^{m=l} |\hat{f}_l^m|^2} \qquad\qquad 5.8$$

The feature vector is formed (signature of the range model) by concatenating these amplitudes. The zero-order component is omitted:

$$f = (\|f_{l=1}\|, \|f_{l=2}\|, \dots, \|f_{l=i}\|, \dots, \|f_{l=N}\|) \qquad\qquad 5.9$$

The Euclidean distance is used to compare two signatures. By assigning value one to non-object parts; the same information is included to all range models which is a neutral element, besides in comparison step the zero-order coefficient is ignored. Zero-order coefficient is related with the sphere shaped basis shown in Figure 5-3. Rotations around the z-axis become invariant with this representation. The global characteristic of range models captured by the camera is described.
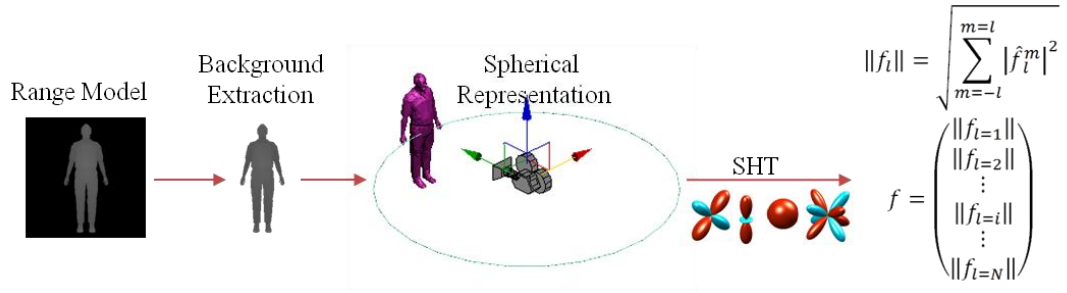
Figure 5-4. Main steps for computing our spherical harmonics descriptor for range models.
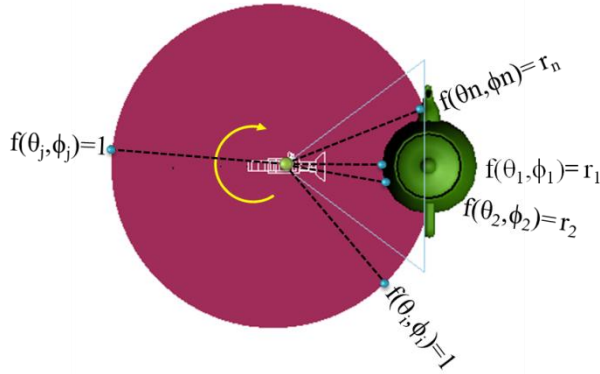
Figure 5-5 : *(Top view)* Description of the world with respect to the camera. Spherical function $f(\theta, \phi)$ is normalized such that maximum extend is equal to one.

5.6 Comparison with Related Techniques

According to our knowledge, this thesis is the first attempt to test range model retrieval among the range model database. Previously addressed 3D similarity retrieval approaches usually assume that the complete geometric information of an object is available. This information contains explicit neighboring relations between points, surface patches or polygons, as in the case of mesh representations. The methods, which are suitable in utilizing it for similarity retrieval of point clouds, can only be tested for range data. Besides the SHT on ray based concentric spheres (*classical SHT*) proposed by Funkhouser *et al*. [121] and Vranic [127] , *D2* distribution [23] is an appropriate shape descriptor among the mesh descriptors. It is proposed by Osada et al. [23], and corresponds to the distribution of the distances (Euclidean) between object points that are selected randomly (Figure 5-6, middle). This method can be applicable to every 3D geometric representation and it is invariant to transformations. Besides, in [23], it is shown that *D2* shape distribution is robust to noise,

small cracks and holes. Osada et al. argue that robustness is satisfied with the random selection strategy.

We tested *D2* shape distribution by our database; besides, we also tested our modified version, namely *Salient-D2* that has a modification in the point selection strategy. Instead of random point selection, we impose saliency constraint in *Salient D2*. Points which are informatively salient are selected and the *D2* distribution is evaluated among them. On any surface, salient points are the ones having high or low curvature values compared to their local neighborhood (Figure 5-6, right). Surface curvature at point *p* is computed using the formula [140]:

$$c = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} \ where\ \lambda_0 \leq \lambda_1 \leq \lambda_2 \qquad 5.10$$

by comparing the three eigenvalues obtained from Principle Component Analysis (PCA) of the local neighborhood of the surface point *p*.
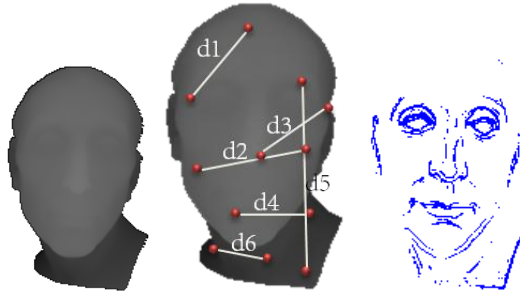
Figure 5-6 *(Left to right)* Range model, sample D2 distribution evalution with six distances, salient points (curvature values are above a threshold value) of the range model

5.7 Experimental Results

For testing the proposed description method (*Lossless SHT*), we build our own database. It contains 544 range models divided into 18 classes. Representative models for each class are shown in Figure 5-7. Complete database is presented in the Appendix. We collect 3D models from Princeton Shape Benchmark, AIM@SHAPE repository, NTU shape database, and Konstanz University database. Then, by use of a COTS computer graphics software, we obtain range images of these models, as if they are acquired by a scanner. The main reason for generating our own database is to ensure the diversity. Despite the publicly available

range databases, they are not suitable for testing similarity retrieval methods due to their sizes (within class and overall) and their diversities.
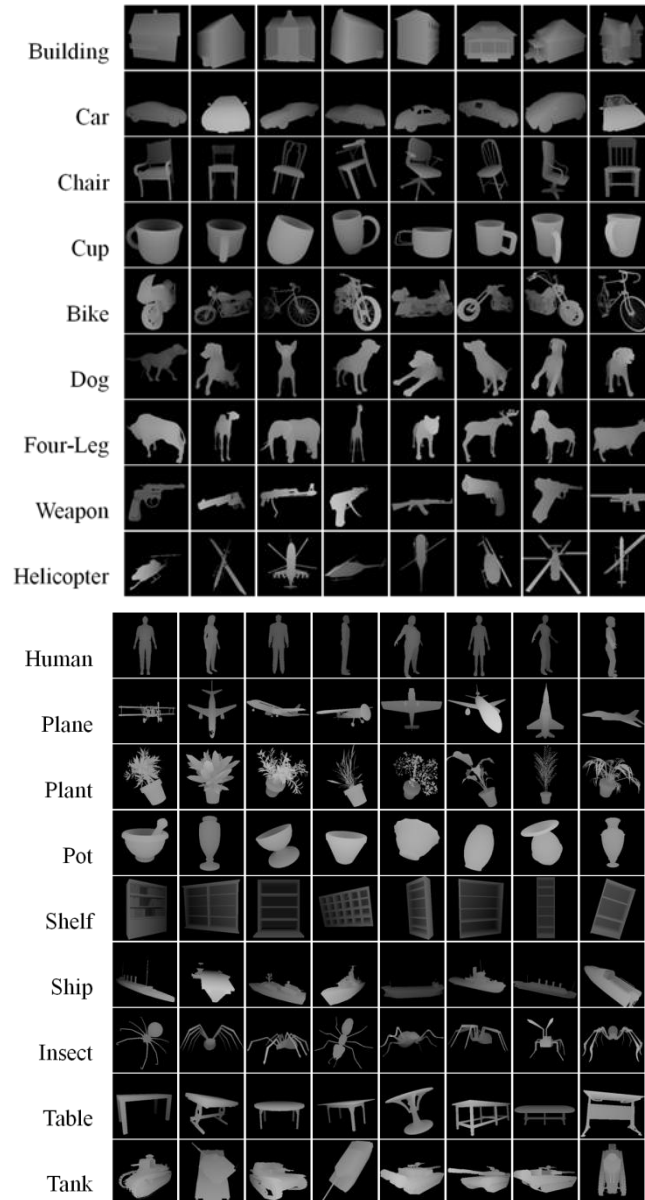
Figure 5-7. Representative models of the database.

Although our database consists of toy data, there are two reasons for this obligation: *i) the number of intra class size, ii) diversity in viewing direction*. Databases in shape similarity retrieval applications have significant effect; so for obtaining more accurate performance results, algorithms should be tested on a large database having high diversity. For example, different types of *ships (including huge sized ones), helicopters (including military ones), animals*

(including wild ones) etc. and their different viewings of them should be included in the database. Real scanned database containing such kind of data is not available in today's possibilities. Despite the synthetic structure, our database is a challenging one. First, intra class similarity is quite low (mainly due to varied viewing directions). Secondly, for some classes, interclass similarity is considerably high (i.e. cup- pot, dog-fourleg).

Our range images have a size of *256x256* and we discretize the space into a *512 x 512* grid. Totally, *266,144* points are defined on the sphere. We compute the spherical harmonic expansion coefficients using *S2KIT* (http://www.cs.dartmouth.edu/~geelong/sphere). We use *l=256,* so we have a signature of length *255*. In the offline phase, we extract the signatures of the all database models, when the query is presented online phase takes place. Signature of the query is evaluated and compared with the signatures of the database models using L_2 norm.

In the firt stage of the classical SHT, a model is translated so that its center of mass coincides with the coordinate origin. The the model is normalized and the radius is discretized into 32 levels; and *l=32* is used. The signature is of length 1024.

To construct the *D2* and the *D2-Salient* descriptors we utilize several sampling densities such as 1000, 1500, 2500 and 10^4 point pairs. *Jensen-Shannon* Divergence *(JSD),* which is a symmetric version of *Kullback-Leibler* Divergence, is used in comparing the distributions.

We use "precision-recall" curve, First-Tier (FT) and Second-Tier (ST) to present the retrieval performance. Precision is defined as the ratio of the number of shapes retrieved correctly over the total number of retrieved shapes. Recall is defined as the ratio of retrieved shapes over the total number of relevant shapes in the database. First-tier is the percentage of the matches belonging to same class of the query which appears at the top N matches where N is the number of items in the query's class. Second-tier is the percentage of the top relevant matches (items that are in the same class with the query) appear in the top 2*N matches. Figure 5-9 shows average precision curve for the proposed method, classical SHT and the *D2* shape distribution. A random experiment is also conducted in order give a different idea about the performance of the algorithm. *First-Tier* and *Second-Tier* percentages are presented in Table 5.

88
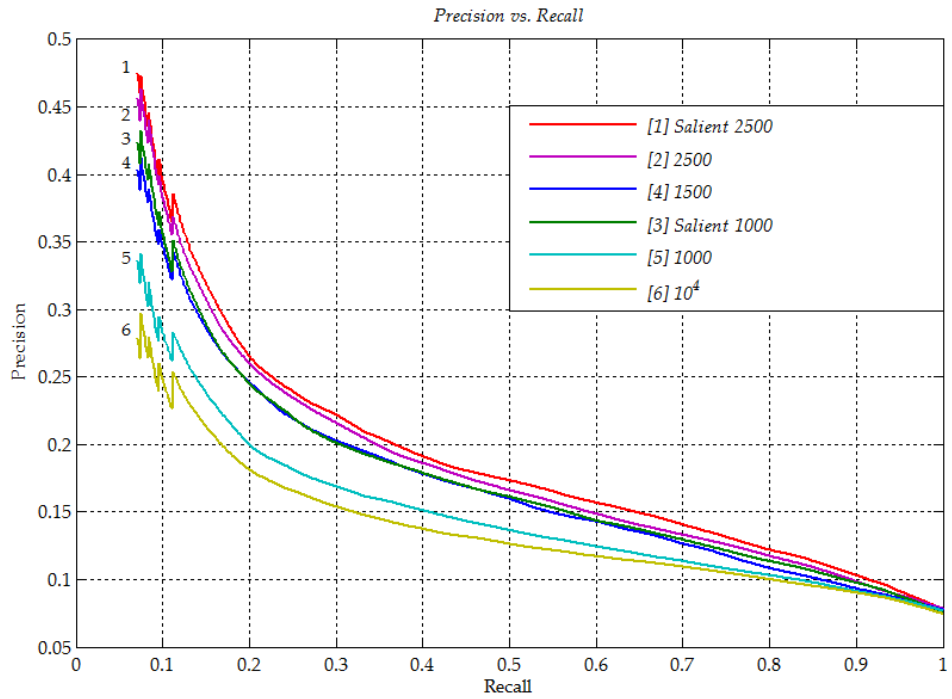
Figure 5-8. Average Precision-Recall curve of *D2* distributions
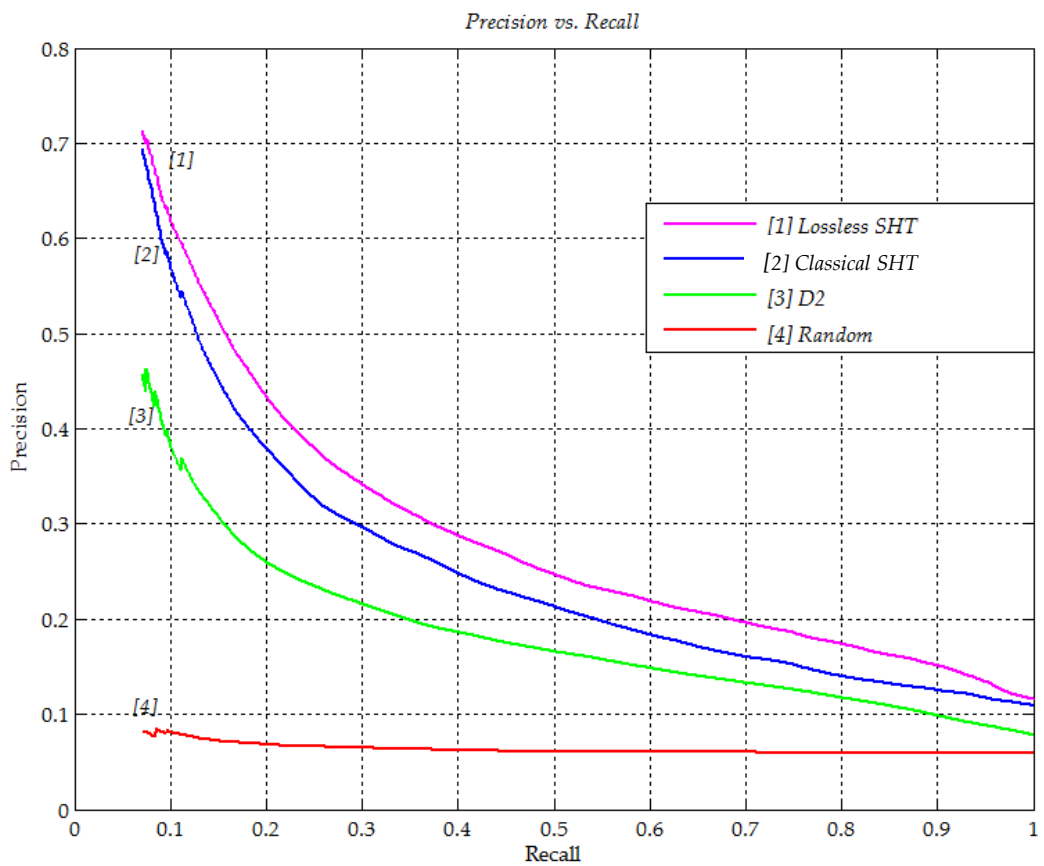
Figure 5-9. Average Precision-Recall curve of Lossless SHT, classical SHT, and *D2* distribution and a random retrieval experiment

The *D2* distribution yields better performance by increasing sampling density up to a point. Further increase in sampling size decreases the discriminative power of the method. In the limit, the distances between all combinations of the point pairs will be included in the distribution. Perhaps some of them will contribute more than once. The best performance among *D2* distribution belongs to *Salient-D2* distribution with 2500 point pairs. The proposed modification in *D2* distribution improved the performance slightly for 2500 point pairs. However, for 1000 pairs, *Salient-D2* distribution obviously has better retrieval characteristic than standard *D2* distribution. For time and space savings, such a modification could be utilized.

Classical SPH perform better than D2 distribution and slightly worse than the proposed method. On the other hand, our proposed method, *Lossless SHT*, has a much better average precision-recall curve while the descriptor size is smaller than the classical SHT. Random case experiment is a good criterion for realizing the complexity and the size of the database. The *First-Tier* and *Second-Tier* measures of *Salient-D2* and *Lossless SHT* for each class are shown in Table 1. *Lossless SHT* has higher performance in both FT and ST percentages for 13 classes among the 18 classes. On the average, *Lossless SHT's* FT is 26,5% and ST is 32,8% whereas *Salient-D2* distribution's FT is 15% and ST is 22,9%. Sample retrieval results of three queries are given in Figure 5-10 and Figure 5-11. In our method, false matches belong to geometrically similar models, such as retrieving gun in case of human query, retrieving helicopter for spider query and retrieving pottery in querying cup sample. They can be considered as "*good false matches*", whereas these good false matches are not presented in retrieval results of *D2* distribution method.

5.8 Conclusion

We propose a global description method that relies on *SHT* for similarity retrieval of range models. Despite the challenges, which are due to the view dependent structure of the data, the proposed technique achieves a Second Tier measure up to *87,1%* . On the other hand, the popular *D2* distribution can achieve a maximum of *53,2%* Second Tier performance. The performance difference is more evident in First Tier characteristics. Proposed *Lossless SHT* obtain a maximum of *61,9%* First Tier performance whereas *D2* distribution remains at *29,6%*. Also, Precision-Recall curve of the *Lossless SHT* shows a better characteristic.

Increased complexity in the database and the 2.5D data type challenges, which are not present in complete 3D mesh retrieval research, degrades the retrieval success compared to the ideal case. Nevertheless, experimental results are encouraging. Our method *Lossless SHT* shows satisfying performances regarding the *"good false matches"* and the performance of the other methods as well as the random case experiment.

Table 5 Retrieval results for both Salient-D2 distribution and our Lossless SHT

	Salient D2		Lossless SHT	
	First Tier	**Second**	**First Tier**	**Second**
Human	29,6%	50,6%	61,9%	87,1%
Building	15,7%	24,7%	51,5%	74%
Cycle	23,2%	38,9%	20,2%	26,5%
Spider	15,2%	25,3%	36,6%	53%
Fourleg	24,1%	35,6%	17,1%	26,8%
Pottery	16,5%	23,1%	32,2%	47,7%
Helicopter	16,2%	26,7%	24,4%	36,4%
Ship	19,6%	30,6%	21,4%	31,6%
Tank	28,4%	47,3%	19,5%	28,7%
Dog	14,7%	23,7%	33,9%	44,3%
Gun	14%	24,8%	13,8%	22,4%
Car	22,8%	37,7%	27,2%	40%
Cup	27%	42,6%	42,4%	68%
Table	18%	26,6%	15,4%	23,4%
Plane	12,4%	22,5%	26,9%	39,4%
Plant	18,8%	31,6%	20,7%	29,3%
Chair	31,9%	53,2%	34,6%	51%
Shelf	28,3%	42,9%	38,7%	46,4%
Average	**15,0%**	**22,9%**	**26,5%**	**32,8%**
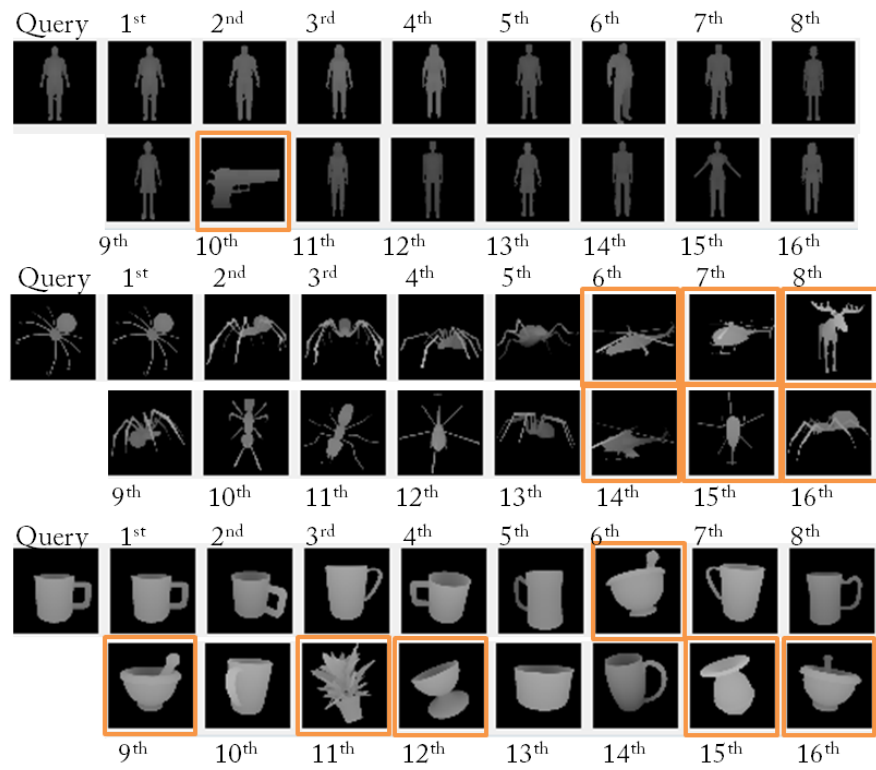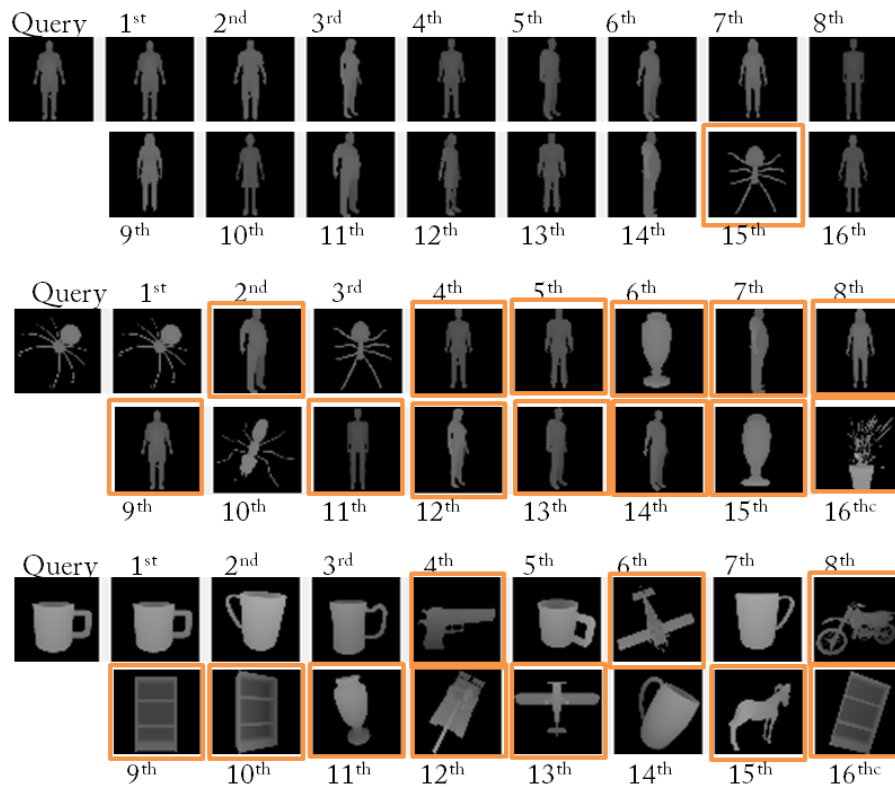
Figure 5-10. Sample retrieval results of three queries belonging to three different classes human, spider, and cup respectively. First 16 matches are shown. Queries are also included in the database, so first match is always the query itself. *(Top)* Retrieval results of *Lossless SHT*, *(Bottom)* retrieval results of *classical SHT*.

92
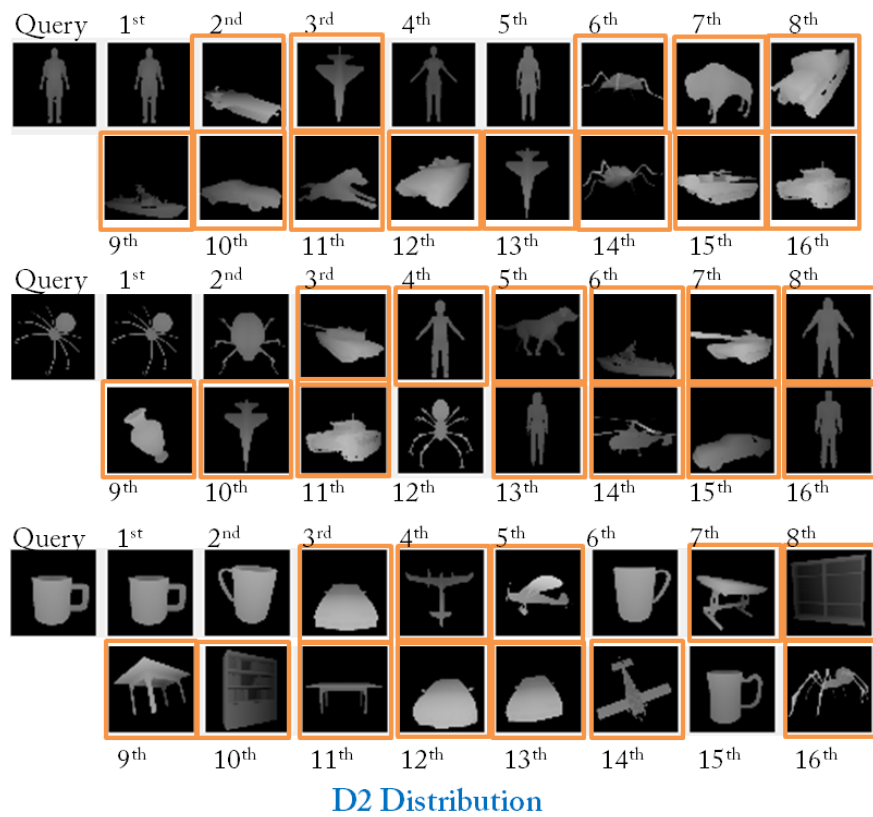
Figure 5-11. Sample retrieval results of *D2* description with 2500 point pairs. Three queries belonging to three different classes human, spider, and cup respectively. First 16 matches are shown. Queries are also included in the database, so first match is always the query itself.

CHAPTER 6

ANALYSIS OF HEAT DIFFUSION ON RANGE IMAGES

6.1 Introduction

Some feature detectors and descriptors for recognizing 3D objects are explained in Chapter 2. Feature based descriptors are classified into two sub-categories: *global descriptors* and *local descriptors* [8]. Recently a third class of 3D descriptors, multi-scale descriptors, is proposed [128, 129, 136, 141, 142]. The descriptor, Heat Kernel Signature (HKS) or Auto-Diffusion function, is based on the heat diffusion process on surfaces. The heat diffusion process is governed by the heat kernel, *H(t)*.

Dissipation of heat over the time from a surface point onto the rest of the shape captures information about the shape of neighborhood for that point. The heat kernel provides a multi-scale descriptor by the time parameter *t*. As times getting larger, the heat dissipates from a point gradually to a larger neighborhood of the point. The information about larger neighborhoods of the point *x* is gathered over a longer period. Similarly, the behavior of heat diffusion over short time reflects local shape features. Besides the multi-scale property of HKS, it is argued in [128] the HKS preserves all of the information about the intrinsic geometry of the shape. It can also be argued that HKS describes an object both from local and global aspects at different time instants. Due to these properties, heat diffusion process is receiving more interest in shape analysis of 3D mesh models [128, 129, 136, 141-146].

Sharma and Horaud [144] used eigenvalues and eigenvectors of the discrete diffusion operator for point matches. Rustamov [147] evaluate Laplace Beltami operator for a fixed point and construct the feature vector of that point using eigenfunctions of the Laplace Beltrami operator. Later Sun et al. [128] proposed using the diagonal of the heat kernel as a

local descriptor and called it as Heat Kernel Signature (HKS). For each point x on the shape, its heat kernel signature is an n-dimensional descriptor formed by evaluating auto-diffusion value at n different time instants. Local and global information about the neighborhood of the point is captured by this description. Dey et al. utilize HKS with persistent homology for shape retrieval. A scale invariant version of HKS, which is called SI-HKS, is proposed by Bronstein et al. [136, 141]. Mahmoudi used the distribution of the diffusion distances as a global descriptor [22]. For a fixed point on surface S, diffusion distance between this point and all other points are calculated and histogram of this distribution is used as a descriptor.

Besides the increasing reputation of heat diffusion descriptor among 3D mesh models, the theoretical and the practical analysis of heat diffusion on range data is missing in literature. This chapter analyses heat diffusion process on range images. Challenges and some experimental results are presented. Also some future directions are proposed

6.2   The 3D Heat Equation

Suppose there is a function of space and time T(x,t) describing the temperature at a given surface location x at time t. The value of this function will change over time as heat spreads throughout the surface. This change depends on the shape of the surface and the heat sources on it. Moreover, initial heat distribution and boundary conditions have effects. This shape dependent character of heat diffusion makes it a state-of-the-art method in shape description.

The 3D heat equation is given as follows:

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} - \frac{1}{\alpha}\frac{\partial T}{\partial t} = 0 \qquad\qquad 6.1$$

Equation 6.1 is called as *heat equation* or *diffusion equation*. It is generally written in the following format:

$$\left(\frac{\partial}{\partial t} - \Delta_M\right)T(x,y,z,t) = 0 \qquad\qquad 6.2$$

The solution of $T(x,y,z,t)$ (partial differential equation solution) depends on the boundary conditions and the initial conditions, $T(x,y,z,0)$. Δ_M is the *Laplace–Beltrami operator and* will be denoted by L after this.

6.3   Graph Laplacians

In the discrete case, one can think of a 3D surface as graph. Surface points $X_i \epsilon \mathbb{R}^d, i = \{1,2,..n\}$ are represented with nodes. Hence, using K-nearest neighbor a connected graph can be obtained. Several graph Laplacian operators are proposed for the discrete case [148-151]:

Combinatorial Laplacian (L) is defined as:

$$L = D\text{-}W,$$

6.3

where W is the adjacency matrix and D is the diagonal degree matrix. The weight matrix W is defined in several ways, which will be addressed later. The "degree matrix" D is written as:

$$D_{ii} = \sum_{i=1}^{n} w_{ij}$$

6.4

The *normalized Laplacian* (L_N) is defined as follows:

$$L_N = I - D^{-1/2} W D^{-1/2}$$

6.5

The *random walk Laplacian* (L_R) is defined by the following relation:

$$L_R = I - D^{-1} W$$

6.6

Finally, *Mesh Laplacian* used in [128] is defined as:

$$L = A^{-1} W,$$

6.7

where A is a diagonal matrix where element A_{ii} represents the area associated with vertex i.

Moreover, the adjacency matrix W is defined in several ways:

1. Uniform weighting or umbrella operator is

$$W = \begin{cases} W_{ij} = 1 & \textit{if there is an edge between i and j} \\ W_{ij} = 0 & \textit{if there is no edge between i and j} \\ W_{ij} = 0 & i = j \end{cases}$$

This weighting only describes the topological properties of the connectivity (mesh, point cloud, etc.)

2. Cotangent weighting (Figure 6-1)

$$W = \begin{cases} W_{ij} = \dfrac{1}{2\left(\cot \alpha_{ij} + \cot \beta_{ij}\right)} & \textit{if there is an edge between i and j} \\ W_{ij} = 0 & \textit{if there is no edge between i and j} \\ W = 0 & i = j \end{cases}$$
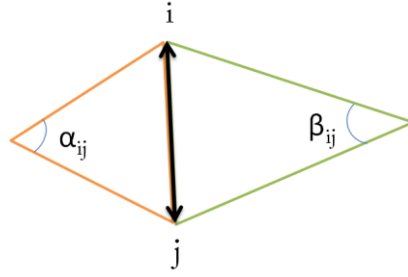
Figure 6-1. Angles used in cotangent weighting scheme for Laplacian construction

3. Distance based weighting

$$W = \begin{cases} W_{ij} = e^{-\frac{d_{ij}^2}{\sigma^2}} & \text{if there is an edge between } i \text{ and } j \\ W_{ij} = 0 & \text{if there is no edge between } i \text{ and } j \\ W_{ij} = 0 & i = j \end{cases} ,$$

where d_{ij} is the Euclidean, or preferably geodesic, distance between the nodes (vertices) i and j.

Properties of Graph Laplacian

- L is a symmetric semi-definite positive matrix $L = \mathbb{U}\Lambda\,\mathbb{U}^T$

- It has eigenvalues such that $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n$ and eigenvectors $u_1 = 1, u_2, \dots u_n$ which are orthogonal. Additinonally, $\sum_{k=2}^{n} u_{ik} = 0, \forall k\epsilon\{2, \dots, n\}$

- Since eigenvectors form a basis, L can be written as

$$L = \sum_{k=2}^{n} \lambda_k\, u_k u_k^T \tag{6.8}$$

Consider the heat equation in discrete case $\left(\frac{\partial}{\partial t} - L\right) T(t) = 0$, with $T(t) = \{T_1(t),\ T_2(t),\dots,\ T_n(t)\}$, which is a vector indexed by the nodes of the graph. The solution in the discrete case is:

$$T(t) = H(t)T(0) \tag{6.9}$$

where H denotes the discrete heat operator: $H(t) = e^{-t\mathcal{L}}$, and T(0) is the initial heat distribution at graph nodes, and for simplicity, it is usually given with a unit heat at single node $T(0) = (0, \dots\dots, T_i = 1, \dots\dots, 0)$. Starting with this initial condition, the heat distribution at t, $T(t) = \left(T_1(t), \dots\dots, T_i(t), \dots\dots, T_n(t)\right)$, is given by the *i-th* column of the heat operator.

Computation of Heat Kernel

$$H(t) = e^{-t\mathcal{L}} = e^{-t\mathbb{U}\Lambda\,\mathbb{U}^T} = \mathbf{U}e^{-t\Lambda}\mathbf{U}^T$$

$$e^{-t\Lambda} = Diag[\ e^{-t\lambda_1}\ \dots\ e^{-t\lambda_n}\ ]$$

$$H(t) = \mathbf{U}\ Diag[\ e^{-t\lambda_1}\ \dots\ e^{-t\lambda_n}\ ]\quad \mathbf{U}^T$$

The heat kernel has the eigenvalues $1 = e^{-t0} > e^{-t\lambda_1} \geq \dots \geq e^{-t\lambda_n}$. As it can be observed from the above equation, the eigenvectors are same as the Laplacian matrix.

$$H(t) = \sum_{k=2}^{n} e^{-t\lambda_k} u_k u_k^T \qquad 6.10$$

In order to compute the heat kernel, we only need the Graph Laplacian matrix and its Singular Value Decomposition. To be more specific, an entry of the heat operator, that is the amount of heat at vertex v_i and at time t, starting from a heat distribution located at a single vertex v_j, is given by the following equation:

$$h(i,j,t) = \sum_{k=2}^{n} e^{-t\lambda_k} u_{ik} u_{jk}^T \qquad 6.11$$

The heat kernel contains redundant information and all the information is contained at the diagonal terms [143]. The diagonal terms of the heat operator ($h(i,i,t)$) correspond to the *auto-diffusion function* [129] or *Heat Kernel Signature (HKS), [128].* HKS function values at each vertex on a 3D mesh model for four different time instances is shown in Figure 6-2.
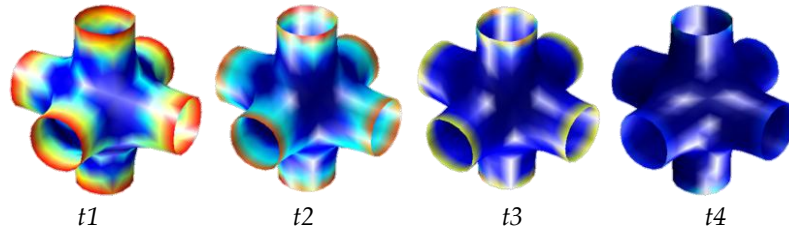
t1	*t2*	*t3*	*t4*

Figure 6-2. HKS function on a 3D mesh model for four different time instances where t1<t2<t3<t4. The function values increase from blue to yellow and to red. The mapping is consistent across the shapes.

The distance between two graph vertices, which is as an Euclidean metric, is called diffusion distance and is given as

$$d_t^2(i,j) = h(i,i,t) + h(j,j,t) - 2h(i,j,t) \qquad 6.12$$

6.4   Heat Diffusion and Scale Space Theory

Scale space representation of an image is first proposed by Iijima in 1959 [152, 153]. Later, scale-space get attention with Witkin [154]. After one year in 1984, Koenderink [155] stated that Gaussian scale-space representation of an image is equivalent to the solution of the heat diffusion equation [156] . The intensity values are considered as the initial temperature distribution, image is interpreted as a planar surface and the heat distribution at different

times are observed by changing the variance of the Gaussian kernel. In other words, the standard deviation of the Gaussian kernel is equal to $\sigma = 2\sqrt{t}$.

6.5   Heat Kernel on Range Images

Constructing graph Laplacian on a range image is challenging. This is due to the fact that point connectivity is not defined explicitly as in mesh representation. The first way to construct adjacency relation is using pixel neighboring. This representation ignores the true geometry and directs the diffusion of heat to semantically meaningless directions. This case can be observed in Figure 6-3. The first two images represent the HKS function values with a RGB representation at a particular time instant on a complete mesh, whereas the third image shows a range image obtained from the corresponding mesh. The next one (Figure 6-3d) represents HKS values again with a RGB representation at a particular time instant which is obtained by using pixel neighboring relations. Heat diffuses from fingers to body which will not be the case in reality. This is due to the self occlusions. In order to overcome this problem, an occlusion detection method is proposed.

6.5.1 Occlusion Detection

For detecting the occlusions, we generate mesh structure from range images (see Figure 6-3 and Figure 6-4). For this purpose, two filters are used:

- Compare the length of the triangles with mesh resolution given by angular resolution of the range image and the distance of the object
- Check the side ratio; an optimal mesh should have equilateral sides

After occlusion detection, range image is represented with disconnected regions (meshes). Under these conditions, the graph Laplacian properties will no longer be valid. The following two theorems prove this argument:

a.   b.   c.   d.   e.   f.
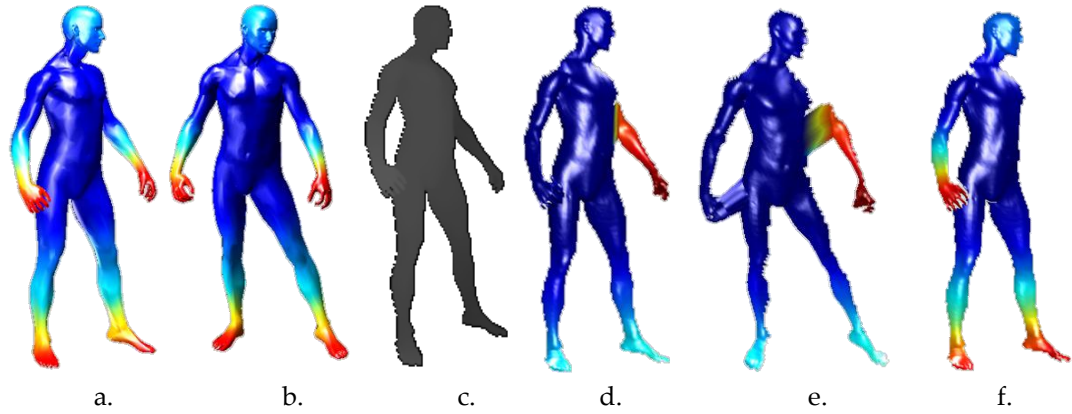
Figure 6-3. a. HKS function on a 3D complete mesh model ([157]), b. The same function on the same mesh is represented in a different view to give an additional information, c. A range image of that model (background is removed), d. HKS function is evaluated on the range data in 'c' without occlusion detection, heat diffuses from right hand of the human model to the body directly, e. Another view of (d), f. HKS function on the occlusion aware mesh model obtained from the range image in 'c', some parts, such as left hand, are removed (The function values increase from blue to yellow and to red. The mapping is consistent across the shapes. The function value is calculated for a small time instant).
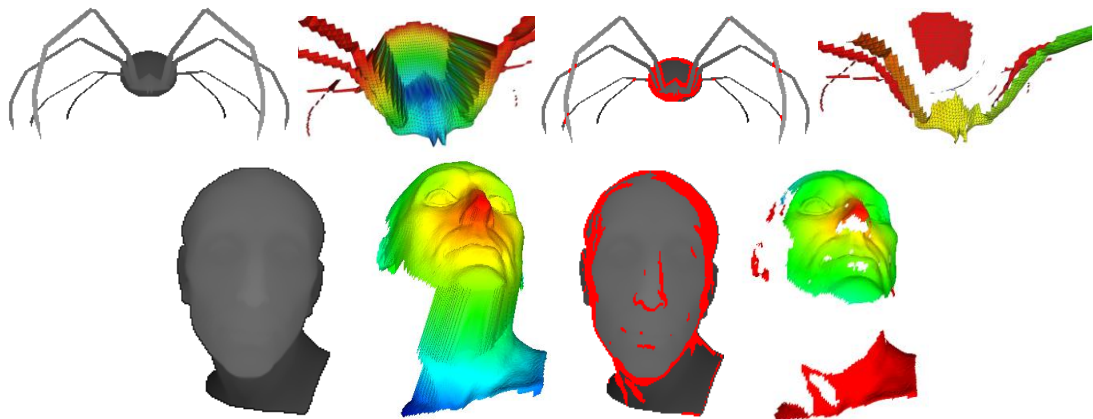
Figure 6-4. Occlusion detection examples.

Theorem 6.1. *(Disjoint Union Spectrum)[158] Given a graph G and its Laplace spectrum L_G has eigenvectors $v_1,...,v_n$ with eigenvalues $\lambda_1,...,\lambda_n$,. Similary another graph H with its Laplace spectrum L_H has eigenvectors $w_1,...,w_n$ with eigenvalues $\mu_1,...,\mu_n$, then the spectrum of the disjoint union of these two graphs L_{G+H} has eigenvectors:*

$$v_1,...,v_n, \ w_1,..., w_n$$

with corresponding eigenvalues:

$$\lambda_1,...,\lambda_n, \mu_1,...,\mu_n.$$

Theorem 6.2. *The multiplicity of the Laplacian eigenvalue 0 is the numberof connected components of the graph[158].*

100

Proofs of Theorem 6.1 and Theorem 6.2 can be reached at [158]. By the help of these theories, we conclude that eigendecomposition of the Laplacian of the union graph will not reflect the connectivity and topological properties in the global sense. Heat kernel of a surface can be inferred from the corresponding graph Laplacian matrix of that surface, if and only if the graph is connected.

6.5.2 Local Description

In this section, local behavior of the heat kernel signature is investigated. For this purpose, some keypoints are located on range images and the regions around the local neighborhood of these keypoints are constructed. For key point selection, we use the previously mentioned curvature maxima defined as:

$$c = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2}, \quad \lambda_0 \leq \lambda_1 \leq \lambda_2,$$

by comparing the three eigenvalues λ_i obtained from Principle Component Analysis (PCA) of the covariance matrix (C) defined on the local neighborhood of the query point p (Figure 6-5).

Figure 6-5 Keypoint selection, range image, surface variance mapping, keypoints as the maxima of the surface variance

After keypoint selection, we use a region growing on the occlusion aware mesh constructed previously starting from the keypoint. This region growing ensures the unwanted short circuited connections. Then, the signature of each keypoint, $p(x)$, with an n-dimensional vector is constructed, similar to the Jian's work [128]:

$$p(x) = c(x)(h (x, x, t1), \ldots , h (x, x, tn)),$$

where $c(x)$ is chosen in such a way that $\| p(x) \|_2 = 1$. This signature captures differential information in a small neighborhood of a keypoint for small t and global information about

the shape for large values of t. As a result, such a signature can be utilized as an multiscale feature descriptor.

Jian et.al [128] compute the difference between two HKS's to see if two points x and x' are matched at the given scales specified by the time interval [$t1$, $t2$],

$$merr = \left(\int_{t1}^{t2} |h_t(x,x) - h_t(x',x')|^2 \, dt \right)^{\frac{1}{2}}.$$

This measure (*merr*) should be zero in order to match x and x' exactly. However, in practice, matching is achieved approximately and it should be selected accordingly. Jian points that $h_t(x,x) = \sum_{k=2}^n e^{-t\lambda_k} u_{xk} u_{xk}$ decays exponentially as t increases that makes the difference at large scales negligible compared to those at small scales. He addresses the problem by scaling signature $h_t(x,x)$ by $\int_M h_t(x,x)dx$ where is the entire manifold, M, for all t. This heuristic ensures that the differences between two signatures at different timescales contributes almost equally. Another problem that Jian points is related to the temporal domain. For small t values, $h_t(x,x)$ function variance is large, whereas it decays as t increases. This result is obtained since the HKS signature $h_t(x,x)$ is easily affected from the local neighborhood and is not stable. Hence, Jian scale the temporal domain logaritmically and obtain the following final measure:

$$d(x,x') = \left(\int_{t1}^{t2} \left(\frac{|h_t(x,x) - h_t(x',x')|}{\int_M h_t(x,x)dx} \right)^2 \, dlogt \right)^{\frac{1}{2}}.$$

During our experiments, we use the same keypoint description and matching strategy [128]. Figure 6-6 shows the matching result of a two "Homer" models. There is small rotation, four degrees about the z-axis, between the models. Matching result is expected to be more accurate due to this small rotaion, but many false matches are resulted, since, in the presence of a boundary, the HKS signature is significantly affected. When three specific keypoints p1, p2, and p3 are examined, their multiscale HKS signature are resulted to be dissimilar for p1 and p2 and distance between p1 and p3 is smaller than the distance between p1 and p2.
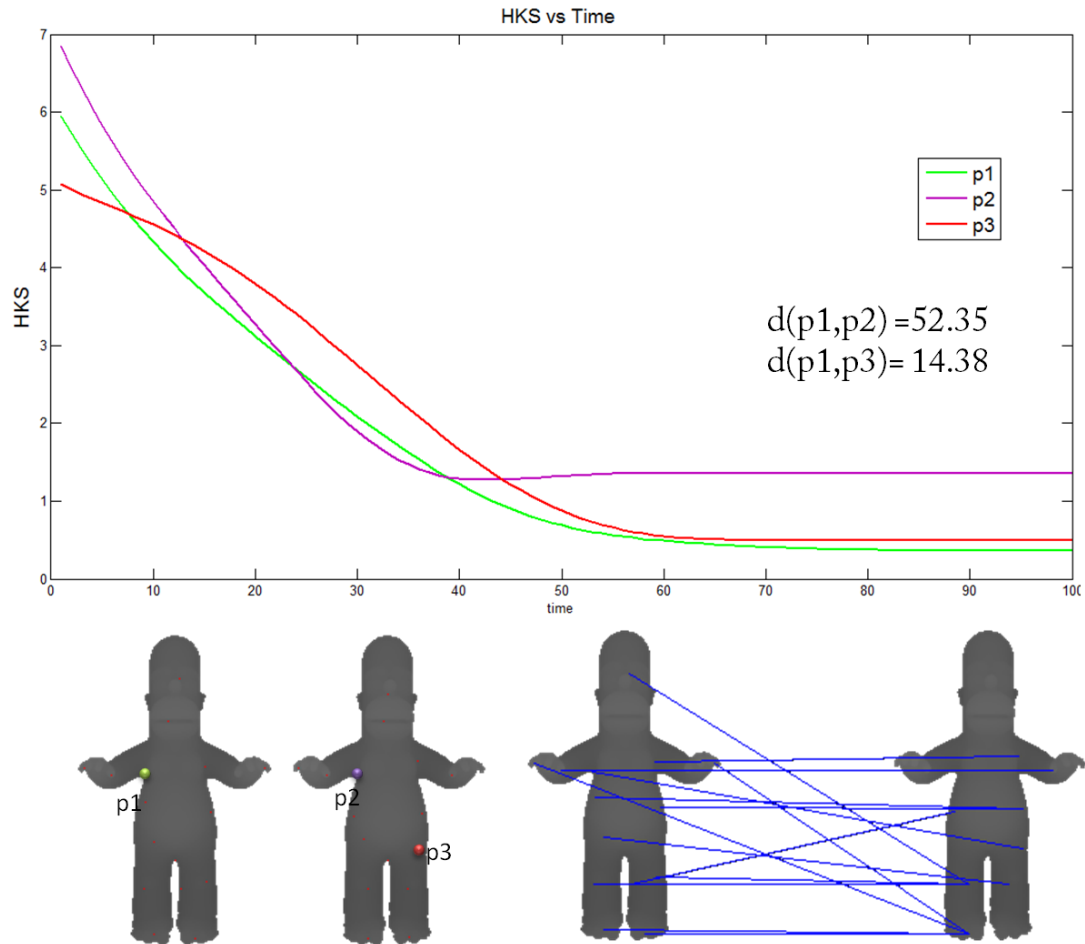
Figure 6-6. Two models are matched using local HKS signature. Last figure represents the matching result. There is a rotation of 4 degrees between two models. Three specific keypoints, p1, p2, and p3 are examined in detail. Keypoints p1 and p2 are expected to match; however, p1 matched with p3. Besides, a small rotation between the two models many false matched are presented, since the HKS behavior is significantly affected with the changing boundary.

6.6   Conclusion and Future Directions

A recent approach, heat diffusion, in shape analysis with its multi-scale character has been increasingly popular for 3D shape recognition. The approach describes the behavior of heat diffusion on the surface overtime. The attractive property of the HKS descriptor is its multi-scale property. However, utilization of HKS descriptor on the range data for similarity retrieval and correspondence matching purposes introduces some challenges. These are summarized as follows:

- The heat kernel defines a shape uniquely [143] and heat kernel of two shapes will be same, if and only if they are isometric [143]. However, range data representations

are view dependent, the same heat kernel cannot be obtained for different views of a particular object. Even, a similar behavior of HKS values of a particular surface point cannot be obtained due partially available shape information. This observation is due to the fact that the same point can take part in two different surfaces considering the viewing direction. The points can be at the boundary at some viewing direction or can reside in the middle of a surface for another direction. The multi-scale property of the HKS will be limited in such cases. The suggestion is to restrict the time parameter for smaller times. It should noted that in that case, only local surface properties can be gathered from the HKS descriptor. In addition to this proposal, also boundary points should be avoided for obtaining feature descriptors.

- The heat kernel on a discrete surface can be evaluated utilizing the eigenfunctions of the graph Laplacian. This theorem holds, if and only if the graph is connected. Since the mesh representation of a range image might contain disconnected parts, spectrum of the corresponding graph Laplacian cannot be used in analyzing the heat diffusion process on range images. Each part should be analyzed separately. A bag-of-words approach can be utilized after this step. This proposition can be applicable, if the segmentation is consistent among the shapes. For example, if all of the human models in the database are segmented into legs, arms, heads and body then the HKS description in each part can be utilized. In that case, some more segmentation for some of the parts would be required.

- HKS descriptor is dependent on the global scale of the shape [136]. However, the global information is missing in range data. Evaluation the HKS descriptor on local surface patches seems to be the easiest suggestion for this challenge. However, the keypoint repeatability cannot be achieved for all times. In such cases, local patches that are formed around a keypoint will be different which results significantly different signatures. This is shown by an example in the previous section. One suggestion for overcoming this challenge would be using the occlusion aware mesh representation of range data. A range data can be described with the HKS description evaluated on the largest connected part of it (Figure 6-3f). The proposal can be feasible, if the viewing directions do not vary significantly.

CHAPTER 7

SUMMARY, CONCLUSIONS, AND FUTURE DIRECTIONS

7.1 Summary

The motivation of semantically describing and representing objects arise from the fact that there is a significant need for recognizing, organizing, classifying and searching the *content* of visual data. Since capturing, displaying and storing the 3D media increased rapidly and 3D data takes place in our daily life, such as TV (3DTV), laptops, chemistry (e.g., protein modeling), archeology (museum data), medicine, geography, military, industry (CAD), computer games, architecture, medical surgery, virtual reality programs, education and entertainment.

In this thesis we studied 3D semantic analysis of range data. Under this broad title we address segmentation of range scenes, correspondence matching of range images and the similarity retrieval of range models. Inputs are considered as single view depth images.

First, possible research topics related to 3D semantic analysis are introduced. Problem variety and the distinction depending on the 3D data type are stated. Later, 3D representation formats and the differences between them are presented. Challenges, namely representation diversity, alignment, description requirements, similarity measure, registration and segmentation are defined. Performance evaluation metrics in similarity retrieval research are introduced. Classification of similarity retrieval methods is proposed and a comprehensive literature survey is presented.

Next chapter begins with planar structure detection in range scenes. Problem definition, proposals, modifications on available plane detection methods and results are presented. Later, the adaptations of the well-known segmentation methods for segmenting real range scenes, where the inputs are the Time-of-Flight Camera outputs, are presented. Integration of the color with the spatial information and its effects on the performance of the segmentation methods are evaluated. Utilization local surface properties such as curvatures, surface variances, and normals with region growing are presented. Finally, a novel algorithm to segment 3D point cloud (obtained via TOF camera) into objects by using the spatial information is presented. This segmentation proposal starts with estimating the major plane, which represents the surface that objects can be placed. Afterwards, 3D points are projected onto the major plane and a saliency map is generated by using density estimation method. The local peaks in the saliency map represent the probable object locations. Finally, points which are closer to the local peaks than a certain threshold are segmented into objects.

Correspondence matching in range images is investigated in the following chapter. We proposed a novel local range image matching method which combines 3D surface properties with the 2D scale invariant feature transform. Local surface properties that are utilized in this work are explained and feature extraction methodology is described. Finally matching results and comparison to the previous studies are presented.

Next, our proposal for retrieving similar models where the query and the database both consist of only range models is presented. The challenges and the differences between the similarity retrieval of range images and the similarity retrieval of 3D mesh data are presented. The shortcomings in representing the range models with local descriptor are then stated. Our method relies on the well known Spherical Harmonics Transform. Steps in constructing the lossless spherical function that represents the range object is explained. Comparative performance results of the proposed method are presented. Tests are performed on a large database having high diversity and it is shown that performance of the proposed method is superior to the performance of popular D2 distribution.

Finally, heat diffusion process on range images is analyzed in the last chapter. Challenges and some experimental results are presented. Also some future directions are proposed.

7.2 Conclusion

The performance of the proposed Recursive Hough Transform in finding planar structures in range images is shown to be superior to the classical Hough Transformation. While the proposed modification in Hough Transform introduces additional complexity in space and time, it is quite successful in detecting planar structures in range image like data. This is mainly due to eliminating the fine parameter discretization ambiguity that is presented in the classical way. Moreover, RANSAC performance is improved slightly by introducing constraints in the initial estimation. Also, the HK segmentation method is applied for finding planar regions in TOF data. It is shown to be inefficient since the curvature estimation of a point by utilizing the local neighborhood is excessively noise sensitive. Furthermore, we utilize classical Hough Transformation after segmenting the scene. Each segment is considered as a planar region. The experimental results indicate that, while the time and space complexities is lower than that of proposed Recursive Hough Transformation, the performance is directly related with the segmentation quality. Recursive Hough Transformation could achieve superior performance on the airborne LIDAR data, because the piecewise linearity assumption pretty holds for that kind of data, such as roofs, roads, lands.

Various segmentation algorithms, namely K-Means, Mean-Shift, Mixture of Gaussians, and region growing by local surface properties are adapted for segmenting real TOF data. The low spatial resolution of the TOF camera decreases the accuracy of the local surface property estimation process and makes it sensitive against noise. Moreover, false measurements due to the multiple reflections might result in significant variance in the estimated normal values within a small surface patch. This variance leads to oversegmented or undersegmented regions. Furthermore, color-based segmentation algorithms are ineffective, when the objects in the scene have similar color values with the background or consist of different colored patches. 3D information or edge based segmentation methods also fail due to the continuous and smooth transitions, originated from the multiple reflections, between object surfaces and background surfaces on which the objects are placed.

The proposed algorithm for segmenting range scenes is attractive, because it overcomes most of the aforementioned shortcomings. The proposed algorithm exploits the fact that

many objects stand orthogonal to the ground plane due to gravity and the projection of 3D points onto this plane could be equivalent to Fisher Linear Discriminant methodology. The projection of the 3D points is followed by a kernel density estimation process in which a saliency map is generated. Salient regions which represent the most probable object locations have high values in the generated map. Then the points with a probability value larger than some certain threshold are assigned to the closest local maximums segmented into objects. The experimental results show that the locations of the objects are determined quite accurately. Compared to the tested methods proposed algorithm is less sensitive to the noise, less parameter dependent and leads more accurate segmentation results.

We also developed a technique for correspondence matching in depth data. There has been a little work in this area unlike the 2D case. The methods used in registering 3D point clouds are utilized, such as ICP, for finding correspondences in range images. Such global methods ignore shape details and gather the main properties of objects; thus sensitive to outliers. Besides, the need for segmentation is obvious. We proposed a local descriptor that puts more emphasis on the details which is more suitable for matching. The two recent aforementioned studies, which also utilize local descriptors require preprocessing step in removing problematic boundary regions. Such preprocessing is not needed in our proposal. Besides, the matching performance in complicated scenes with the presence of other objects, clutter and occlusion are not evaluated in those studies. In order to demonstrate the versatility of the proposed method in the cluttered scenes, we conducted several tests with complicated scenarios. The proposed local surface description method does not require any initial segmentation step; it can also handle affine transformations up to a scale. The experimental results indicate that the proposed approach improve the performance of two recent methods from the literature. Moreover, clutter and occlusion do not affect the efficiency of the proposed method significantly.

Next, our proposal for similarity retrieval of range images utilize one of the main challenges of range imaging, namely view dependency, as an advantage. Although, there exists a satisfying number of 3D mesh similarity retrieval methods, most of them are not suited for utilizing them in retrieving range models. Because, in contrary to the mesh data type, range images do not contain complete geometric information of the object, they are view dependent, occlusions are present, and transformations such as scale differences are not

easily handled. Our proposal relies on a global description method *Spherical Harmonics Transform*. Although SHT is not a novel concept in shape retrieval research for 3D complete models, we utilize it for 2.5D range images by representing the models in a reciprocal world observed from the camera. The difference, as well as the advantage of our algorithm, is being information lossless. In other words, the available shape information is completely exploited for obtaining the descriptor, whereas other mesh retrieval applications utilizing SHT "approximates" the shape that yields information loss. The descriptor is invariant to rotations about z-axis. The proposed technique is tested on a large database having high diversity and its performance is superior to the performances of the popular *D2* distribution and the classical SHT.

Finally, heat diffusion on range images is analyzed. It is shown that self occlusions and varying boundary conditions are the major challenges. Occlusion detection introduces disconnected parts where the spectrum of the graph laplacian fails in analyzing the heat diffusion process on range images. The heat diffusion analysis on the local basis is more challenging since boundary conditions effect the descriptor significantly.

7.3 Future Directions

While this dissertation addresses some of the problems in analysis of range data, many issues left open for future research. Some of them, which are certainly incomplete, are as follows;

- The proposed methods' performances on single view range images are evaluated. Utilizing multi-view depth data should improve the performances. The proposed segmentation method should certainly benefit from this additional information.
- Local surface properties can be employed for planar structure detection. The information of normal vector, when estimated precisely, gives valuable clues for planarity. The angle between the normal vectors of two neighboring points should be zero if they are lying on a planar surface.
- Corresponding matching and similarity retrieval proposals are tested on toy data. This is because of the low signal to noise ratio of available TOF cameras, and the necessity in ensuring the diversity of the database for similarity retrieval. As the

technology improves, quality and the quantity of the real scanned data will be improved. Therefore, the performances should be evaluated for such data.

- The ratio of scale to the resolution is not constant in range imaging. The effects of scale changes are different from their 3D mesh counterparts. Fine details disappear as the objects become distant from the camera. Although absolute scale invariance cannot be reached due to the nature of the data, some negative effects can be reduced by resampling the objects according to their distances to the camera.

- We utilize "shape index" which is a curvature based feature for correspondence matching. The integration of 2D features could improve the matching performance. This can be utilized if the range data is accompanied with the corresponding color image. Recently high level sensors [14] provide such data.

- SIFT keypoint selection and description is the only employed method for correspondence matching between "shape index" mapped images. The performance of other keypoint detectors and feature descriptors remain an open issue that needs to be examined.

- Regarding the 3DTV archive systems, timing considerations might be critical. The analysis of time complexity of online phase in similarity retrieval approach is certainly in demand.

BIBLIOGRAPHY

[1]   J. C. Richard and J. F. Patrick, "A survey of free-form object representation and recognition techniques," *Comput. Vis. Image Underst.,* vol. 81, pp. 166-210, 2001.

[2]   T. Funkhouser, "Lecture Notes on Computer Graphics, http://www.cs.princeton.edu/courses/archive/spring06/cos426,", last accessed on 09/05/2008.

[3]   A. A. Alatan, Y. Yemez, U. Güdükbay, X. Zabulis, K. Müller, C. E. Erdem, C. Weigel, and A. Smolic, "Scene representation technologies for 3DTV : A survey," *IEEE transactions on circuits and systems for video technology,* vol. 17, pp. 1587-1605, 2007.

[4]   L. Piegl and W. Tiller, *The NURBS Book*. New York, NY: Springer-Verlag, 1997.

[5]   P. In Kyu, Y. Il Dong, and L. Sang Uk, "Constructing NURBS surface model from scattered and unorganized range data," in *Second International Conference on 3-D Digital Imaging and Modeling, 1999.*, 1999, pp. 312-320.

[6]   Wikipedia, "Wikipedia the free encyclopedia." vol. 2008.

[7]   J. W. H. Tangelder and R. C. Veltkamp, "A survey of content based 3D shape retrieval methods," in *Shape Modeling Applications*, 2004, pp. 145-156.

[8]   J. W. H. Tangelder and R. C. Veltkamp, "A survey of content based 3D shape retrieval methods," *Multimedia Tools and Applications,* vol. 39, pp. 441-471, 2008.

[9]   C. B. Akgül, "Density Based Shape Descriptors and Similarity Learning for 3D Object Retrieval," 2007.

[10]   P. N. Shilane, "Shape Distinction For 3D Object Retrieval," 2008.

[11]   S. Goodall, "3-D Content-Based Retrieval and Classification with Applications to Museum Data." vol. Ph.D. Thesis: University of Southampton, 2007.

[12]   M. Kazhdan, "Shape representations and algorithms for three-dimensional model retrieval," Princeton University, Ph.D. Thesis, 2004, p. 120.

[13]   N. Iyer, S. Jayanti, K. Lou, Y. Kalyanaraman, and K. Ramani, "Three-dimensional shape searching: state-of-the-art review and future trends," *Geometric Modeling and Processing 2004,* vol. 37, pp. 509-530, 2005.

[14]   Microsoft, "Kinect Sensor, http://www.xbox.com/en-SG/kinect,", last accessed on 14/07/2011.

[15]   L. I. Smith, "A tutorial on Principal Components Analysis," 2002.

[16]   K. Pearson, "On lines and planes of closest fit to systems of points in space," *Philosophical Magazine,* vol. 2, pp. 559-572, 1901.

[17]   B. Benjamin, K. Daniel, S. Dietmar, S. Tobias, and V. Dejan, "An Experimental Comparison of Feature-Based 3D Retrieval Methods," in *Proceedings of the 3D Data Processing, Visualization, and Transmission, 2nd International Symposium*: IEEE Computer Society, 2004.

[18]   D. Helin, S. Bulent, and Y. Yucel, "Transform-Based Methods for Indexing and Retrieval of 3D Objects," in *Proceedings of the Fifth International Conference on 3-D Digital Imaging and Modeling*: IEEE Computer Society, 2005.

[19]   D. V. Vranic, "DESIRE: a composite 3D-shape descriptor," in *IEEE International Conference on Multimedia and Expo, ICME* 2005, p. 4.

[20]   M. M. Bronstein and A. M. Bronstein, "Shape Recognition with Spectral Distances," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 33, pp. 1065-1071.

[21]   C. B. Akgul, B. Sankur, Y. Yemez, and F. Schmitt, "3D Model Retrieval Using Probability Density-Based Shape Descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 31, pp. 1117-1133, 2009.

[22]   M. Mona and S. Guillermo, "Three-dimensional point cloud recognition via distributions of geometric distances," *Graph. Models,* vol. 71, pp. 22-31, 2009.

[23]   R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, "Shape distributions," *ACM Trans. Graph.,* vol. 21, pp. 807-832, 2002.

[24]   J. Assfalg, M. Bertini, A. D. Bimbo, and P. Pala, "Content-Based Retrieval of 3-D Objects Using Spin Image Signatures," *IEEE Transactions on Multimedia,* vol. 9, pp. 589-599, 2007.

[25]   F. Roli, S. Vitulano, G. Antini, S. Berretti, A. Del Bimbo, and P. Pala, "Curvature Correlograms for Content Based Retrieval of 3D Objects," in *Image Analysis and Processing – ICIAP 2005*. vol. 3617: Springer Berlin / Heidelberg, 2005, pp. 859-866.

[26]   B. Alberto Del and P. Pietro, "Content-based retrieval of 3D models," *ACM Trans. Multimedia Comput. Commun. Appl.,* vol. 2, pp. 20-43, 2006.

[27]   E. Wahl, G. Hillenbrand, and G. Hirzinger, "Surflet-pair-relation histograms: a statistical 3D-shape representation for rapid classification," in *Fourth International Conference on 3-D Digital Imaging and Modeling, 3DIM* 2003, pp. 474-481.

[28]   R. Ohbuchi and T. Furuya, "Scale-weighted dense bag of visual features for 3D model retrieval from a partial view 3D model," in *IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)*, 2009, pp. 63-70.

[29]   H. Günter, L. Bastian, L. Paul, and S. Bernt, "3D Object Recognition from Range Images using Local Feature Histograms," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, pp. 394-394.

[30]   M. Ankerst, G. Kastenmüller, H. P. Kriegel, and T. Seidl, "3D Shape Histograms for Similarity Search and Classification in Spatial Databases," in *Proceedings of the 6th International Symposium on Advances in Spatial Databases*: Springer-Verlag, 1999.

[31]   J. W. H. Tangelder and R. C. Veltkamp, "Polyhedral model retrieval using weighted point sets," in *Shape Modeling International*, 2003, pp. 119-129.

[32]   L. Yi, W. Xu-Lei, W. Hua-Yan, Z. Hongbin, and Q. Hong, "Learning Robust Similarity Measures for 3D Partial Shape Retrieval," *Int. J. Comput. Vision,* vol. 89, pp. 408-431.

[33]   D. C. Nicu, M. F. Demirci, S. Deborah, S. Ali, J. D. Sven, and B. K. Paul, "3D Object Retrieval using Many-to-many Matching of Curve Skeletons," in *Proceedings of the International Conference on Shape Modeling and Applications 2005*: IEEE Computer Society, 2005.

[34]   R. R. Mauro, P. Giuseppe, S. Michela, and S. Dietmar, "Spectral-Driven Isometry-Invariant Matching of 3D Shapes," *Int. J. Comput. Vision,* vol. 89, pp. 248-265.

[35]   D. Mateus, R. Horaud, D. Knossow, F. Cuzzolin, and E. Boyer, "Articulated shape matching using Laplacian eigenfunctions and unsupervised point registration," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, 2008, pp. 1-8.

[36]   B. Amberg, R. Knothe, and T. Vetter, "SHREC'08 entry: Shape based face recognition with a Morphable Model," in *IEEE International Conference on Shape Modeling and Applications, SMI* 2008, pp. 253-254.

[37] S. Lee, S. Yoon, I. Yun, D. Kim, K. Lee, and S. Lee, "A New 3-D Model Retrieval System Based on Aspect-Transition Descriptor," in *Computer Vision – ECCV 2006*. vol. 3954: Springer Berlin / Heidelberg, 2006, pp. 543-554.

[38] M. F. Bayramoglu and A. O. Yilmaz, "A Hilbert space of probability mass functions and applications on the sum-product algorithm," in *5th International Symposium on Turbo Codes and Related Topics*, 2008, pp. 338-343.

[39] A. David and V. Sergei, "k-means++: the advantages of careful seeding," in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* New Orleans, Louisiana: Society for Industrial and Applied Mathematics, 2007.

[40] H. Permuter, J. Francos, and I. H. Jermyn, "Gaussian mixture models of texture and colour for image database retrieval," in *IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP)*, 2003, pp. III-569-72 vol.3.

[41] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, pp. 603-619, 2002.

[42] S. Jianbo and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, pp. 888-905, 2000.

[43] P. J. Besl and H. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 14, pp. 239-256, 1992.

[44] G. C. Sharp, S. W. Lee, and D. K. Wehe, "ICP registration using invariant features," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 24, pp. 90-102, 2002.

[45] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," in *The Third International Conference on 3-D Digital Imaging and Modeling*, 2001, pp. 145-152.

[46] "SHREC, Shape Retrieval Contest, http://www.aimatshape.net/event/SHREC", last accessed on 10/08/2011.

[47] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser, "The Princeton Shape Benchmark," in *Shape Modeling Applications*, 2004, pp. 167-178.

[48] K. Järvelin and J. Kekäläinen, "IR evaluation methods for retrieving highly relevant documents," in *23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Athens, Greece, 2000, pp. 41-48.

[49] B. K. P. Horn, "Extended Gaussian images," *Proceedings of the IEEE,* vol. 72, pp. 1671-1686, 1984.

[50] S. B. Kang and K. Ikeuchi, "The complex EGI: a new representation for 3-D pose determination," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 15, pp. 707-721, 1993.

[51] E. Paquet and M. Rioux, "Nefertiti: a query by content software for three-dimensional models databases management," in *International Conference on Recent Advances in 3-D Digital Imaging and Modeling*, 1997, pp. 345-352.

[52] E. Paquet, M. Rioux, A. Murching, T. Naveen, and T. Ali, "Description of shape information for 2-D and 3-D objects," *Signal Processing: Image Communication,* vol. 16, pp. 103-122, 2000.

[53] E. Paquet and M. Rioux, "The MPEG-7 standard and the content-based management of three-dimensional data: a case study," in *IEEE International Conference on Multimedia Computing and Systems*, 1999, pp. 375-380 vol.1.

[54] N. Marcin and K. Reinhard, "3D zernike descriptors for content based shape retrieval," in *Proceedings of the 8th ACM Symposium on Solid Modeling and Applications* Seattle, Washington, USA, 2003.

[55] S. Dietmar and V. V. Dejan, "3D Model Retrieval with Spherical Harmonics and Moments," in *Proceedings of the 23rd DAGM-Symposium on Pattern Recognition*: Springer-Verlag, 2001.

[56] F. A. Sadjadi and E. L. Hall, "Three-Dimensional Moment Invariants," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. PAMI-2, pp. 127-136, 1980.

[57] A. Mihael, K. Gabi, ller, K. Hans-Peter, and S. Thomas, "Nearest Neighbor Classification in 3D Protein Databases," in *Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology*: AAAI Press, 1999.

[58] A. E. Johnson and M. Hebert, "Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 21, pp. 433-449, 1999.

[59] A. Mian, "Representations and Matching Techniques for 3D Free-form Object and Face Recognition," in *School of Computer Science and Software Engineering*. vol. Doctor of Philosophy: The University of Western Australia, 2006, p. 268.

[60] A. Ceyhun Burak, S. Bülent, Y. Yücel, and S. Francis, "Density-based 3D shape descriptors," *EURASIP Journal on Advances in Signal Processing,* pp. 209-209, 2007.

[61] D.-Y. Chen, X.-P. Tian, Y.-T. Shen, and M. Ouhyoung, "On Visual Similarity Based 3D Model Retrieval," *Computer Graphics Forum,* vol. 22, pp. 223–232, 2003.

[62] P. Papadakis, I. Pratikakis, T. Theoharis, and S. Perantonis, "PANORAMA: A 3D Shape Descriptor Based on Panoramic Views for Unsupervised 3D Object Retrieval," *International Journal of Computer Vision,* vol. 89, pp. 177-192, 2010.

[63] L. Zhouhui, G. Afzal, and S. Xianfang, "Visual Similarity Based 3D Shape Retrieval Using Bag-of-Features," in *Proceedings of the 2010 Shape Modeling International Conference*: IEEE Computer Society.

[64] P. Daras and A. Axenopoulos, "A Compact Multi-view Descriptor for 3D Object Retrieval," in *Seventh International Workshop on Content-Based Multimedia Indexing, CBMI*, 2009, pp. 115-119.

[65] H. Masaki, S. Yoshihisa, K. Taku, and L. K. Tosiyasu, "Topology matching for fully automatic similarity estimation of 3D shapes," in *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*: ACM, 2001.

[66] T. Zaharia and F. Preteux, "3D shape-based retrieval within the MPEG-7 framework," in *Nonlinear Image Processing and Pattern Analysis*, 2001, pp. 133-145.

[67] J. J. Koenderink, *Solid shape*: MIT Press, 1990.

[68] D. V. Vranic and D. Saupe, "3D Shape Descriptor Based on 3D Fourier Transform," in *Proceedings of the EURASIP Conference on Digital Signal Processing for Multimedia Communications and Services (ECMCS)* 2001, pp. 271-274.

[69] S. H. Jiao and R. Goutte, "Invariant descriptor of Fourier-Mellin for indexing by contents a database of three dimensional images," in *Signal Processing, 2002 6th International Conference on*, 2002, pp. 858-860 vol.1.

[70] S. H. Jiao and R. Goutte, "Invariant descriptor of Fourier-Mellin for indexing by contents a database of three dimensional images," in *6th International Conference on Signal Processing*, 2002, pp. 858-860 vol.1.

[71] F. Stein and G. Medioni, "Structural indexing: efficient 3-D object recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 14, pp. 125-145, 1992.

[72] C. Chin Seng and J. Ray, "Point Signatures: A New Representation for 3D Object Recognition," *International Journal of Computer Vision,* vol. 25, pp. 63-85, 1997.

[73]   A. Frome, D. Huber, R. Kolluri, T. Bülow, and J. Malik, "Recognizing Objects in Range Data Using Regional Point Descriptors," in *European Conference on Computer Vision (ECCV)*, 2004, pp. 224-237.

[74]   G. Ran and C.-O. Daniel, "Salient geometric features for partial shape matching and similarity," *ACM Trans. Graph.*, vol. 25, pp. 130-150, 2006.

[75]   P. Shilane and T. Funkhouser, "Distinctive regions of 3D surfaces," *ACM Trans. Graph.*, vol. 26, p. 7, 2007.

[76]   H. Chen and B. Bhanu, "3D free-form object recognition in range images using local surface patches," *Pattern Recognition Letters.*, vol. 28, pp. 1252-1262, 2007.

[77]   A. Flint, A. Dick, and A. Hengel, "Thrift: Local 3D Structure Recognition," in *9th Biennial Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications*, 2007, pp. 182-188.

[78]   S. Shalom, L. Shapira, A. Shamir, and D. Cohen-or, "Part Analogies in Sets of Objects," Eurographics Workshop on 3D Object Retrieval, 3DOR, Crete, Greece, 2008, pp. 33-40.

[79]   S. Lior, S. Ariel, and C.-O. Daniel, "Consistent mesh partitioning and skeletonisation using the shape diameter function," *The Visual Computer: International Journal of Computer Graphics* vol. 24, pp. 249-259, 2008.

[80]   R.-C. Salvador, G. S. Linda, and M. Marina, "A New Paradigm for Recognizing 3-D Object Shapes from Range Data," in *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*: IEEE Computer Society, 2003.

[81]   M. Pauly , R. Keiser , and M. Gross, "Multi-scale Feature Extraction on Point-Sampled Surface," in *In Proceedings of Eurographics*, 2003.

[82]   R. Gal, A. Shamir, and D. Cohen-Or, "Pose-Oblivious Shape Signature," *IEEE Transactions on Visualization and Computer Graphics,* vol. 13, pp. 261-271, 2007.

[83]   B. Matei, Y. Shan, H. S. Sawhney, Y. Tan, R. Kumar, D. Huber, and M. Hebert, "Rapid Object Indexing Using Locality Sensitive Hashing and Joint 3D-Signature Space Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28, pp. 1111-1126, 2006.

[84]   L. Xinju and I. Guskov, "3D object recognition from range images using pyramid matching," in *IEEE 11th International Conference on Computer Vision, ICCV* 2007, pp. 1-6.

[85]   H. Huy Tho and D. Gibbins, "Multi-scale feature extraction for 3d surface registration using local shape variation," in *23rd International Conference Image and Vision Computing New Zealand, IVCNZ* 2008, pp. 1-6.

[86]   U. Castellani, M. Cristani, S. Fantoni, and V. Murino, "Sparse points matching by combining 3D mesh saliency with statistical descriptors," *Computer Graphics Forum,* vol. 27, pp. 643-652, 2008.

[87]   M. Dong and Y. Chen, "Salient region detection and feature extraction in 3D visual data," *International Conference on Image Processing,* pp. 185-188, 2008.

[88]   T.-W. R. Lo and J. P. Siebert, "Local feature extraction and matching on range images: 2.5D SIFT," *Computer Vision and Image Understanding,* vol. 113, pp. 1235-1250, 2009.

[89]   A. Mian, M. Bennamoun, and R. Owens, "On the Repeatability and Quality of Keypoints for Local Feature-based 3D Object Retrieval from Cluttered Scenes," *International Journal of Computer Vision,* vol. 89, pp. 348-361.

[90]   R. Benlamri, "Range image segmentation of scenes with occluded curved objects," *Pattern Recognition Letters,* vol. 21, pp. 1051-1060, 2000.

[91]   CSEM, "SR3000 Users Manual," October 2006.

[92]   S. Á. Guðmundsson, "Robot Vision Applications using the CSEM SwissRanger Camera," Informatics and Mathematical Modelling, Technical University of Denmark,DTU, 2006.

[93]   "Camera Calibration Toolbox for Matlab, http://www.vision.caltech.edu/ bouguetj/calib_doc/," Last accessed on 01/11/2008.

[94]   P. Hough, "Method and Means for Recognizing Complex Patterns," United States Patent Office, 1962.

[95]   T. Zaharia and F. J. Preteux, "3D-shape-based retrieval within the MPEG-7 framework," in *Proc. SPIE Conference 4304 on Nonlinear Image Processing and Pattern Analysis XII*, San Jose, CA, 2001, pp. 133-145.

[96]   T. Zaharia and F. Preteux, "Shape-based retrieval of 3D mesh models," in *IEEE International Conference on Multimedia and Expo, ICME*, 2002, pp. 437-440 vol.1.

[97]   M. Hebel and U. Stilla, "Automatic Registration of Laser Point Clouds of Urban Areas," in *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* Munich, Germany, 2007.

[98]   K. Okada, S. Kagami, M. Inaba, and H. Inoue, "Plane Segment Finder: Algorithm, Implementation and Applications," *International Conference on Robotics and Automation*, pp. 2120-2125, 2001.

[99]   S. Augusto and T. Stefano, "Detection and characterisation of planar fractures using a 3D Hough transform," *Elsevier, Signal Processing*, vol. 82, pp. 1269-1282, 2002.

[100]   R. Roncella and G. Forlani, "Extraction of planar patches from point clouds to retrieve dip and dip direction of rock discontinuities," in *ISPRS Workshop Laser scanning*, Volume XXXVI, Netherlands, 2005.

[101]   H. Cantzler, B. F. Robert, and D. Michel, "Quality Enhancement of Reconstructed 3D Models Using Coplanarity and Constraints," in *Proceedings of the 24th DAGM Symposium on Pattern Recognition*: Springer-Verlag, 2002.

[102]   N. Andreas, chter, and H. Joachim, "Towards semantic maps for mobile robots," *Robot. Auton. Syst.*, vol. 56, pp. 915-926, 2008.

[103]   R. Schnabel, R. Wahl, and R. Klein, "Efficient RANSAC for Point-Cloud Shape Detection," *Computer Graphics Forum*, vol. 26, pp. 214-226, 2007.

[104]   T. Emanuele and V. Alessandro, *Introductory Techniques for 3-D Computer Vision*: Prentice Hall, 1998.

[105]   X. Cao and N. Shrikhande, "Quadric surface fitting for sparse range data," in *IEEE International Conference on Systems, Man, and Cybernetics*, 1991, pp. 123-128.

[106]   S. Flöry, "Fitting curves and surfaces to point clouds in the presence of obstacles," *Computer Aided Geometric Design*, vol. 26, pp. 192-202, 2009.

[107]   A. Gruen and D. Akca, "Least squares 3D surface and curve matching," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 59, pp. 151-174, 2005.

[108]   S. Ariel, S. Lior, and C.-O. Daniel, "Mesh analysis using geodesic mean-shift," *Visual Computing*, vol. 22, pp. 99-108, 2006.

[109]   T. Rabbani, "Automatic Reconstruction of Industrial Installations Using Point Clouds and Images." vol. PhD: Delft University of Technology, 2006.

[110]   H. Hugues, D. Tony, D. Tom, M. John, and S. Werner, "Surface reconstruction from unorganized points," *SIGGRAPH Comput. Graph.*, vol. 26, pp. 71-78, 1992.

[111]   M. Pauly , M. Gross, and L. P. Kobbelt, "Efficient simplification of point-sampled surfaces," in *Proceedings of the conference on Visualization '02* Boston, Massachusetts: IEEE Computer Society, 2002.

[112] T. K. Dey, G. Li, and J. Sun, "Normal estimation for point clouds: a comparison study for a Voronoi based method," in *Point-Based Graphics, 2005. Eurographics/IEEE VGTC Symposium Proceedings*, 2005, pp. 39-46.

[113] O. Schall, A. Belyaev, and H. P. Seidel, "Robust filtering of noisy scattered point data," in *Point-Based Graphics, 2005. Eurographics/IEEE VGTC Symposium Proceedings*, 2005, pp. 71-144.

[114] W. E. Snyder and A. Cowart, "An Iterative Approach to Region Growing Using Associative Memories," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. PAMI-5, pp. 349-352, 1983.

[115] "Expectation Maximization, http://opencv.willowgarage.com/wiki/ MachineLearning#Expectation-Maximization," , last accessed on 01/07/2009.

[116] J. A. Bilmes, "A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models," University of California at Berkeley 1998.

[117] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics,* vol. 7, pp. 179-188, 1936.

[118] O. D. Richard, E. H. Peter, and G. S. David, *Pattern Classification (2nd Edition)*: Wiley-Interscience, 2000.

[119] B. Bustos, D. Keim, D. Saupe, T. Schreck, and D. Vranic, "Feature-based similarity search in 3D object databases," *ACM Comput. Surv.,* vol. 37, pp. 345-387, 2005.

[120] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, "Rotation invariant spherical harmonic representation of 3D shape descriptors," in *Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing* Aachen, Germany: Eurographics Association, 2003.

[121] F. Thomas, M. Patrick, K. Michael, C. Joyce, H. Alex, D. David, and J. David, "A search engine for 3D models," *ACM Trans. Graph.,* vol. 22, pp. 83-105, 2003.

[122] A. Mademlis, P. Daras, D. Tzovaras, and M. G. Strintzis, "3D object retrieval using the 3D shape impact descriptor," *Pattern Recognition,* vol. 42, pp. 2447-2459, 2009.

[123] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision,* vol. 60, pp. 91-110, 2004.

[124] G. L. David, "Object Recognition from Local Scale-Invariant Features," in *Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2*: IEEE Computer Society, 1999.

[125] M. Novotni and R. Klein, "3D zernike descriptors for content based shape retrieval," in *Proc. of the 8th ACM symp. on Solid Modeling and Applications* Seattle, Washington, USA: ACM, 2003.

[126] D. V. Vranic, D. Saupe, and J. Richter, "Tools for 3D-object retrieval: Karhunen-Loeve transform and spherical harmonics," in *IEEE Fourth Workshop on Multimedia Signal Processing*, 2001, pp. 293-298.

[127] D. V. Vranic, "An improvement of rotation invariant 3D-shape based on functions on concentric spheres," in *International Conference on Image Processing, ICIP*, 2003, pp. III-757-60 vol.2.

[128] S. Jian, O. Maks, and G. Leonidas, "A concise and provably informative multi-scale signature based on heat diffusion," in *Proceedings of the Symposium on Geometry Processing* Berlin, Germany: Eurographics Association, 2009.

[129] K. Gębal, J. A. Bærentzen, H. Aanæs, and R. Larsen, "Shape Analysis Using the Auto Diffusion Function," *Computer Graphics Forum,* vol. 28, pp. 1405-1413, 2009.

[130] S. Biasotti, D. Giorgi, M. Spagnuolo, and B. Falcidieno, "Reeb graphs for shape analysis and applications," *Theoretical Computer Science,* vol. 392, pp. 5-22, 2008.

[131]   H. Sundar, D. Silver, N. Gagvani, and S. Dickinson, "Skeleton based shape matching and retrieval," in *Shape Modeling International*, 2003, pp. 130-139.

[132]   N. D. Cornea, D. Silver, X. Yuan, and R. Balasubramanian, "Computing hierarchical curve-skeletons of 3D objects," *The Visual Computer,* vol. 21, pp. 945-955, 2005.

[133]   P. Daras and A. Axenopoulos, "A Compact Multi-view Descriptor for 3D Object Retrieval," in *Content-Based Multimedia Indexing, 2009. CBMI '09. Seventh International Workshop on*, 2009, pp. 115-119.

[134]   R. Ohbuchi, K. Osada, T. Furuya, and T. Banno, "Salient local visual features for shape-based 3D model retrieval," in *IEEE International Conference on Shape Modeling and Applications, SMI* 2008, pp. 93-102.

[135]   B. Bustos, D. Keim, D. Saupe, and T. Schreck, "Content-Based 3D Object Retrieval," *IEEE Computer Graphics and Applications,* vol. 27, pp. 22-27, 2007.

[136]   A. M. Bronstein, M. M. Bronstein, L. J. Guibas, and M. Ovsjanikov, "Shape google: Geometric words and expressions for invariant shape retrieval," *ACM Trans. Graph.,* vol. 30, pp. 1-20, 2011.

[137]   G. Stavropoulos, P. Moschonas, K. Moustakas, D. Tzovaras, and M. G. Strintzis, "3-D Model Search and Retrieval From Range Images Using Salient Features," *Multimedia, IEEE Transactions on,* vol. 12, pp. 692-704.

[138]   T. K. Dey, K. Li, C. Luo, P. Ranjan, I. Safa, and Y. Wang, "Persistent Heat Signature for Pose-oblivious Matching of Incomplete Models," *Computer Graphics Forum,* vol. 29, pp. 1545-1554, 2010.

[139]   D. M. Healy, D. N. Rockmore, P. J. Kostelec, and S. Moore, "FFTs for the 2-Sphere-Improvements and Variations," *Journal of Fourier Analysis and Applications,* vol. 9, pp. 341-385, 2003.

[140]   P. Mark, G. Markus, and P. K. Leif, "Efficient simplification of point-sampled surfaces," in *Proceedings of the conference on Visualization '02* Boston, Massachusetts: IEEE Computer Society, 2002.

[141]   M. M. Bronstein and I. Kokkinos, "Scale-invariant heat kernel signatures for non-rigid shape recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),*, 2010 pp. 1704-1711.

[142]   R. Dan, M. B. Michael, M. B. Alexander, and K. Ron, "Volumetric heat kernel signatures," in *Proceedings of the ACM workshop on 3D object retrieval* Firenze, Italy: ACM.

[143]   V. Amir, B.-C. Mirela, and G. Craig, "A multi-resolution approach to heat kernels on discrete surfaces," *ACM Trans. Graph.,* vol. 29, pp. 1-10, 2010.

[144]   A. Sharma and R. Horaud, "Shape matching based on diffusion embedding and on mutual isometric consistency," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2010, pp. 29-36.

[145]   M. Ovsjanikov, Q. Mérigot, F. Mémoli, and L. Guibas, "One Point Isometric Matching with the Heat Kernel," *Computer Graphics Forum,* vol. 29, pp. 1555-1564.

[146]   U. Castellani, A. Ulas, V. Murino, M. Bellani, G. Rambaldelli, M. Tansella, and P. Brambilla, "Selecting Scales by Multiple Kernel Learning for Shape Diffusion Analysis," in *Proceedings of the Third International Workshop on Mathematical Foundations of Computational Anatomy - Geometrical and Statistical Methods for Modelling Biological Shape Variability*, pp. 148-158.

[147]   M. R. Raif, "Laplace-Beltrami eigenfunctions for deformation invariant shape representation," in *Proceedings of the fifth Eurographics symposium on Geometry processing* Barcelona, Spain: Eurographics Association, 2007.

[148] B. Mikhail, S. Jian, and W. Yusu, "Discrete laplace operator on meshed surfaces," in *Proceedings of the 24th Annual Symposium on Computational Geometry*, College Park, MD, USA: ACM, 2008.

[149] R. Martin, W. Franz-Erich, and P. Niklas, "Laplace-Beltrami spectra as 'Shape-DNA' of surfaces and solids," *Comput. Aided Des.*, vol. 38, pp. 342-366, 2006.

[150] K. Polthier and U. Pinkall, "Computing Discrete Minimal Surfaces and Their Conjugates," *Experimental Mathematics*, vol. 2, pp. 15-36, 1993.

[151] A. Marc and W. Max, "Discrete Laplacians on general polygonal meshes," *ACM Trans. Graph.*, vol. 30, pp. 1-10.

[152] T. Iijima, "Basic theory of pattern observation," *Papers of Technical Group on Automata and Automatic Control, IECE, Japan (in Japanese)*, 1959.

[153] J. Weickert, S. Ishikawa, and A. Imiya, "Linear Scale-Space has First been Proposed in Japan," *Journal of Mathematical Imaging and Vision,* vol. 10, pp. 237-252, 1999.

[154] P. W. Andrew, "Scale-space filtering," in *Proceedings of the Eighth international joint conference on Artificial intelligence - Volume 2* Karlsruhe, West Germany: Morgan Kaufmann Publishers Inc., 1983.

[155] J. J. Koenderink, "The structure of images," *Biological Cybernetics,* vol. 50, pp. 363-370, 1984.

[156] C. Anderson, T. Ralph, and V. Luiz, "Discrete Scale Spaces via Heat Equation," in *Proceedings of the XIV Brazilian Symposium on Computer Graphics and Image Processing*: IEEE Computer Society, 2001.

[157] M. Ovsjanikov, A. M. Bronstein, M. M. Bronstein, and L. J. Guibas, "Shape Google: a computer vision approach to isometry invariant shape retrieval," in *IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)*, 2009, pp. 320-327.

[158] R. Boulet, "Disjoint unions of complete graphs characterized by their Laplacian spectrum," *Electronic Journal of Linear Algebra,* pp. pp. 773-783., December 2009.

[159] "Geometric & Intelligent Computing Laboratory, Drexel University, http://www.designrepository.org/,", last accessed on 05/12/2008.

[160] "3D CAD Browser 3D Models, http://www.3dcadbrowser.com/", last accessed on 05/12/2008.

[161] "McGill University 3D Shape Benchmark, http://www.cim.mcgill.ca/~shape/benchMark/," Last accessed on 05/12/2008.

[162] "Stanford Computer Graphics Laboratory data archives, http://graphics.stanford.edu/data", last accessed on 05/12/2008.

[163] "3D Models, Plugins, Textures, and more at Turbo Squid, http://www.turbosquid.com/", last accessed on 05/12/ 2008.

[164] "3DXtras.com - Download Absolutely Free 3D Models, http://www.3dxtras.com/," , last accessed on 05/12/2008.

[165] H. Richard and Z. Andrew, *Multiple view geometry in computer vision*: Cambridge University Press, 2000.

[166] Y. A. Çengel, *Heat transfer: A Practical Approach*: WBC McGraw-Hill, 1998.

APPENDIX

EXTRA INFORMATION

A.1 3D Shape Databases and Contests

Sources of 3D models for retrieval people are research groups of various universities and commercial repositories. Princeton Shape Retrieval and Analysis Group provide 1184 mesh models for scientific usage which are used throughout the thesis proposal. They also have an online 3D search engine, Princeton Shape Benchmark (PSB), and aiming to provide standard models and evaluation tools for shape retrieval [47]. PSB supplying models freely with four classifications which divides the models into a training set and test set with fine grained classes. The base classification spans a large variety of classes including animals, plants, airplanes, furniture, and vehicles. Other classifications provides coarser version of the base classification. Similarly, National Taiwan university computer department research group supply a public large database with 10910 3D .obj file [61]. However, a small number of them are classified.

National Design Repository, Geometric and Intelligent Computing Laboratory at Drexel University [159] supplies 3D CAD and solid models of (mostly) mechanical/machined engineering designs. 3D CAD Browser [160] provides over 7600 3D models mostly CAD models in 3ds, dxf, dwg, obj, lwo, iges formats. These are not classified, but public databases.

AIM@SHAPE repository [46], McGill 3D Shape Benchmark [161], Stanford Computer Graphics Laboratory data archives [162], Turbo Squid [163] which is a commercial database

of 3dsmax shape objects, and 3dxtras [164], are the leading 3D model suppliers. Almost all of these repositories supply watertight models, some of them offer range data and 2D samples.

AIM@SHAPE organizing a 3D shape retrieval evaluation event called SHREC since 2006. There were seven tracks in 2007, namely watertight models, partial matching, Protein models, CAD models, Relevance feedback, Similarity measures, and 3D face models. SHREC is a scientific contest and works that are submitted are evaluated against each other within each track.

A.2 Surface Fitting

Considering the piecewise continuous nature of the surfaces appearing in the typical scenes, we fit quadratic surfaces to local patches. A quadric surface equation can be written by F(x,y,z)=0 where implicit function F(x,y,z) is equal to:

$$ax^2 + by^2 + cz^2 + 2exy + 2fyz + 2gxz + 2lx + 2my + 2nz + d = 0 \qquad \text{A.1}$$

Each point $(x_i \in S, i = 1,2,...,n)$ in the local patch should satisfy the following equation in the noise free case:

$$\underbrace{\begin{bmatrix} x_1^2 & y_1^2 & z_1^2 & x_1y_1 & y_1z_1 & x_1z_1 & x_1 & y_1 & z_1 & 1 \\ x_2^2 & y_2^2 & z_2^2 & x_2y_2 & y_2z_2 & x_2z_2 & x_2 & y_2 & z_2 & 1 \\ & & & & \vdots & & & & & \\ x_n^2 & y_n^2 & z_n^2 & x_ny_n & y_nz_n & x_nz_n & x_n & y_n & z_n & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a \\ b \\ c \\ 2e \\ 2f \\ 2g \\ 2l \\ 2m \\ 2n \\ d \end{bmatrix}}_{v} = 0 \qquad \text{A.2}$$

$$Av = 0$$

However, as evident not all the input points will lie on the surface, implicit function will not vanish and result in some error F(x,y,z)=ε for those points. Hence, for fitting the "best" surface on the local patch, a minimization should be performed on F(x,y,z) such that $\| \sum_{x,y,z} F(x,y,z) \|^2$ is minimum. The trivial solution of this minimization problem is *a=b=c=d=e=....=n=0*. Therefore, a constraint on the polynomial coefficients can be introduced such that

$$a^2 + b^2 + c^2 + d^2 + e^2 + f^2 + g^2 + l^2 + m^2 + n^2 = 1 \qquad \text{A.3}$$

Now, the minimization problem can be solved using least squares subject to $v^T v = 1$.

$$\min(\| F(x,y,z) \|^2) = \min(\| Av \|^2)$$
$$= \min((Av)^T(Av))$$
$$= \min(v^T A^T A\, v) \qquad subject\ to\ v^T v = 1$$

Remembering the method of Lagrange multiplier in optimization, this equation states that the solution is the eigenvector corresponding to the smallest eigenvalue of the $A^T A$ [104], i.e.

$$v \simeq eigenvector(\lambda_0), where\ \lambda_0 \leq \lambda_1 \dots \leq \lambda_9 \qquad \text{A.4}$$

Before evaluating the surface parameters normalization on the input points should be performed. This step is necessary due to magnitude differences between the 3D point coordinates. Consider the row of the matrix A in Equation A.2, let (x,y,z)=(100,100,1) then A becomes a ill-conditioned matrix and we cannot trust the solution [4]. Hence, we have to normalize the coordinates of the input points such that their order of magnitudes is similar [165].

The normalization step is as follows:

- Translate points such that their mean after the translation is equal to zero
  $\bar{X} = X - mean(X)$, $(\bar{Y} = Y - mean(Y)$, $\bar{Z} = Z - mean(Z)$,

- Scale the points such that their variance is equal to one
  $\tilde{X} = \frac{\bar{X}}{std(\bar{X})}$, $\tilde{Y} = \frac{\bar{Y}}{std(\bar{Y})}$, $\tilde{Z} = \frac{\bar{Z}}{std(\bar{Z})}$, std: standard deviation

Above operations can be expressed by matrix operations:

$$\begin{bmatrix} \tilde{X} \\ \tilde{Y} \\ \tilde{Z} \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1}{std(\bar{X})} & 0 & 0 & -\frac{mean(X)}{std(\bar{X})} \\ 0 & \frac{1}{std(\bar{Y})} & 0 & -\frac{mean(Y)}{std(\bar{Y})} \\ 0 & 0 & \frac{1}{std(\bar{Z})} & -\frac{mean(Z)}{std(\bar{Z})} \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{T} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \qquad \text{A.5}$$

Then, if we rewrite the previous equations in the following form:

$$\underbrace{\underbrace{[x\ \ y\ \ z\ \ 1]}_{P^T}\ T^T}_{\tilde{P}^T}\ \underbrace{T^{-T}\underbrace{\begin{bmatrix} a & e & g & l \\ e & b & f & m \\ g & f & c & n \\ l & m & n & d \end{bmatrix}}_{V}T^{-1}}_{V'}\ T\underbrace{\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}}_{\underset{\tilde{P}}{P}} = 0$$

$$V = T^T V' T$$

We can recover the real quadric patch coefficients as in Equation A.6. Now, we can use the points lying on the surface, instead of using the original point cloud. This goal can be achieved by projecting the input points on the surface, but the procedure is complex. Instead

---

[4] http://numericalmethods.eng.usf.edu

of projection, we preserve the x and y coordinates of the input points and obtain z coordinates from the surface equation. When we put x and y values in the Equation A.1 we obtain a second order equation in the form of:

$$Az^2 + Bz + C = 0$$

We solve the equation easily and obtain two roots such that: $z_{1,2} = \frac{-B \mp \Delta}{2A}$ where $\Delta = B^2 - 4AC$. We are expecting that solution of this equation is real. However, in most of the cases it might be complex. This can be explained by an example in 2D:

Let the function to be estimated be $y^2 - x = 0$. The estimation includes five parameters {a, b, c, d, e} which are the coefficients of the following equation $ay^2 + bx^2 + cx + dy + e = 0$. Let the estimated parameters be {a, b, c, d, e} = {1, 0,-1, 0, -2}, due to errors, that is $y^2 - x = 2$. Now, we want to calculate y coordinates by putting x coordinates into the equation. If we put x=1 in the original equation, we can obtain two real roots y=1, and y=-1; however, if we put the same x in the estimated equation, we obtain complex conjugate y values (Figure A-1). This is due to the estimation errors, which cannot be avoided. Hence, we decided to use the norm of the complex solution as the z coordinate in our original problem in 3D.
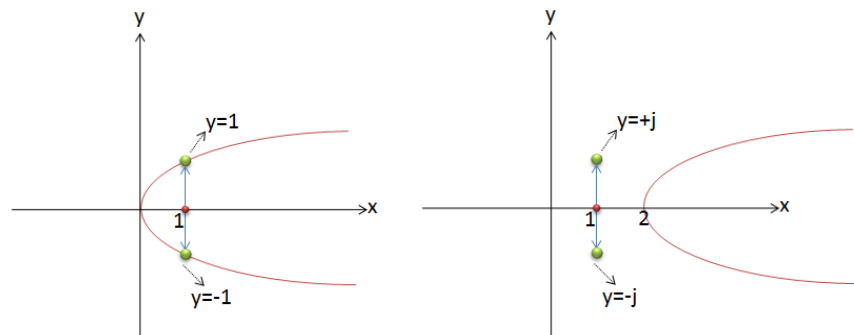

Figure A-1. Projections of the scene points on the estimated surface is not unique, since estimated surface may not be defined on the given coordinates. In our case, z is not defined on the given (x, y) pair.

This solution can only be used for recovering smooth surfaces, but if we want to analyze the local surface properties, points lying on the estimated surfaces should be estimated properly. By experiencing the difficulties in this approach, we focuse on another model for fitting surfaces to 3D point clouds. Instead of F(x,y,z)=0 equation, we used z=f(x,y) for surface fitting. Idea is similar to the previous one; the explicit equation is:

$$ax^2 + by^2 + cxy + dx + ey + f = z \qquad \text{A.7}$$

Each point $(x_i \in S, i = 1,2,\dots,n)$ in the local patch should satisfy the following equation in the ideal case:

$$\underbrace{\begin{bmatrix} x_1^2 & y_1^2 & x_1y_1 & x_1 & y_1 & 1 \\ x_2^2 & y_2^2 & x_2y_2 & x_2 & y_2 & 1 \\ & & \vdots & & & \\ x_n^2 & y_n^2 & x_ny_n & x_n & y_n & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a \\ b \\ c \\ d \\ e \\ f \end{bmatrix}}_{v} = \underbrace{\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}}_{Z} \qquad \text{A.8}$$

$$Av = Z$$

The optimal solution in the least squares sense is given by the following equation [104]:

$$v = (A^T A)^{-1} A^T z \qquad \text{A.9}$$

In [104], it is suggested to use pseudoinverse of $(A^T A)$. However, since we cannot apply the normalization process that is used in the previous fitting problem, the solution is noise sensitive. The only normalization is using zero mean points.

A.3   Heat Diffusion

The main reference in this section is [166].

Definitions:

- Specific heat (C) is defined as the energy required to raise the temperature of a unit mass of a substance by one degree.
- Let the amount of heat transferred during the process is denoted by Q.
- The amount of heat transferred per unit time is called heat transfer rate, and is denoted by $\dot{Q}$.
- The rate of heat transfer per unit area normal to the direction of heat transfer is called heat flux $\dot{q} = \frac{\dot{Q}}{A}$
- The heat equation is derived from ***Fourier's law :*** *the flow rate of heat energy through a surface is proportional to the negative temperature gradient across the surface*

$$Rate\ of\ heat\ conduction \ \propto \frac{(Area)(Temperature\ Difference)}{Thickness} \qquad \text{A.10}$$

$$\dot{Q}_{cond} = kA\frac{(T_1 - T_2)}{\Delta x} = -kA\frac{\Delta T}{\Delta x} \qquad \text{A.11}$$

in the limit

124

$$\dot{Q}_{cond} = -kA\frac{dT}{dx},$$

A.12

where k is the thermal conductivity and A is the area (see Figure A-2). Heat conduction direction is proportional to the temperature gradient in that direction.
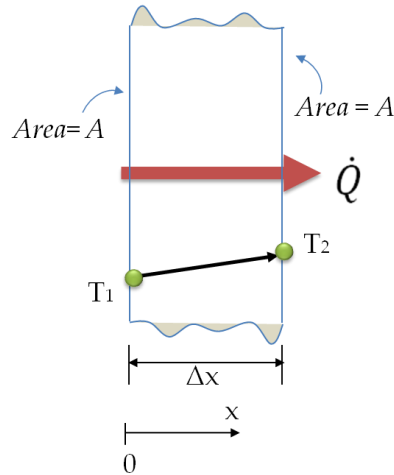

Figure A-2. Heat transfer parameters

A.3.1 The First Law Of Thermodynamics :

The net change (increase or decrease) in the total energy of the system during a process is equal to the difference between the total energy entering and the total energy leaving the system during that process.

$$\underbrace{E_{in} - E_{out}}_{\substack{Net\ Energy \\ Transfer}} = \underbrace{\Delta E_{system}}_{\substack{Change\ in\ internal \\ energy}}$$

A.13

If the time derivative is taken then the above equation becomes:

$$\underbrace{\dot{E}_{in} - \dot{E}_{out}}_{\substack{Rate\ of\ net\ Energy \\ Transfer}} = \underbrace{\frac{dE_{system}}{dt}}_{\substack{Rate\ of\ change\ in \\ internal\ energy}}$$

A.14

A.3.2 One Dimensional Heat Conduction
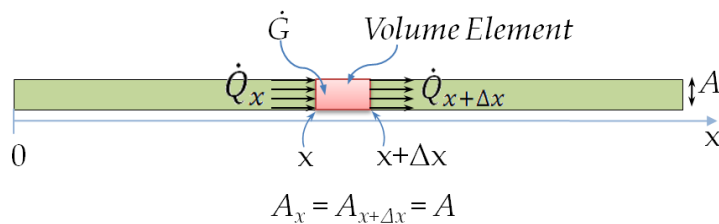

Figure A-3. Heat transfer parameters in one dimension.

125

First law of thermodynamics:

$$\begin{pmatrix} Rate\ of\ heat \\ conduction \\ at\ x \end{pmatrix} - \begin{pmatrix} Rate\ of\ heat \\ conduction \\ at\ x + \Delta x \end{pmatrix} + \begin{pmatrix} Rate\ of\ heat \\ generation\ inside \\ the\ element \end{pmatrix} = \begin{pmatrix} Rate\ of\ change\ of \\ the\ energy\ content \\ of\ the\ element \end{pmatrix}$$

$$\dot{Q}_x - \dot{Q}_{x+\Delta x} + \dot{G}_{element} = \frac{\Delta Element}{\Delta t} \tag{A.15}$$

$$\Delta Element = E_{t+\Delta t} - E_t = mC(T_{t+\Delta t} - T_t) = \rho A \Delta x C(T_{t+\Delta t} - T_t) \tag{A.16}$$

Assume no heat source inside, that is $\dot{G}_{element} = 0$.

$$\dot{Q}_x - \dot{Q}_{x+\Delta x} = \rho A \Delta x C \frac{(T_{t+\Delta t} - T_t)}{\Delta t} \tag{A.17}$$

Divide both sides of Equation 6. 7 by $\Delta x A$

$$-\frac{1}{A}\frac{\dot{Q}_{x+\Delta x} - \dot{Q}_x}{\Delta x} = \rho C \frac{(T_{t+\Delta t} - T_t)}{\Delta t} \tag{A.18}$$

Taking the limit as $\Delta x \to 0$ and $\Delta t \to 0$

$$\lim_{\Delta x \to 0} \frac{\dot{Q}_{x+\Delta} - \dot{Q}_x}{\Delta x} = \frac{\partial \dot{Q}}{\partial x} = \frac{\partial}{\partial x}\left(-kA\frac{dT}{dx}\right) \tag{A.19}$$

Combining Equation A.18 and Equation A.19

$$-\frac{1}{A}\frac{\partial}{\partial x}\left(-kA\frac{\partial T}{\partial x}\right) = \rho C \frac{\partial T}{\partial t} \tag{A.20}$$

Assuming constant k (thermal conductivity)

$$\frac{\partial^2 T}{\partial x^2} = \frac{1}{\alpha}\frac{\partial T}{\partial t} = 0 \tag{A.21}$$

This result can be imported to the 3D heat equation as follows:

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} - \frac{1}{\alpha}\frac{\partial T}{\partial t} = 0 \tag{A.22}$$
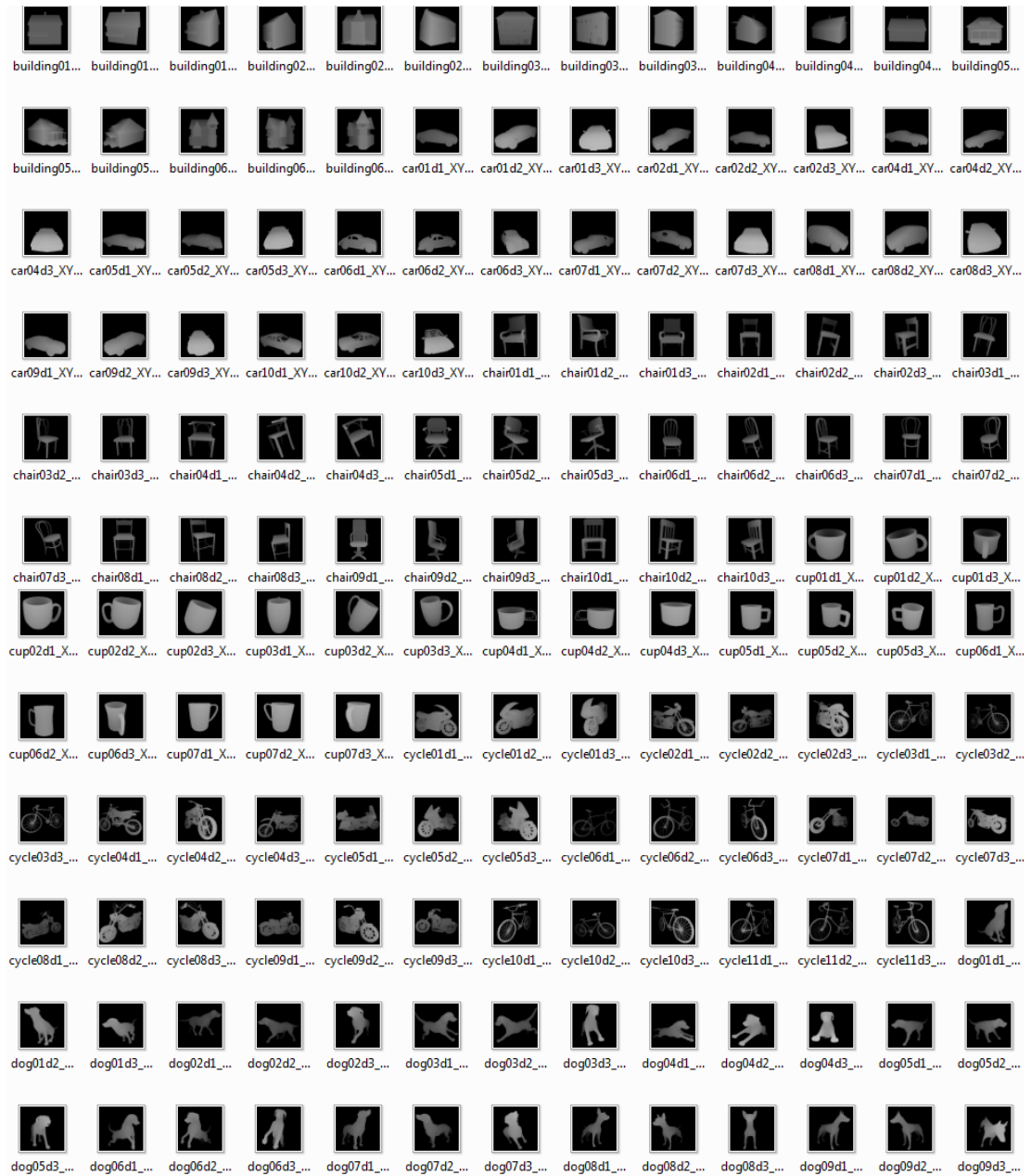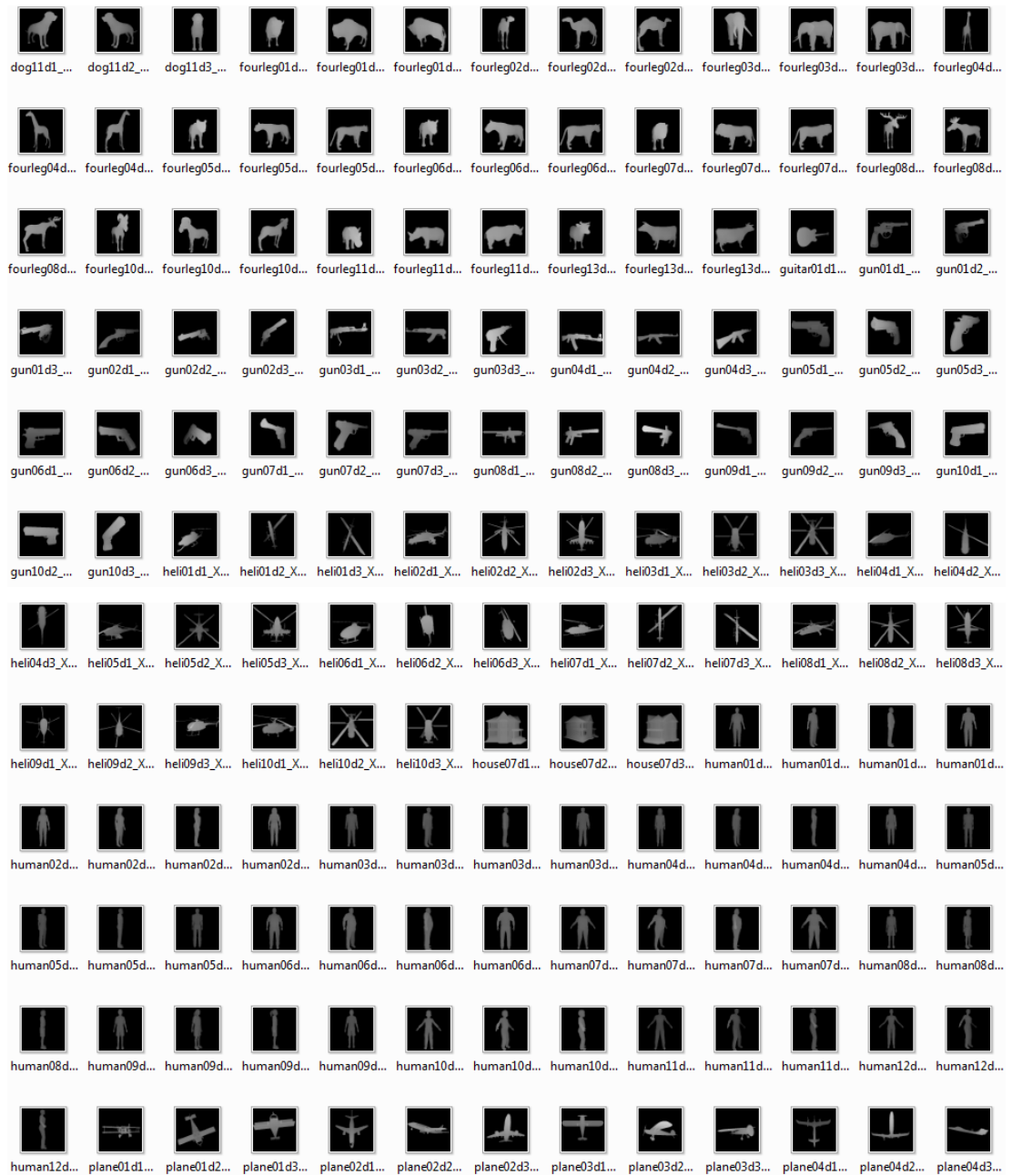
A.4 Database

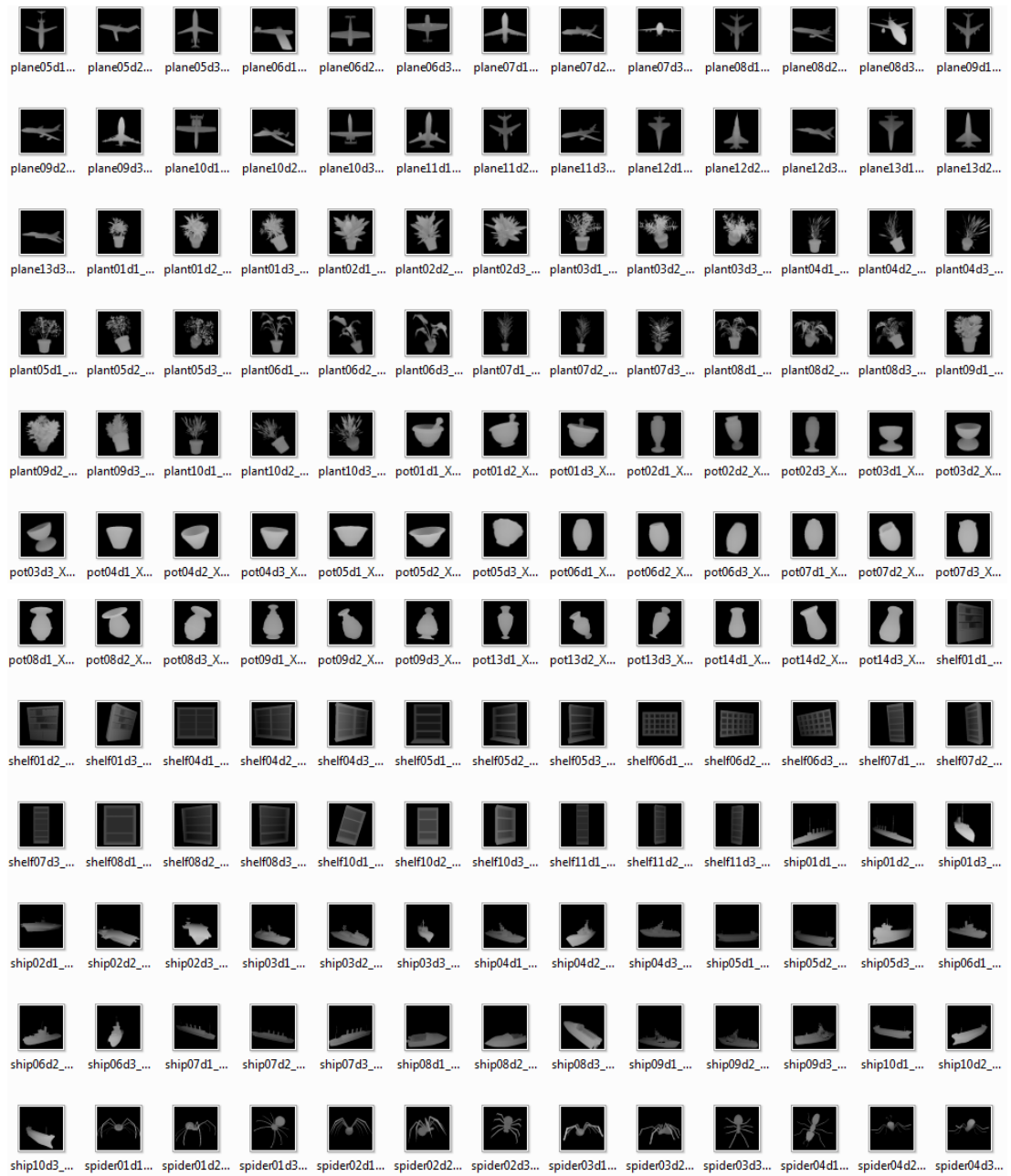
Figure A-4. Database part 1

Figure A-5. Database part 2

Figure A-6. Database part 3

Figure A-7. Database part 4

VITA

Personal Information
Place and Date of Birth: Ankara, Turkey | 1980
Web Address: www.eee.metu.edu.tr/~neslihan
e-mail: neslihan@eee.metu.edu.tr

Education
September 2011 (Expected) Doctor of Philosophy in Electrical and Electronics Engineering
Middle East Technical University, Ankara, Turkey
Thesis Title: "Semantic Information Retrieval from 3D Data"
Advisor: Prof. Dr. A. Aydın Alatan

July 2005 Master of Science in Electrical and Electronics Engineering
Middle East Technical University, Ankara, Turkey
Thesis Title: "Recursive Shortest Spanning Tree Algorithms for Image
Segmentation"
Advisor: Assos. Prof. Dr. Cüneyt Bazlamaçcı

June 2002 Bachelor of Science in Electrical and Electronics Engineering
Middle East Technical University, Ankara, Turkey

Work Experience
Dec 2002 - Sep 2009 *Research and Teaching Assistant*
Department of Electrical and Electronics Engineering

Middle East Technical University
Aug 2002 - Dec 2002 *Digital Design Engineer*
ASELSAN Military Electronics Industries Inc., Ankara

SUMMER 2001 *Intern*, ASELSAN Military Electronics Industries Inc., Ankara

SUMMER 2000 *Intern*, ASELSAN Military Electronics Industries Inc., Ankara

Publications
_ Bayramoglu N., Alatan A. A., "Range Image Recognition: Segmentation, Matching and Similarity Retrieval", journal paper (in preperation)
_ Bayramoglu N., Alatan A. A., "Occlusion Aware Heat Kernel Signature on Range Images", (in preperation)
_ Bayramoglu N., Alatan A. A., "Lossless Description of 3D Range Models", (accepted to SPIE, Electronic Imaging 2012, Visual Information Processing and Communication)

_ Bayramoglu N., Alatan A. A., "Shape Index SIFT: Range Image Recognition Using Local Features", 20*th* International Conference on Pattern Recognition (ICPR'10), Istanbul, Turkey, 2010

_ Akman O., Bayramoglu N., Alatan A. A., Jonker P., "Utilization of Spatial Information for Point Cloud Segmentation", 3DTV Conference 2010, Tampere, Finland, 2010

_ Bayramoglu N., Akman O., Alatan A. A., Jonker P., "Integration of 2d Images And Range Data For 3D Object Segmentation And Recognition", 12*th* International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines (CLAWAR), Istanbul, Turkey, 2010

_ Bayramoglu N., Alatan A. A., "Range image recognition using invariant local features", IEEE 18th Conference on Signal Processing and Communications Applications, SIU 2010, Diyarbakır, Turkey, 2010 (in Turkish)

_ Bayramoglu N., Alatan A. A., "Segmentation driven semantic information inference from 2.5D data", IEEE 17th Conference on Signal Processing and Communications Applications, SIU 2009, Antalya, Turkey, 2009 (in Turkish)

_ Bayramoglu N., Bazlamacci C.F., "Recursive Shortest Spanning Tree Algorithms for Image Segmentation", Proc. of the IEEE Int. Conference on "Computer as a Tool" (EUROCON 2005) Belgrade, Serbia and Montenegro, 2005

Honors and Awards
2006 Graduate Courses Performance Award
Graduate School of Natural and Applied Sciences,

Middle East Technical University
2002 Senior Design Project Award: 2nd Place
Department of Electrical and Electronics Engineering,
Middle East Technical University

1998 Ranked 11th in
Nationwide University admission examination among 1.3 million Exam Takers

Languages
Turkish: Mothertongue
English: Good
German: Elementary

Computer Skills
C++, Matlab, Python, Excel, Word, PowerPoint, LaTEX