Comparison of 3D Local and Global Descriptors for Similarity Retrieval of Range Data

Neslihan Bayramoğlu[a,*], A. Aydın Alatan[b]

[a] *Center for Machine Vision Research, University of Oulu, Finland*
[b] *Dept. of Electrical and Electronics Eng., Middle East Technical University, Turkey*

Abstract

Recent improvements in scanning technologies such as consumer penetration of RGB-D cameras, lead obtaining and managing range image databases practical. Hence, the need for describing and indexing such data arises. In this study, we focus on similarity indexing of range data among a database of range objects (range-to-range retrieval) by employing only single view depth information. We utilize feature based approaches both on local and global scales. However, the emphasis is on the local descriptors with their global representations. A comparative study with extensive experimental results is presented. In addition, we introduce a publicly available range object database which is large and has a high diversity that is suitable for similarity retrieval applications. The simulation results indicate competitive performance between local and global methods. While better complexity trade-off can be achieved with the global techniques, local methods perform better in distinguishing different parts of incomplete depth data.

Keywords: range data retrieval, local descriptors, global descriptors, similarity indexing, single view depth data description

*Corresponding author
Email addresses: nyalcinb@ee.oulu.fi (Neslihan Bayramoğlu), alatan@eee.metu.edu.tr (A. Aydın Alatan)

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 improved scanning technologies, increased processing power, increased storage capabilities, and the progress in visualization and printing technologies, as well as the consumer penetration of Kinect sensor and *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.

We distinguish similarity retrieval, instance recognition and classification (category recognition) problems. These problems are all considered under the broad title of *"object recognition"* usually. Classification task is defined as determining the category name of a new observation *(query)* based on the training set of objects with known class memberships (e.g. query is a member of *"dog"* class). Supervised learning methods are usually employed which gather features from a training set and obtain a representative descriptor of a class. Instance recognition (verification) on the other hand does not utilize class labels but searches for specific objects learned beforehand among the new observation data (e.g. query does not contain *"Pluto"* but contains *"Scooby-Doo"*). Number of objects can be learned during the training phase. Conversely, similarity retrieval (indexing) applications usually do not employ supervised learning techniques. Main motivation in this strategy is to ensure the scalability of the retrieval application. Scalability guaranties that new objects can be introduced into the database easily without performing complex/manual data labelling and re-training the system. Similarity retrieval applications search for data in the database that are similar to the query and these applications are not skilled to convey other semantic information (e.g. query is most similar with the database object obj_i, then obj_j, obj_k, ...). Therefore, classification, verification and indexing methods operate on different domains and serve distinctly.

Searching geometrically similar examples of a complete 3D mesh model (query)

among a database of complete 3D mesh models is called *3D-to-3D retrieval*. In a similar manner, searching a range model among a database of complete 3D mesh models is called *range-to-3D retrieval*. In literature, 3D-to-3D retrieval is studied extensively, whereas there are relatively limited studies on range-to-3D retrieval research. On the other hand, searching a range image among a database of range images, *range-to-range retrieval*, is not studied much. Here, it should be noted that we are aware of the recent studies employing depth data in similarity retrieval applications. However, in those studies, either multiple views of query and/or database models are employed [1, 2, 3] or additional information such as texture, colour or intensity is incorporated [4]. On the other hand, this work considers the problem of retrieving objects based solely on the depth information of a single view among a database of similar data; supplementary information such as texture, colour, intensity, etc. is not incorporated.

The frequency of occurring range data in daily life are more frequent relative to other 3D representations. The widespread usage of range data is due to new generation depth sensors such as Kinect [5] which is affordable and has a real-time nature. Therefore, we focus on *range-to-range retrieval* in this work and present a comparative study. While the input data type is range image we limit ourselves to isolated objects. We have not focused on segmenting objects in range scenes; several approaches exist ([6],[7]) and their performances probably have effects here which should be studied in addition.

This paper extends our prior work [8] and explores several other feature based approaches for range model retrieval. In this study, global and local feature extraction methods are proposed, evaluated, and compared on our database which is suitable for testing similarity retrieval methods. The paper is organized as follows. Next section gives a brief summary on the 3D shape retrieval literature. Features that are based on local surface properties employed in this study are then presented in Section 3. The following section describes global features along with our lossless description technique. Section 5 describes the database and presents experimental results. Finally, Section 6 summarizes and concludes the paper.

2. Related Work

3D object description is treated in 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* [1]). Vision side is mainly interested in instance recognition, point correspondences, and registration. 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 [9], 3D Zernike moments [10], shape histograms [11], spherical harmonics [12, 13, 14, 15], shape distributions [16], and diffusion distances [17] can be listed. On the other hand, shape spectrum [18], splashes and 3D curves [19], point signatures [20], spin images [21], local feature histograms [22], multi-scale features [23], auto diffusion function/ heat kernel signatures *(HKS)* [24, 25], and 3D histogram of oriented gradients (3DHOG) [26] 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 [27, 28], skeletons [29], curve-skeletons [30] 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 [31], compact multi-view descriptor *(CMVD)* [32], bag-of-features SIFT (BF-SIFT) [33], and panoramic views [34] are view based studies. The literature certainly contains many other studies and we refer the reader to Guo et al. [35], Tangelder and Veltkamp [36] and Bustos et al. [37] for detailed surveys.

[1] `http://www.aimatshape.net/event/SHREC/`

The usage of range data became widespread after the release of new generation depth cameras and range scanners such as Kinect [5] and research on 3D range object recognition in computer vision side has accelerated afterwards. Early studies on range image analysis [20, 19, 21, 38], address the surface matching problem. Complete object models which are constructed initially are searched in a partial scene by matching points. These *instance recognition* methods are also utilized in surface alignment. Later, *range-to-3D retrieval* methods are presented [39]. Finally, RGB-D data are utilized for *instance recognition* and *classification* [40, 41, 42, 43]. These RGB-D cameras have significant advantages compared to laser scanning devices: *i)* they can operate in real time (up to 30 Hz), *ii)* they are affordable, and *iii)* depth data is synchronized with the color information.

In this paper we focus on similarity retrieval of single view depth data instead of employing complete 3D models, multiple views, or additional information such as texture, colour, intensity. The motivation for obtaining a range image retrieval system could be due to the some new paradigms, such as 3DTV archive systems, 3D range object databases or LIDAR systems. There are some similarities between range image similarity retrieval and partial matching or range-to-3D retrieval research; however, the differences are quite crucial. In partial matching research [44, 45], 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 [46, 32, 33] among a database consisting of complete 3D models. In this case, database models are viewed from several directions 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 3D similarity retrieval approaches is that single view of a query is used instead of multiple views. The descriptors are usually obtained from 2D image features.

If the database images and the query are both range images, then partial matching and view based approaches become deficient, since 3D geometry of the

object is not explicitly given in the range image representation. Although range images contain 3D information, they are different from complete 3D models. These differences are due to *i) self-occlusion* (Figure 1), *ii) transformations* and *iii) view dependent partial geometry* (Figure 2). An exaggerated example of a self-occlusion is presented in Figure 1 where two distinct objects have almost same range image representations. Usually self-occluded regions are formed around salient regions. These salient regions are generally considered as the informative parts of objects. The surface represented by a range image containing self occluded regions might be different than the actual model. This situation can be observed in a hand model shown in Figure 2. In cases where self-occlusions are present, local descriptors which are extracted around such regions (e.g. fingers) might be misleading. Therefore, the description obtained for the same region in another view probably will be different. Global descriptor might also be misleading in some situations as shown in Figure1. Translations and rotations are other sources for information losses in range imaging. Sampling density varies as the objects are subject to such transformations. As a result, fine details disappear as the objects become distant from the camera.

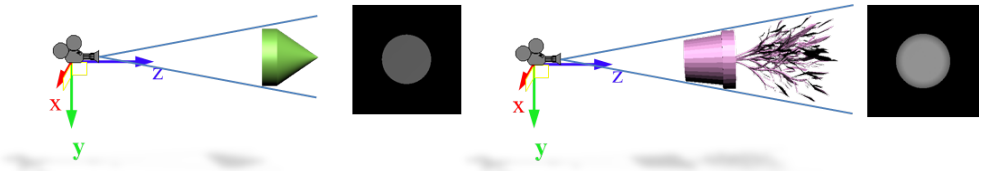

Figure 1: An extreme example of self occlusion. Distinct objects may result in similar range images depending on the viewing direction.

In this paper, local and global 3D features are employed for similarity retrieval of range models among a set of range models within an unsupervised framework. Well-known and simple descriptors which can be applied both locally and globally are selected. Local features are combined by the *"Bag of Features"* (BoF) method. This framework also allows employing more complicated surface descriptors such as the ones implemented in the Point Cloud Library (PCL) [2]. This study adapts

[2] `http://pointclouds.org/`

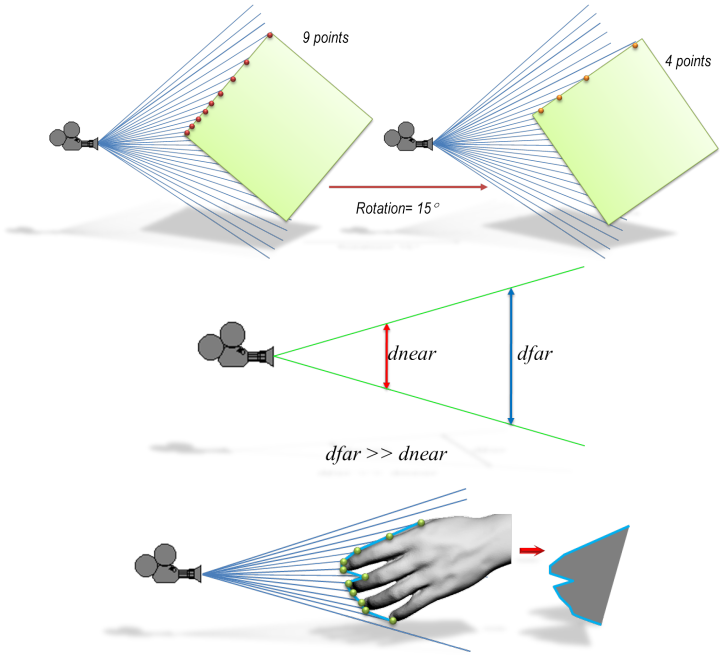

Figure 2: Objects shape information captured by range sensors depend on the viewing direction. *(Top)* Sampling density changes as objects experience rotations and *(Middle)* translations. *(Bottom)* Self occlusions is another challenge in range imaging. Details and/or discriminative regions may lost due to occlusions.

image based features such as Scale Invariant Feature Transform *(SIFT)*[47] and Speeded-Up Robust Features *(SURF)* [48] for range images by utilizing shape index mapping. We investigate the performances by utilizing depth information only. Incorporating other object attributes such as color, texture, and scene semantics probably yield better results in indexing similarities. However, we aim to compare depth-only features for retrieval applications rather than improving the indexing performance.

3. Local Description

Since 3D shapes have insufficient features and keypoint repeatability is not satisfied, local descriptors in 3D shape analysis are considered to be less discriminative and far from being robust [49]. However, any experimental evaluation

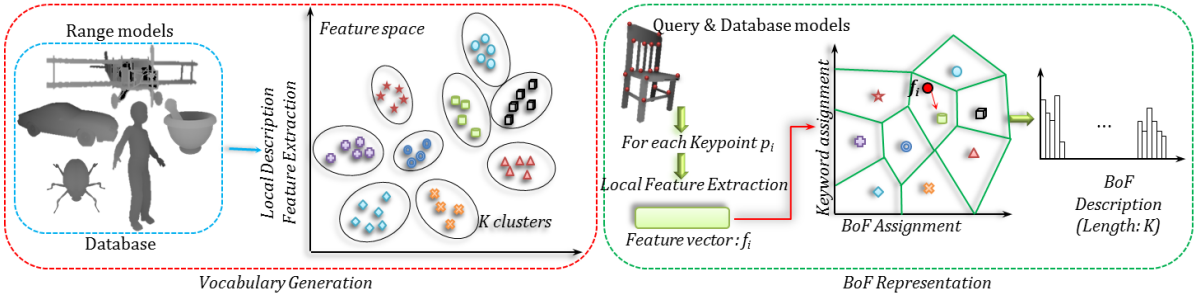

Figure 3: *Bag of Features* framework. In the off-line phase, vocabulary is generated by simply clustering local features (*n-dimensional* vectors) into k groups. The set of range objects that are used in generating the vocabulary does not have strict constraints. It could be any combination of objects that can provide adequate representative samples for local geometry. Secondly, database objects' features are assigned to the closest vocabulary cluster and counted to form a histogram which eventually forms BoF representations. During test time, BoF descriptors of query and database objects are compared and sorted according to the employed distance measure.

for comparing local and global descriptors for depth data are demanding. In this study, we select *"spin images"* [21], adopted *"D2 distributions"* [16] and *"3D moment invariants"* [50] for local description, and utilized *"shape index"* mapping for extracting image based local features. These features are then combined using the *"bag of features"* method [51] to represent range images. Similarity is computed as the distance between the corresponding bags of features representations.

3.1. Bag of Features

In the BoF approach, there are three main steps: i) feature detection and description, ii) construction of visual vocabulary(dictionary), iii) matching (see Figure 3). The main goal in the feature detection is to find keypoints encapsulating significant information that are also robust across transformed versions of the image. Feature descriptors which are usually represented as vectors carry local information in the neighbourhood of each keypoint. Visual *vocabulary* (dictionary) is built by clustering all the extracted features from a dataset of images. The selection of the number of clusters *(k)* is empirical, although it

is critical. Obtaining BoF representation of database images is the next stage. First, features are extracted from an image. After that features are assigned to the closest cluster (word) in the vocabulary. Then, the count of each word that appears in the image is used to form the BoF representation of the image. When a query is placed, firstly the BoF representation is constructed. Then the BoF representation of the query image and BoF representations of the database images are compared and matched.

3.2. Descriptors

Spin Images:. The spin image descriptor [21] is one of the well-known object-centered surface descriptor that can be used both locally and globally. It is a two-dimensional histogram of the spatial distribution of neighbouring points around a keypoint. The local coordinate system is computed at a point using its position and surface normal. The positions of other surface points with respect to the local coordinate system are then described by two parameters (α, β) (Figure 4). The perpendicular distance to the surface normal $\overrightarrow{n_x}$ is defined as the α *(radial)* coordinate and the signed perpendicular distance to the tangent plane defined by surface normal and the surface point p_x is defined as β *(elevation)* coordinate. These (α, β) parameters are computed and a histogram representation is obtained for points residing in the *support region*. The support region is defined as the maximum allowed distance from the point of interest. The histogram obtained in object-centered coordinate system can be represented as a 2D image *(spin image)* and it is utilized as the descriptor. The control on the amount of local information can be adjusted by varying the support region parameter. In the limit, all points are included in the descriptor and a global description can be obtained. Utilizing object-centered coordinate systems makes the spin images descriptor rotation and translation invariant.

3D moment invariants:. Moments, especially *Hu* moments, are popular tools in 2D object recognition. Also, their 3D counterparts [50] are proposed. Hu moments are scalar quantities used to describe the distribution of object points.

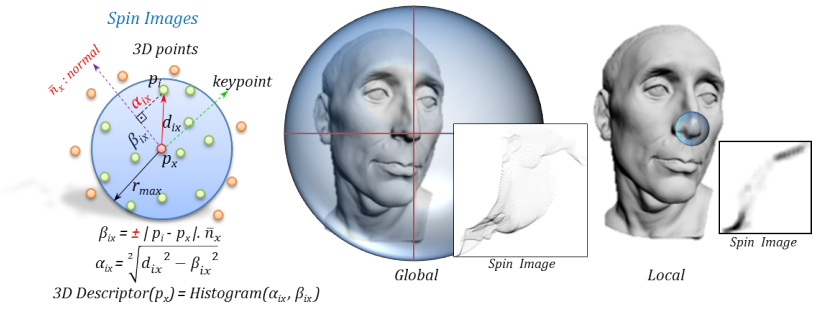

Figure 4: *(Left)* Spin image calculation. *(Middle)* Sample spin image obtained around the tip of the nose globally. *(Right)* Same point is used for describing the surface locally by decreasing the support region r_{max}.

They are simple descriptors allowing translation and rotation invariant computation for 3D models. Low order moments capture coarser shape information whereas high orders define the details. In order to obtain a translation invariant descriptor central moments $\mu_{k\ell m}$ of order *"$k\ell m$"* are employed which are defined as follows:

$$\mu_{klm} = \sum\sum\sum_{\forall p \in R} (x - \overline{x})^k (y - \overline{y})^\ell (z - \overline{z})^m \tag{1}$$

where $(\overline{x}, \overline{y}, \overline{z})$ is the centroid of the local region R and *p(x,y,z)* is the 3D point coordinates in $\mathbb{R}^3$. Following this, translation and rotation invariant second order moments are defined as:

$$\begin{aligned} J_1 &= \mu_{200} + \mu_{020} + \mu_{002} \\ J_2 &= \mu_{200}\mu_{020} + \mu_{200}\mu_{002} + \mu_{020}\mu_{002} - \mu_{110}^2 - \mu_{101}^2 - \mu_{011}^2 \\ J_3 &= \mu_{200}\mu_{020}\mu_{002} + 2\mu_{110}\mu_{101}\mu_{011} - \mu_{002}\mu_{110}^2 - \mu_{020}\mu_{101}^2 - \mu_{200}\mu_{011}^2 \end{aligned} \tag{2}$$

In this study, these three invariants are concatenated into a compact feature vector $f = (J_1, J_2, J_3)$ to form the final descriptor.

D2 distributions:. Osada et al. [16] propose a method for describing 3D shapes as a probability distribution sampled from a function and have experimented five shape functions measuring global geometric properties of an object. Shape

distributions are easily computed and similarity between the objects can be measured using metric distances. One example of a shape distribution which is called *"D2"* represents the distribution of Euclidean distances between pairs of object points selected randomly. In this study we limit the selection of points to a local region in order to obtain a local descriptor (Figure 5a). Moreover, in the following section, we also employ the D2 distribution as a global descriptor by releasing this constraint. The motivation of employing this early descriptor in this work is due its simplicity and yet its efficiency in describing 3D shapes in a pose independent way.

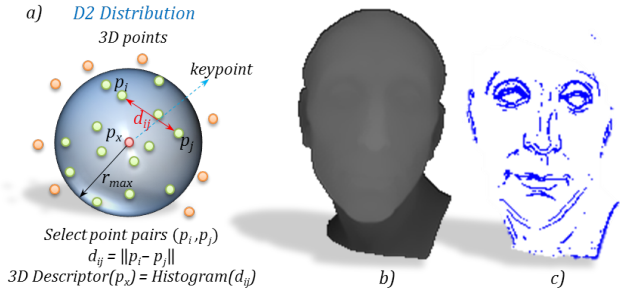

Figure 5: a) D2 Distribution method, b) Sample range image c) Salient regions

Image based features based on shape index mapping:. There are two key features for 3D surfaces: orientation and curvature. A 3D point can be described by its minimum and maximum curvatures (principal curvatures) or some other functions of these principal curvatures κ_1, κ_2. *"Shape Index"* (SI) [52] is one of a such measure with an appealing property of being scale, translation, and rotation invariant. The curvature values on a surface can be obtained robustly by fitting a quadric surface to the local patch. Then the shape index at point p is calculated using principal curvatures as follows:

$$SI(p) = \frac{1}{2} - \frac{1}{\pi} \arctan \frac{\kappa_1(p) + \kappa_2(p)}{\kappa_1(p) - \kappa_2(p)} \tag{3}$$

where $\kappa_1 \geq \kappa_2$. Distinct surfaces correspond to a unique shape index value in the $[0, 1]$ interval except planar surfaces. Principal curvatures vanishes ($e.g.: \kappa_1 = \kappa_2 = 0$) on planar points and shape index become indeterminate. Representing

surface points in range images by their corresponding shape index values generates a new mapping from 3D point coordinates *(x,y,z)* to shape index domain [0,1]. Image representation of this transformation is called *"shape index mapping"*. Such a mapping makes a strong emphasis on the points where the surface deviate from being smooth. The effect can be observed even for small changes due to equation's non-linear nature (Equation 3). In Figure 6, three examples of shape index mapping are shown. Compared to depth maps, surface characteristics and shape details are more significant in shape index mapped images. Therefore, utilization of feature extraction methods based on shape index images becomes more feasible. In this study, (SIFT) [47] and (SURF) [48]; which are popular descriptors because of their accomplished performances; are employed as image based features .

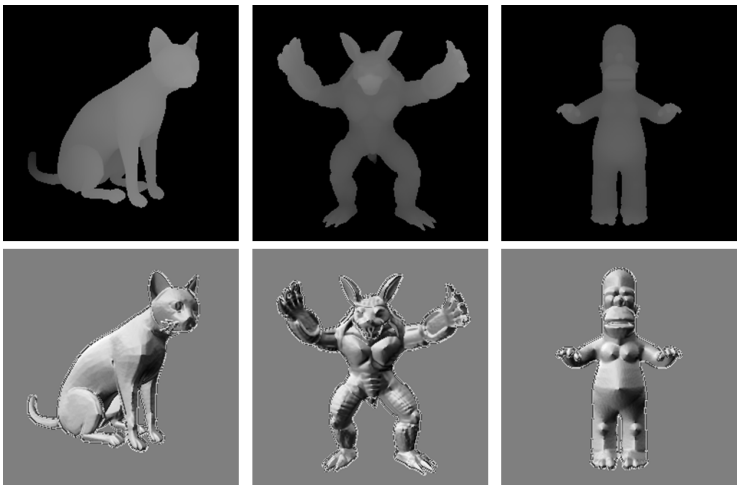

Figure 6: Shape index mapping of range images. *(Top)* Examples of range images, and *(bottom)* corresponding shape index mapped images.

3.3. Keypoint Detection

Identifying salient points on a 3D surface is more challenging than detecting keypoints in images, since images have richer and distinct features. Therefore, in images, keypoint repeatability can be achieved relatively easily. Nevertheless, there exists number of studies for keypoint detection on 3D surfaces [53, 54].

However, none of these methods are able to ensure the distinctiveness criterion on surface points since the criterion highly depends on the descriptor. Therefore, in addition to the classical way of detecting image based keypoints we also locate interest points by regular sampling on the 2D range image domain. Sparse sampling would yield insufficient descriptive information whereas fine sampling would yield gathering the descriptor of a specific feature for multiple times. In the latter case, similar descriptors will be merged in the BoF framework whereas distinct features will be clustered into different *"words"*. Therefore, our implementation uses a fine sampling to ensure distinctiveness. On the other hand, we also utilize SIFT and SURF keypoints for shape index mapped images.

4. Global Descriptors

In this section, global methods for describing single view depth models are presented. Here, previously employed local descriptors are utilized for global description. Besides our previously proposed *spherical harmonics transform (SHT)* based descriptor is also presented. Although *SHT*, is not a new concept in shape retrieval research, we propose to utilize it for depth data by representing the models in a world-oriented manner.

Local to Global:. In this study, *Spin images*, *3D moment invariants* and *D2 distribution* descriptors are also utilized for global description. This is achieved by extending their supporting regions to cover the entire shape. Similar computations as explained in the previous section are then followed. Besides, a modification to D2 distributions based on a saliency constraint is proposed.

4.1. Constrainted D2 Distribution

In [16], it is shown that *D2* distribution is robust to noise, small cracks and holes. Osada et al. [16] argue that the robustness is satisfied due to the random selection strategy. We tested ordinary *D2* distribution descriptor on our database as a global descriptor, besides we also tested our modified version *(Salient D2)*. The modification is the point selection strategy. Instead of random

point selection, we impose saliency constraint. Points which are informatively salient are selected and the *D2* distribution is evaluated among them. We define salient points as the ones having high or low curvature values compared to their local neighbourhood. Surface curvature estimated at point p is computed using the formula: $c_p = \lambda_0/(\lambda_0 + \lambda_1 + \lambda_2)$ where $\lambda_0, \lambda_1, \lambda_2$ are the three eigenvalues satisfying $\lambda_0 \leq \lambda_1 \leq \lambda_2$ obtained from Principle Component Analysis *(PCA)* of the local neighbourhood of the surface point p.

4.2. 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 $f(\theta, \varphi)$ and $g(\theta, \varphi)$ in this space is defined as follows:

$$\langle f, g \rangle = \int_0^{\pi} f(\theta, \varphi)\overline{g(\theta, \varphi)} sin\theta \, \mathrm{d}\theta \mathrm{d}\varphi \tag{4}$$

The Spherical Harmonics Y_ℓ^m of degree ℓ and order m ($\mid m \mid \leq \ell$) form an orthonormal basis in this space. In Figure 7, visual representation of spherical harmonics $Real\{Y_\ell^m\}^2$ is shown up to degree 3. They are related with the associated Legendre polynomials P_ℓ^m as follows:

$$Y_\ell^m(\theta, \varphi) = \underbrace{\sqrt{\frac{(2\ell+1)}{4\pi}\frac{(\ell-m)!}{(\ell+m)!}}}_{K_{\ell m}} P_\ell^m(\cos\theta)\, e^{im\varphi} \tag{5}$$

$$P_\ell^m(x) = \frac{(-1)^m}{2^\ell \ell!}(1-x^2)^{m/2}\ \frac{d^{\ell+m}}{dx^{\ell+m}}(x^2-1)^\ell \tag{6}$$

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

$$f(\theta, \varphi) = \sum_{\ell=0}^{\infty} \sum_{m=-\ell}^{\ell} f_\ell^m\, Y_\ell^m(\theta, \varphi) \tag{7}$$

where expansion coefficients $\hat{f}_\ell^m$ are projections of the function $f(\theta, \varphi)$ on the basis functions. They can be obtained utilizing the inner product (Equation 4)

defined in this space as follows:

$$\hat{f}_\ell^m = \int_0^\pi \int_0^{2\pi} f(\theta,\varphi)\, K_{\ell m} P_\ell^m(\cos\theta)\, e^{im\varphi} sin\theta \mathrm{d}\varphi \mathrm{d}\theta \tag{8}$$

If the function $f(\theta,\varphi)$ 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, $f(\theta,\varphi)$, sampled in an equiangular grid $(2B \times 2B)$ with a sum of $4B^2$ points, expansion coefficients $\hat{f}_\ell^m$ are obtained follows [55]:

$$\hat{f}_\ell^m = \frac{\sqrt{2\pi}}{2B} \sum_{j=0}^{2B-1} \sum_{k=0}^{2B-1} w_j f(\theta_j,\varphi_k) P_\ell^m(cos\theta) e^{-im\varphi} \tag{9}$$

The coefficients $\hat{f}_\ell^m$ is equal to zero for $\ell \geq B$ 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

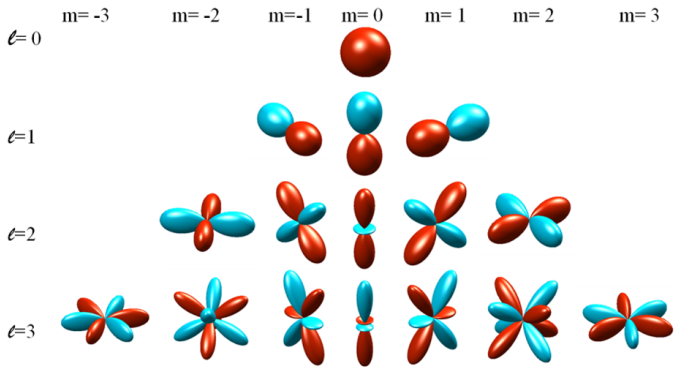

Figure 7: Visual representation of spherical harmonics up to degree 3. $Real\{Y_l^m\}^2$ is plotted, positive and negative portions are coloured with red and blue respectively

4.2.1. Spherical Harmonics in Shape Analysis

Vranic et al. proposed to use Spherical Harmonics Transform in 3D model retrieval [14]. They describe the shape as a spherical function $f(\theta, \varphi)$, where the origin is selected as center of the mass of the model. The value of the function $f(\theta, \varphi)$ 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. The descriptors are then compared using L_1 norm. Their method has two disadvantages. First one is the necessity of a pose normalization step. Vranic propose a modified Principal Component Analysis (PCA) method for this purpose. Secondly, the functional representation proposed by Vranic [14] 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 [13]. 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:

$$\sqrt{\sum_{m=-\ell}^{m=\ell} |\hat{f}_\ell^m|^2} = \sqrt{\sum_{m=-\ell}^{m=\ell} |\hat{f}_{\ell,ROTATED}^m|^2} \tag{10}$$

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 et al. [15] 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 in [15] that this method outperforms rotation invariant spherical harmonics descriptor based on binary voxel grid [13].

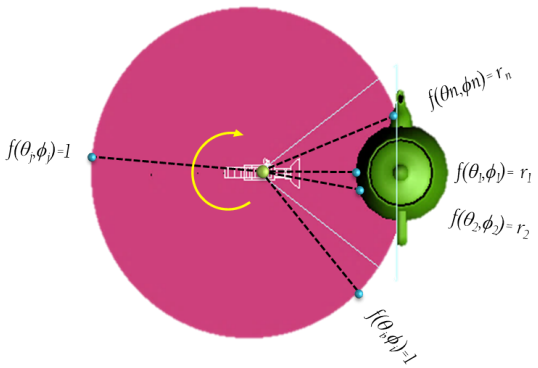

Figure 8: Description of the world with respect to the camera *(Top view of a "tea pot" scene)*. Spherical function f is normalized such that maximum extend is equal to one.

Moreover, Kazhdan et al. [12] 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 such as in a very recent study [26] as it is employed to represent 3D histogram of oriented gradients (3DHOG) for obtaining rotation-invariant 3D descriptors.

4.2.2. Lossless SHT

Obviously, SHT can describe functions defined on two-sphere effectively. 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. In this study, we formulate a lossless representation of depth data as follows: A range image can be represented with a spherical function with ray casting strategy. Besides, all available information is preserved with this representation. Instead of describing the shape, we describe the world captured from the camera (Figure 8). The main steps for computing our spherical harmonics descriptor for range models is shown in Figure 9.

Firstly, background is removed. The origin of the coordinate frame is assigned as the center of the camera. Spherical coordinates is used and the 3D space is discretized according to the resolution of the input image. The surface $f(\theta, \varphi)$ is

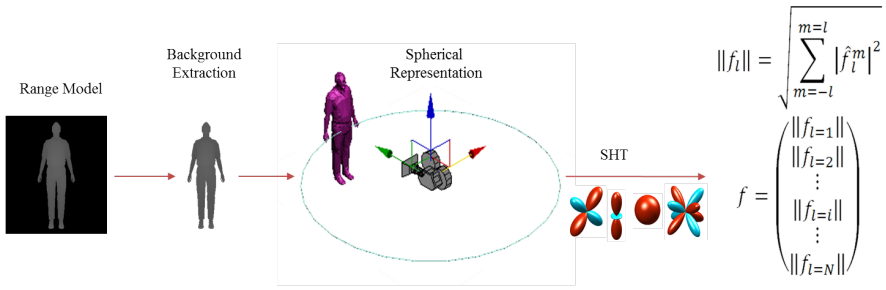

Figure 9: Main steps for computing our spherical harmonics descriptor (*Lossless SHT*) for range models.

initialized with zeros. The coordinates of data points are 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 $f(\theta, \varphi)$. The maximum ray length is determined and the function is normalized such that the function has the maximum value of one. For non-object parts of $f(\theta, \varphi)$ the value one is assigned (Figure 8). With the use of SHT, the function is expressed as a finite weighted summation of the basis functions. Utilizing the rotation invariant property of the spherical harmonics, the amplitude of the coefficients within each frequency band (ℓ) is computed. The feature vector is formed (signature of the range model) by concatenating these amplitudes. The zero-order component is omitted:

$$f = (\|f_{\ell=1}\|, \|f_{\ell=2}\|, \cdots \|f_{\ell=N}\|) \tag{11}$$

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 3. Rotations around the z-axis become invariant with this representation. The global characteristic of range models captured by the camera is described.

5. Evaluation

5.1. Database

To test and compare range data descriptors, we have built our own database. It is publicly available and can be downloaded from here [3]. It contains 545 range models divided into 18 classes. Our range images have a size of 256×256 pixels. Each pixel p_i is associated with a 3D point coordinate $p_i = (x_i, y_i, z_i)$. Representative models for each class are shown in Figure 10. We collected 3D mesh models from Princeton Shape Benchmark [4], AIM@SHAPE repository [5], NTU shape database [6], and Konstanz University database [7]. Then, by the use of a COTS computer graphics software, we obtain depth views 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 such as [4], most of them have some limitations for testing similarity retrieval methods. These are due to their sizes (within class and overall), their diversities, and due to object imperfections. Although our database consist of toy data, we have higher diversity and larger number of object instances. In addition, a toy dataset can be considered to be more convenient for comparison purposes such that performance measure of a method, when tested on a toy dataset, is independent of object segmentation errors and independent of sensor measurement errors.

In shape similarity retrieval applications, choice of databases has 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 should also be included

[3] http://www.ee.oulu.fi/~nyalcinb/sub/shape.html

[4] http://shape.cs.princeton.edu/benchmark/

[5] http://www.aimatshape.net/resources

[6] http://3d.csie.ntu.edu.tw/~dynamic/database/index.html

[7] http://www.informatik.uni-konstanz.de/en/saupe/research/finished-projects/3d-model-retrieval/

in the database. Databases comprised of real life scannings do not contain such kind of data because of the limitations in today's possibilities. Despite the synthetic structure, our database is a challenging one. First, intra class similarity is quite high (mainly due to varied viewing directions). Secondly, the interclass similarity is also high for some of the classes (i.e. cup- potter, dog-fourleg, spider-helicopter, human-gun, etc.). For simplicity, we are restricting the "similarity" criterion according to class labels. However, other similarity arguments can also be adopted depending on the application such as accepting a specific chair and a specific table as similar.

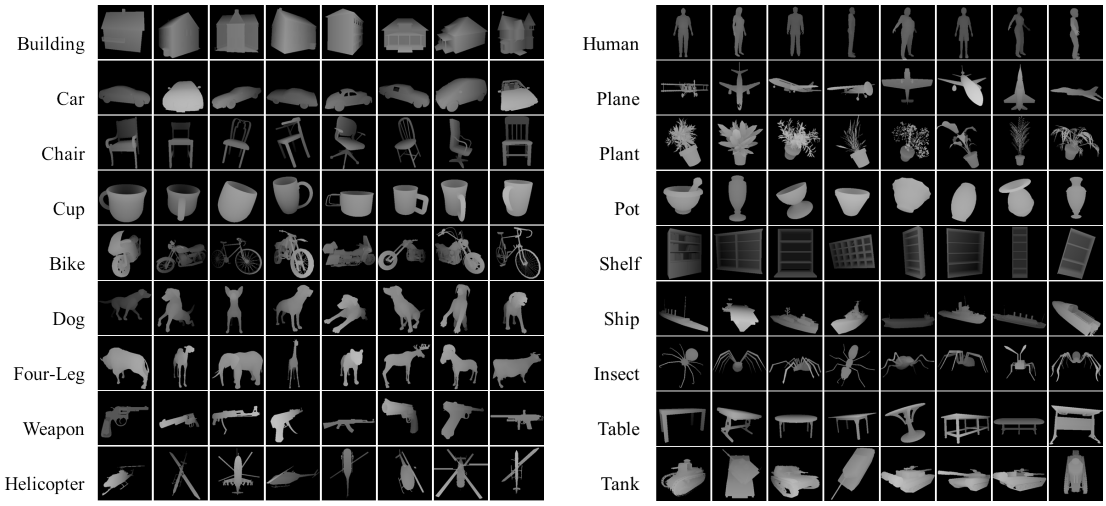

Figure 10: Exemplar depth map images from our database.

5.2. Performance Measures

In the procedure of retrieval, in response to a given set of users queries, an algorithm searches the benchmark database and returns an ordered list of responses called the ranked list. The number of retrieved elements can be as high as the size of the database. The importance is given to the relevant matches appearing at the top locations of the lists. Therefore, evaluation of retrieval methods is not exactly the same as in a typical classification system. The most commonly used statistics for measuring the performance of retrieval algorithms is

the Precision-Recall values. It is almost a standard procedure in shape retrieval and enables an objective comparison of different methods. Precision is defined as the ratio of the number of shapes retrieved correctly over the total number of retrieved shapes (Equation 12). Recall is defined as the ratio of retrieved shapes over the total number of relevant shapes in the database (Equation 12):

$$precision = \frac{\text{number of relevant items retrieved}}{\text{number of retrieved items}}$$
$$recall = \frac{\text{number of relevant items retrieved}}{\text{total number of relevant items in the database}} \tag{12}$$

Recall is also called *sensitivity*. Other measures used in information retrieval evaluation such as *E-Measure*, *F-Measure*, *First Tier*, and *Second Tier* share the similar idea with precision and recall. That is, they check the ratio of retrieved elements in the query's class that also appear within the top matches. Therefore, we utilize the widely used precision-recall (PR) curves in this study to present the retrieval performance.

5.3. Experimental Results

Local Descriptors. In order to investigate the effects of parameter selection and to present a fair comparison we performed several experiments. Effects of vocabulary size in the Bag-of Features framework, support region, and keypoint sampling density is analyzed. Precision-Recall curves of these experiments are shown in Figures 11,12, 13 and 14. Vocabulary sizes of 10, 30, and 100 are evaluated. Local windowing is utilized for defining the support region. Window sizes of 5×5, 15×15, and 35×35 are employed. And finally, keypoint sampling density in the uniform selection strategy is analyzed. Step sizes of 5, 10, and 15 pixels between neighbouring keypoint locations are evaluated. Increasing step size further yields empty keypoint sets for some of the input data. One particular parameter is evaluated at a time and remaining ones are fixed. Vocabulary Size of 30, support window size of 15×15 and step size of 5 in the keypoint

selection is used as default parameters if they are fixed. *L2* norm is employed in comparing the BoF descriptors.

In the calculation of *D2 descriptor*, the number of points selected randomly among the support region is also adjusted. Number of samples corresponding to the support region is noted in Figure 11. Histograms are calculated with 8 bins. In calculating the *spin images* descriptors, spherical coordinate space (α, β) is discretized into 10 bins. Shape index mapping is evaluated with SIFT *(128 dimensional)* and SURF *(64 dimensional)* features. Here, in addition to the uniform keypoint sampling, SIFT and SURF keypoint detection strategies are also utilized. SIFT keypoints are computed at local maxima and minima of difference-of-Gaussian (DoG) images whereas SURF uses a hessian based blob detector to find interest points.

In the vocabulary generation stage of our Bag of Features framework, we used the complete database. Similarly, during test time, every object in the database is queried and similar objects are searched and indexed in the complete database one by one. Therefore, we tested each method with 545 queries and the first retrieved result (the most similar object in the database with the query) of indexing is always the query itself. Precision-Recall values are reported based on the mean values of the 545 tests.

Based on the experiments, it can be concluded from the Precision-Recall curves (Figure 11) of *D2 Distribution* descriptor that small support size degrades the performance. Similarly, utilizing quite large image patches does not increase the performance as expected. On the other hand, increasing the vocabulary size enhances the performance of the *D2 Distribution* descriptor up to a point along the retrieval dimension. However, after that point precision clearly decreases. In the keypoint sampling experiments it is evident that fine sampling provides better description.

In the *Spin Images* description tests (Figure 12), effect of support size changes shows similar attitudes with the *D2 Distribution* descriptor. While small support region is not capable of representing surface properties well enough, descriptor that utilizes large regions looses it discriminative power. Higher vocabulary size

and fine keypoint sampling again improves the indexing performance.

3D Moment invariants descriptor presents the most inferior retrieval performance (Figure 13). In contrast to previous ones, fine keypoint sampling and large vocabulary do not enhance their retrieval results. Moments are not unique descriptors, that is different shape geometries (different point distributions) can yield exactly same descriptor values. Therefore, the limiting factor in this representation does not lie in the BoF's parameter selection but lies in the unsatisfactory way of description.

Performance of *Shape Index mapping* approach along with image based features are given in Figure 14. As expected, SIFT and SURF provides similar performances in all settings as they have very similar feature description methods. Support region size in the uniform sampling strategy is fixed and utilized as 16×16, whereas support sizes are decided according to the selected scale within the SIFT and SURF keypoint detection routines. We compare keypoint detection strategies in the last graph of Figure 14. Best performed uniform sampling (Grid Keypoints) and SIFT and SURF keypoint detection performances are plotted. Fine uniform sampling is superior to SIFT and SURF keypoint detectors. As noted previously, 3D shapes have insufficient features and keypoint repeatability is not satisfied. Therefore, classical feature detection techniques are incompetent in capturing salient locations, even though *Shape Index Mapping* provides fair saliency amplification.

Finally, Figure 17 gives an overall comparison of local descriptor performances in BoF framework. BeThe noise model is generated using a zero-mean Gaus sian distribution.st performing settings are selected for each method for fair comparison. Indexing performance of *Moments* are far behind the remaining ones. Other methods showed similar PR characteristics with a fine tuning of parameters in BoF, as well as the parameters that take place in the descriptor calculation.

Global Descriptors. We employ *Spin images*, *3D moment invariants* and *D2 distribution* descriptors also for global description by extending their supporting

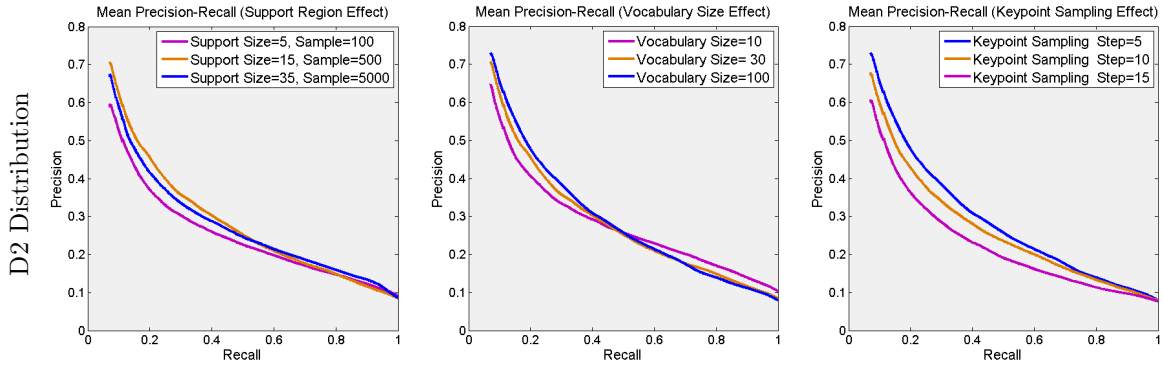

Figure 11: Performance Measures of Local D2 Distribution Descriptor with different parameter settings. Best viewed in color.

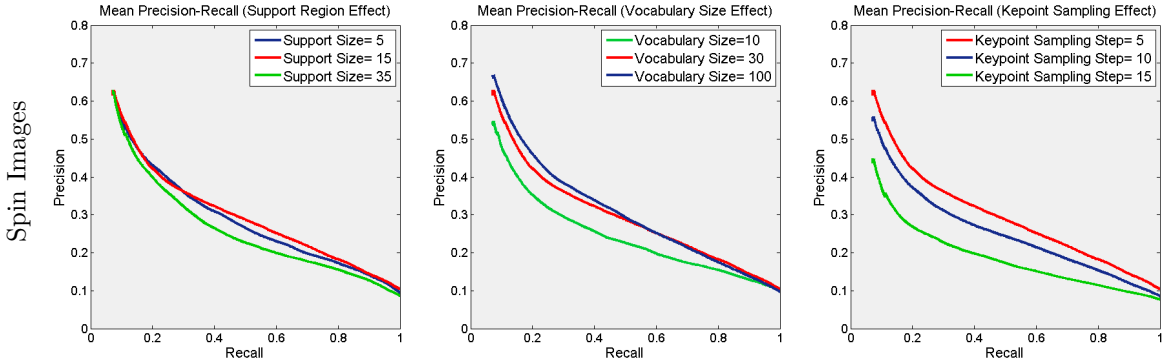

Figure 12: Performance Measures of Local Spin Image Descriptor with different parameter settings. Best viewed in color.

regions to cover the entire shape. In addition, we tested spherical harmonics transform using concentric spheres with ray casting [15] and our Lossless SHT method for global description.

To construct the global *D2 distributions* and the global *D2-Salient* distributions descriptors we utilize several sampling densities such as 1000, 1500, 2500 and 10^4 point pairs. Point pairs are selected randomly among the entire object. *Jensen-Shannon Divergence* (JSD), which is a symmetric version of *Kullback-Leibler Divergence*, is used in comparing distributions. Figure 15 presents corresponding PR curves. The global *D2 distribution* yields better

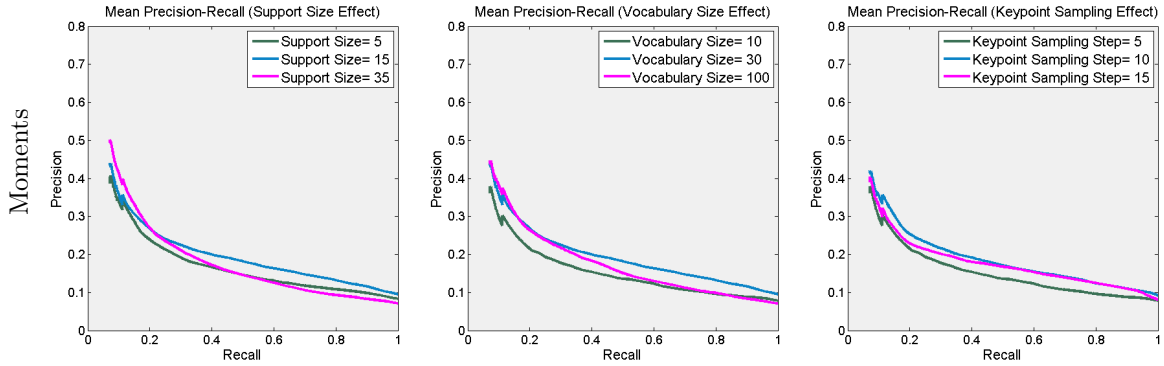

Figure 13: Performance Measures of Local Moments Descriptor with different parameter settings. Best viewed in color.

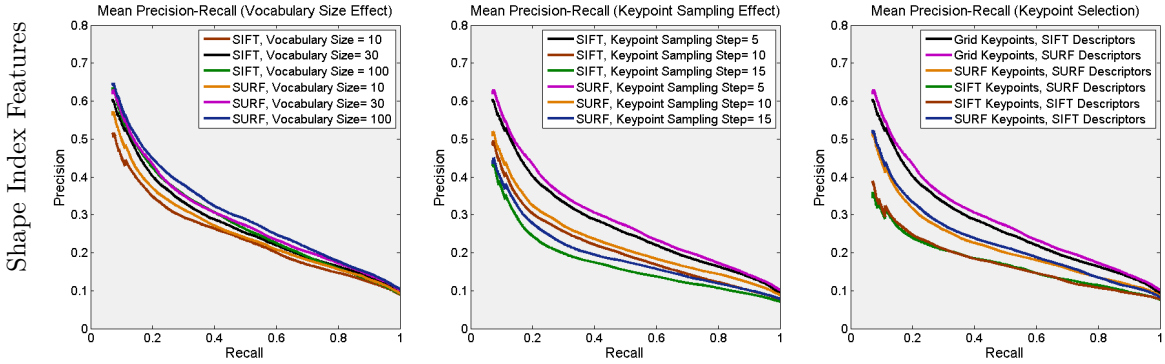

Figure 14: Performance Measures of Local Features on Shape Index Mapped Images with different parameter settings. Best viewed in color.

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 are included in the distribution. Perhaps in that case some of the point pairs contribute to the distribution more than once. The best performance among those tests belongs to the 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.

We tested *Spin Image descriptor* in the global scale with different α and β step sizes. Descriptor is constructed around the center point by first evaluating the normal direction and the tangent plane of the surface at this point. Figure 16 compares different α, β settings. Increasing the Spin Image size by dividing the spherical space into finer partitions does not increase the description performance. Although higher resolution descriptor provides more information about the geometry of the shape it can not handle interclass similarity.

We compute the spherical harmonics expansion coefficients using *S2KIT* [8]. In the *Lossless SHT* calculation we discretize the 3D space represented by (r, θ, φ) into a $(1, 512, 512)$ grid. Totally, $262,144$ points are defined on the sphere. We use $l = 256$, so we have a signature (descriptor) of length 255 as we use ray casting method using only one radius in a occlusion free representation. In the offline phase, we extract the signatures of all database models similarly, 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 *L2* norm.

In the first stage of the classical *SHT*, range model is translated so that its center of mass coincides with the coordinate origin. In contrast to Lossless SHT, such representation could contain self occlusions w.r.t. the origin. Therefore, concentric spheres should be utilized in classical SHT calculation. First, size of the model is normalized to unit sphere and the radius is discretized into 32 levels; and $l = 32$ is used. Eventually, the classical SHT descriptor become a 1024 length vector (32×32). Although concentric spheres are utilized, classical way of representing range models using SHT is lossy due to the finite r discretization.

Moment descriptors are easy to calculate and does not require parameter tuning. Figure 18 gives an overall comparison of global descriptors. Again, whenever applicable, best performing parameters are selected. This time D2 distribution descriptor shows the worst performance. Classical *SHT* perform better than *D2 distribution* and slightly worse than the proposed method. On the other hand,

[8] `http://www.cs.dartmouth.edu/~geelong/sphere`

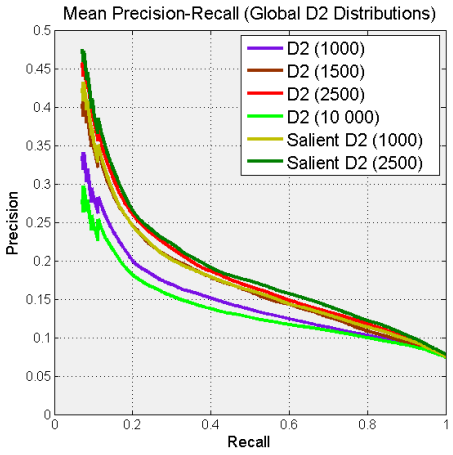

Figure 15: Comparison of mean Precision-Recall curves of global D2 Disribution and Salient D2 Distribution descriptors for different point samples.

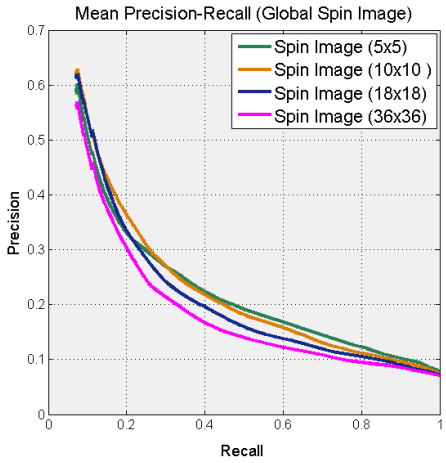

Figure 16: Comparison of mean Precision-Recall curves of global Spin Images descriptor for different α and β discretization.

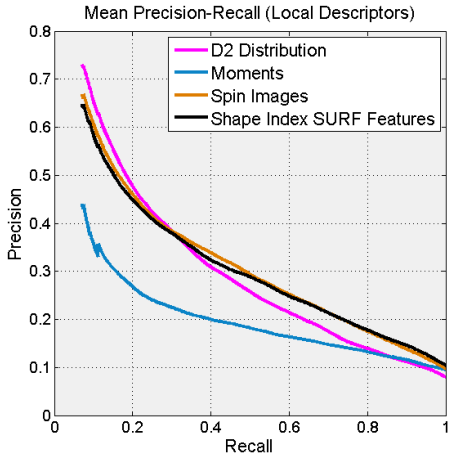

Figure 17: Comparison of local methods in BoF frame work. Best performing settings are selected for each method.

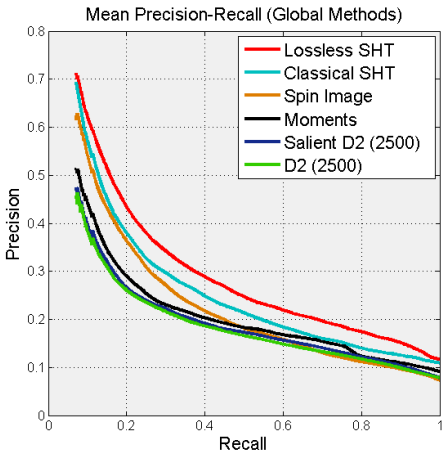

Figure 18: Comparison of global methods.

our proposed method, *Lossless SHT*, has the best mean precision-recall curve while the descriptor size is smaller than the classical *SHT*. Besides, computational complexity of SHT is much more complex. In the classical method, number of

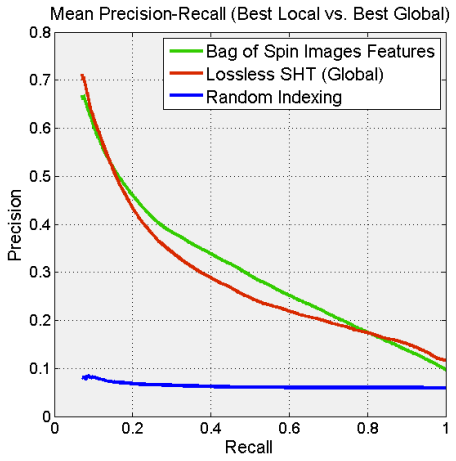

Figure 19: Comparison of best performing global method (*Lossless SHT*) and best performing local method (*Spin Images*). A random retrieval performance is also plotted for comparison.

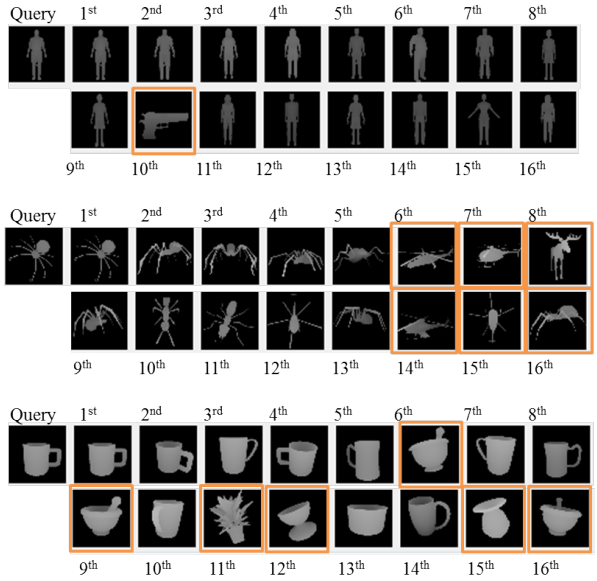

Figure 20: Sample retrieval results of three queries belonging to three different classes human, spider, and cup respectively using best global method (*Lossless SHT*). First 16 matches are shown. Queries are also included in the database, so first match is always the query itself.

transformations is equal to the number of concentric spheres (radius discretization) whereas in our Lossless SHT only one transformation is performed. Besides, our method has a lossless structure, available shape information is completely

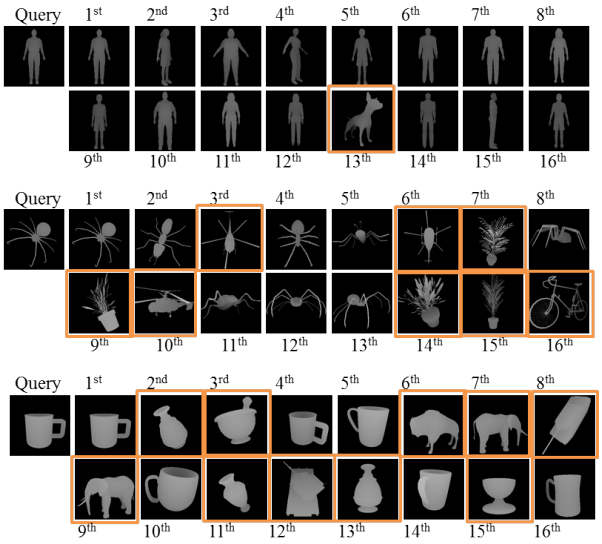

Figure 21: Sample retrieval results of three queries belonging to three different classes human, spider, and cup respectively using best local method (*Bag of Spin Images Features*). First 16 matches are shown. Queries are also included in the database, so first match is always the query itself.

included in obtaining the descriptor whereas classical way "approximates" the shape geometry which could be utilized in shape reconstruction.

We compare best local method (BoF Spin Images) and best global method (Lossless SHT) in Figure 19. Both approaches present very similar performances in indexing range models in our database. Sample search results of first 16 matches are also given in Figures 20 and 21 for randomly chosen three queries (specific id's in the database: human01d1, spider01d3, and cup05d1). False matches are marked in the figure. These particular retrieval performances of both methods are also similar except the cup search with BoF Spin Images. Lossless SHT retrieves more accurate objects for the query. In Figure 19 indexing performance of a random experiment is also shown in order give an idea to the reader about performances of the analyzed approaches and the size of the database.

In Figures 20 and 21, there are many objects which are visually and geomet-

rically similar to the query but are not listed in the same class with the query are retrieved and marked as false matches. We define those matches as *"good false matches"*. Retrieving helicopter for spider query and retrieving pottery in querying cup sample can be considered as *"good false matches"*. If those good false matches are counted as true positives then PR curves would move up to a higher level.

Finally, in the following experiments, we compare performances of the two winning methods from local and global descriptors on noisy and partially occluded models.

Noise. The noise model is generated using a zero-mean Gaussian distribution as the noise of Kinect-like sensors consist of such signals. We add noise to the depth value of each sample with standard deviations (σ_d) of 0.001, 0.01, 0.1, and 0.5 proportional to its depth as follows:

$$Z_{noisy} = Z + Z \times \mathcal{N}(0, \sigma_d) \tag{13}$$

Figure 22 shows the performances of Lossless-SHT and BoF Spin Images under noise. At low noise levels with $\sigma_d = 0.001$ and 0.001, for Lossless SHT, the area under PR curve seems to be same with a different curve. On the other hand, at high noise levels with $\sigma_d = 0.1$ and 0.5, the performance is clearly decreased. For BoF Spin Images, retrieval performance is gradually decreasing with the increasing noise. However, we believe that the performance decrease in our global method is also due to the overfitting problem which is more likely at high noise levels.

Occlusion. For experimenting the occlusion effect, we applied simple masks onto the objects. Rectangular masks of sizes 21×21, 61×61, and 101×101 centered at image centers are employed as shown in Figure 23. Figure 24 presents retrieval performances for occluded models. Self occlusions are already present in depth data and objects are partial. Introducing additional occlusion affected the performance of global descriptor more than the local descriptor

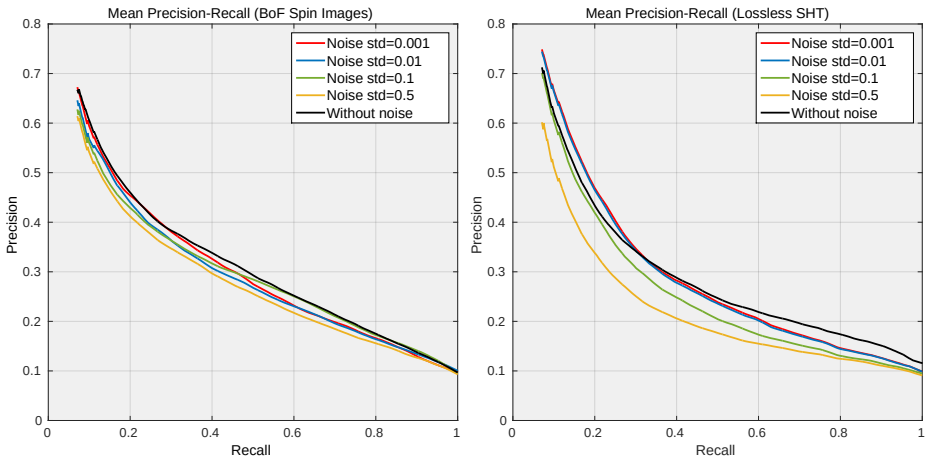

Figure 22: Noise effect. (*Left*) BoF Spin Images, (*right*) Lossless SHT.

Figure 23: Rectangular masks are employed to simulate occlusion. Original model is shown first and occluded model with masks sizes of 21×21, 61×61, and 101×101 are shown sequentially.

as expected. Occlusions just in the center of the model with sharp borders must have introduced high frequency components in the spherical harmonics transformation which could be one of the reasons for the performance reduction.

5.4. Summary and Discussions

We present here a systematic evaluation and comparison of local and global shape descriptors. For local description, we evaluate Spin Images, 3D moment invariants, D2 distributions and image based features (SIFT, SURF) on shape index maps. We observe that the changes in BoF parameters (keypoint sampling, support size, dictionary size) imply considerable variations in the performance. In most settings, finer keypoint localization exhibit higher performances. Best performance is obtained by Spin Images whereas the 3D moment invariants has the worst performance compared to others. Moments capture coarser details

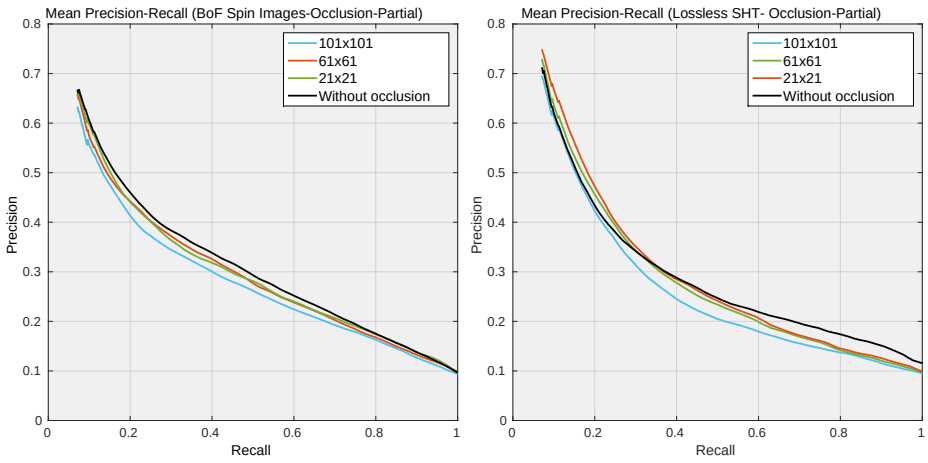

Figure 24: Occlusion effect. (*Left*) BoF Spin Images, (*right*) Lossless SHT.

hence, they are not powerful in distinguishing shapes. Although spin images do not generate unique descriptors, they have a higher discrimination power both in local and global configuration. Image based featured on shape index mapping achieves a descriptive power similar to spin images at the cost of additional processing to obtain mapping. For time crucial applications and large datasets global descriptors is the best choice. Because, memory and computation requirements are high for local descriptors.

In this work, local features are also adapted for global description. Spin images descriptor when it is utilized to describe objects globally performed slightly worse than its local counterpart. Conversely, Moments has a better performance on the global scale. D2 distributions descriptor performs almost same in both cases. Among the global approaches, our Lossless SHT method has the best PR curve. Global descriptors have rather simple formulation and they can be extended with additional features easily. However, global techniques suffer from generalizing class specific signatures and suffer from overfitting.

Computational complexity. Execution time comparison for Bag of Spin Images and Lossless SHT is given in Table 1. Although, our codes are not optimized for running time, the table gives a rough idea about computation times. Since BoF framework needs a dictionary prior to the stage of descriptor computation,

complexity for local descriptors is higher than their global counterparts. In addition, after building the dictionary, feature vectors from local regions are handled again to form a histogram representation by finding the closest *"word"* from the *"dictionary"*. The timing for construction of dictionary and BoF representations are directly proportional to the number of keypoints selected from the models. Therefore, we present two exemplar timing for two different keypoint sampling. Similarly, memory requirements are again related with BoF parameters, dictionary size, and the length of the descriptor vectors. Therefore, local descriptors are more space demanding than global ones. In Table 1, we report running times for the complete database; including the small overhead to compute PR curves and indexing. Similarity indexing is performed using the efficient Fast Library for Approximate Nearest Neighbors (FLANN) from the OpenCV library.

Table 1: Computation times

Bag of Spin Images		
Stage	**Keypoint step =1 Time (in sec)**	**Keypoint step =5 Time (in sec)**
Building Dictionary	3113.36	148.13
BoF Representation	1650.51	79.79
Indexing	0.65	0.71
PR curve	0.03	0.03
Total	**4764.58**	**228.66**

Lossless SHT	
Stage	**Time (in sec)**
Descriptor Extraction	152.24
Indexing	3.14
PR curve	0.03
Total	**155.41**

Advantages and disadvantages. In the light of the experimental results, we present a comparison table for local and global descriptors in Table 2. It summaries the *pros* and *cons* of the approaches for analysis of depth data. Robustness to noise and computational complexities favors global descriptors but local descriptors achieves better retrieval performance. Although local descriptors require additional effort for tuning the parameters, they are more discriminative than global descriptors based on the tested data, although the results could vary in some other data sets. On the other hand, scalability in

Table 2: Comparison of local and global descriptors for representing depth data.

	Local Descriptors	**Global Descriptors**
Pros	More discriminative	Robust to noise
	Better suited to handle clutter and occlusions	Simple to construst
		Scalable
Cons	Sensitive to noise	Overfitting
	Laborious parameter tuning	Sensitive to clutter and occlusion
	Computationally complex	

case of additional data can be handled easily with global description. Whereas, additional data could require constructing new dictionaries within the bag of features framework for local descriptors.

6. Conclusion

To conclude, similar performances could be achieved both with local and global feature extraction but slightly better retrieval performance is achieved with the local ones. Local description strategy provides flexibility in handling different scale properties therefore, tolerate incomplete information of depth data. However, computational complexity of local description is more complex than global approaches due to laborious parameter tuning. A disadvantage of global methods could be due to overfitting problem. In other words, they tend to describe objects in detail which sometimes makes feature vectors of similar objects draw apart in the descriptor space. On the other hand, local descriptors extract a summary from the query therefore, they have an increased capacity to generalize data. This could lead to underfitting problems if the number of "words" in the BoF framework is not tuned properly.

Finally, we would like to point out that global and local descriptors can be merged to achieve a higher performance as they are complementary. Future

work includes evaluating different strategies for combining descriptors and test the performance gain. Future work would also test the robustness of merged descriptors in case of noise and occlusion.

References

[1] Y. Wang, J. Feng, Z. Wu, J. Wang, S.-F. Chang, From low-cost depth sensors to cad: Cross-domain 3d shape retrieval via regression tree fields, in: Computer Vision–ECCV 2014, Springer, 2014, pp. 489–504.

[2] M. Bae, I. Park, Content-based 3d model retrieval using a single depth image from a low-cost 3d camera, The Visual Computer 29 (6-8) (2013) 555–564.

[3] J. Machado, A. Ferreira, P. B. Pascoal, M. Abdelrahman, M. Aono, M. El-Melegy, A. Farag, H. Johan, B. Li, Y. Lu, et al., Shrec'13 track: retrieval of objects captured with low-cost depth-sensing cameras, in: Proceedings of the Sixth Eurographics Workshop on 3D Object Retrieval, 2013, pp. 65–71.

[4] K. Lai, L. Bo, X. Ren, D. Fox, A large-scale hierarchical multi-view rgb-d object dataset, in: Robotics and Automation (ICRA), 2011 IEEE International Conference on, IEEE, 2011, pp. 1817–1824.

[5] Kinect, Kinect - xbox.com (Jan. 2013).
URL `http://www.xbox.com/en-US/Kinect/`

[6] C. Zhang, L. Wang, R. Yang, Semantic segmentation of urban scenes using dense depth maps, Computer Vision–ECCV 2010 (2010) 708–721.

[7] O. Akman, N. Bayramoglu, A. Alatan, P. Jonker, Utilization of spatial information for point cloud segmentation, in: 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2010, 2010, pp. 1 –4.

[8] N. Bayramoğlu, A. Alatan, Lossless description of 3d range models, Proc. of SPIE-IS&T Electronic Imaging Vol 8305 (2012) 83050B.

[9] E. Paquet, M. Rioux, A. Murching, T. Naveen, A. Tabatabai, Description of shape information for 2-d and 3-d objects, Signal Processing: Image Communication 16 (1) (2000) 103–122.

[10] M. Novotni, R. Klein, 3d zernike descriptors for content based shape retrieval, in: Proceedings of the eighth ACM symposium on Solid modeling and applications, ACM, 2003, pp. 216–225.

[11] M. Ankerst, G. Kastenmüller, H. Kriegel, T. Seidl, 3d shape histograms for similarity search and classification in spatial databases, in: Advances in Spatial Databases, Springer, 1999, pp. 207–226.

[12] M. Kazhdan, T. Funkhouser, S. Rusinkiewicz, Rotation invariant spherical harmonic representation of 3d shape descriptors, in: Proceedings of the 2003 Eurographics/ACM SIGGRAPH symposium on Geometry processing, Eurographics Association, 2003, pp. 156–164.

[13] T. Funkhouser, P. Min, M. Kazhdan, J. Chen, A. Halderman, D. Dobkin, D. Jacobs, A search engine for 3d models, ACM Transactions on Graphics (TOG) 22 (1) (2003) 83–105.

[14] D. Vranic, D. Saupe, J. Richter, Tools for 3d-object retrieval: Karhunen-loeve transform and spherical harmonics, in: Multimedia Signal Processing, 2001 IEEE Fourth Workshop on, IEEE, 2001, pp. 293–298.

[15] D. Vranic, An improvement of rotation invariant 3d-shape based on functions on concentric spheres, in: Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on, Vol. 3, IEEE, 2003, pp. III–757.

[16] R. Osada, T. Funkhouser, B. Chazelle, D. Dobkin, Shape distributions, ACM Transactions on Graphics (TOG) 21 (4) (2002) 807–832.

[17] M. Mahmoudi, G. Sapiro, Three-dimensional point cloud recognition via distributions of geometric distances, Graphical Models 71 (1) (2009) 22–31.

[18] T. Zaharia, F. Preteux, 3d-shape-based retrieval within the mpeg-7 framework, in: Photonics West 2001-Electronic Imaging, International Society for Optics and Photonics, 2001, pp. 133–145.

[19] F. Stein, G. Medioni, Structural indexing: Efficient 3-d object recognition, IEEE Trans. Pattern Anal. Machine Intell 14 (2) (1992) 125–145.

[20] C. Chua, R. Jarvis, Point signatures: A new representation for 3d object recognition, International Journal of Computer Vision 25 (1) (1997) 63–85.

[21] A. Johnson, M. Hebert, Using spin images for efficient object recognition in cluttered 3d scenes, Pattern Analysis and Machine Intelligence, IEEE Transactions on 21 (5) (1999) 433–449.

[22] G. Hetzel, B. Leibe, P. Levi, B. Schiele, 3d object recognition from range images using local feature histograms, in: IEEE Conference onComputer Vision and Pattern Recognition, 2001., Vol. 2, IEEE, pp. II–394.

[23] M. Pauly, R. Keiser, M. Gross, Multi-scale feature extraction on point-sampled surfaces, in: Computer graphics forum, Vol. 22, Wiley Online Library, 2003, pp. 281–289.

[24] J. Sun, M. Ovsjanikov, L. Guibas, A concise and provably informative multi-scale signature based on heat diffusion, in: Computer Graphics Forum, Vol. 28, Wiley Online Library, 2009, pp. 1383–1392.

[25] K. Gbal, J. Bærentzen, H. Aanæs, R. Larsen, Shape analysis using the auto diffusion function, in: Computer Graphics Forum, Vol. 28, Wiley Online Library, 2009, pp. 1405–1413.

[26] K. Liu, H. Skibbe, T. Schmidt, T. Blein, K. Palme, T. Brox, O. Ronneberger, Rotation-invariant hog descriptors using fourier analysis in polar and spherical coordinates, International Journal of Computer Vision 106 (3) (2014) 342–364.

[27] M. Hilaga, Y. Shinagawa, T. Kohmura, T. Kunii, 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, pp. 203–212.

[28] S. Biasotti, D. Giorgi, M. Spagnuolo, B. Falcidieno, Reeb graphs for shape analysis and applications, Theoretical Computer Science 392 (1) (2008) 5–22.

[29] H. Sundar, D. Silver, N. Gagvani, S. Dickinson, Skeleton based shape matching and retrieval, in: Shape Modeling International, 2003, IEEE, 2003, pp. 130–139.

[30] N. Cornea, D. Silver, X. Yuan, R. Balasubramanian, Computing hierarchical curve-skeletons of 3d objects, The Visual Computer 21 (11) (2005) 945–955.

[31] D. Chen, X. Tian, Y. Shen, M. Ouhyoung, On visual similarity based 3d model retrieval, in: Computer graphics forum, Vol. 22, Wiley Online Library, 2003, pp. 223–232.

[32] P. Daras, A. Axenopoulos, A compact multi-view descriptor for 3d object retrieval, in: Content-Based Multimedia Indexing, 2009. CBMI'09. Seventh International Workshop on, IEEE, 2009, pp. 115–119.

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

[34] P. Papadakis, I. Pratikakis, T. Theoharis, S. Perantonis, Panorama: A 3d shape descriptor based on panoramic views for unsupervised 3d object retrieval, International Journal of Computer Vision 89 (2) (2010) 177–192.

[35] Y. Guo, M. Bennamoun, F. Sohel, M. Lu, J. Wan, 3d object recognition in cluttered scenes with local surface features: A survey, Pattern Analysis and Machine Intelligence, IEEE Transactions on 36 (11) (2014) 2270–2287.

[36] J. Tangelder, R. Veltkamp, A survey of content based 3d shape retrieval methods, Multimedia Tools and Applications 39 (3) (2008) 441–471.

[37] B. Bustos, D. Keim, D. Saupe, T. Schreck, Content-based 3d object retrieval, Computer Graphics and Applications, IEEE 27 (4) (2007) 22–27.

[38] A. Mian, M. Bennamoun, R. Owens, Three-dimensional model-based object recognition and segmentation in cluttered scenes, Pattern Analysis and Machine Intelligence, IEEE Transactions on 28 (10) (2006) 1584–1601.

[39] H. Dutagaci, A. Godil, C. Cheung, T. Furuya, U. Hillenbrand, R. Ohbuchi, Shrec 2010-shape retrieval contest of range scans, in: Eurographics Workshop on 3D Object Retrieval, 2010.

[40] L. Bo, X. Ren, D. Fox, Depth Kernel Descriptors for Object Recognition, in: IROS, 2011.

[41] L. Bo, X. Ren, D. Fox, Unsupervised feature learning for rgb-d based object recognition, ISER, June.

[42] L. Alexandre, 3d descriptors for object and category recognition: a comparative evaluation, in: Workshop on Color-Depth Camera Fusion in Robotics at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vilamoura, Portugal, 2012.

[43] W. Wohlkinger, A. Aldoma, R. Rusu, M. Vincze, 3dnet: Large-scale object class recognition from cad models, in: Robotics and Automation (ICRA), 2012 IEEE International Conference on, 2012, pp. 5384 –5391.

[44] R. Gal, D. Cohen-Or, Salient geometric features for partial shape matching and similarity, ACM Transactions on Graphics (TOG) 25 (1) (2006) 130–150.

[45] P. Shilane, T. Funkhouser, Distinctive regions of 3d surfaces, ACM Transactions on Graphics (TOG) 26 (2) (2007) 7.

[46] W. Wohlkinger, M. Vincze, Shape-based depth image to 3d model matching and classification with inter-view similarity, in: Intelligent Robots and

Systems (IROS), 2011 IEEE/RSJ International Conference on, 2011, pp. 4865 –4870.

[47] D. Lowe, Distinctive image features from scale-invariant keypoints, International journal of computer vision 60 (2) (2004) 91–110.

[48] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, Speeded-up robust features (surf), Computer vision and image understanding 110 (3) (2008) 346–359.

[49] A. Bronstein, M. Bronstein, M. Ovsjanikov, Feature-based methods in 3d shape analysis, 3D Imaging, Analysis and Applications (2012) 185–219.

[50] D. Xu, H. Li, Geometric moment invariants, Pattern Recognition 41 (1) (2008) 240 – 249.

[51] J. Sivic, A. Zisserman, Video google: A text retrieval approach to object matching in videos, in: Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on, IEEE, 2003, pp. 1470–1477.

[52] N. Bayramoglu, A. A. Alatan, Shape index sift: Range image recognition using local features, in: Pattern Recognition (ICPR), 2010 20th International Conference on, IEEE, 2010, pp. 352–355.

[53] A. Mian, M. Bennamoun, R. Owens, On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes, International Journal of Computer Vision 89 (2) (2010) 348–361.

[54] S. Holzer, J. Shotton, P. Kohli, Learning to efficiently detect repeatable interest points in depth data, in: Computer Vision ECCV 2012, Vol. 7572 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2012, pp. 200–213.

[55] D. Healy, D. Rockmore, P. Kostelec, S. Moore, Ffts for the 2-sphere-improvements and variations, Journal of Fourier Analysis and Applications 9 (4) (2003) 341–385.