

# 3D alakfelismerés részleges pontfelhőkből

Zoltan Rozsa<sup>1</sup> and Tamas Sziranyi<sup>1,2</sup>

<sup>1</sup> Faculty of Transportation Engineering and Vehicle Engineering, Budapest University of Technology and Economics, Hungary

<sup>2</sup> Research Institute for Computer Science and Control (MTA SZTAKI), Budapest, Hungary

---

## Abstract

*3D-ben az alakfelismerés nagy kihívást jelent, kiváltképpen ha pontfelhőkkel dolgozunk. A felismerés részleges információkból egy másik kihívást jelentő probléma. Azonban a tárgyak felismerésére részleges 3D-s adatokból nagy szükség van az ipar különböző területein, mint automatizálás, megfigyelés, gép-gép és gép-ember robot rendszerek. Ez a tanulmány egy olyan rendszer létrehozásának követelményeit vizsgálja, ami részleges pontfelhőkből képes az alakfelismerésre, és megoldást is ajánl. Ez a megoldás magában foglalja a lokális skála becslést, kulcspont detekciót és egy lokális struktúra definícióját. A módszer hatékonyságát 3D-s adatbázisokon és Kinect pontfelhőkön történő tesztelés támasztja alá.*

---

## 1. Introduction

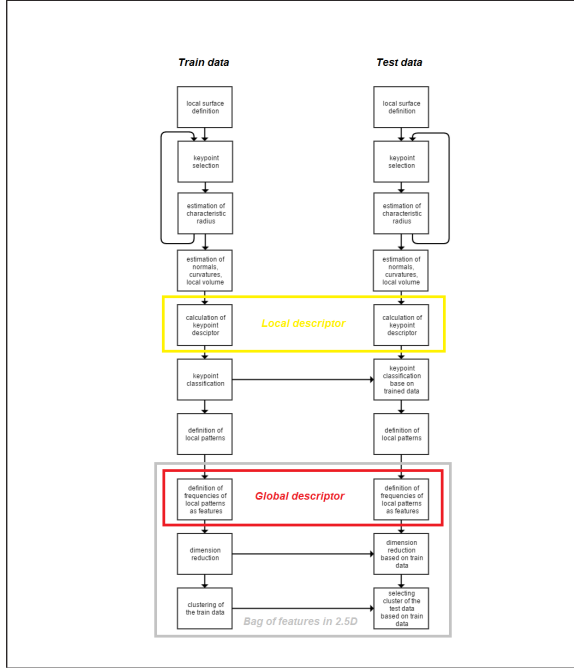
Point clouds have the potential of fusing them, creating even full 3D range images compared to depth ones. SLAM methods like dense scanning with Kinect<sup>1</sup> or monocular streaming<sup>2</sup> already provide the possibility of 3D scanning and reconstruction with cloud registration based on vision or surface descriptors<sup>3,4</sup> and aligning algorithms e.g. ICP<sup>5</sup>. However, in real scenarios of industry, observation with fixed camera position is lot more frequent (which still can have the advantage of merging clouds of moving objects changing over the time) than the possibility of full 3D reconstruction. Following the practical needs we will work on single-view/close view point cloud scanning methods, why we should critically reconsider the requirements and possibilities for this 2.5D scanning. One can distinguish dense and sparse clouds, in the following we concentrate the previous case (especially considering the cloud merging possibilities). The main research direction in this field is the object recognition and pose estimation, and there are many systems dealing with occlusion and clutter<sup>6,7,8</sup>. These pipelines has three major steps, which are keypoint detection, local surface description and hypothesis verification. A survey about 3D features and shape recognition can be found in<sup>9</sup>. Searching for keypoints with similar features is an expensive but successful way of recognizing a given object, but it is less prosperous if we look for a given type of objects which have distinct local features. Beside that applying the state of the art preprocessing (statistical outlier removal, smoothing, up-

sampling with MLS<sup>10</sup>) background<sup>11</sup> foreground subtraction<sup>12</sup> and segmentation<sup>13,14</sup> techniques the input clouds can be free of occlusion and clutter containing mainly the separated object parts.

Our goal was to perform object categorization based on partial point clouds. Some of the earlier SHREC<sup>†</sup> challenges are close related to our aim, however we did not intend to use any texture,<sup>15</sup> color<sup>16</sup> information (because it restricts the recording device) or adjust the method to rigid or non-rigid<sup>17</sup> shapes. In addition to the above we did not want our method to depend on meshing of objects<sup>18</sup> (because mesh generation step of point cloud processing can be more computationally intensive than working directly on the clouds and it results in the same recognition problems but with derived information) or relate it to 3D computer models<sup>19</sup> (which have very distinct features compared to real clouds), so our intent is categorization instead of retrieval. Finally the main difference between our objective is the size of the visible object part. Basically we investigate the problem of category recognition from as small part as possible, or what size is sufficient to obtain correct results and what requirements have to be fulfilled compared to full 3D shape or 2.5D (with high object content) recognition. Whether based only on an arm or a bust part a man can be recognized? We think that it has to be based on 'semi-global' description of

---

<sup>†</sup> <http://www.itl.nist.gov/iad/vug/sharp/contest/2015/Range/>



**Figure 1:** Flow chart of general partial view based object categorization

the object, because on one side the using of a local surface descriptor with too large radius to encode the small object parts may average the local features; on the other side the global structure characterized by global descriptors cannot be represented in the local structure.

## 2. Proposed system

In the following we define the construction elements and requirements for building a general system for partial view based object categorization. Figure 1 shows the general structure of the proposed system.

### 2.1. Data requirements

There are two important criteria of input data in order to proper operation. The scanned object part has to be segmented and the cloud density has to be sufficient (suitable to search local radius on detail levels). Obviously each sub-process (surface fitting, normal estimation, etc.) requires a minimum input point number. If we want to identify objects from 2.5D views, training with 3D clouds can lead to incorrect results if the radius of the examination is not carefully chosen: the point density and keypoint definition over- or under-sample the salient characteristic details. Hidden points (which are accounted in a full 3D but invisible in real 2.5D scanning) on the 2.5D view can influence both the keypoint finding and the descriptor in the full 3D model.

### 2.2. Characteristic scale

In 3D scale is one of the most important information which can result in quick and appropriate decision if one want to differentiate objects with different sizes. However if we only know partial cloud we do not know the size of the whole object so relative scale is not achievable, but absolute scale of the details can be well estimated. The solution to this problem lies in the definition of keypoints. We used the Harris 3D<sup>20</sup> keypoint detector for finding the salient feature points; however, we work on point clouds and not on meshes. Keypoints in our case defined as points at which the minimum of the principal Harris curvature has large value (this means both curvature of them are big, it is corner like point<sup>21</sup>). If we find one keypoint by assigning to it a characteristic radius, the next one can only be outside of the sphere defined by the radius. So instead of global scales we characterize local ones with this radius. The local scale can be estimated through finding the radius of Optimal neighborhood<sup>22</sup> by local shape variation<sup>23</sup>, or simply calculating a feature with length dimension e.g. from the curvature.<sup>24</sup> shows that the characteristic scale of complex and partially random structures may not always be identical to the optimum neighborhood size of covariance features. We defined our characteristic radius from the principal Harris curvatures as it is already calculated in the keypoint selection stage and it is always positive. However it was not self-evident how to do that. We used the following relations, starting from a case of a sphere:

$$\rho = \frac{1}{k_{p1}} = \frac{1}{k_{p2}} = \frac{1}{k_p} \quad (1)$$

where  $\rho$  is the radius of the sphere and  $k_{p1} = k_{p2} = k_p$  are the principal curvatures of it.

$$\lambda_1 = \lambda_2 = 2k_{p2}^2 \quad (2)$$

where  $\lambda_1 = \lambda_2$  are the eigenvalues of the Harris matrix (principal Harris curvatures). Equation 2 can be deduced from the definition of 3D Harris and by calculating the principal curvatures of the fitted surface operator in<sup>20</sup> in the origin with the formulas of elementary differential geometry.  $\rho$  is the value what we intend to use as corresponding radius of the keypoint. However, in general case  $k_{p1} \geq k_{p2}$ , so we can define two radii instead of one. The obvious choice of the characteristic radius corresponds to the smaller curvature  $k_{p2}$ , because we sorted the points to descending order as function of it. So searching for less relevant keypoints to smaller curvature we get monotonically increasing characteristic radius.

$$\rho_1 = \frac{2}{\sqrt{\lambda_1}} \quad (3)$$

This local radius determines only the number of keypoints we will find, but there are two important things in the definition of this feature. First it should be repeatable in order to locate similar structures in different clouds; second, it is recommended that locations of higher curvature should be more relevant as keypoints (also to be avoided of finding many keypoints in a flat region).

### 2.3. Local descriptors

Local surface descriptors have a wide range of literature but relatively few of them can be applied in point cloud processing, they are rather applicable to depth image<sup>25 26</sup> or mesh<sup>27 28</sup> type range images. A comparative evaluation of applying some of them for categorization purposes can be found in<sup>29</sup>. ThrIFT descriptor is a good feature of the local 3D structure, also can be used for recognition in true 3D structures. The descriptor constructs a 1D weighted histogram according to surface normal angles of neighboring points, which is defined as the angle between two normal vectors calculated with different window size at a given point<sup>30</sup>.

$$\cos(\theta_s) = \frac{n_{small} \cdot n_{large}}{\|n_{small}\| \|n_{large}\|} \quad (4)$$

Another local surface descriptor is the so called Point Feature Histogram (PFH) proposed by Rusu et al.<sup>31</sup> and<sup>32</sup>. For all point pairs in the local neighborhood after selecting the  $p_s$  source and  $p_t$  target points with  $n_s$  and  $n_t$  associated normal vectors<sup>33</sup> a unique Darboux frame is defined in  $p_s$  origin with the basis vectors:

$$\begin{aligned} u &= n_s \\ v &= u \times \frac{p_t - p_s}{\|p_t - p_s\|} \\ w &= u \times v \end{aligned} \quad (5)$$

This descriptor uses four measures to accumulate the neighboring points into a 16 bin histogram. Later FPFH (Fast Point Feature Histogram) was proposed in order to reduce computational complexity<sup>34</sup>. In our local descriptor keypoints are characterized by Point Feature Histogram to maintain maximum discriminative power. PFH is density invariant. Components of the local descriptor we used:

- PFH with 8 bin using the measures<sup>33</sup> :

$$\begin{aligned} \alpha &= w \cdot n_t \\ \phi &= u \cdot \frac{p_t - p_s}{\|p_t - p_s\|} \\ \theta &= \arctan(w \cdot n_t, u \cdot n_t) \end{aligned} \quad (6)$$

- Surface normal angle calculated at the keypoint
- Modified shape index value:

$$I_p = \frac{1}{2} - \frac{1}{\pi} \arctan \frac{k_{p1} + k_{p2}}{k_{p1} - k_{p2}}, \quad (7)$$

$$I_{pmod} = \left| \frac{1}{2} - I_p \right| \quad (8)$$

where  $I_p$  is the shape index and  $k_{p1} \geq k_{p2}$ <sup>35</sup>. The modified value is a mapping from [0 1] interval to [0 0.5] in order to retain only relative orientation information between the curvatures (the original shape index is influenced by definition of the local reference frame as well).

- Characteristic radius calculated from the maximum of Harris principal curvature

$$\rho_2 = \frac{2}{\sqrt{\lambda_2}} \quad (9)$$

- Volume of the convex hull determined by the neighboring points

Thus our descriptor has 12 dimensions. Why did we choose these features? The shape index instantly assigns a type (spherical cup or cap, rut or ridge, saddle) to the keypoint itself by examining the proportion and relative orientation of its curvatures. Complementing this PFH is the basis of our descriptor as it is density invariant generalization of the curvature. We gathered information about the point, the local surface (defined by the search radius) and using surface normal angle we store outside of the given neighborhood too.

- Storing a characteristic *radius value* we preserve the scale information of the curvatures which is neglected in the shape index calculation. Here we calculate  $\rho_2$  by Equation 9 instead of  $\rho_1$  (Equation 3), because in optimal case we found mainly significant keypoints. Based on our definition the significance of a keypoint is inverse proportional to  $\rho_1$ , so in terms of  $\rho_2$  larger variance is expected.
- Finally, as we are aware of corresponding *local volume* was not included in any earlier descriptor as feature, but we found it discriminative. Subjecting the standardized features to Principal Component Analysis (PCA) the average loading of the original features constructing the first four component (which has about 75 % share of the sum of eigenvalues) of the features in the transformed space is about 5-12 %. So we concluded they are linearly independent and each of them has significance.

### 2.4. Global descriptors

In object categorization global descriptors have significant role. Local descriptors are adequate for local surface matching, from which real objects are built up. So instead of exhaustive search of all the local features, matching the descriptors of the whole object is done. These can be classified into four main categories: histogram (distribution) based, transform based, 2D view based and graph based ones<sup>36</sup>. One of the histogram based approaches is the descriptor of shape distributions proposed by<sup>37</sup>. Histograms based on measured shape functions as distance (between a surface point and the center of mass of the model, between two surface points), angle (of three surface points), area (of the triangle formed by three surface points) and volume (of the tetrahedron defined by four surface points) are appropriate to distinguish broad categories or for pre-classification step. Another distribution based technique is the so called Shape context<sup>38</sup> which is similarly defined as its 2D version.<sup>39</sup> For one point it is a histogram of the relative coordinates of the remaining points. According to<sup>40</sup> the 3D shape contexts descriptor is less efficient than the other currently available

methods, its indexing is not straightforward, and the dissimilarity measure obtained from the method does not obey the triangle inequality. Other representatives of this category are Extended Gaussian Images (EGI) or orientation histogram<sup>41</sup> (which is the mapping of objects' normals to the Gaussian sphere considering the area of surface) and Viewpoint Feature Histogram (VFH)<sup>42</sup>. The application of previous one is pose determination because it is not invariant to rotation, its main disadvantage is noise sensitivity. VFH is also used for pose estimation, so it is designed not to be rotation and translation invariant. In<sup>43</sup> VFH was extended to the so called clustered version of it (CVFH), which is less sensitive to missing parts by applying the original method on the connected components after smooth region segmentation.

The transform based methods provides pose invariance and compact representation by transforming the geometric information into different domains. Organized structure (spherical grid or 3D voxel) is required to accomplish it. For example 3D Fourier Transform<sup>44</sup> requires voxelization using the bounding cube, what is outlier sensitive. This category also covers method like Angular Radial Transform (ART)<sup>45</sup>, Spherical Harmonics (SH)<sup>46</sup> and Spherical Trace Transform (STT). The STT is proved to be one the best transformed base algorithms in terms of retrieval accuracy<sup>47</sup>.

In case of 2D view-based global descriptors the 3D surface is transformed into a set of 2D projections, and the goal is the computation of 2D image descriptors on each view. These methods can rely the Bag-of-Feature of local visual features like SIFT<sup>48</sup> or on 2D global descriptor like Zernike moments which is among the most successful representations<sup>49</sup>. One problem of rendering 3D into 2D views is the theoretically infinite possible number of views<sup>50</sup>.

In case of graph based methods a graph is built out of the surface which is transformed to a numerical description. Reeb graphs are defined over the object surface at multiple levels of resolution of Reeb functions like curvature, height or integrated geodesic distance. The choice of function is determinative and it is not applicable to all classes of shapes.<sup>51</sup> Sundar et al.<sup>52</sup> propose a technique comparing 3D objects with the help of skeletal graph matching.

In general consequently from their definition the global methods are either not applicable to partial matching (because global attributes are not present) or requires exhaustive search (subgraph matching of graph based methods)<sup>53</sup>. There were attempts to single view based categorization like<sup>54</sup> or<sup>55</sup> but encoding global information does not allow the recognition of partial objects. We propose the use of descriptor which is based on local patterns instead of points or surface patches, so it stores semi-global information. Among others we tested a method combines the advantages of distribution and graph based approaches by well defined similarity of the local patterns (colored subgraphs), more detailed description can be found in the next section.

### 3. Main Steps in the Procedure

As we mentioned earlier we examine the criteria of successful object recognition, so we propose a general idea of achieving that instead of defining a strict structure of elements. We outline a robust and scale-independent methodology for the different kind of sensors, shapes, circumstances. The description of the tested methods can be found in this section.

#### 3.1. VFH

The traditional approach of categorizing partial objects is based on Clusters. This makes VFH<sup>42</sup> sensitive to missing parts. CVFH<sup>43</sup> is less sensitive to these defects, it is able to recognize theoretically if one cluster match is found. However it is obvious there are limitations of missing points and percentage of visible object part. We tested VFH recognition evolution on our database, because it is efficient to compute based on viewpoint direction and surface shape component.

#### 3.2. Local Pattern definition

Our approach for realizing category recognition of very partial clouds is based on local patterns which encode the semi-global structure of the object. Theoretically only the size of local patterns limits the minimum size of the object from which it is recognizable. Another advantage of using structural information is that the whole object can be build up from local patterns, so missing parts causes less trouble. Systems we tested uses the keypoint detection based on our characteristic radius definition. They can be divided into two main parts. The first one deals with the definition of local patterns.

##### 3.2.1. Local 3D surface definitions around points

The first substep is the definition of local surface, it will be represented all the neighboring points within a predefined radius (it depends on the density of the cloud, but it is worth to be choose as small radius as possible because it predetermines the scale of the finest feature we detect and also the precision of the estimated normal vector).

##### 3.2.2. Finding Keypoints

After fitting a surface to the local neighborhood we searched for points with the help of Harris operator. The high eigenvalues of Harris matrix indicates keypoints. Based on the proposed methods in<sup>8</sup>, one can compare 3D point clouds to that of 2.5 single view scanning data. We use one of the ideas of this paper in our 2.5D - 2.5D comparison case, boundary points and its neighborhood in a given radius should not be picked up as keypoints because of unknown environment.

### 3.2.3. Curvature and scale calculus

From the fitted surface for describing the local features in the 3D point clouds, we are to use differential geometry on the estimated surface, getting 3D curvature and second order functions in the local neighborhood<sup>23</sup>. Local scale can be defined in various ways mentioned earlier.

### 3.2.4. Local normal vectors and descriptors

The estimation of normal vectors happens through performing PCA on the coordinates. If we deal with point cloud computed from one camera view and the viewpoint is known we simply orient these vectors to the camera direction. In case of unknown camera-point of scanning (fused point-clouds from different views) we use *Normal orientation propagation* for finding the direction of normal vectors of local surfaces<sup>56</sup>. When Harris eigenvalues, principal curvatures and normal vectors have been computed as described in the previous section can be calculated.

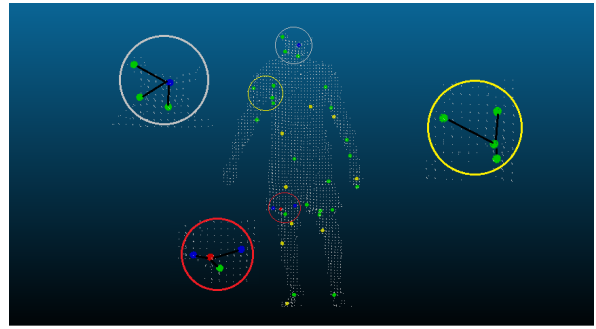
### 3.2.5. Classification of Keypoints with K-means

When the descriptors are determined for all the keypoints of the training set, after transforming the measurements into an orthogonal space with PCA and choosing the necessary dimension, clustering of the keypoints is done. Naturally, the number of appropriate clusters depends on the dataset.

### 3.2.6. Local pattern of neighboring Keypoints

Three different system is tested, which are separated in this substep.

- Bag of Keypoints (BoK): Simply the keypoints themselves (with the inherent local surface information) were chosen as local features.
- Bag of Graphs (BoG): In case of each keypoint, local patterns are defined as colored graphs with a center of  $p$  keypoint and ordered (according to cluster number, not distance) list of 3 closest keypoints. Number of keypoints constituting the different subgraphs have to be identical in favor of simple comparison and avoiding computational requirements of subgraph matching problem. It was chosen to be 4 because volume can be assigned to 4 points. So one more feature is added to this graph descriptor. From the four points of a graph a tetrahedron volume is calculated, considering all graphs in the training set these volumes are clustered and the fifth feature of this local pattern will be this cluster number of the graph volume. Figure 2 illustrates the definition of the local patterns. Semi-global information is stored.
- Based on the idea of Global Structure Histogram (GSH)<sup>54</sup>, we defined Keypoint based Global structure Histogram (further referred as KGH), where keypoint-cluster pair distances formed global patterns.



**Figure 2:** Definition of local patterns illustrated on a Kinect point cloud

### 3.2.7. Classification issues of local patterns

It is hard to define similarity between the local patterns (formed by colored subgraphs) because of the cardinality. In case of high keypoint cluster number (which is needed because distinct features are present) or high graph volume cluster number (which is needed because similar features are present in different scales) there are too many patterns to deal with. The great number of possible feature clusters can be partly manageable if we use dimensionality reduction (e.g. PCA) and statistical grouping as the Bag of Features.

## 3.3. Bag of Features in 2.5D

The second big part of the system tries to solve the matching of the found local patterns. These steps are taken in case of each subsystem introduced in previous.

### 3.3.1. Local patterns as features

The statistics of these local patterns can be a global descriptor of the category, where partial views give back the partial of this statistic. Counting the numbers of local patterns in the training measurements from all these features gives our final descriptor.

### 3.3.2. Dimension reduction with PCA

Unfortunately, as we mentioned earlier the number of these features can be very high to deal with. In case of one measurement one can count a very sparse vector. That is why PCA is used for dimension reduction on the measurements of frequency of features of the training set.

### 3.3.3. Clustering of features by K-means

After dimension reduction clustering of training measurements happens in order to determine cluster centers to which test measurements will be compared.

## 3.4. Experimental conditions and test environment

For testing the systems part of<sup>57</sup> dataset were used representing the models by the vertices of the meshes. In case of



this dataset this was possible because of large point (vertex) density without the simulation of 3D scanning. Compared to other meshed model databases which represents planes with only boundary points. From the dataset we collected categories with comparable sizes (human, chair, table, angular machine part, quadruped), resized the objects to real life measure and finally downsized the object numbers of categories to the amount where affiliation of objects can be identified by human. From the meshed models we left only the vertices to get full 3D point clouds and we generated the one-view 2.5D point-cloud from the 3D cloud by using Hidden Point Removal (HPR) <sup>58</sup> from 6 different viewpoints for each object. Figure 3 shows the original point cloud and two clouds from our dataset (HPR operator applied) for one object. For a sample object a more (Figure 4/b) and a less (Figure 4/c) informative cloud is presented here for illustrating the variety and challenge of the database.

#### 4. Evaluation

In order to simple comparison of the methods in the decision stage only one cluster center was chosen to represent a category. The final decision was made based on L1 distance, which is recommended for global recognition pipeline using VFH descriptor as it is more robust to occlusion than L2. Table 1 shows the overall recognition rate is about 80% for the different methods on the set, these views covered maximum 45 % percent of the full 3D cloud. VFH has the lowest score on the test of full views, this shows this views very already challenging. The follow up tables (Table 2, 3) were made by top-to-bottom exploring sequence. 10 visible keypoints stage corresponds about 10 % while 25 keypoints stage covered cc. 20 % of the whole 3D cloud. From these one can conclude that even in the very beginning of shape exploration we can get useful predictions, from almost all methods. KGH is the least suitable for this. As it was expected, from the stored global structure information very few percent present in the very partial view. To reach these scores the KGH descriptor was constructed as a 64x15 'matrix' by 8 keypoint cluster. In case of BoK 60 keypoint cluster were defined and in case of BoG from 6 keypoint cluster and 5 graph volume cluster 1032 type graph were defined in the training set, which was reduced to a 65 dimension descriptor. We experienced with BoK and BoG that if we deal with categories of very different average keypoint numbers it is worth to normalize the descriptors in the beginning of exploration. By this category recognition can be improved by after a high number of keypoints the normalization can be omitted. One reason of the phenomenon is that reaching a stage where many keypoints are seen (which only arises in specific categories) we can keep this information to help the decision. Later, toward to the whole shape (but not knowing the whole size), the decision is consolidating (see Figure 5), and we reach a comparable good result to that of the other methods of limited view of the whole shapes (<sup>54,55</sup>).

Figure 5 shows the class variation of successful categorizations as a function of percent of full 3D point cloud. When only few percent of the object is visible the hesitation between categories is significant, but as it (and the number of local patterns) grows as the hesitation decrease. Horizontal lines in each category (except for table which resulted poor categorization rate) indicating the decision did not change after a given point. By contrast in Figure 6 which diagram shows curves with false final decision; consolidated horizontal lines are much less frequent.

The evolution of the classification process is characteristic for the quality of the possible decision: average *category change* of stages resulted in final correct / incorrect decision are significantly different. This change average is 0.41 / 0.46 at 10 keypoints stage; 1.3 / 1.6 at 25 keypoints stage; and 2.8 / 3.9 at stage of 55 keypoints.

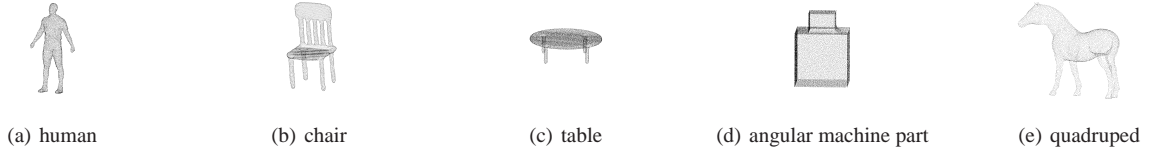
#### 5. Conclusions

We compare here the average efficiency for the different methods:

- For cc. 10 % of visible points:  
VFH: 57 % KGH: 32 % BoK: **64** % BoG: 54 %,
- For cc. 20 % of visible points:  
VFH: 60 % KGH: 44 % BoK: **68** % BoG: **66** %.
- For full one-view (max. 45% of points):  
VFH: 64 % KGH: 89 % BoK: **89** % BoG: **83** %.

This shows that GSH like descriptor (KGH) may also characterize well the full one-views with its global information as it was shown in <sup>54</sup>, but our novel proposed methods (BoK and BoG) based on scale selected keypoints have similar efficiency at the full one-view case. Looking at the case of partial body scans the novel methods outperforms the others; addressing this limited view is our main challenge. The presented subgraph based new object description is able to characterize objects from partial information and inherits the possibility of the semantic information based object categorization which is our future research goal. The method has good prediction (early recognition) results on very partial 2.5D views. We expect even better results in big databases, which will be tested in the near future and developing decision method based on tendency of object exploring is also planned. The new method introduces local descriptors, like:

- local scale, independently of the whole shape and size,
- keypoints tailored by local scale,
- local histogram data,
- clustered keypoints,
- local graph descriptors with scale dependence in radius,
- volume of the local graph,
- evolution tendency based on the detected features as additional feature for endurance test.


**Figure 3: Original full 3D point clouds**
**Table 1: Categorization results (in%) on the full one-view test clouds (2.5D, max. 45% of real 3D points)**

 Viewpoint Feature Histogram (VFH)<sup>43</sup>, Keypoint based Global struct.Hist.(KGH)<sup>54</sup>, new: Bag of Keypoints (BoK), Bag of Graphs (BoG)

Query / Re- sult	Human				Chair				Table				Angular machine part				Quadruped			
	VFH	KGH	BoK	BoG	VFH	KGH	BoK	BoG	VFH	KGH	BoK	BoG	VFH	KGH	BoK	BoG	VFH	KGH	BoK	BoG
Human	28	82	<b>88</b>	67	17	6	6	17	0	6	0	11	0	6	0	0	55	0	6	5
Chair	0	0	0	0	89	<b>100</b>	<b>100</b>	94	0	0	0	6	0	0	0	0	11	0	0	0
Table	6	0	0	0	44	17	33	28	22	<b>72</b>	67	67	0	11	0	5	28	0	0	0
Angular machine part	0	0	0	0	0	6	0	6	0	0	0	0	83	94	<b>100</b>	94	17	0	0	0
Quadruped	0	6	11	6	0	0	0	0	0	0	0	0	0	0	0	0	<b>100</b>	94	89	94

**Table 2: Categorization results (in%) on the single view in stage of 10 visible keypoints / partial shape**

Query / Re- sult	Human				Chair				Table				Angular machine part				Quadruped			
	VFH	KGH	BoK	BoG	VFH	KGH	BoK	BoG	VFH	KGH	BoK	BoG	VFH	KGH	BoK	BoG	VFH	KGH	BoK	BoG
Human	61	0	56	<b>67</b>	17	6	11	11	0	94	0	10	0	0	0	6	22	0	33	6
Chair	0	0	6	11	83	72	<b>94</b>	72	0	28	0	11	0	0	0	0	17	0	0	6
Table	6	0	44	33	72	11	22	11	17	<b>89</b>	17	33	0	0	11	17	5	0	6	6
Angular machine part	0	0	0	0	22	22	11	6	0	78	0	5	61	0	<b>89</b>	<b>89</b>	17	0	0	0
Quadruped	22	0	16	39	17	17	6	6	0	83	11	28	0	0	0	16	61	0	<b>67</b>	11

**Table 3: Categorization results (in%) on the single view in stage of 25 visible keypoints / partial shape**

Query / Re- sult	Human				Chair				Table				Angular machine part				Quadruped			
	VFH	KGH	BoK	BoG	VFH	KGH	BoK	BoG	VFH	KGH	BoK	BoG	VFH	KGH	BoK	BoG	VFH	KGH	BoK	BoG
Human	56	0	39	<b>61</b>	11	83	11	6	0	17	0	6	0	0	0	11	33	0	50	16
Chair	0	0	0	5	89	<b>100</b>	94	89	0	0	0	6	0	0	6	0	11	0	0	0
Table	6	0	28	17	50	17	33	11	16	<b>72</b>	28	50	0	11	11	16	28	0	0	6
Angular machine part	0	0	0	0	16	50	11	17	0	0	0	0	67	50	<b>89</b>	78	17	0	0	5
Quadruped	16	0	11	44	6	72	0	0	0	28	0	6	0	0	0	0	78	0	<b>89</b>	50



Figure 4: Point clouds after HPR

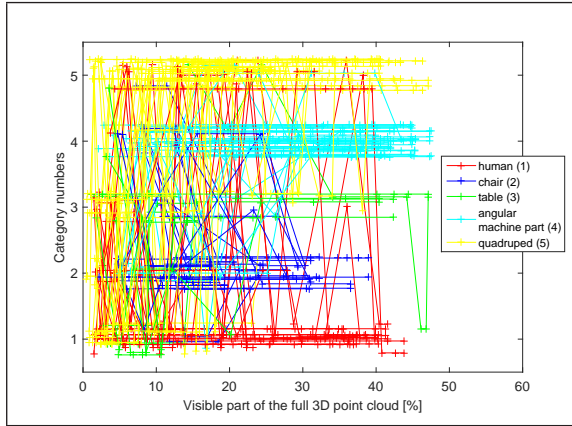


Figure 5: Evolution of recognition stages resulted correct categorization as function of visible part of the full 3D cloud, BoG method

Acknowledgements

This work has been supported by the Hungarian Research Fund, OTKA #106374.

References

1. R. A. Newcombe and et al., “Kinectfusion: Real-time dense surface mapping and tracking,” in *IEEE ISMAR*. IEEE, October 2011. 1
2. J. Engel, T. Schöps, and D. Cremers, “LSD-SLAM: Large-scale direct monocular SLAM,” in *European Conference on Computer Vision (ECCV)*, September 2014. 1
3. R. Kidson, D. Stanimirovic, D. Pangercic, and M. Beetz, “Elaborative evaluation of rgb-d based point cloud registration for personal robots,” in *ICRA 2012 Workshop on Semantic Perception and Mapping for Knowledge-enabled Service Robotics*, St. Paul, MN, USA, May 14–18 2012. 1
4. R. Hänsch, T. Weber, and O. Hellwich, “Comparison of 3D interest point detectors and descriptors for point cloud fusion,” *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 57–64, Aug. 2014. 1
5. P. Besl and N. D. McKay, “A method for registration of 3-d shapes,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 14, no. 2, pp. 239–256, Feb 1992. 1

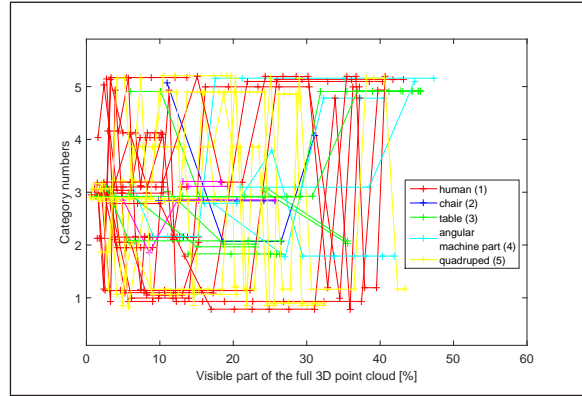


Figure 6: Evolution of recognition stages resulted incorrect categorization as function of visible part of the full 3D cloud, BoG method

6. Y. Guo, F. Sohel, M. Bennamoun, J. Wan, and M. Lu, “Rops: A local feature descriptor for 3d rigid objects based on rotational projection statistics,” in *Proc. 1st Int. Conf. Commun., Signal Process., Appl.*, 2013, pp. 1–6. 1
7. E. Rodolà, A. Albarelli, F. Bergamasco, and A. Torsello, “A scale independent selection process for 3d object recognition in cluttered scenes,” *International Journal of Computer Vision*, vol. 102, no. 1-3, pp. 129–145, 2013. 1
8. 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, no. 2-3, pp. 348–361, 2010. 1, 4
9. Y. Guo, M. Bennamoun, F. Sohel, M. Lu, and J. Wan, “3d object recognition in cluttered scenes with local surface features: A survey,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, no. 11, pp. 2270–2287, Nov 2014. 1
10. R. Rusu and S. Cousins, “3d is here: Point cloud library (pcl),” in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, May 2011, pp. 1–4. 1
11. K. Greff, A. Brandao, S. Krauß, D. Stricker, and E. Clua, “A comparison between background subtraction algorithms using depth images with a consumer camera,” *International Conference on Computer Vision Theory and Applications - VISAPP 2012*, 2 2012, poster. 1
12. K. Litomisky and B. Bhanu, “Removing moving objects from point cloud scenes,” in *Advances in Depth Image Analysis and Appl.*, ser. Lecture Notes in Computer Science, X. Jiang, O. Bellon, D. Goldgof, and T. Oishi, Eds. Springer, 2013, vol. 7854, pp. 50–58. 1
13. C. Benedek, D. Molnár, and T. Szirányi, “A dynamic MRF model for foreground detection on range data se-



- quences of rotating multi-beam Lidar,” in *International Workshop on Depth Image Analysis (WDIA)*, ser. Lecture Notes in Computer Science, Tsukuba City, Japan, 2012, vol. 7854, pp. 87–96. 1
14. A. Nguyen and B. Le, “3d point cloud segmentation: A survey,” in *Robotics, Automation and Mechatronics (RAM), 2013 6th IEEE Conference on*, Nov 2013, pp. 225–230. 1
  15. A. Giachetti and et al., “Retrieval of Non-rigid (textured) Shapes Using Low Quality 3D Models,” in *Eurographics Workshop on 3D Object Retrieval*, I. Pratikakis, M. Spagnuolo, T. Theoharis, L. V. Gool, and R. Veltkamp, Eds. The Eurographics Association, 2015. 1
  16. Y. Gao and et al., “3D Object Retrieval with Multimodal Views,” in *Eurographics Workshop on 3D Object Retrieval*, I. Pratikakis, M. Spagnuolo, T. Theoharis, L. V. Gool, and R. Veltkamp, Eds. The Eurographics Association, 2015. 1
  17. Z. Lian and et al., “Non-rigid 3D Shape Retrieval,” in *Eurographics Workshop on 3D Object Retrieval*, I. Pratikakis, M. Spagnuolo, T. Theoharis, L. V. Gool, and R. Veltkamp, Eds. The Eurographics Association, 2015. 1
  18. P. B. Pascoal and et al., “Retrieval of Objects Captured with Kinect One Camera,” in *Eurographics Workshop on 3D Object Retrieval*, I. Pratikakis, M. Spagnuolo, T. Theoharis, L. V. Gool, and R. Veltkamp, Eds. The Eurographics Association, 2015. 1
  19. A. Godil and et al., “Range Scans based 3D Shape Retrieval,” in *Eurographics Workshop on 3D Object Retrieval*. The Eurographics Assoc., 2015. 1
  20. I. Sipiran and B. Bustos, “Harris 3d: A robust extension of the Harris operator for interest point detection on 3d meshes,” *Vis. Comput.*, vol. 27, no. 11, pp. 963–976, Nov. 2011. 2
  21. A. Kovacs and T. Sziranyi, “Harris function based active contour external force for image segmentation,” *Pattern Recognition Letters*, vol. 33, no. 9, pp. 1180 – 1187, 2012. 2
  22. M. Weinmann, B. Jutzi, S. Hinz, and C. Mallet, “Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers,” *[ISPRS] Journal of Photogrammetry and Remote Sensing*, vol. 105, pp. 286 – 304, 2015. 2
  23. H. T. Ho and D. Gibbins, “Multi-scale feature extraction for 3d models using local surface curvature,” in *Digital Image Computing: Techniques and Applications (DICTA), 2008*, Dec 2008, pp. 16–23. 2, 5
  24. R. Blomley, M. Weinmann, J. Leitloff, and B. Jutzi, “Shape distribution features for point cloud analysis - a geometric histogram approach on multiple scales,” *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, pp. 9–16, Aug. 2014. 2
  25. E. R. do Nascimento, G. L. Oliveira, A. W. Vieira, and M. F. Campos, “On the development of a robust, fast and lightweight keypoint descriptor,” *Neurocomputing*, vol. 120, pp. 141 – 155, 2013, image Feature Detection and Description. 3
  26. 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, no. 12, pp. 1235 – 1250, 2009, special issue on 3D Representation for Object and Scene Recognition. 3
  27. J. Knopp, M. Prasad, G. Willems, R. Timofte, and L. Van Gool, “Hough transform and 3d surf for robust three dimensional classification,” in *Proceedings of the 11th European Conference on Computer Vision: Part VI*, ser. ECCV’10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 589–602. 3
  28. F. Tombari, S. Salti, and L. Di Stefano, “A combined texture-shape descriptor for enhanced 3d feature matching,” in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, Sept 2011, pp. 809–812. 3
  29. L. A. 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, October 2012. 3
  30. A. Flint, A. Dick, and A. V. D. Hengel, “Thrift: Local 3d structure recognition,” in *Digital Image Computing Techniques and Applications, 9th Biennial Conference of the Australian Pattern Recognition Society on*, Dec 2007, pp. 182–188. 3
  31. R. Rusu, Z. Marton, N. Blodow, and M. Beetz, “Learning informative point classes for the acquisition of object model maps,” in *Control, Automation, Robotics and Vision, 2008. ICARCV 2008. 10th International Conference on*, Dec 2008, pp. 643–650. 3
  32. R. B. Rusu, Z. C. Marton, N. Blodow, and M. Beetz, “Persistent point feature histograms for 3d point clouds,” in *In Proceedings of the 10th International Conference on Intelligent Autonomous Systems (IAS-10, 2008)*. 3
  33. R. B. Rusu, “Semantic 3d object maps for everyday manipulation in human living environments,” Ph.D. dissertation, Computer Science department, Technische Universitaet Muenchen, Germany, October 2009. 3
  34. R. B. Rusu, N. Blodow, Z. C. Marton, and M. Beetz, “Aligning point cloud views using persistent feature histograms,” in *Proceedings of the 21st IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, september 2008, p. 22a–26. 3
  35. T. Zaharia and F. Preteux, “Three-dimensional shape-based retrieval within the MPEG-7 framework,” in *Proceedings SPIE Conference on Nonlinear Image Pro-*

- cessing and Pattern Analysis XII, vol. 4304, Jan. 2001, pp. 133–145. 3
36. R. Paulsen and K. Pedersen, *Image analysis: 19th Scandinavian Conference, SCIA 2015, Copenhagen, Denmark, June 15-17, 2015. Proceedings*. Springer, 2015. 3
  37. R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin, “Shape distributions,” *ACM Transactions on Graphics*, vol. 21 (4), pp. 807–832, 2002. 3
  38. M. Körtgen, G.-J. Park, M. Novotni, and R. Klein, “3d shape matching with 3d shape contexts,” in *The 7th Central European Seminar on Computer Graphics*, april 2003. 3
  39. S. Belongie, J. Malik, and J. Puzicha, “Shape matching and object recognition using shape contexts,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 4, pp. 509–522, Apr 2002. 3
  40. Zhang L., da Fonseca M., Ferreira A., “Survey on 3d shape descriptors,” Fundação para a Ciência e a Tecnologia, Lisboa, Portugal, Tech. Rep. Technical Report, DecorAR (FCT POSC/EIA/59938/2004), 2007. 3
  41. B. K. P. Horn, “Extended gaussian images,” in *Proc. of the IEEE*, 72, 1984, p. 1671–1686. 4
  42. R. B. Rusu, G. Bradski, R. Thibaux, and J. Hsu, “Fast 3d recognition and pose using the viewpoint feature histogram,” in *Proc. of the 23rd IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2010. 4
  43. A. Aldoma, N. Blodow, D. Gossow, S. Gedikli, R. Rusu, M. Vincze, G. Bradski, “Cad-model recognition and 6 dof pose estimation using 3d cues,” in *ICCV 2011, 3D Representation and Recognition (3dRR11)*, 2011, pp. 585–592. 4, 7
  44. D. V. Vrančić and D. Saupe, “3d shape descriptor based on 3d fourier transform,” in *Proc. of the EURASIP Conference on Digital Signal Processing for Multimedia Communications and Services (ECMCS 2001)*, september 2001. 4
  45. J. Ricard, D. Coeurjolly, and A. Baskurt, “Generalizations of angular radial transform for 2d and 3d shape retrieval,” *Pattern Recognition Letters*, vol. vol. 26, no. 14, pp. 2174–2186, 2005. 4
  46. M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz, “Rotation invariant spherical harmonic representation of 3d shape descriptors,” in *Proc. 2003 Eurographics/ACM SIGGRAPH Symp. Geometry Processing*, 2003, pp. 156–164. 4
  47. D. Zarpalas, P. Daras, A. Axenopoulos, D. Tzovaras, M.G. Strintzis, “3d model search and retrieval using the spherical trace transform,” *EURASIP J. Advances in Signal Processing*, 2007. 4
  48. R. Ohbuchi, K. Osada, T. Furuya, T. Banno, “Salient local visual features for shape-based 3d model retrieval,” in *Proc. IEEE Shape Modeling International (SMI)*, 2008. 4
  49. H. Chen, B. Bhanu, “3d free-form object recognition in range images using local surface patches,” *Pattern Recognition Letters*, vol. 8(10), pp. 1252–1262, 2007. 4
  50. H. Dutagaci, A. Godil, B. Sankur, and Y. Yemez, “View subspaces for indexing and retrieval of 3d models,” *CoRR*, vol. abs/1105.2795, 2011. 4
  51. T. Tung, F. Schmitt, “The augmented multiresolution reeb graph approach for content-based retrieval of 3d shapes,” *Int. J. Shape Modeling*, vol. vol. 11, no. 1, 2005. 4
  52. *Skeleton Based Shape Matching and Retrieval*, 2003. 4
  53. Johan W. H. Tangelder, Remco C. Veltkamp, “A survey of content based 3d shape retrieval methods,” *Multimed Tools Appl*, vol. 39, p. 441–471, 2008. 4
  54. M. Madry, C. Ek, R. Detry, K. Hang, and D. Kragic, “Improving generalization for 3d object categorization with global structure histograms,” in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, Oct 2012, pp. 1379–1386. 4, 5, 6, 7
  55. M. Madry, H. Afkham, C. Ek, S. Carlsson, and D. Kragic, “Extracting essential local object characteristics for 3d object categorization,” in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, Nov 2013, pp. 2240–2247. 4, 6
  56. H. Hoppe, T. DeRose, T. Duchamp, J. McDonald, and W. Stuetzle, “Surface reconstruction from unorganized points,” *SIGGRAPH Comput. Graph.*, vol. 26, no. 2, pp. 71–78, Jul. 1992. 5
  57. D. Giorgi, S. Biasotti, and L. Paraboschi, “Shape retrieval contest 2007: Watertight models track,” 2007. 5
  58. S. Katz, A. Tal, and R. Basri, “Direct visibility of point sets,” in *ACM SIGGRAPH 2007 Papers*, ser. SIGGRAPH '07. New York, NY, USA: ACM, 2007. 6