HAL
open science

# 3D Keypoints Detection for Objects Recognition

Ayet Shaiek, Fabien Moutarde

## To cite this version:

Ayet Shaiek, Fabien Moutarde. 3D Keypoints Detection for Objects Recognition. 16th International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV'2012), Jul 2012, Las Vegas, United States. hal-00741499

**HAL Id: hal-00741499**

**https://minesparis-psl.hal.science/hal-00741499v1**

Submitted on 12 Oct 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# 3D Keypoints Detection for Objects Recognition

**Ayet Shaiek**[1], **and Fabien Moutarde**[1]

[1]Robotics laboratory (CAOR) Mines ParisTech 60 Bd St Michel, F-75006 Paris, France

**Abstract -** *In this paper, we propose a new 3D object recognition method that employs a set of 3D local features extracted from point cloud representation of 3D views. The method makes use of the 2D organization of range data produced by 3D sensor. A detector of 3D interest points requires the expression of the local surface variation around points. In our case, we opted for a curvature-based approach. We test six methods which combine principles curvatures values under the form of: 1) a measure of the Shape Index (SI), 2) a measure of a Quality Factor (FQ), 3) a map of Shape Index (SI) and curvedness(C), 4) a map of Gaussian (H) and Mean (K) curvatures, 5) a combination of 3 and 4 (SC_HK) and 6) a combination of 5 and 4(SC_HK_FQ). For each extracted point, a local description using the point and its neighbors is done by combining the shape index histogram and the normalized histogram of angles between normals. This local surface patch representation is used to find the correspondences between a model-test view pair. Performance evaluation of the detectors in terms of stability and repeatability shows the robustness of the proposed detectors to viewpoint variations. Experimental results on the Minolta data set are presented to demonstrate the efficiency of the proposed approach in view based object recognition.*

**Keywords:** 3D keypoints detector, Mean Curvature, Gaussian Curvature, Shape index, Curvedness, Normals Histogram.

## 1 Introduction

3D Object class detection and recognition has become an extremely active research theme over the last decade, due to good success of object recognition techniques in the 2D field, and to the promising reliability of the new 3D acquisition techniques. 3D recognition, however, conveys several issues related to class variability, partial information, as well as scales and viewpoints differences are encountered. As previous works in the 2D case have shown, local methods perform better than global features to partially overcome those problems. Global features need the complete, isolated shape for their extraction. Examples of global 3D features are volumetric part-based descriptions [1]. These methods are less successful when dealing with partial shape and intra-class variations while remaining partially robust to noise, clutter and inter-class variations. Several 3D categorization methods based on local features have already been proposed, like tensors [2] and integral shape descriptors [3]. Point-of-interest (POI) detection, widely used in 2D image analysis, is also extended to 3D and therefore many recent researches have investigated in finding 3D interest-points detectors and descriptors. For example, the Harris detector has been extended to three dimensions, first in [4] with two spatial dimensions and time, then in [5] which discuss variants of the Harris measure and recently in [6] where a 3D-SURF adaptation is proposed.

Regarding descriptors of local 3D features, in addition to 3D-SURF, we can mention Spin Images [7] which records a spatial histogram of the 3D model's spatial occupancy by the remaining points w.r.t the current point.

3D Keypoint approaches can be classified into two main categories: fixed scale category and scale invariant category. In the Scale invariant category, we mention the 3D SURF and KPQ Scale Invariant (KPQ-SI) [8]. In the fixed scale approach, we find for example, the Local Surface Patches (LSP) [9] and the KeyPoint Quality (KPQ) [8]. Our proposed method belongs to the second category and we aim to detect salient and repeatable keypoints under viewpoint variation. We propose to use a measure of curvature in the line of Chen and Bhanu's work [9] and construct a patch labeling to classify different surface shapes [9,10]. Most 3D object recognition methods doing surface shape classification use mean-gaussian curvatures (HK) or shape index-curvedness (SC) values. In [11], authors present a comparison of the two approaches to show the qualitatively different classification and the impact of thresholds and noise levels.

In this paper, we propose a new method that combines criteria to extract invariant 3D feature points directly from a point cloud, using differential measures. The complete recognition system with detection, description and matching phases is introduced in §2. The proposed methods are evaluated and compared in §3.

# 2 Methodology

## 2.1 Subdivision of 3D Points Cloud into Local Patches

As we address a recognition scenario wherein only 2.5 views are matched, we deal with some views of the models from specific viewpoints. In the work presented here, we exploit the lattice structure provided by the range image. First, we search the coordinates of the maximum and minimum points at x-axis and y-axis in the sample, and build a bounding box based on the two limit points. After, our process visits all the delimited points, determines their neighbourhood patch and computes the measure of saliency over the local patch. In our approach, we consider a rectangular region around the point, with a span in the x and y direction, and we threshold the distance between neighbour points. The x and y spans are chosen adaptively for covering a proportion $r_1$ of the bounding box dimensions, so as to make our method robust to different spatial samplings, and to scaling. An advantage of subdividing the point cloud in local regions is to avoid mutual impact between them.

## 2.2 Keypoint Detectors

The aim of this step is to pick out a repeatable and salient set of 3D points.

Principal curvatures correspond to the eigenvalues of the Hessian matrix and are invariant under rotation. Hence, we propose to use local curvatures which can be calculated either directly from first and second derivatives, or indirectly as the rate of change of normal orientations in a local context region. The usual pair of Gaussian curvature K and mean curvature H only provides a poor representation, since the values are strongly correlated. Instead, we use them in composed form with curvature based quantities.

### 2.2.1 Shape Index

This detector type was proposed in [9], and uses the shape index ($SI_p$) for feature point extraction. It is a quantitative measure of the surface shape at a point p, and is defined by (1),

$$SI_p = \frac{1}{2} - \frac{1}{\pi} arctg\left(\frac{k_p^1 + k_p^2}{k_p^1 - k_p^2}\right) \qquad (1)$$

With this definition, all shapes are mapped into the interval [0, 1] where every distinct surface shape corresponds to a unique value of SI (except for planar surfaces, which will be mapped to the value 0.5, together with saddle shapes). Larger shape index values represent convex surfaces and smaller shape index values represent concave surfaces. The main advantage of this measure is the invariance to orientation and scale. A point is marked as a feature point if its shape index $SI_p$ satisfies (2) within point neighbours,

$$I_p = \text{max of shape indexes and } I_p >= (1+\boldsymbol{\alpha}) * \mu ;$$
$$\text{Or}$$
$$I_p = \text{min of shape indexes and } I_p <= (1-\boldsymbol{\beta}) * \mu ; \qquad (2)$$

where $\mu$ is the mean of shape index over the SI point neighbours values and $0<=\boldsymbol{\alpha}$ , $\boldsymbol{\beta}<=1$. In above expression (2), $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ parameters control the selection of feature points. In figure 1 is illustrated the range image of one model and its shape index image. In the depth map, the darker the pixel, the farther the real point is from camera. On the other hand, in the shape index image brighter pixels correspond to the greatest values of SI (i.e domes and ridge) and darker ones represent rut or cup surfaces. We denote this detector by « SI ».

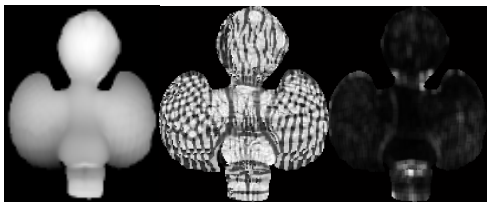

**Fig. 1.** On left, range image of the angel model; in the middle, Shape index Image; on right, Factor quality image

### 2.2.2 Factor Quality

The second detector we have implemented is based on a keypoint quality measure introduced by Mian et al. and used for ranking keypoints after the detection process [10]. We associate at each point k a quality measure $Q_k$ is given by (3),

$$Q_k = \frac{1000}{r^2} \sum |K| + \max(100K) + |\min(100K)| + \max(10\, k_p^1) + |\min(10 k_p^2)| ;\ K = k_p^1 k_p^2 \qquad (3)$$

where $k^1_p$ and $k^2_p$ are maximum and minimum principal curvatures, respectively. Summation, maximum and minimum values are calculated over the point neighbours. Absolute values are taken so that positive and negative curvatures do not cancel each other; positive and negative values of curvatures are equally descriptive. We compute the maximum value maxFQ of the quality factor over all the points. A threshold equal to maxFQ/ $r_2$ is chosen to select keypoints corresponding to the higher values. We then perform a connected component analysis to group neighboring points. Final keypoints are centers of connected components. In figure 1, the map of factor quality values of the angel model is shown. Brighter pixels correspond to the highest values of FQ, and are located in descriptive regions within important shape variation. We denote this detector by « FQ ».

### 2.2.3 HK and SC Classification

The idea here is to build shape classification space using the pair mean-Gaussian curvatures (HK) or the pair shape index - curvedness (SC).

Typically, for HK classification, we use the type function $T_p$ used in LSP descriptor [9] that associates to each couple of H and K values a unique type value (4),

$$T_p = 1 + 3\left(1 + \mathrm{sgn}_{\varepsilon_H}(H)\right) + \left(1 - \mathrm{sgn}_{\varepsilon_K}(K)\right)$$

$$\mathrm{sgn}_{\varepsilon_X}(X) \begin{cases} +1 & if\ X > \varepsilon_X, \\ 0 & if\ |X| \leq \varepsilon_X, \\ -1 & if\ X < \varepsilon_X \end{cases} \qquad (4)$$

where $\varepsilon_H$ and $\varepsilon_K$ are two thresholds over the H and K. Nine region types are defined (figure 2).

In the shape index-curvedness (SC) space, S defines the shape and C defines the degree of curvature and is the square-root of the deviation from flatness. Similarly to HK representation, the continuous graduation of SI subdivides surface shapes into 9 types. Planar surfaces are classified using the C value.

We define a type function $S_p$ (5) that associates a unique type value to each couple of SI and C values (i.e values between 0.8125 and 0.9375 correspond to dome and $S_p = 7$).

$$\begin{cases} S_p = 0 \text{ if } C \leq \varepsilon_C \\ \text{else} \\ S_p \in [1,8]\ ;\ SI \in [0,1]. \end{cases} \qquad (5)$$

For both classifications, salient regions are selected as those of one of the 5 following types: dome, trough, spherical, saddle rut and saddle ridge regions. More details are given in [11, 12]. The HK and SI classifications of surface are illustrated in figure 2.
We denote the two detectors by « HK » and « SC ».

| Mean curvature $H$ | Gaussian curvature $K$ | | |
|---|---|---|---|
| | $K > 0$ | $K = 0$ | $K < 0$ |
| $H < 0$ | Peak $T_p = 1$ | Ridge $T_p = 2$ | Saddle ridge $T_p = 3$ |
| $H = 0$ | None $T_p = 4$ | Flat $T_p = 5$ | Minimal $T_p = 6$ |
| $H > 0$ | Pit $T_p = 7$ | Valley $T_p = 8$ | Saddle valley $T_p = 9$ |

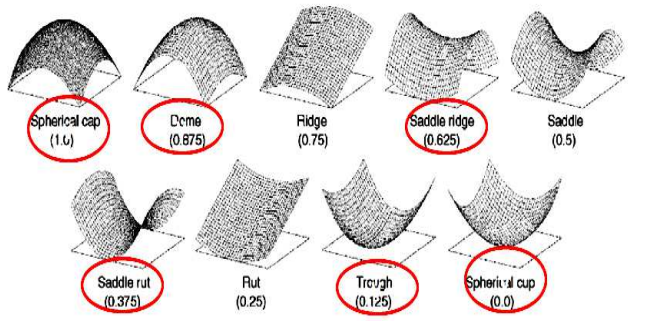


**Fig. 2.** HK classification (from [9]) on first column and SI classification [12] on second column.

### 2.2.4 Combinaison of Criteria

Theoretically, the two classifications HK and SC should provide the same result; therefore we suggest combining the two criteria to increase reliability. In fact, our result will be validated with two measures of keypoints detection. After labeling points with a pair of value ($T_p$,$S_p$), points with salient type pair, are selected, in other words, if the two labels correspond to the same of the 5 salient region types previously mentioned. Then, points with the same pair value are grouped using the connected- component labeling. Connectivity is carried out by checking the 8-connectivity of each point. Finally, the centers of the connected component are selected as keypoints. We call the detector combining the two criteria « SC_HK ».

We also propose further combination by ranking the selected keypoints according to their factor quality value. Therefore, we compute the maximum value maxFQ of quality factor values over the selected keypoints and only points with FQ value superior to maxFQ/ $r_2$ are finally selected as keypoint. We call this last detector « SC_HK_FQ ».

## 2.3 Keypoint Descriptors

After keypoints detection step, a 3D descriptor is computed around each selected 3D interest point. In the case of range data, the dominant orientation at a point is the direction of the surface normal at that point. For selected keypoints, we compute the shape index values and the angles θ between the reference surface normals at the feature point and the neighbour's ones. The reference normal of a keypoint is obtained by averaging the normals of points belonging to the connected component associated to the feature point. We

suggest comparing two ways to cumulate the shape index values and the cosine of the angle values:

- Combined descriptor: we form a 2D histogram by accumulating points in particular bins along two axes which relates the shape index value and the cosine of the angle to the 2D histogram bin. One axis of this histogram is the shape index which is in the range [0, 1]; the other is the cosine of the angle (cos θ) between the surface normal vectors and one of its neighbours.

- Concatenate descriptor: we cumulate shape index and cosine of the angle values into 1D histogram.

Two different spans are used to bin the cosine axes since more informative values appear when neighbour normal direction is near the orthogonal direction of the reference normal. Therefore, the span is smaller in the interval corresponding to near orthogonal directions.

## 2.4 Matching and Recognition

We are validating the proposed detector and descriptor using a view matching approach. Given a test object, we compute a measure of similarity between descriptors extracted on the test view and those of the models in database.

### 2.4.1 Hash table building

To speed up the comparison process, we use the mean and standard deviation of shape index of the neighbors around the feature point to index a hash table and insert the corresponding hash bin the information (model ID, 2D histogram, surface type, the centroid). For each model object, we repeat the same process to build the model database. For a test object, we index each keypoint and compute histogram similarity.

Histogram-based local descriptors are often compared by bin-to-bin metrics, especially the $\chi^2$ distance. Hence, for each histogram Q {qi} from test view, we find the best matching descriptor V {vi} from database view using the $\chi^2$ − divergence (2). Two keypoints are matched according to their histogram distance and their type of surface.

$$\chi_2(Q,V) = \sum_i \frac{(q_i - v_i)^2}{(q_i + v_i)} \qquad (6)$$

### 2.4.2 Geometric Constraints

A set of nearest neighbors is returned after histogram matching. The potential corresponding pairs are filtered and grouped based on the geometric constraints in equation (7) below, where $d_{S_1,S_2}$ and $d_{M_1,M_2}$ are Euclidean distance between centroids of two surface patches. For two correspondences $C_1 = \{S_1|M_1\}$ and $C_2 = \{S_2|M_2\}$ where S means test surface patch and M means model surface patch, they should satisfy (7) if they are consistent corresponding pairs.

Given a list of corresponding pairs, the grouping procedure for every pair in the list is as follows:

- Initialize each pair of a group.
- For every group, add other pairs to it if they satisfy (7).
- Repeat the same procedure for every group. Select the group which has the largest size.

$$d_{C_1,C_2} = \left| d_{S_1,S_2} - d_{M_1,M_2} \right| < \varepsilon_1 \qquad (7)$$

# 3 Experimental results

## 3.1 Data and Parameters

We performed our experiments on real range data from the Minolta data set [13]. There are 16 objects in our database with a total of 348 frames (figure 3). The numbers of feature points detected from these range images vary from 4 to 250, depending on the viewpoint and the complexity of input shape. To every feature is assigned a 11x19-dimensional signature.

The parameters of our approach are: $r_1$=3%, $r_2$ =20, $\alpha$ =0.45, $\beta$= 0.25, $\varepsilon_H = 0.003$, $\varepsilon_K = \varepsilon_H$, $\varepsilon_C = \varepsilon_H$, $\varepsilon_1 = 5$.

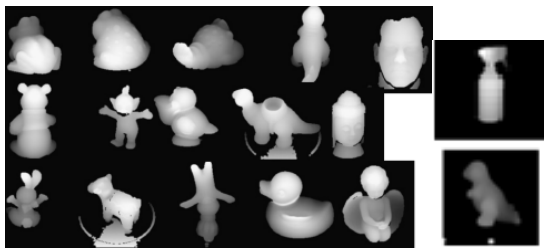

**Fig. 3.** Range images of the 16 objects from the Minolta Database

## 3.2 Keypoint Stability

To evaluate detector performance, we propose the use of the absolute repeatability which is the number of repeatable keypoints of one view in another view of the same object [14]. A keypoint is said to be repeatable if:

$$\underbrace{\left\| R_{ms} K_m^i + t_{ms} \right.}_{\text{Model keypoint rotated and translated}} - \underbrace{\left. K_s^j \right\|}_{\text{Scene keypoint}} < \underbrace{\varepsilon}_{\text{Repeatability threshold}} \quad (8)$$

[14]

Hence to measure the repeatability of detected keypoints between different views/scales, we consider two views: view 1 and view 2 of the same object. As we know the real transformation T (rotation or scaling) between the two views, we compute the distance to the nearest neighbor between

positions of every keypoint detected in view 1 after the application of the transformation T and keypoint detected in view 2. We plot the average of the repeatability measures between different pairs of views in Minolta dataset. Figure 4 illustrates the six plots of keypoint repeatability of the 9 objects, respectively for SC_HK_FQ detector (SC_HK_FQ), SC_HK detector (SC_HK), SC detector (SC), HK detector (HK), FQ detector (FQ) and SI detector (SI). The y-axis shows the percentage keypoints of the transformed views which could find a corresponding keypoint in the initial view within the distance shown on the x-axis. Results show that SC_HK_FQ and SC_HK have almost the same behavior and outperform the four other detectors. FQ has clearly the lowest repeatability. The repeatability reaches 80% at a nearest neighbor distance of ~0.7% of the average diagonal distance for SC_HK_FQ and SC_HK, 70% for HK, SC and SI, and 60% for FQ.

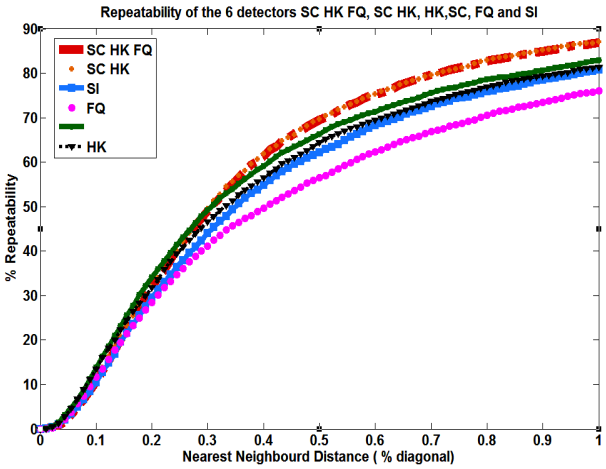


**Fig. 4.** Keypoint repeatability between different views for the six detectors: *SC_HK_FQ, SC_HK, SI, SC, HK and FQ*

Furthermore, visual comparison of keypoint positions detected with SC_HK_FQ, SC_HK, HK and SC detector is shown in figure 5. It reveals that the final selected points are quite well localized. The combining process (figure 6) allows a better feature point filtering than SC or HK alone where false detected point in both are eliminated and points with correct surface type remain. Figure 7 illustrates the relative stability of keypoint's positions detected with SC_HK_FQ detector when varying viewpoints for the same object. Clearly, we recover almost same keypoint positions in the different views, which qualitatively illustrate the stability of our keypoints.

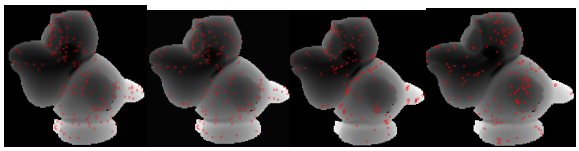

**Fig. 5.** Positions of detected keypoint on bird model with: SC_HK_FQ in first column, SC_HK in second column, HK in third colon and SC in fourth column

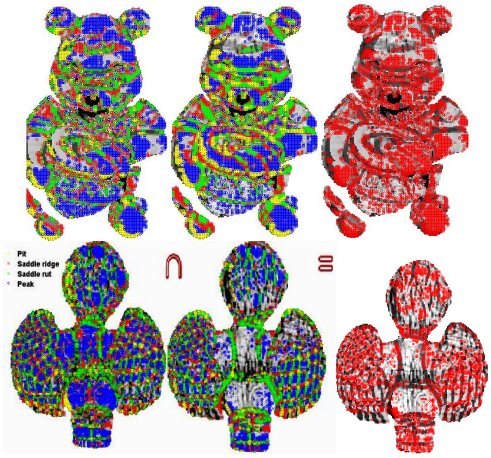

**Fig. 6.** On the left, SC classification; in the middle, HK classification; and on the right, combination of SC and HK (surface types: *pit, saddle ridge, saddle rut and peak*)

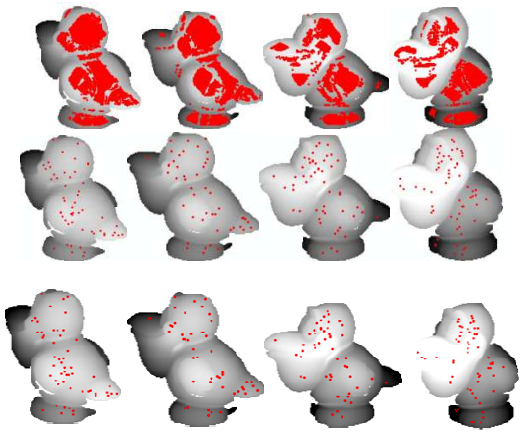

**Fig. 7.** Illustration of detectors stability, showing positions of detected keypoints (shown in *red*) for the views (100°, 120°, 140° and 180°) in bird model. Ligne 1: result of the connected components processes. Ligne 2, result of SC_HK detector and ligne 3, result of SC_HK_FQ detector.

### 3.3 Impact of Noise in the detection

A good detector is able to extract local features in the original surface as well as in the noisy data. In order to evaluate the repeatability of the feature points, white Gaussian noise with standard deviation σ ranging from 0.1 to 1.2 was

added to the 3D surfaces. When noise is introduced to the point clouds, the details of the shape are less visible. As a result, here would be fewer features points detected in noisy images compared to the original one. However, it is important that the local keypoints detected in the original surface will present in the noisy data.

It is to notice that the tested database already contains noise since it represents real object captures. Figure 8 shows the features extracted from the 'dough' model for different levels of noise for the two detectors SC_HK_FQ and SC_HK. It can be seen from the figure that a large portion of local features from the original model are presented in the noisy versions. For example, there are still many feature points lying around salient structures such as the nose, eyes, crest and arms even in the noisiest surface. With the noise level of σ= 0.36, most of the keypoints in the original image appears in the noisy version. A quantitative evaluation of the repeatability of the features for four different 3D models with a total of 72 views is shown in Figure 9. At the noise level of σ= 0.06, nearly all of the features in the original model (98%) can be detected in the noisy surface.

Even when the standard deviation of the noise goes to σ= 0.6, about 80%of the original features repeat in the noisy data.

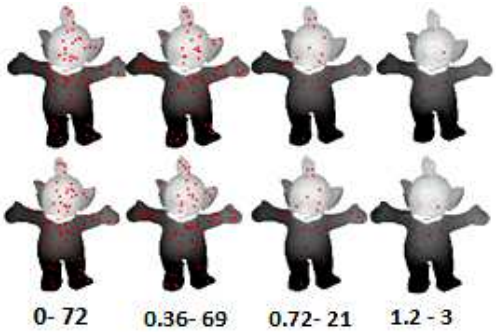


Figure 8. Features detected from the 'Dough' model with different noise levels, SC_HK_FQ detector with in ligne 1 and SC_HK detector in ligne 2. In the third ligne, it's indicated the couple: noise level – number of keypoints in each column.

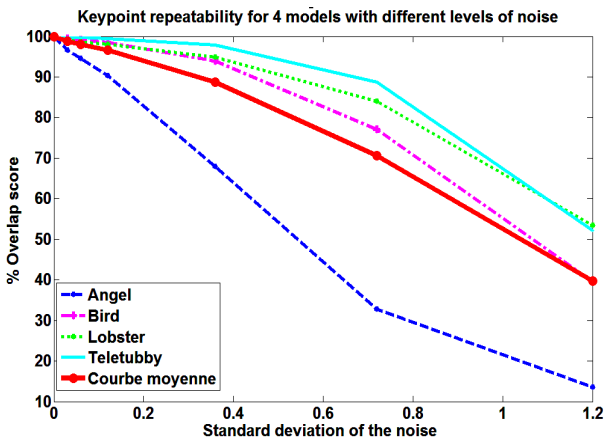


Figure 9. Repeatability of the features for 4 different models in different noise conditions.

### 3.4 Matching Result

We present the test protocol for recognition in table1. We carry out two experiments, in the first one; we choose manually N test views per object and use the remaining views for the training stage. The descriptor used for this experiment is the concatenate version. The same evaluation is done in experiment 2 with the *combined* descriptor. The two experiments are carried out using the 6 detectors. The results are shown in table 2. The overall recognition rate is quite promising for the SC_HK_FQ method in comparison to the other results, with 96.4%. This rate is achieved using the combined version of the descriptor which suggests that it is more descriptive then the concatenate version. We notice here that the computation time when matching the combined feature is more important (more bins to compare), which can be an inconvenient when dealing with real time application.

**Table1**. Test protocol for object recognition

| View/Per object | Experiment 1 with concatenate descriptor | Experiment 2 with combined descriptor |
|---|---|---|
| Test | 4 views(num 80, 90, 100, 107) for orangedino and 3 views (20°, 180°and 300°) for 8 other object | same as experiment 1 |
| Training | The 32 remaining views for orangedino and 15 remaining views for the others | same as experiment 1 |

**Table2**. Recognition rates for the 6 methods

| | SC_HK_FQ | SC_HK | SC | HK | SI | FQ |
|---|---|---|---|---|---|---|
| Exp 1 | 82.1% | 75% | 82.1 % | 82.1 % | 35.1 % | 64.2 % |
| Exp 2 | **96.4%** | 82% | 92.8 % | 92.8 % | 64.2 % | 64.2 % |

## 4 Conclusions and Perspectives

In this paper, a comparison of six proposed detectors based on curvature is presented. Our principal contribution is the idea of combining criteria for detection, and proposing a new 3D object recognition method that employs a set of 3D local features (3D keypoints, or "points-of-interest", POI) extracted from point cloud representation of 3D views. Furthermore, a quantitative evaluation of the stability of obtained keypoints under viewpoint change on real-world

depth images has shown promising results, with 80% close repeatability obtained by combining SC (shape curvedness) and HK criteria. The original combination process using SC_HK_FQ detector seems to provide a pertinent description of the local surface typology. For the moment, measures of curvatures are calculated at a constant scale level, the feature's scale is still ambiguous. To overcome this fact, we propose to search for features at different scale levels for a future work.

Regarding the 3D keypoint descriptor, we compare two descriptors that encode the occurrence frequency of shape index values vs. the cosine of the angle between the normal of reference feature point and that of its neighbours. Results show that the combined version is more efficient than the concatenate one.

As for the overall performance of the proposed methods for object recognition, we obtain best recognition rate for the SC_HK_FQ method, with 96.4% on 9 objects from real-world Minolta public dataset.

# 5 References


[1] Medioni, G.G. and François, A.R.J. "3-D structures for generic object recognition," Computer Vision and Image Analysis, 1, 1030 (2000).

[2] Bowyer, K.W. Chang, K. and P. Flynn, "A survey of 3D and multi-modal 3D+ 2D face recognition," Notre Dame Department of Computer Science and Engineering Technical Report (2004).

[3] Li, X. and Guskov, I. "Multi-scale features for approximate alignment of point-based surfaces," in Proceedings of the third Eurographics symposium on Geometry processing, 217 (2005).

[4] Paul S., Saad A. and Mubarak S., "A 3-dimensional SIFT descriptor and its application to action recognition", Proceedings of the 15th International Conference on Multimedia, 357–360 (2007).

[5] Fredrik V., Klas N., Mikael K. "Point-of-Interest Detection for Range Data", ICPR IEEE, 1-4 (2008).

[6] Jan K., Mukta P., Geert W., Radu T., Luc van G., "Hough Transform and 3D SURF for robust three dimensional classification," Proceedings of the European Conference on Computer Vision, (2010).

[7] Johnson, A.E., Hebert, M., "Using spin images for efficient object recognition in cluttered 3d scenes," IEEE PAMI 21, 433-449 (1999).

[8] Mian, A. Bennamoun, M. and Owens, R. "On the Repeatability and Quality of Keypoints for Local Feature-based 3D Object Retrieval from Cluttered Scenes," International Journal of Computer Vision, 89 (2), 348-361 (2010).

[9] Chen, H. and Bhanu, B. "3D free-form object recognition in range images using local surface patches," Pattern Recognition Letters, 28(10), 1252-126 (2007).

[10] Erdem A., Omer E., Ilkay U. "Scale-space approach for the comparison of HK and SC curvature descriptions as applied to object recognition". ICIP, 413-416 (2009).

[11] H. Cantzler, R. B. Fisher, "Comparison of HK and SC curvature description methods" In Conference on 3D Digital Imaging and Modeling, 285-291 (2001).

[12] J. Koenderink and A. J. Doorn. "Surface shape and curvature scale ", *Image Vis. Comput.*, vol. 10, no. 8, pp. 557–565, (1992).

[13] http://cheepnis.cse.nd.edu/~flynn/3DDB/3DDB/RID/index.htm

[14] Samuele S., Federico T., and Luigi Di S. "A Performance Evaluation of 3D Keypoint Detectors," 3DIMPVT, (2011).