

Shape Index SIFT: Range Image Recognition Using Local Features

Neslihan Bayramoğlu, A. Aydın Alatan
 Department of Electrical and Electronics Engineering
 METU
 Ankara, Turkey
 {neslihan, alatan}@eee.metu.edu.tr

Abstract—Range image recognition gains importance in the recent years due to the developments in acquiring, displaying, and storing such data. In this paper, we present a novel method for matching range surfaces. Our method utilizes local surface properties and represents the geometry of local regions efficiently. Integrating the Scale Invariant Feature Transform (SIFT) with the shape index (SI) representation of the range images allows matching of surfaces with different scales and orientations. We apply the method for scaled, rotated, and occluded range images and demonstrate the effectiveness it by comparing the previous studies.

Keywords—range image recognition; scale invariance; local description; shape index;

I. MOTIVATION

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

II. RELATED WORK

The problem of 3D shape matching and retrieval is studied extensively in the past [1, 2]. There are two main approaches for shape identification: *global description* and *local description* [1]. First type of approach requires complete models and most of the studies from the literature are focused on global approach. In the second type, descriptors are obtained and evaluated on some local keypoints and utilized for complete or partial matching. However, there are also some “partial” matching methods that also require global models [3,4]. For local descriptions, *spin images* [5], *splashes and 3D curves* [6], *point signatures* [7], *local feature histograms* [8], and *regional point descriptors* [9] can be listed as remarkable

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

A. Global vs. Local Descriptions

Firstly, it should be noted that global description of range object is only possible after segmentation from the scene, but segmentation is a very challenging task; besides object structure is still partially available even after segmentation. Thus, global methods are likely to fail for the applications based on 3D video. Local description could be more promising than global description for such problems in many aspects: segmentation is avoided, complete models are not necessary, occlusion and self occlusion can be handled more easily. Moreover, in 2D object detection and matching, local feature extraction methods [10] are shown to perform notably better. There is also an important difference between these two approaches; while the global methods ignore shape details and gather the main properties of objects, the local descriptors put more emphasis on the details. As a result, the global methods perform better in similarity search, while local methods are more suited for matching, verification and identification.

B. Local Descriptors for Range Data

In range image recognition, surface properties, such as *normal* or *curvatures*, are widely used [6,8,12]. The early work *splashes and 3D curves* [6] makes use of distribution of normal vectors in the neighborhood of a keypoint. Although splash technique is a promising approach the representation is complex. Similarly, *point signatures* [7] obtain a contour by intersecting the surface with a sphere centered at a keypoint. Then, a plane is fitted to these contour points and translated to the keypoint. After this fitting, a 1D signature, which is the signed distances of contour points to the fitted plane, is utilized as the descriptor. However, both of these

methods do not address the scale and rotation invariance problem. Another well-known effort, namely *spin images* [5], are widely used as local shape description method and applicable for scenes containing occlusions and clutter. It is a 2D histogramming around a keypoint in which spatial configuration is represented with a cylindrical coordinate system. However it has the following disadvantages: it is not scale invariant, it requires high storage and results multiple ambiguous correspondences, since spin images of near points are similar; i.e. spin images are not unique [11]. Yet another similar histogramming technique is *3D shape contexts* [9]. In that approach, spatial distributions of the surface points are accumulated utilizing their spherical coordinates. The same arguments for the spin images also hold for this description as well.

C. Shape Index as a Local Descriptor

There are two recent studies that use *shape index* within the context of local description. In the first study [12], the authors identify keypoints, where the shape index values are extremum. Around these keypoints, 2D histogramming is formed, in which one of the dimensions belongs to the shape index values of the neighboring pixels, whereas the other one denotes the angles between the normal of the keypoint and the normal of the neighboring pixels (SI-Normal-Hist). In their work, scale invariance problem is not addressed. In a different recent effort, namely *2.5D Scale Invariant Feature Transform (2.5D SIFT)* [13], keypoint detection is achieved on 2D range images similar to the Lowes work [10]. However, an extra preprocessing step is required. After this step, using the histogram of the shape index values and the range gradient orientations around these keypoints, a description is obtained. Both of the methods use isolated objects in which problematic boundary regions can be eliminated by simple thresholding however it is not applicable in range scenes due to occlusion. Besides, the matching performance in complicated scene with the presence of other objects, clutter and occlusion are not evaluated for these studies. This paper presents a local range image matching method which combines 3D surface properties with the 2D scale invariant feature transform (Figure 1). In the next section, local surface properties that are proposed in this paper are explained. Next, feature extraction methodology is stated and finally matching results and comparison to the previous studies are presented.

III. LOCAL SURFACE PROPERTIES

There are two key features of the 3D surfaces: *normal vector* and *curvature*. A 3D point can be described by its minimum and maximum curvatures (*principal curvatures*) or some functions of these principal curvatures (κ_1 , κ_2) at the point of interest. One of these measures is the *Shape Index (SI)* which is introduced by Koenderink [14]. Firstly, it is used to obtain global description, named as *shape spectrum*

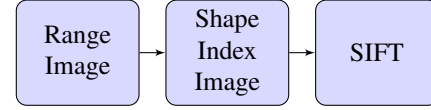


Figure 1. Flowchart of the proposed method

[15]. Later, as mentioned previously, it is used both for local description and keypoint selection [12, 13]. The curvature values on 3D surfaces can be obtained robustly after fitting a quadric surface to a local patch. Then the shape index is calculated using principal curvatures as follows:

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

The most appealing property of shape index is being scale, translation, and rotation invariant. Moreover, shape index makes strong emphasis on points where surfaces deviate from being smooth, even for small changes. Thus, representing the range image with its shape index values which are mapped to 0-255 makes an lighting effect. Shape details become clearly visible (Figure 3). Therefore, utilization of SIFT-type features becomes more feasible.

IV. KEYPOINT SELECTION AND FEATURE EXTRACTION

SIFT is an invariant 2D local descriptor which is robust to some specific transformations [10]. SIFT has three main steps: *scale space construction*, *keypoint detection*, and *feature extraction*. Scale space is constructed by taking the difference of the Gaussian (DoG) blurred images at different scales [10]. Next, keypoints are detected as the local extremum of the DoG images across scales. A gradient orientation histogram is computed in the neighborhood of the keypoint as the feature vector [10]. In this study, SIFT keypoints and the features are computed by using the publicly available software¹.

V. EXPERIMENTAL RESULTS

Experiments are performed on range images from the Ohio State University range database² and range images formed by combining models collected from AIM@SHAPE repository³ and ISDB database.

Table 1 shows the comparison of the proposed method with 2.5D SIFT [13] and SI-Normal-Hist. [12]. The experiments are performed on the same data (Figure 2). Based on the experimental results, the number of matches and correct matches is obtained more than three times of the number of matches that are obtained by [13]. Although limited comparison with the LSP is achieved due to the data availability, the proposed method shows better performance. The reason for this superior performance could be due the

¹<http://www.vlfeat.org/~vedaldi/>

²<http://sampl.ece.ohiostate.edu/data/3DDB/RID/minolta/>

³<http://shapes.aim-at-shape.net/>

Table I
LOCAL DESCRIPTOR MATCHING RESULTS FOR PROPOSED METHOD,
SI-NORMAL-HIST., AND 2.5D SIFT

Data (RotationAngle)/ Method	Total Matches	Correct Matches	False Matches
Lobster(20°)			
Proposed	33	19	14
SI-Normal-Hist[12]	30	13	17
Pooh (20°)			
Proposed	37	27	10
2.5DSIFT [13]	9	8	1
Cow (15°)			
Proposed	37	37	0
2.5D SIFT [13]	10	10	0
Duck (20°)			
Proposed	25	19	6
2.5D SIFT [13]	6	4	2
Lobster2 (20°)			
Proposed	32	20	12
2.5D SIFT [13]	8	8	0
Bird(20°)			
Proposed	37	30	7
2.5D SIFT [13]	9	8	1

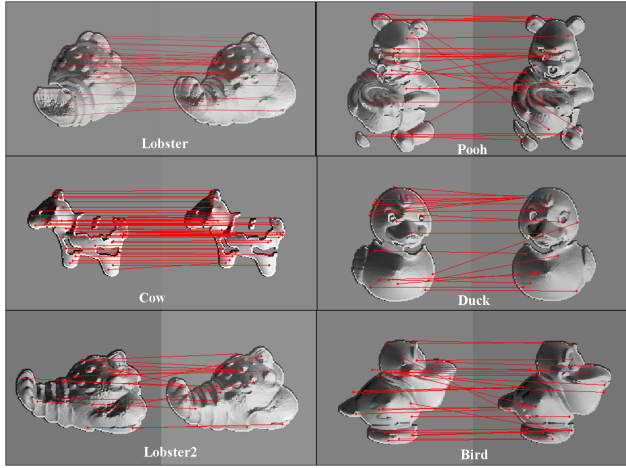


Figure 2. Matched pairs obtained by the proposed method are shown on the SI images.

fact that SIFT keypoint selection and feature extraction on SI images is more effective than selecting the keypoints as SI extremum and describing features as SI histogram around the keypoints. Former method is also more effective than applying SIFT directly to the range image. This result can also be explained as follows: SI comprises the understanding of the neighbor geometry whereas range image pixel only indicates its depth. Besides informative regions are visually strengthen by the nonlinearity behavior of the SI function. Additionally, SIFT features are considerably discriminative

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

VI. CONCLUSION

The proposed local surface description method does not require any initial segmentation step; it can also handle affine transformations up to a scale. The experimental results indicate that the proposed approach improve the performance of two recent methods from the literature. Moreover, clutter and occlusion do not affect the efficiency of the proposed method significantly.

ACKNOWLEDGEMENT

This research is supported by The Scientific and Technological Research Council of Turkey (TÜBİTAK).

REFERENCES

- [1] J. W. H. Tangelder and R. C. Veltkamp, "A survey of content based 3D shape retrieval methods" *Multimedia Tools and Applications*, vol. 39, pp. 441-471, 2008.
- [2] B. Bustos, D. Keim, D. Saupe, T. Schreck, and D. Vranic, "Feature-based similarity search in 3D object databases," *ACM Comput. Surv.*, vol. 37, 345-387, 2005.
- [3] P. Shilane and T. Funkhouser, "Distinctive regions of 3D surfaces," *ACM Trans. Graph.*, vol. 26, p. 7, 2007.
- [4] S. Shalom, L. Shapira, A. Shamir, and D. Cohen-or, "Part Analogies in Sets of Objects," *Eurographics Workshop on 3D Object Retrieval, 3DOR 2008*, Crete, Greece, pp. 33-40.
- [5] A. E. Johnson and M. Hebert, "Using Spin Images for Efficient Object Recognition in Cluttered 3D Scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, pp. 433- 449, 1999.
- [6] F. Stein and G. Medioni, "Structural indexing: efficient 3-D object recognition," *Patt. Anal. and Machine Intell. IEEE Transactions on*, vol. 14, pp. 125-145, 1992.
- [7] C. Chin Seng and J. Ray, "Point Signatures: A New Representation for 3D Object Recognition," *Int. J. Comput. Vision*, vol. 25, pp. 63-85, 1997.
- [8] H. Günter, L. Bastian, L. Paul, and S. Bernt, "3D Object Recognition from Range Images using Local Feature Histograms," in *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2001, pp. 394-394.
- [9] A. Frome, D. Huber, R. Kolluri, T. Blow, and J. Malik, "Recognizing Objects in Range Data Using Regional Point Descriptors," in *in Proceedings of the European Conference on Computer Vision (ECCV, 2004*, pp. 224-237

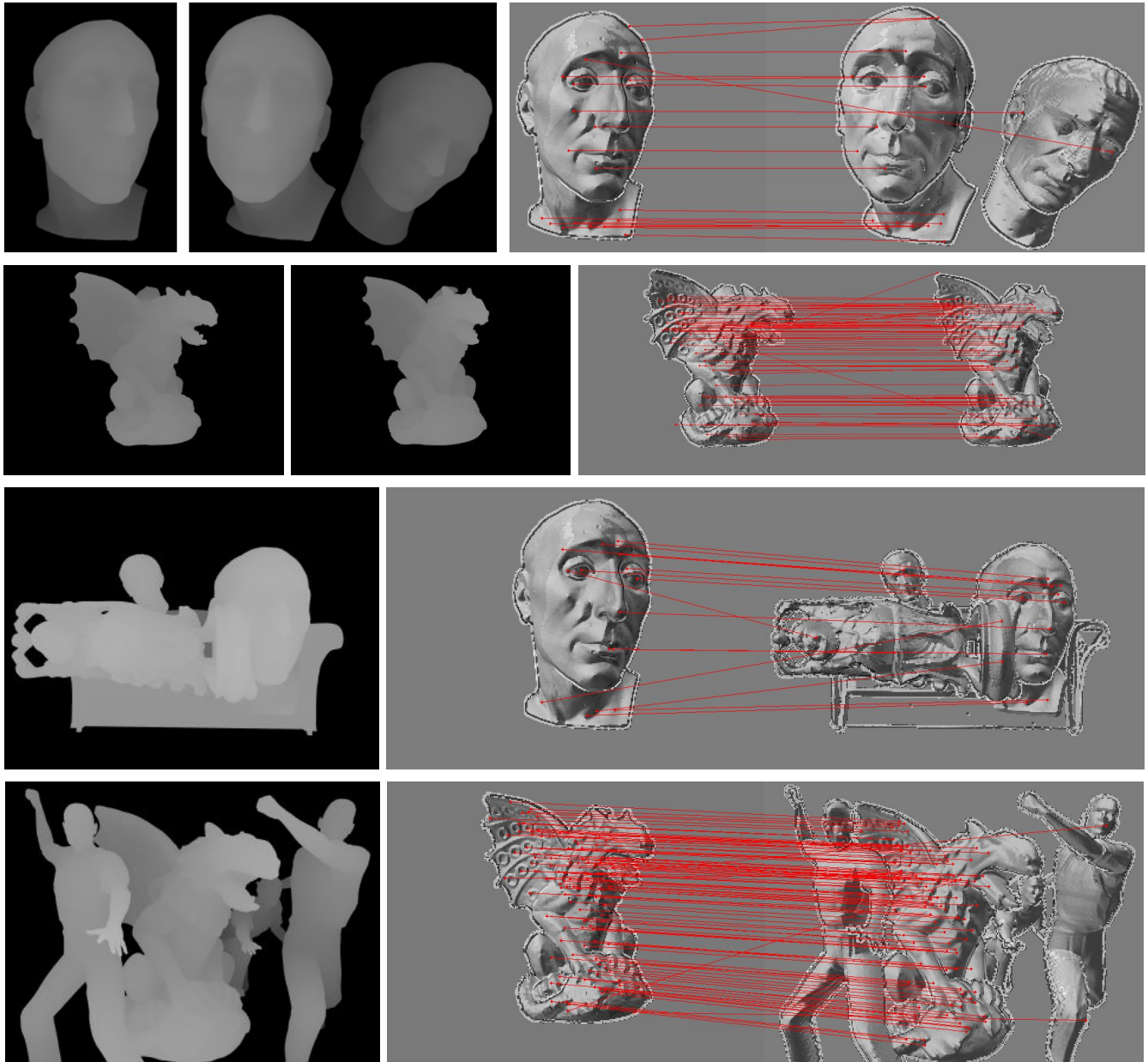


Figure 3. First two rows: (left to right) Query range images, target range images, combined shape index images with matched pairs. Rotated query objects are included in the target images. Last two rows: (left to right) Target range images, combined shape index images with matched pairs. Scaled and occluded query objects are included in the target images.

- [10] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91-110, 2004.
- [11] A. Mian, "Representations and Matching Techniques for 3D Free-form Object and Face Recognition," PhD Thesis, The University of Western Australia, 2006.
- [12] H. Chen and B. Bhanu, "3D free-form object recognition in range images using local surface patches," *Pattern Recognition Letters*, vol. 28, pp. 1252-1262, 2007.
- [13] T.-W. R. Lo and J. P. Siebert, "Local feature extraction and matching on range images: 2.5D SIFT," *Computer Vision and Image Understanding*, vol. 113, pp. 1235-1250, 2009.
- [14] J. J. Koenderink, *Solid shape*: MIT Press, 1990.
- [15] T. Zaharia and F. J. Preteux, "3D-shape-based retrieval within the MPEG-7 framework," *SPIE Conf. Nonlinear Img. Proc. and Patt. Ana.*, CA, 2001, pp. 133-145.