Eurographics Workshop on 3D Object Retrieval (2017)
I. Pratikakis, F. Dupont, and M. Ovsjanikov (Editors)

GSHOT: a Global Descriptor from SHOT to Reduce Time and Space Requirements

C. M. Mateo and P. Gil and F. Torres

University of Alicante, Physics, Systems Engineering and Signal Theory Department, Spain

Abstract

This paper presents a new 3D global feature descriptor for object recognition using shape representation on organized point clouds. Object recognition applications usually require significant speed and memory. The proposed descriptor requires 57 times less memory and it is also up to 3 times faster than the local feature descriptor in which it is based. Experimental results indicate that this new 3D global descriptor obtains better matching scores in comparison with known state-of-the-art 3D feature descriptors on two standard benchmark dataset.

1. Introduction

There are several state-of-the-art feature descriptors for object recognition. However, they do not completely resolve the object recognition problem, especially when faced with hard problems such as texture-less objects, noise or missing parts of the objects in the acquisition process. Moreover, they are far from being efficient.

In general, surface recognition of point clouds from RGBD sensor is usually carried out using global or local approaches. An example of global approach is the work commented in [ABG*11],in which a unique and repeatable descriptor with a single signature is able to describe an entire point cloud. Those works are based on the previous idea shown in [RBTH10], where a descriptor is used to represent object surfaces with regards to viewpoint. A year before, the same author [RBB09] presented a multi-signature local descriptor built using a reference frame for each point in the point cloud. More recently, an other local approach using a different reference frame was proposed in [STD14].

The recognition process not only depends on the use of descriptors but also on the used matching machine, or classifier. Currently, the most common techniques in the literature for point clouds recognition are k Nearest Neighbor algorithms(kNN) [ML14], Support Vector Machine (SVM) [CV95] and discriminative Random Regression Forests [Bre01]. In this work, we use the first two. Aside from the matching method or adopted descriptor, the recognition process is dependent on the dataset used in the training phase. There are wide number of works about object recognition using 3D point clouds from RGBD sensors [RBTH10, ABG*11, STD14]. Moreover, the number of datasets for evaluating recognition methods is gradually being increased [SMKF04, LBRF11, SSN*14, JKJ*13].

The core idea behind our approach is to extend the descriptor presented in [STD14], transforming it into a global descriptor. This approach improves results in terms of temporal and spatial performance. Moreover, the new descriptor increases the recognition rates using SVM and kNN as a classifier when the classification is made over well segmented point clouds.

The paper is organized as follows: section 2 discusses related works about the most common descriptors in the literature; section 3 presents a new approach called GSHOT, based on the descriptor Signature of Histogram of OrienTations (SHOT) to eliminate the dependence of the viewpoint; section 4 presents a complete evaluation analysis including a comparison with other descriptors throughout experiments using two datasets. The results prove the effectiveness of our approach; and section 5 provides the conclusions of this work.

2. Descriptors for object recognition

The use of normal-based features as an object recognition approach has become in a classic strategy for describing point clouds but this one has still issues to resolve. Two different types of descriptors have been analyzed to be compared with the new proposed descriptor. They are the descriptors which use Darboux reference frame and those of other which use eigenvectors from covariance matrix as reference frame. The descriptors based on Darboux reference frame compute a tuple $\langle\alpha,\phi,\theta\rangle$ for each relationship among the points of a same neighborhood area. Each tuple represents the relationship among one and all normal vectors in the neighborhood, as follows,

$$\alpha = \mathbf{v}^T\mathbf{n}_j,\ \phi = \mathbf{n}_j^T\frac{p_i - p_j}{d_i},\ \theta = \arctan(\mathbf{w}^T\mathbf{n}_i, \mathbf{n}_i^T\mathbf{n}_j) \qquad (1)$$

where $\mathbf{n}_i$ is the normal vector to the tangent plane of the underlying

DOI: 10.2312/3dor.20171053

surface to the point p_i, $d_i = \|p_i - p_j\|_2$, $\mathbf{v}$ is the director vector from point p_i to point p_j and $\mathbf{w}$ is a vector perpendicular to $\mathbf{n}_i$ and $\mathbf{v}$. In contrast, other descriptors use eigenvectors from the covariance matrix defined as equation 2.

$$M = \frac{1}{\sum_{i=d_i}^{R}(R-d_i)} \sum_{i=d_i}^{R} (R-d_i)(p_i-p)(p_i-p)^T, \qquad (2)$$

where R is the radius used to determine the neighborhood, and d_i is the maximum distance between a point p_i and the centroid p. Consequently, some of the most relevant descriptors in the literature which use those reference frames, are the followings:

- Viewpoint Feature Histogram (VFH) [RBTH10] is a global extension of the Simple Point Feature Histogram (SPFH) [RBB09]. VFH is a histogram with two components; one represents the Darboux's angles considering each cloud point and the centroid, and the other represents the angles among each point's normal and the director vector determined by the centroid and viewpoint. Moreover, the distance from each point to the centroid is included in this second component.
- Clustered Viewpoint Feature Histogram (CVFH) [ABG*11] splits an object into a set of smooth and continuous regions or clusters. Then, parts of object such as edges, ridges and other discontinuities are not considered to be used for the recognition process because they are usually affected by noise. Therefore, the object shape is only described from a VFH descriptor computed for each cluster.
- Signature of Histogram of OrienTations (SHOT) [STD14] is a local feature descriptor which uses the eigenvector of a covariance matrix as reference frame. Generally, it uses a spherical grid partitioned into $d = 32$ sectors as follows: 2 divisions for elevation, 8 for azimuth and 2 for radial. The descriptor assesses the differences between the normal vector at each point of the surface (within the local reference frame) and the normal vector at the center of the local reference frame. This difference is computed by dot-product and interpolated into one of the $b = 11$ classes (bins), so the dimension signature is $d * b = 352$.

3. An approach to make a Global SHOT (GSHOT)

SHOT descriptor achieves good results in terms of accuracy but its computation time is too long as was proved in [MGT16]. This fact is due to its local character inasmuch as SHOT computes for each keypoint (point of interest) in the point cloud P a signature. The proposed descriptor GSHOT computes a unique signature for a whole point cloud, extending SHOT to be a global descriptor (Figure 1). In order to compare GSHOT and SHOT under equal condition, the same number of divisions of the spherical grid presented in [STD14] has been used in this work.

This work proves that GSHOT not only reduces the time requirements of the original method but also improves its results in well segmented point clouds (see Section 4) in comparison with SHOT. Another advantage of GSHOT is that it avoids the requirement of fixing a radius size R for computing the descriptor signatures because this one is auto-computed from the outset. R is computed as the maximum distance between the centroid c of the point cloud and its farthest point. Once, both c and R of the sphere are calculated, the global reference frame localized in the centroid c with a radius R is obtained by applying EVD for the matrix M of equation (2). Therefore, GSHOT is calculated as a unique SHOT for an input point cloud. Although, GSHOT is tested with point clouds well segmented, we use to remove noise (outliers) the statistical method described in [Sch05]. The procedure for GSHOT is detailed in the Algorithm 1. GSHOT has been implemented from SHOT in Point Cloud Library (PCL) [AMT*12].

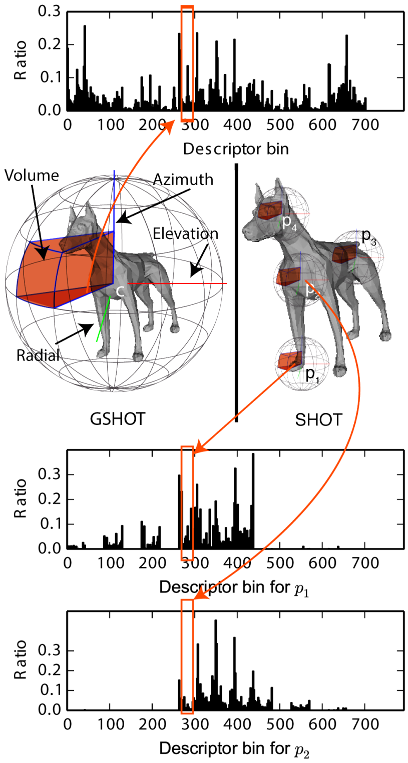

Figure 1: *Comparative of the reference frames for SHOT and GSHOT descriptors, using 792 bins with 12 divisions for azimuth, 6 for elevation and 1 for radial and 11 bins per histogram.*

4. Experiments

RGBD object dataset [LBRF11] and the dataset of Princeton Shape Benchmark (PSB) [SMKF04] were chosen among other datasets mentioned in state-of-art in order to test GSHOT in this work. PSB is composed of 1814 instances of CAD-objects (50% for training and 50% for testing) whereas RGBD is composed of 300 real household objects where each of them has 540 views (image, depth and point clouds) generated using 3 camera positions and 180 object poses. PSB was selected because it allows us to test GSHOT using CAD-models for avoiding problems such as acquisition noise, holes due to lack of points in the scanned surface or disparity in the texture or viewpoint, that are caused in the acquisition and capture processes with sensors. This fact adds difficulty in evaluating the capacity of GSHOT and its goodness for recognition processes.

Algorithm 1 To compute GSHOT descriptor

```
procedure GSHOTEstimation(P)
    $c \leftarrow [0,0,0]^T$
    for $i \leftarrow 1, |P|$ do
        $c \leftarrow c + p_i,\ p_i \in P$
    end for
    $c \leftarrow \frac{c}{|P|}$
    $R \leftarrow \{l \mid l = \|c - p_i\|_2 \wedge l > \|c - p_j\|_2 \wedge i \neq j\}$
    $[\mathbf{x}, \mathbf{y}, \mathbf{z}] \leftarrow$ getRF(P, $\{c\}$, 1, $R$)
    return SHOTEstimation(P, $[\mathbf{x}, \mathbf{y}, \mathbf{z}]$, $R$)
end procedure

procedure getRF(P, P*, $i$, $R$)
    $c \leftarrow p_i \in P^*$
    for $j \leftarrow 1, |P|$ do
        $q \leftarrow p_j \in P$
        $d_j \leftarrow \|c - q\|_2$
        $M \leftarrow M + (R - d)(c - q)(c - q)^T$
    end for
    $M \leftarrow \mathbf{VDV}^{-1}, \mathbf{V} = [\mathbf{x}^+, \mathbf{y}^+, \mathbf{z}^+]$
    $//Disambiguateaxes$
    $S_x^+ \leftarrow \{i : d_j \leq R \wedge (p_j - p) \cdot \mathbf{x}^+ \geq 0\}$
    $S_z^+ \leftarrow \{i :\ d_j \leq R \wedge (p_j - p) \cdot \mathbf{z}^+ \geq 0\}$
    $S_x^- \leftarrow \{i : d_j \leq R \wedge (p_j - p) \cdot \mathbf{x}^- > 0\}$
    $S_z^- \leftarrow \{i :\ d_j \leq R \wedge (p_j - p) \cdot \mathbf{z}^- > 0\}$
    $\mathbf{x} \leftarrow x^+$ if $|S_x^+| \geq |S_x^-|$ else $\mathbf{x}^-$
    $\mathbf{z} \leftarrow z^+$ if $|S_z^+| \geq |S_z^-|$ else $\mathbf{z}^-$
    $\mathbf{y} \leftarrow \mathbf{z} \times \mathbf{x}$
end procedure

procedure SHOTEstimation(P, $[\mathbf{x}, \mathbf{y}, \mathbf{z}]$, $R$)
    for $i \leftarrow 1, |P|$ do
        $p'_i \leftarrow [\mathbf{x}, \mathbf{y}, \mathbf{z}] \times p_i$
        quantize $p'_i$ wrt the spatial grid        ▷ which means estimate the sphere volume where lies $p'_i$
        $\theta \leftarrow \mathbf{n}_i \cdot \mathbf{z}$
        quantize $\theta$ wrt the shape histogram bins
        quadrilinear interpolation to accumulate $p'_i$        ▷ according to the authors of [STD14]
    end for
    normalize the descriptor to Euclidean norm 1
end procedure
```

Furthermore, to test GSHOT with real objects from RGBD sensors, 29 household object from RGBD dataset was randomly chosen (75% for training and 25% for testing). Note that RGBD dataset was not adapted to be used, but PSB CAD-objects was transformed to point clouds from their original formats as polygonal meshes to be able to apply GSHOT and the rest of descriptors of experiments. To do this, we have simulated a laser beam with a resolution of 1 mm for sampling each instance of object and thus a point cloud without acquisition noise can be obtained. Âă

We have used the PSB dataset organized in 7 classes with a Support Vector Machine (SVM) [CV95] and the RGBD dataset with a K Nearest Neighbors(kNN) [ML14]. We have tested, the proposed descriptor with the two classifiers to prove the non-dependency of them. For finding the best parameter γ and the constant C of SVM classifier, a grid search technique based on cross-validation with the training set was performed. This search for the best parameters tuning was carried out for all descriptors, not just GSHOT. This way, kNN has been adjusted to $k = 16$, being that the used datasets gather a large number of samples.

In general terms, the experimental results are illustrated using "Area Under the Curves" (AUC) of "Receiver Operating Characteristic" (ROC) and "Precision-Recall" (PR).

In Experiment 1, Figure 2 shows how GSHOT works with respect to each object class. The animal 90.8% and furniture 89.1% objects class obtain the best results, in contrast to the household 61.9% and non-class 61.3% objects class, that keep the worst results. This fact may be due to their geometries. The first classes represent instances in which all of them have a similar geometry, e.g. all instances of table have more or less the same geometry. In the second classes, there is a lot of variability in the geometry of the different instances of a same object class. This issue occurs in the non-class, e.g they can be wheels or slot machines. Additionally, in the case of household, this fact happens because the objects are almost flat (without volume) as scissors.

In experiment 2, we have compared the goodness of GSHOT (Figure 3) versus other descriptors aforementioned in Section 2 for both PSB and RGBD datasets. First, to highlight that ROC-PSB of the figure shows as GSHOT improves the success ratio with respect to its local implementation SHOT, GSHOT reaches precision rates of 88% versus 86% of both VFH and SHOT, and 87% of CVFH. Moreover, the ROC-PSB for GSHOT is always over the curves for the rest of descriptors aforementioned in Section 2. As expected, PR-PSB shows as the GSHOT precision decreases in a similar way to the rest of descriptors but its average value is a little better than the others when the recall is high. Only when the recall level is low, CVFH is slightly better 91% versus 90% for GSHOT. Second, note that ROC-RGBD shows how GSHOT keeps a good behaviour although CVFH has a better result. This fact could be as a consequence of how CVFH computes its descriptor for each stable region, as was explained in section 2. Other reason is that CVFH is more robust to noise present in real scenes as gathered in RGBD dataset. Summarizing, in this last case, CVFH is favored against GSHOT because the point clouds are not closed and they usually present occlusions caused by the camera's viewpoint.

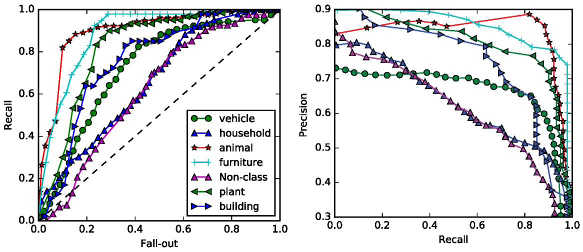

Figure 2: *(Left) ROC and (Right) PR to determine the behaviour of GSHOT to classify the 7 classes of the PSB dataset.*

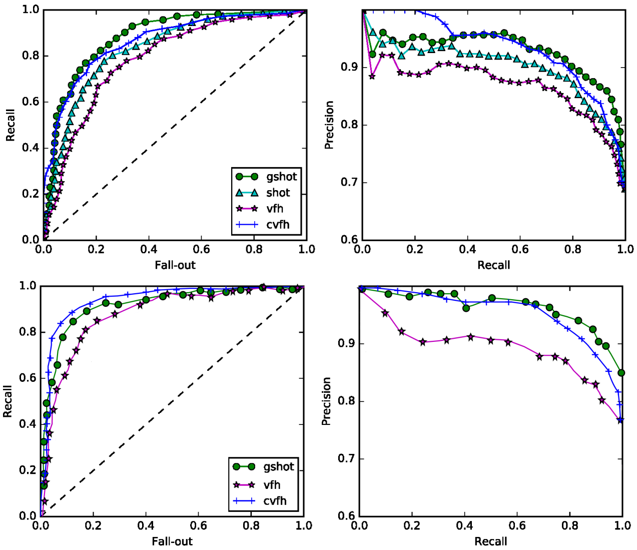

Figure 3: *ROC and PR curves: (Top) PSB dataset. (Bottom) RGBD dataset.*

Additionally, using PSB dataset, a brief analysis of presented descriptor in terms of both time and space complexity was also performed as shown in Table 1. As observed, GSHOT not only reaches better success ratio than SHOT but also it improves SHOT ×57 in space and ×3 in runtime. Besides, it accomplishes similar performance results of VFH at the same time that it achieves better results in terms of classification as shown Figure 3(Top). GSHOT also has less computational cost than CVFH and its precision is better than CVFH for CAD-Objects and only slightly worse for objects from real scenes. Note that the time retrieves in each experiment is dependent on the amount of points within the point cloud. The average of point per cloud for all the object instances in the PSB dataset is ∼200 thousand.

Table 1: *The average of timing and spatial performance using an Oracle Grid Engine with 26 nodes composed of 2 Intel XEON X5600 hexacore and 48GB of RAM*

Descriptor	*Mean time(s)*	*Space requirements(KB)*
CVFH	3.7 ± 0.1	5.27 ± 3.17
GSHOT	2.4 ± 0.2	4.36 ± 0.51
SHOT	7.8 ± 0.5	246.82 ± 44.41
VFH	2.1 ± 0.04	2.55 ± 0.52

5. Conclusions

This work has presented a new global descriptor (GSHOT) based on the modification SHOT. GSHOT enhances the performance of current descriptor methods at the same time that it gets excellent results in phase of the classification using two well-known datasets such as PSB and RGBD. GSHOT not only improves the runtimes and memory required regarding to SHOT and other global descriptors, but also it increases the success rate in comparison with those descriptors.

Acknowledgment

This work was supported by the European FEDER funds, the Spanish Ministry of Economising and the Valencia Regional Government through the research projects DPI 2015-68087R and PROMETEO 20/3/085, as well as the pre-doctoral grant BES-2013-062864

References

[ABG*11] Aldoma A., Blodow N., Gossow D., Gedikli S., Rusu R. B., Vincze M., Bradski G.: CAD-model recognition and 6 DOF pose estimation using 3D cues. In *Proc. ICCV Workshop on 3D Representation and Recognition (3DRR)* (2011), pp. 585–592. 1, 2

[AMT*12] Aldoma A., Marton Z.-C., Tombari F., Wohlkinger W., Potthast C., Zeisl B., Rusu R. B., Gedikli S., Vincze M.: Tutorial: Point cloud library: Three-dimensional object recognition and 6 dof pose estimation. *IEEE Robotics & Automation Magazine 19*, 3 (2012), 80–91. 2

[Bre01] Breiman L.: Random forests. *Machine learning 45*, 1 (2001), 5–32. 1

[CV95] Cortes C., Vapnik V.: Support-vector networks. *Machine learning 20*, 3 (1995), 273–297. 1, 3

[JKJ*13] Janoch A., Karayev S., Jia Y., Barron J. T., Fritz M., Saenko K., Darrell T.: A category-level 3d object dataset: Putting the kinect to work. In *Consumer Depth Cameras for Computer Vision*. Springer, 2013, pp. 141–165. 1

[LBRF11] Lai K., Bo L., Ren X., Fox D.: A large-scale hierarchical multi-view rgb-d object dataset. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on* (2011), IEEE, pp. 1817–1824. 1, 2

[MGT16] Mateo C., Gil P., Torres F.: Visual perception for the 3d recognition of geometric pieces in robotic manipulation. *The International Journal of Advanced Manufacturing Technology 83*, 9-12 (2016), 1999–2013. 2

[ML14] Muja M., Lowe D. G.: Scalable nearest neighbor algorithms for high dimensional data. *IEEE Transactions on Pattern Analysis and Machine Intelligence 36*, 11 (2014), 2227–2240. 1, 3

[RBB09] Rusu R. B., Blodow N., Beetz M.: Fast Point Feature Histograms (FPFH) for 3D registration. In *IEEE International Conference on Robotics and Automation* (2009), pp. 3212–3217. doi:10.1109/ROBOT.2009.5152473. 1, 2

[RBTH10] Rusu R. B., Bradski G., Thibaux R., Hsu J.: Fast 3D recognition and pose using the viewpoint feature histogram. In *IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings* (2010), pp. 2155–2162. doi:10.1109/IROS.2010.5651280. 1, 2

[Sch05] Robust filtering of noisy scattered point data. *Proceedings Eurographics/IEEE VGTC Symposium Point-Based Graphics, 2005.* (2005), 71–144. URL: http://ieeexplore.ieee.org/document/1500321/, doi:10.1109/PBG.2005.194067. 2

[SMKF04] Shilane P., Min P., Kazhdan M., Funkhouser T.: The princeton shape benchmark. In *Shape modeling applications, 2004. Proceedings* (2004), IEEE, pp. 167–178. 1, 2

[SSN*14] Singh A., Sha J., Narayan K. S., Achim T., Abbeel P.: Bigbird: A large-scale 3d database of object instances. In *Robotics and Automation (ICRA), 2014 IEEE International Conference on* (2014), IEEE, pp. 509–516. 1

[STD14] Salti S., Tombari F., Di Stefano L.: SHOT: Unique signatures of histograms for surface and texture description. *Computer Vision and Image Understanding 125* (2014), 251–264. URL: http://dx.doi.org/10.1016/j.cviu.2014.04.011, doi:10.1016/j.cviu.2014.04.011. 1, 2, 3