PRINCIPAL CURVATURE OF POINT CLOUD FOR 3D SHAPE RECOGNITION

Justin Lev[*†], *Joo Hwee Lim*[⋆†], *Nizar Ouarti*[*‡]

[*] IPAL, CNRS UMI 2955, Singapore
[⋆] Institute for Infocomm Research (I2R), A-STAR, Singapore
[†] Université Grenoble Alpes (UGA), France
[‡] Université Pierre et Marie Curie (UPMC), France

ABSTRACT

In the recent years, we experienced the proliferation of sensors for retrieving depth information on a scene, such as LIDAR or RGBD sensors (Kinect). However, it is still a challenge to identify the meaning of a specific point cloud to recognize the underlying object. Here, we wonder if it is possible to define a global feature for an object that is robust to noise, sampling and occlusion. We propose a local measure based on curvature. We called it *Principal Curvature* because rather than using the Gaussian curvature we keep the information of the two principal curvatures. In our approach, this local information is then aggregated as histograms that are compared with a Chi-2 metric. Results show the robustness of the method particularly when only few points are available. This means that our approach can be very suitable to match objects even with a limited resolution and possible occlusions. It could be particularly adapted to recognize objects with LIDAR inputs.

Index Terms— Object recognition, Point cloud, Principal curvature, Histograms.

1. INTRODUCTION

According to Dali (1956), "*The least one can ask of a sculpture is to stay still*". In this study, we reinterpret this idea by determining an invariant measure of shape not dependent on the point of view of the observer.

With the proliferation of sensors for retrieving depth information on a scene, such as LIDAR [1, 2, 3] or RGBD with a kinect [4, 5, 6, 7, 8, 9, 10, 11], it has become common to record 3D points. However, in fields such as robotics and autonomous cars, it becomes critical to indentify the meaning of these point clouds. In order to extract a true semantic points clouds, different processing levels are involved. The first level of analysis is the segmentation (ground, wall, objects, etc.), another one is the recognition of the objects in the scene that was previously segmented. In this study, we will focus on the aspect of object recognition.

{justin.lev, nizar.ouarti}@ipal.cnrs.fr, joohwee@i2r.a-star.edu.sg

The major problem of the recognition of 3D objects, is that a previously scanned object can be seen, in a real scene, in different poses and scales [12]. Furthermore, the recording of the object can be noisy or some points could not be necessarily detected by the sensors. Another challenge could be the occlusion of some parts.

Is it possible to create a robust measure to recognize objects, a specific signature of an object, despite a fickle the signal due to the real world conditions?

Tabia et al. [13] exploit an image-oriented approach, the principle is to record the image depth from multiple poses and then extract descriptors that the authors claim to be pose-invariant. Graph approaches are based on the topology underlying the different vertices that compose the object [14, 15]. They are also immune to non-rigid deformation because the topological relationships among graph vertices are preserved. A disadvantage for such approach is that we must build a mesh before we can run these methods. Moreover, it is not always a simple task to determine an accurate mesh from a point cloud, especially when the point cloud is noisy. This approach also becomes unreliable when data are missing (hole in the points cloud), resulting in missing vertices in the mesh.

In this study we propose a new local measure based on principal curvatures and that is robust to scale, rotation, translation but also to random downsampling and occlusions. The originality of our study is to be practical with unstructured point clouds (a mesh is not needed) and without color information ie our performances are based solely on the 3D geometry.

2. OUR APPROACH

Our approach is based on a local measure of the curvature that we wanted to connect to human perception. Some physiological data suggest that visual perception in Humans is sensitive to 3D local features [16]. In this study we decided to aggregate these local information to obtain a global signature of the shape inspired by the Gestalt[1]. Indeed, one of the

[1] "An organized whole that is perceived as more than the sum of its parts.", Oxford English Dictionary.

essential points of the Gestalt is to consider that the perception of the overall shape emerges from individual stimuli.

In this context the individual stimuli are the curves and signature of the overall shape referred to a histogram. By comparing a histogram with different recorded histograms, it is possible to recognize a given shape.

For a point cloud V with N points, $V = \{v_i\}_{i\in[1;N]}$, we proposed to apply the following transformation. We call S the surface underlying by V.

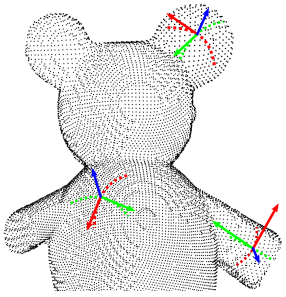


Fig. 1: Principal curvatures r/g/b: 1st/2nd/normal directions.

2.1. Preprocessing

To obtain a more robust results, our first step consists in a normalization of the point cloud V:

$$\forall i \in [1;N],\ \tilde{v_i} = \alpha \cdot \frac{v_i - \mu_V}{\sqrt{\sum_i \|v_i\|_2^{\,2}}}, \quad \mu_V = \frac{1}{N} \cdot \sum_{i=1}^{N} v_i \quad (1)$$

Then, the barycenter of V is set at the origin and $\|V\|_\infty = \alpha$ (with α set to 100).

From now onward, V and $\{v_i\}_i$ will refer to the centered and normalized data.

2.2. Neighgourhood choice

To compute the curvatures of each points v_i of V, we define for every of them a neighbourhood B_i of size r_λ such as:

$$B_{i,\lambda} = B(v_i, r_\lambda) = \{p \in \mathbb{R}^3 \,|\, \|v_i - p\|_2 \leqslant r_\lambda\} \quad (2)$$

$$\text{with}: \quad r_\lambda = \lambda \cdot \underbrace{\max_{i,j}(\|v_i - v_j\|_2)}_{M}, \quad \lambda \in [0;1]. \quad (3)$$

Thus, our neighbourhood radius is proportional to the maximum length M of the object.

For a given λ value, we gather in $V_{i,\lambda}$ all points belonging to both V and $B_{i,\lambda}$: $V_{i,\lambda} = V \cap B_{i,\lambda}$. If the number of points in $V_{i,\lambda}$ is under a threshold τ ($\tau = 6$) then the point is skipped. Indeed, the computation requires a minimum numbers of point which can affect the accuracy of the computation.

2.3. Local and adapted basis

In order to determine an adapted local basis of origin $v_{i,\lambda}$, another step is the PCA processing [17, 18] on the neighbourhood $V_{i,\lambda}$ which materialize the surface $S_{i,\lambda}$. After a change of basis to this fitted one, called $\mathcal{B}_{\text{fit}}$, we can locally describe the surface $S_{i,\lambda}$ with:

$$z(x,y) = J_{B,n}(x,y) + O(\|(x,y)^{n+1}\|_2), \quad (4)$$

$$\text{with} \quad J_{B,n}(x,y) = \sum_{p=1}^{n}\left(\sum_{q=0}^{p} B_{p-q,q}\, x^{p-q} y^q\right). \quad (5)$$

In $\mathcal{B}_{\text{fit}}$ we select the polynomial function $J_{A,n}$ of degree n ($n = 2$ in our case) with a surface $Q_{i,\lambda}$ closest (in the least mean square sense) to $S_{i,\lambda}$:

$$\varepsilon = \sum_{p\in S_i} \left(J_{A,n}(p) - z(p)\right)^2 \quad (6)$$

Finally, a Taylor development of this approximation can provide the normal and the origin of this new Monge frame, $\mathcal{B}_{\text{Monge}}$, that is obtain thanks to the shape operator also called "Weinggarten map". Then, it is possible to have access to the first and the second principal directions ($\overrightarrow{d_{i,\lambda,1}}$ and $\overrightarrow{d_{i,\lambda,2}}$). The first and the second principal curvature $k_{i,\lambda,1}$ and $k_{i,\lambda,2}$ ($k_{i,\lambda,1} \geqslant k_{i,\lambda,2}$) are also available for the point v_i. We represent these vectors on figure 1.

2.4. Principal curvatures histogram

The key novelty of our approach is to preserve the data related to $k_{i,\lambda,1}$ and $k_{i,\lambda,2}$. For this reason we did not combine them like in *Gaussian curvature* with

$$k_{gaussian,i,\lambda} = k_{i,\lambda,1} \cdot k_{i,\lambda,2}\,. \quad (7)$$

Each curvature are processed *independently*. In our case, the first and the second curvature are associated to a specific distributions $K_{\lambda,1} = \{k_{i,\lambda,1}\}_i$ and $K_{\lambda,2} = \{k_{i,\lambda,2}\}_i$, which provide the histograms $H_{\lambda,1}$ and $H_{\lambda,2}$ in an interval $I = [x_{min}; x_{max}]$. For points which curvature are beyond the extrema of I, we reduce their value to the closest border: x_{min} or x_{max}. A normalisation of the histograms is performed by the number of points used to compute the curvature.

Next step is the concatenation of $H_{\lambda,1}$ and $H_{\lambda,2}$: the resulting histogram H_λ is constituted by *all the information related to the curvature* of the object, as it signature.

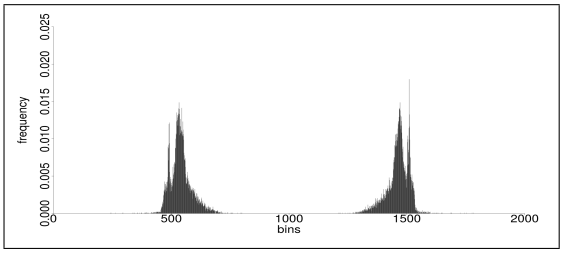


Fig. 2: H_λ, $\lambda = 6\%$

2.5. Recognition test

We will take advantage of these histograms to compare objects two by two. The estimation of the histograms similarity of two distinct objects A and B is based on Chi-2 distance:

$$d_{Chi2}(A,B) = \sum_j \frac{(H_A(j) - H_B(j))^2}{H_A(j) + H_B(j)} \tag{8}$$

To recognize an object, we compute the Chi-2 distance with each element in the reference dataset. We use a *Winner Take All* strategy so that the unknown object is assimilate to the one with the smallest distance.

However, before using the computation of the Chi-2 distance d_{Chi2}, we need to ensure the following:

1. $\sum_j H_A(j) = 1$ and $\sum_j H_B(j) = 1$, these conditions are obtained by normalisation.

2. Supposing that B is our reference object and A the query object (eq. 8), it requires that:

   $$\exists\, \eta \in \mathbb{R}^{+*} \quad | \quad \forall\, j \in [1;b] \quad \eta \leqslant H_B(j)$$

   This implies each bin to hold a sufficient number of element to consider H_B as a valid distribution. To overcome this problem, we aggregate the bins j with it neighbour bin $j+1$ when $H_B(j) < \eta$. To guarantee the comparability (same number of bins), the same aggregation has to be done to the histogram H_A. After such transformation, one might note $d_{Chi2}(A,B) \neq d_{Chi2}(B,A)$: the final bins will depend on the aggregate bins which will change with the reference we chose.

The implementation was realised in C++ with the CGAL library [19, 20] and in R[21].

3. MATERIAL AND METHODS

3.1. Dataset of scanned objects

For preliminary test, we will use a small dataset which is initially consists of 18 scanned objects collected from various sources [22, 23, 24, 25, 26, 27, 28]. They are isolated objects and represent various things such as toy animals, sculptures or decorative objects.

The different files were processed in order to remove the information related to the surface (reconstructed faces, mesh, color). Our objective is to be as close as possible from raw data arise from a 3D scanner.

One has to note that our dataset have been scanned with a high resolution and in very good conditions (nearly no noise). These properties are important given that we wanted to observe the impact when noise and downsampling is introduced.

3.2. Different transformations of the objects

To explore the robustness of our approach in conditions that are realistic with real data coming from a 3D scanner we apply some transformations.

Translation A translation is performed to the point cloud by a random vector $\overrightarrow{v}$.

Rotation A random rotation in the range $\theta \in [0; 2\pi[$ with a random axis $\overrightarrow{v}$ is performed to the point cloud.

Random downsampling To simulate resolution restriction during data acquisition, we performe a random downsampling by removing p percent of the points, $p \in \{10, 20, \cdots, 80, 90\}\%$.

Noise We corrupt the points by adding noise. This noise follows a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ with $\mu = 0$. σ is considered relatively to a ratio of r_λ (cf. eq.3), $\sigma = r_\lambda \cdot \{5, 10, \cdots, 45, 50\}\%$.

Occlusion Our points cloud is separated in two parts by a random plane. we compute the plane in order to obtain parts with $\{10, 20, 30, 40, 50\} \pm 5\%$ of the initial number of points. We also retain complementaries part, so we also have the one with $\{50, 60, 70, 80, 90\} \pm 5\%$ of points.

To conclude, for each object from the dataset, 62 transformations (2 translations, 2 rotations, 18 resamplings, 20 noises and 20 occlusions) were applied for total of 1116 objects to be recognize.

3.3. Principal curvature versus Gaussian curvature

One salient aspect of our algorithm is to use histogram of principal curvatures. For this reason we wanted to characterise this specific feature by comparing the effect on the accuracy of recognition when Gaussian Curvature is used instead. It means that the same algorithm is applied in the case of Gaussian Curvature but with a variant when the histogram are computed.

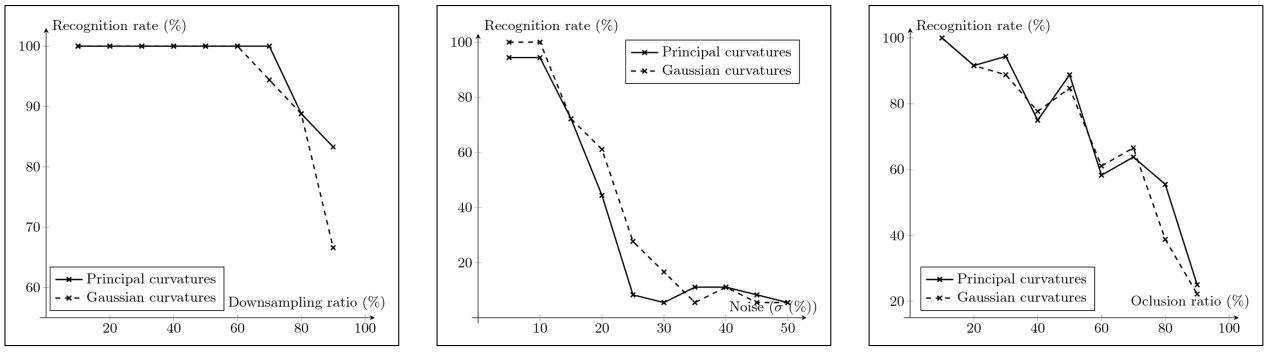
4. EXPERIMENTAL RESULTS

The overall success of recognition with a *Winner take All* strategy is a 70% recognition rate.

Translation and Rotation Invariance In the case of a rotation or a translation of the object, a success rate of 100% is obtained, with an average d_{Chi2} distance inferior to 10^{-2}.

Downsampling We observe (fig.3a) that our measure is robust to downsampling. The performance are very stable even after removing 80% of the original points.

We note that our signature (using principal curvatures) gives a significantly better recognition rate than the one



(a) depending on re-sampling ratio (b) depending on noise (with σ value) (c) depending on occlusion ratio

Fig. 3: Recognition rate

using the Gaussian curvature, particularly when only few points are available.

Noise robustness with variation of σ Our method is more sensitive to the noise (fig.3b). If σ exceeds 15% of r_λ (cf. eq.3), the results begin to be less accurate and fall suddenly when σ is above 20% of r_λ. Unlike the downsampling case, the result is now slightly better by using the Gaussian curvature. But both curve decrease quickly when σ is higher than 10% of r_λ.

occlusion robustness The recognition rate is linearly related to the number of points available for the processing (fig.3c). We think this result is encouraging for our next phases. Here again, our approach provide in average better results than the classic use of the Gaussian curvature.

5. DISCUSSION

5.1. Interpretation of results

The results regarding the translations and rotations are good as expected. Indeed, our calculation method of the curvature on a point is based on the neighbour V_i (cf.2). The computation is related to the relative position of neighbours hence it is not dependant of the basis in which the coordinates are projected. Consequently, our measure is totally invariant in term of translation and rotation: we can do 3D acquisitions of an object from different poses and that will not affect the recognition.

The algorithm is also very robust to downsampling. It is surprising to obtain a recognition rate superior to 80% for a downsampling of 90% of the points.

Indeed, in our approach only few points are required to be able to compute an accurate curvature. It is less the case with the Gaussian curvature variant where it can be seen that with few points (90% lost) the recognition rate dropped dramatically compared to our method. This can be explained by the fact that our method preserves more information along the two principal component meaning that if one component is lost the second can still be exploited. In the Gaussian curvature approach when one component is lost, it will impact the overall measure.

Regarding the resistance to noise, one can consider the results are good until σ is a value equal to 15% of the neighbourhood size calculation. Here too, the results can be explained easily: a noise on the points of the neighbourhood v_i will substantially modify the approximated surface and thus change the curvature of the point. Compared to the variant based on Gaussian curvature, it is the only perturbation where our approach is not the best.

Finally, we showed a linear relation between the number of points available and the recognition rate. This is a very encouraging results because it is practical to have only 50% of the points and the probability of recognition is still high, around 80%. Compared to Gaussian curvature variant our results are better on average. This could be explained because our Principal Curvature approach captures the specificity specific of a given shape and hence more robust when some points are missing.

6. CONCLUSION

In this study, we characterised the robustness of a new recognition method for 3D point cloud called *Principal Curvature*. This method based on curvature does not require to compute a mesh: a point cloud is sufficient to recognise the underlying object. We showed this method is suitable for real situations where the object is seen in any intended pose, the noise is present, the acquisition points are sparse and some part of the points cloud are missing. Given the properties of a LIDAR, our technique can be very effective with this kind of device.

In future works, we will focus on the robustness of our method by using with a multiscale approach using different radius of neighbourhood. We expect to increase the recognition rate and made our framework more robust to noise. Another prospect is recognition of objects included in a 3D scene that requires 3D segmentation [29].

7. REFERENCES

[1] J Niemeyer, F Rottensteiner, and U Soergel, "Conditional random fields for lidar point cloud classification in complex urban areas," *ISPRS annals of the photogrammetry, remote sensing and spatial information sciences*, vol. 1, no. 3, pp. 263–268, 2012.

[2] Adrien Gressin, Clément Mallet, Jérôme Demantké, and Nicolas David, "Towards 3d lidar point cloud registration improvement using optimal neighborhood knowledge," *ISPRS journal of photogrammetry and remote sensing*, vol. 79, pp. 240–251, 2013.

[3] Norbert Haala, Michael Peter, Jens Kremer, and Graham Hunter, "Mobile lidar mapping for 3d point cloud collection in urban areas?a performance test," *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci*, vol. 37, pp. 1119–1127, 2008.

[4] Kevin Lai, Liefeng Bo, Xiaofeng Ren, and Dieter Fox, "A large-scale hierarchical multi-view rgb-d object dataset," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 1817–1824.

[5] Kevin Lai, Liefeng Bo, Xiaofeng Ren, and Dieter Fox, "Sparse distance learning for object recognition combining rgb and depth information," in *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. IEEE, 2011, pp. 4007–4013.

[6] Manuel Blum, Jost Tobias Springenberg, Jan Wülfing, and Martin Riedmiller, "A learned feature descriptor for object recognition in rgb-d data," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 1298–1303.

[7] Shuai Tang, Xiaoyu Wang, Xutao Lv, Tony X Han, James Keller, Zhihai He, Marjorie Skubic, and Shihong Lao, "Histogram of oriented normal vectors for object recognition with a depth sensor," in *Asian conference on computer vision*. Springer, 2012, pp. 525–538.

[8] Liefeng Bo, Xiaofeng Ren, and Dieter Fox, "Unsupervised feature learning for rgb-d based object recognition," in *Experimental Robotics*. Springer, 2013, pp. 387–402.

[9] Saurabh Gupta, Pablo Arbelaez, and Jitendra Malik, "Perceptual organization and recognition of indoor scenes from rgb-d images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 564–571.

[10] Saurabh Gupta, Ross Girshick, Pablo Arbeláez, and Jitendra Malik, "Learning rich features from rgb-d images for object detection and segmentation," in *European Conference on Computer Vision*. Springer, 2014, pp. 345–360.

[11] Mohamed Yassine Tsalamlal, Paul Issartel, Nizar Ouarti, and Mehdi Ammi, "Hair: Haptic feedback with a mobile air jet," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 2699–2706.

[12] Nizar Ouarti, Bruno Sauvet, Sinan Haliyo, and Stéphane Régnier, "Robposit, a robust pose estimator for operator controlled nanomanipulation," *Journal of Micro-Bio Robotics*, vol. 8, no. 2, pp. 73–82, 2013.

[13] Hedi Tabia, David Picard, Hamid Laga, and Philippe-Henri Gosselin, "Compact vectors of locally aggregated tensors for 3d shape retrieval," in *Eurographics workshop on 3D object retrieval*, 2013.

[14] Alexander Agathos, Ioannis Pratikakis, Panagiotis Papadakis, Stavros J Perantonis, Phillip N Azariadis, and Nickolas S Sapidis, "Retrieval of 3d articulated objects using a graph-based representation.," *3DOR*, vol. 2009, pp. 29–36, 2009.

[15] Jin Xie, Yi Fang, Fan Zhu, and Edward Wong, "Deepshape: Deep learned shape descriptor for 3d shape matching and retrieval," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1275–1283.

[16] Mark Wexler and Nizar Ouarti, "Depth affects where we look," *Current Biology*, vol. 18, no. 23, pp. 1872–1876, 2008.

[17] Frédéric Cazals and Marc Pouget, "Estimating differential quantities using polynomial fitting of osculating jets," *Computer Aided Geometric Design*, vol. 22, no. 2, pp. 121–146, 2005.

[18] Manfredo Perdigao Do Carmo and Manfredo Perdigao Do Carmo, *Differential geometry of curves and surfaces*, vol. 2, Prentice-hall Englewood Cliffs, 1976.

[19] The CGAL Project, *CGAL User and Reference Manual*, CGAL Editorial Board, 4.8.1 edition, 2016.

[20] Marc Pouget and Frédéric Cazals, "Estimation of local differential properties of point-sampled surfaces," in *CGAL User and Reference Manual*. CGAL Editorial Board, 4.8.1 edition, 2016.

[21] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2008, ISBN 3-900051-07-0.

[22] Greg Turk and Marc Levoy, "Zippered polygon meshes from range images," in *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*. ACM, 1994, pp. 311–318.

[23] Brian Curless and Marc Levoy, "A volumetric method for building complex models from range images," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 303–312.

[24] Venkat Krishnamurthy and Marc Levoy, "Fitting smooth surfaces to dense polygon meshes," in *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*. ACM, 1996, pp. 313–324.

[25] Ajmal S Mian, Mohammed Bennamoun, and Robyn Owens, "Three-dimensional model-based object recognition and segmentation in cluttered scenes," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no. 10, pp. 1584–1601, 2006.

[26] Ajmal Mian, Mohammed Bennamoun, and Robyn 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.

[27] Arjun Singh, James Sha, Karthik S Narayan, Tudor Achim, and Pieter Abbeel, "Bigbird: A large-scale 3d database of object instances," in *2014 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2014, pp. 509–516.

[28] Kaleem Siddiqi, Juan Zhang, Diego Macrini, Ali Shokoufandeh, Sylvain Bouix, and Sven Dickinson, "Retrieving articulated 3-d models using medial surfaces," *Machine vision and applications*, vol. 19, no. 4, pp. 261–275, 2008.

[29] Haziq Razali and Nizar Ouarti, "Lidarbox: A fast and accurate method for object proposals via lidar point clouds for autonomous vehicules," in *Image Processing (ICIP), 2017 24th IEEE International Conference on*. IEEE, 2017, pp. –.