

Doppler Beam Sharpening for 3D Object Detection

Mato Gudelj[#], Michael Meyer^{*}, Sven Tomforde[^], Johannes Betz[#]

[#]Technical University of Munich, Germany, ^{*}Cruise Munich, Germany, [^]Kiel University, Germany
mato.gudelj@tum.de, michael.meyer@getcruise.com, st@informatik.uni-kiel.de, johannes.betz@tum.de

Abstract—In this paper we investigate the use of Doppler Beam Sharpening (DBS) as a preprocessing step for radar data in an autonomous driving 3D detection pipeline to enhance the effective angular resolution. Due to the assumption of a static scene, the proposed method applies DBS together with a moving point filtering approach to reduce preprocessing artifacts for dynamic objects. We derive an optimal correction region to apply DBS selectively at angles where it provides a resolution advantage over the base angular resolution. We show promising quantitative results with an average improvement of 5.2% in average precision (AP) for off-boresight objects and run an ablation study isolating the effects of the individual components of our approach. While our experiments use a conventional Region Proposal Network (RPN) detection model with radar and camera Feature Pyramid Network (FPN) backbones, the approach is applicable to any radar object detection pipeline.

Keywords—3D object detection, Automotive radar, Autonomous driving, Doppler beam sharpening.

I. INTRODUCTION

Radar sensors play a crucial role in autonomous driving systems due to their robustness in adverse weather conditions and ability to directly measure velocity via the Doppler effect. These characteristics are complementary to camera sensors, which can be exploited with radar-camera fusion. However, the limited angular resolution of traditional radar sensors can hinder their effectiveness in precise 3D object detection tasks [1], [2]. To address this limitation, we propose the use of Doppler Beam Sharpening (DBS) as a preprocessing step in a radar-camera based 3D object detection pipeline.

While originally developed for aerial radar, DBS has been explored in the context of automotive radar, with previous work demonstrating qualitative improvements when combining multiple-input multiple-output (MIMO) radar with DBS [3]. This was extended by follow-up work for dynamic scenes [4] and better boresight resolution [5]. However, the integration of DBS into deep learning-based 3D detection pipelines has been less explored in existing literature.

We aim to bridge this gap by evaluating the effect of DBS on 3D detection quality by incorporating it as a radar data preprocessing step. Our main contributions are as follows:

- We propose to use DBS as a preprocessing step in a 3D object detection pipeline to enhance radar angular resolution, particularly for static off-boresight objects.
- We derive an optimal correction region, that allows us to apply DBS only at angles where it brings a resolution advantage over the base angular resolution.

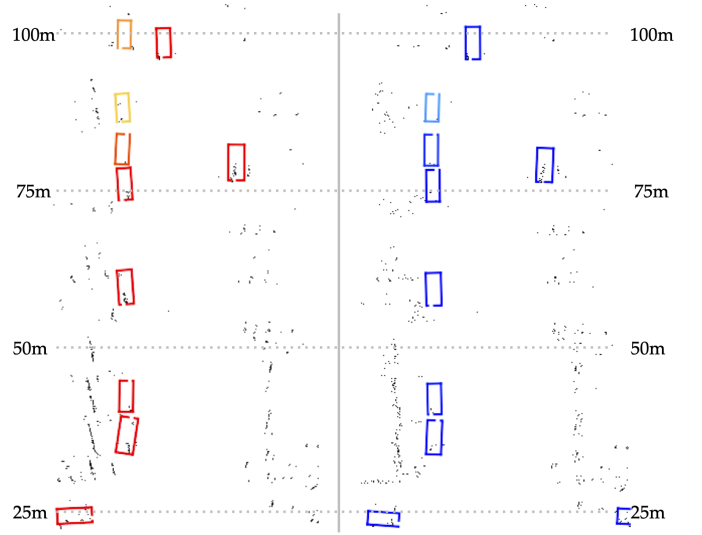


Fig. 1. Example long-range detection results before (left) and after (right) applying DBS. The resulting bounding boxes are better aligned, and the produced point cloud has higher resolution off-boresight.

- We show quantitative improvements to radar-camera 3D object detection performance from DBS and our methods with experiments and an ablation study.

The remainder of this paper has the following structure: Section II discusses related work on radar-camera based object detection and the application of DBS in automotive radar. Section III presents our proposed methods, including a background on DBS, the derivation of the optimal correction region, and the moving point filtering approach. Section IV describes our experimental setup and presents the quantitative results and ablation study. Finally, future research directions are discussed in Section V.

II. RELATED WORK

Camera-based 3D object detection using deep learning has seen significant advancements [6], [7]. However, the performance of camera sensors can be limited in adverse weather conditions and low-light environments. Radar sensors offer a complementary solution, providing robustness in challenging weather conditions and the ability to directly measure object velocity through the Doppler effect [1]. Nevertheless, traditional radar sensors often suffer from limited angular resolution and sparse returns, hindering their effectiveness in precise 3D object detection tasks [2].

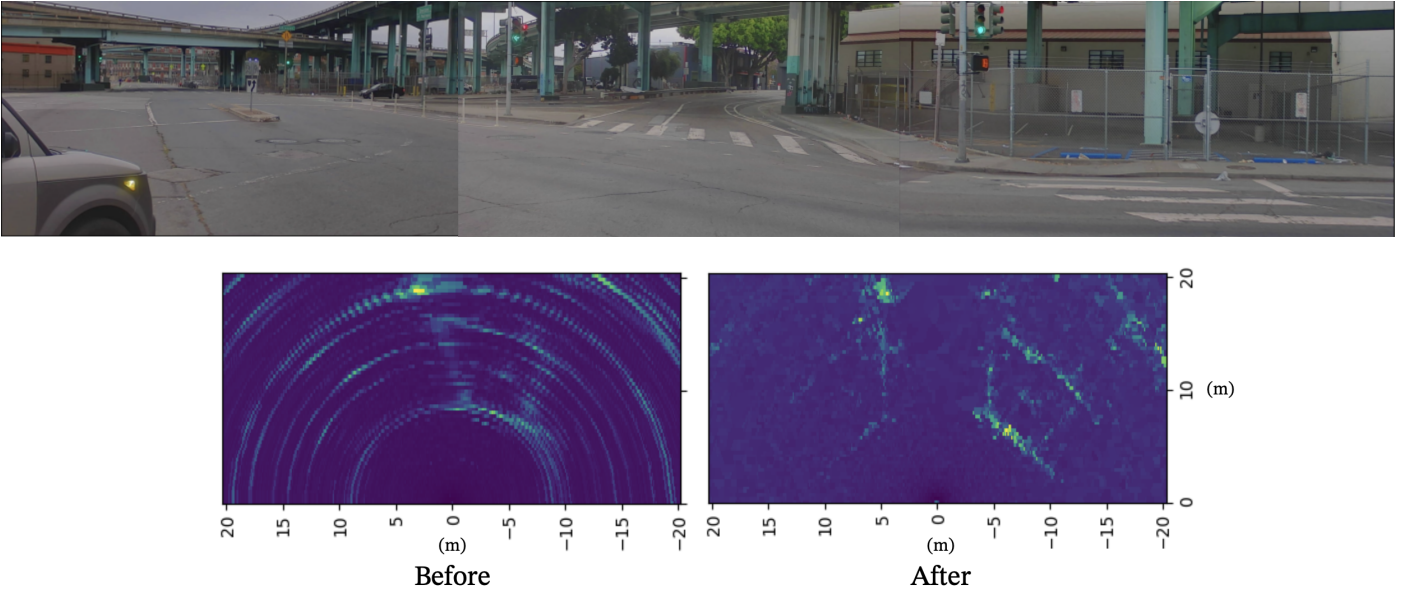


Fig. 2. Raw radar plots before and after applying DBS on an ideal static scene. The first row shows the camera view of the scene, while the second shows the bird’s-eye view (BEV) raw radar. The corrected radar plot displays a sharper outline of the street and right-side fence. This is without further postprocessing described in Sections III-B and III-C.

Radar-camera fusion has been proposed as a solution to overcome the limitations of the individual sensors. Meyer et al. [8] proposed a deep learning approach for 3D object detection on radar point clouds and camera images based on Region Proposal Networks (RPNs) [6]. Nobis et al. [9] investigated fusion strategies for radar and camera, demonstrating improved performance compared to single-sensor approaches, specifically in adverse weather conditions. CramNet [10] is an efficient approach to fuse the sensor readings from camera and radar in a joint 3D space by using attention [11], through a learned ray-constrained cross-attention mechanism. However, all radar-camera association approaches are affected by the relatively low radar resolution. DBS can be used to address the underlying problem by enhancing the angular resolution.

DBS has been previously explored in the context of automotive radar, with studies demonstrating qualitative improvements when combining multiple-input multiple-output (MIMO) 77GHz radar with DBS [3]. This was combined with a multi-target tracker by Cassidy et al. [4] for tracking dynamic actors in non-static scenes. Bekar et al. [5] applied the Burg algorithm to a MIMO-DBS enhanced radar to complement DBS by improving resolution in the forward-looking direction.

However, there is no existing work that investigates the impact of DBS on downstream performance of a deep neural network, including for 3D object detection. Given that radar is used with increasing prevalence in autonomous driving systems to complement the strengths of camera sensors, this presents a large gap in current literature. Methods that exploit the specifics of radar are an under-explored and thus promising direction for future research.

III. METHODS

We integrate DBS as a radar preprocessing step in a radar-camera object detection pipeline, together with two supplementary masking methods that address dynamic scenes and low DBS resolution at boresight (Sections III-B and III-C). In Section IV we evaluate 3D object detection performance of a model with and without DBS, together with an ablation study.

The background on DBS is presented in Section III-A. To minimise performance degradation at boresight due to decreased resolution, DBS is applied to an optimal correction region which we derive in Section III-B. Additionally, we implement a basic method for filtering out moving objects using out of distribution filtering, described in Section III-C.

A. Doppler Beam Sharpening

Doppler Beam Sharpening (DBS) exploits the high range and doppler velocity resolution of a radar system to make up for the inaccurate measurements of the azimuth angle. It works under the assumption of a static scene, where the radar returns are from non-moving objects in world view, e.g. street lights, poles or stationary cars. In this special case, the enhanced azimuth angle can be computed from the measured doppler velocity directly. This enhanced azimuth angle can then be used in conjunction with the range information to compute an updated position of the source of the radar return.

Ignoring elevation, the base relation for deriving DBS is:

$$d = v_e \cdot e(\theta), \quad (1)$$

where d is the processed doppler return representing the radial velocity, v_e is the ego velocity vector and $e(\theta)$ is a unit vector along the line determined by the enhanced azimuth angle θ .

This is the projection of the ego velocity v_e onto the azimuthal line. Rearranging Equation 1 gives a closed form for θ :

$$\theta_d = F(d) = \arccos\left(\frac{d}{|v_e|}\right), \quad (2)$$

with the assumption $|v_e| \neq 0$. The velocities here are not ego-compensated. Since d is a multiple of the velocity bin size v_B (subject to aliasing within the maximum unambiguous velocity range), for a fixed v_e this can be seen as a mapping between the velocity bin and the angle θ . The presence of arccos makes apparent the weak boresight resolution of DBS, as shown in Figure 3. This is further discussed in Section III-B.

Finally, the updated position $[x, y]$ of the point p is calculated using the measured range r :

$$x = r \cos(\theta), \quad y = r \sin(\theta). \quad (3)$$

B. Optimal Correction Region

Doppler correction as outlined in Section III-A has a resolution disadvantage at boresight and at low ego velocities. This is a consequence of the interaction between v_e , v_B and the nonlinearity of arccosine. For low ego velocities, in the extreme case when $v_e \leq v_B$, all θ would have the same value. However, even with a sufficiently large v_e the method still suffers from poor resolution near boresight at $\theta = 0$. This is due to the fact that the bin-angle mapping provided by Equation 2 is non-linear and the bins are concentrated around $\theta = \pi/2$ (see Figure 3). As a result, the correction only brings a resolution improvement for certain triplets (d, v_e, v_B) , and a degradation in resolution for others.

Taking this into account, we can define an *optimal correction region*, which we use to mask out non-beneficial corrections, i.e. those worse than some base angular resolution $1/\Delta$. The region is defined as a predicate, allowing angles θ above a threshold θ_{opt} :

$$\mathcal{J}(\theta) = |\theta| > F(v_B k_{opt}), \quad (4)$$

where F is as defined in Equation 2 and:

$$k_{opt} = \arg \max_k (F(v_B k - v_B) - F(v_B k) < \Delta_k). \quad (5)$$

C. Moving Point Filtering

The main drawback of this approach is the strong assumption of a static scene. On the other hand, most driving scenarios involve both static and dynamic returns. For this reason, to apply the correction method in practice, it is crucial to have a robust way to filter out returns from moving objects. Work by Cassidy et al. [4] applies static thresholding on the Doppler data. Our approach is similar, where we mask out the application of DBS for points with compensated velocities that deviate significantly from the mean.

For a histogram of point velocities $\mathcal{H}$, the resulting indicator filtering function given a point p is simply:

$$\mathcal{I}(p) = |\mu_{\mathcal{H}} - v_p| > k\sigma_{\mathcal{H}}, \quad (6)$$

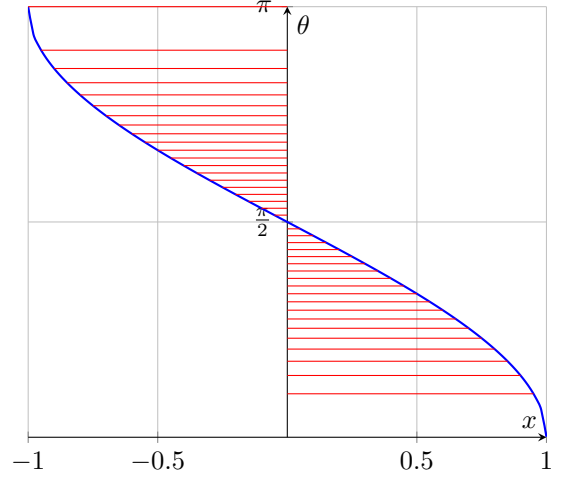


Fig. 3. Plot of the bin-angle mapping provided by Equation 2. The red lines show a mapping of $d/|v_e|$ to θ , for evenly spaced values of d . Static returns for which $d/v_e \approx 1$ are in front of the sensor. For those returns the correction could be very poor, as there are few velocity bins near $\theta = 0$.

where k is a hyperparameter, and $\mu_{\mathcal{H}}$ and $\sigma_{\mathcal{H}}$ are the mean and standard deviation of $\mathcal{H}$ respectively.

IV. EXPERIMENTS

We ran all of our experiments on a conventional RPN detection model with an FPN [7] camera backbone and an FPN radar backbone, which are fused in bird's-eye-view (BEV) [12]. We use a simple model to isolate the effects of the method from interfering approaches, such as learned correspondence matching from [10]. The method is applied to the raw radar data cube, followed by conversion to a point cloud. The input to the radar backbone is an occupancy grid, further augmented with CoordConv [13]. The occupancy grid is two channels, one for the corrected and original point cloud respectively. This is the case in all experiments aside from the ablation study, where we test the benefit of including the original point cloud.

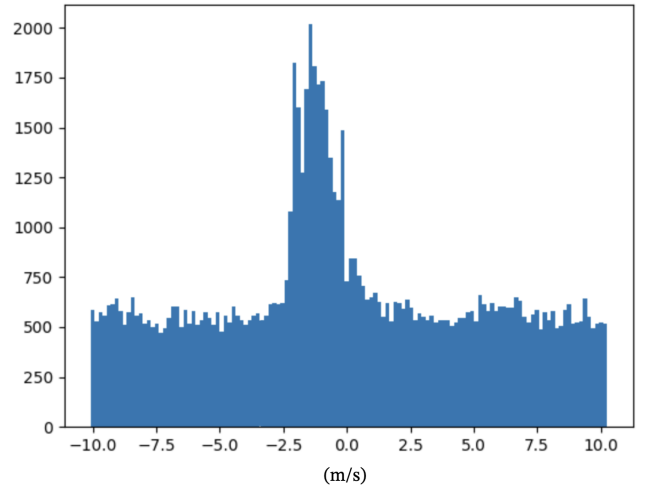


Fig. 4. Returned ego-compensated radar point velocities for a single scene. With our moving point detection method, we mask all points that deviate significantly from the mean. A threshold of 1.5σ is used in all experiments.

Table 1. Performance comparison in average precision (AP) across different ranges, without filtering near-boresight angles.

Method	Baseline (AP)	Ours (AP)
<40m $\uparrow$	72.64	73.86
<70m $\uparrow$	65.61	66.08
<102m $\uparrow$	59.78	60.44

A. 3D Object Detection Results

We ran two sets of evaluations, where one tries to isolate the effects of DBS off-boresight. This is necessary because the majority of objects do not meet the angle criteria necessary for DBS to be active (Section III-B), and would thus dilute the relative improvement for affected objects. Overall results can be found in Table 1, while results off-boresight are presented in Table 2. Both results show an improvement, however, evaluating off-boresight is more representative of the real improvement on affected objects. A direct application of DBS on raw tensor data is shown in Figure 2.

Table 2. Performance comparison in average precision (AP) across different ranges, evaluated on objects > 30 deg off-boresight to isolate the effects.

Method	Baseline (AP)	Ours (AP)
<40m $\uparrow$	50.24	52.18 $\pm$ 0.86
<70m $\uparrow$	42.65	44.92 $\pm$ 0.99
<102m $\uparrow$	40.03	42.54 $\pm$ 0.78

B. Ablation Study

We ran an ablation study to measure the impact of individual methods from Section III. The results are shown in Table 3. A naive application of DBS has a *negative* effect on detection results, due to issues more thoroughly discussed in Sections III-B and III-C. The largest improvement comes from the addition of the Optimal Correction Region (+OCR), as it isolates the effect of DBS only to regions where it has a resolution advantage. Including the unmodified point cloud (+ORG) also leads to an improvement. We hypothesize this is due to the failure of MPF to adequately filter out some moving points, leading to incorrect DBS corrections. Including the original point cloud could allow the network to turn to the original points in these situations.

Table 3. Ablation runs in average precision (AP) across different ranges, same evaluation as in Table 2. MPF is Section III-C, OCR is Section III-B. ORG includes the uncorrected original pointcloud as an extra channel.

Method	Baseline	+DBS	+MPF	+OCR	+ORG
<40m $\uparrow$	50.24	38.55	36.41	50.15	53.18
<70m $\uparrow$	42.65	32.21	32.60	42.48	45.06
<102m $\uparrow$	40.03	30.89	32.05	41.03	42.94

V. CONCLUSION AND FUTURE WORK

In this paper, we investigated the use of Doppler Beam Sharpening (DBS) as a preprocessing step for radar data in an autonomous driving 3D detection pipeline to enhance the effective angular resolution. We derived an optimal correction region to apply DBS only at angles where it provides a resolution advantage over the base angular resolution.

Additionally, we examined a moving point filtering approach to mitigate preprocessing artifacts for dynamic objects.

Our experiments demonstrated an average improvement of 5.2% in AP for off-boresight objects when applying DBS in the optimal correction region along with moving point filtering. The ablation study isolated the effects of each proposed method, highlighting the importance of the optimal correction region in leveraging the benefits of DBS for 3D detection.

Our approach shows potential for enhancing radar-camera 3D object detection with DBS, with several avenues for future research. First, camera-based moving point filtering strategies, e.g. based on optical flow, could be explored to better handle dynamic scenes. Second, higher frequency radars featuring better Doppler resolution and larger unambiguous velocity ranges could further improve the DBS angular resolution and overall detection performance. Third, DBS could be combined with further enhancements to account for the boresight resolution weaknesses [3]–[5]. Moreover, future work could investigate the application of DBS to other deep learning perception tasks, such as semantic segmentation and tracking.

REFERENCES

- [1] J. Dickmann, J. Klappstein, M. Hahn, *et al.*, “Automotive radar the key technology for autonomous driving: From detection and ranging to environmental understanding,” *2016 IEEE Radar Conference (RadarConf)*, pp. 1–6, 2016.
- [2] A. Srivastav and S. Mandal, “Radars for autonomous driving: A review of deep learning methods and challenges,” *IEEE Access*, 2023.
- [3] S. L. Cassidy, S. Pooni, M. Cherniakov, E. G. Hoare, and M. S. Gashinova, “High-resolution automotive imaging using mimo radar and doppler beam sharpening,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 2, pp. 1495–1505, 2023.
- [4] S. Cassidy, E. Hoare, and M. Gashinova, “Doppler beam sharpening for high-resolution imaging in dynamic automotive scenes,” in *2022 19th European Radar Conference (EuRAD)*, 2022, pp. 325–328.
- [5] M. Bekar, C. Baker, and M. Gashinova, “Enhanced angular resolution in automotive radar imagery using burg-aided mimo-dbs approach,” in *2023 20th European Radar Conference (EuRAD)*, 2023, pp. 315–318.
- [6] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” *Advances in neural information processing systems*, vol. 28, 2015.
- [7] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2117–2125.
- [8] M. Meyer and G. Kusch, “Deep learning based 3d object detection for automotive radar and camera,” in *2019 16th European Radar Conference (EuRAD)*, 2019, pp. 133–136.
- [9] F. Nobis, M. Geisslinger, M. Weber, J. Betz, and M. Lienkamp, “A deep learning-based radar and camera sensor fusion architecture for object detection,” in *2019 Sensor Data Fusion: Trends, Solutions, Applications (SDF)*, 2019, pp. 1–7.
- [10] J.-J. Hwang, H. Kretschmar, J. Manela, *et al.*, “Cramnet: Camera-radar fusion with ray-constrained cross-attention for robust 3d object detection,” in *European Conference on Computer Vision*, Springer, 2022, pp. 388–405.
- [11] A. Vaswani, N. Shazeer, N. Parmar, *et al.*, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.
- [12] T. Roddick, A. Kendall, and R. Cipolla, “Orthographic feature transform for monocular 3d object detection,” *British Machine Vision Conference*, 2019.
- [13] R. Liu, J. Lehman, P. Molino, *et al.*, “An intriguing failing of convolutional neural networks and the coordconv solution,” *Advances in neural information processing systems*, vol. 31, 2018.