J. Meas. Sci. Instrum., 2024, 15(2): 276-284. http://jmsi.publish.founderss.cn
DOI: 10.62756/jmsi.1674-8042.2024028 jmsi@nuc.edu.cn

Consecutive-frame latent space normal estimation under sparse point clouds for 4D millimeter-wave radar

WU Yangxu[1,2*], YUAN Xinfang[1,2], CHEN Ping[1,2]

1.School of Information and Communication Engineering, North University of China, Taiyuan 030051, China;
2. State Key Laboratory of Dynamic Measurement Technology, North University of China, Taiyuan 030051, China

***Corresponding author:** WU Yangxu (wuyang@nuc.edu.cn)

Received: April 19, 2024 **Revised:** May 19, 2024 **Accepted:** May 22, 2024

Abstract: Aiming at the sparsity of point cloud data and the low accuracy of spatial alignment exhibited by millimeter-wave frequency-modulated continuous-wave (FMCW) radar in outdoor motion scenarios, a lightweight model for spatial alignment was proposed. This method was specifically tailored for point cloud processing across consecutive multi-frames in outdoor motion scenes captured by millimeter-wave radar. Leveraging spatio-temporal graph neural networks (ST-GNNs), it accurately estimated the hidden spatial normals of adjacent multi-frame point clouds, eliminating the need for position sensors. By transforming radar point cloud data from each frame into a unified observation coordinate system, the method facilitated multi-frame fusion of 4D point clouds and ensured precise scene alignment. Experimental results demonstrated that the proposed approach not only accurately assessed the spatial attitude of 4D point clouds but also effectively corrected and fused the coordinates of each point cloud frame. This enabled precise coordinate alignment during motion and vibration. Furthermore, the algorithm significantly enhanced point cloud imaging density, improved image accuracy and readability, and was capable of imaging both static and dynamic targets. It provided robust support for the application of millimeter-wave radar in outdoor motion scenes.

Key words: millimeter-wave radar; 4D point cloud; spatial alignment; latent space normal estimation; spatio-temporal graph neural networks

0 Introduction

Millimeter-wave radar has been widely used in automatic driving, security monitoring, and environmental sensing because of its good performance under adverse weather conditions[1,2]. Moreover, with the development of millimeter-wave radar hardware performance, radar point cloud, as a vital data format of millimeter-wave radar, has been widely used in advanced sensing tasks such as three-dimensional environment sensing and spatial structure construction.

However, compared with the LiDAR point cloud, the millimeter-wave radar point cloud is limited by antenna size, array size, wavelength limitations, and power factors, and inevitably faces the problem of sparse point clouds. Although the currently advanced pulsed millimeter-wave radar can detect thousands of point clouds within a detection range of 300 m, this achievement is based on sacrificing point cloud confidence, which leads to many ghost targets in the detection scene. Multiple-input multiple-output (MIMO) millimeter-wave radar employs a synthetic aperture approach[3,4], which performs high-resolution virtual aperture imaging in pitch and azimuth and senses the velocity of moving targets, ultimately obtaining a 4D spatial point cloud containing velocity information. But this high resolution comes at the sacrifice of imaging range. The problem not only limits the completeness and accuracy of the data but also poses a significant challenge for further data processing and target identification. Specifically, the sparsity of the millimeter-wave point cloud hinders the visualization of the object's shape, leading to a lack of semantic information. Moreover, due to ground reflection and the multi-path effect, there are a certain number of clutter points within the already scarce point cloud information.

In order to solve the aforementioned problems, performance improvements have been made to both millimeter-wave hardware and imaging algorithms. Currently, the mainstream hardware solution involves implementing MIMO arrays via radar cascading, which expands the virtual aperture and enhances the output of

the effective point cloud[5]. Alternatively, the detection space can be increased by incorporating position sensors and stitching multi-frame point clouds based on the acquired motion and attitude data. Nevertheless, hardware solutions still face significant challenges in further increasing point cloud density due to wavelength and antenna limitations. Additionally, position sensors can introduce perception delays and add complexity to the system.

In terms of imaging algorithms, to address the issue of sparse point clouds, the conventional approach achieves point cloud completion and stitching for large-scale detection scenes through continuous frame fusion[6]. However, this method is only effective in stationary environments. In dynamic environments, the point cloud cannot be accurately spatially corrected due to changes in the coordinate system. The introduction of the spatial normal vector estimation technique to tackle this challenge holds profound significance[7]. This method obtains the spatial coordinate restoration deviation value under the world Cartesian coordinate system by estimating the 3D normal vector for each frame of the point cloud. Subsequently, the coordinate restoration of the single-frame point cloud facilitates the unification of the coordinate system across multiple frames and the stitching of large-scale scenes.

Currently, spatial normal vector estimation is mostly based on the spatial rigid alignment method, which can achieve good results when there are rich point cloud features and high point cloud confidence. However, when using the point cloud obtained from FM continuous wave MIMO millimeter wave radar, the rigid alignment method cannot obtain accurate estimation results due to the sparse point cloud and unstable reflection points. A method is needed that can perform flexible normal vector estimation based on global point cloud features.

With the introduction of the hidden space normal estimation method, it brings unprecedented opportunities for the solution of sparse point cloud spatial normal estimation in millimeter wave radar. The hidden space normal estimation method can output the normal information end-to-end by modeling the input 3D point cloud and establishing the global feature representation of the point cloud. Firstly, the hidden space normal estimation can map the sparse point cloud into a high-dimensional hidden space, which enhances the expression ability and feature resolution of the data, and helps to accurately restore the original data information. Secondly, the hidden space normal estimation method can effectively recover missing information from incomplete data, fill the gaps in the data, and improve the integrity and continuity of the data. More importantly, this method is able to achieve accurate target localization and classification while maintaining the data structure, providing a new solution for target perception in practical application scenarios.

By investigating the hidden space normal estimation method, the method proposed in this paper was expected to provide more reliable and accurate millimeter-wave radar data support for generating high-quality millimeter-wave radar point clouds, particularly in the key areas of target detection, tracking, and identification. A new method was proposed to construct a normal vector correction model between frames using the spatial relationship between adjacent frames, and finally generated high-quality millimeter wave point clouds. Specifically, the innovation of this methodology was mainly reflected in the following two aspects.

1) A new spatial similarity rating method for millimeter-wave radar point clouds was defined, which could be adapted to the task of matching unstructured point clouds in mobile environments.

2) A point cloud spatio-temporal mapping network was constructed, which enabled end-to-end estimation of hidden spatial normals between multiple frames of a millimeter-wave radar point cloud. This method was also tested on real data collected on streets, parking lots, and in indoor and outdoor environments. The proposed method demonstrated better performance with fewer clutter points and denser real point clouds.

The main points and effects of sparse point cloud problem in millimeter wave radar were introduced in detail, and the principle and application of potential space normal estimation method based on spatiotemporal graph network of adjacent frames were discussed in detail. The advantages of the proposed method in improving data processing efficiency and target recognition accuracy were verified by experimental results.

1 Traditional methods

The performance of the automotive radar system is restricted by the limitations imposed by the size of the millimeter-wave radar antenna and the accuracy of signal resolution. Millimeter-wave radar point cloud sparsity remains the main difficulty hindering the expansion of millimeter-wave radar application scenarios. To address this problem, point cloud quality optimization is primarily performed using point cloud generation and

point cloud splicing methods. With the maturity of various generative methods, generating the point cloud using the original sparse point cloud as the basis[8,9] has been widely adopted. Although this method can significantly increase the number of point clouds and provide sufficient data characterization for the target detection tasks, the point clouds generated by this method tend to be detached from the detection process with low confidence. They cannot be applied to autonomous driving scenarios requiring high confidence point clouds. On the other hand, the point cloud splicing approach expands the number of point clouds by collecting and stacking sparse point clouds[10].

Since the point clouds generated by this method all originate from highly confident real detections, it is widely employed in stationary detection environments. However, in dynamic scenes, where each frame possesses a distinct spatial coordinate system, it is necessary to unify these coordinate systems before stacking. To address this problem, the current methods[11-13] calculate the relative spatial relationship between the target point cloud and the test point cloud and utilize the iterative nearest neighbor algorithm to complete the accurate registration of the point cloud pairs to achieve the spatial coordinate alignment. Or the deep neural networks (DNN) network framework is used to register the coordinate system[14]. Although this method can perform the registration of coordinates, it will impact the registration quality when the point cloud quality is poor. To solve this problem, Mikamo M et al.[15] utilized graph neural networks for attitude estimation of the overall point cloud and improved registration accuracy through deep mining of global features. However, this method cannot perform well in long splices and large vibration scenes due to the inability to obtain consecutive features in the front and back frames.

2 Proposed method

Consecutive frames of four-dimensional point clouds were used to perform latent space normal estimation and obtain spatial normal for attitude registration. As shown in Fig. 1, the overall architecture is divided into three parts: 1) Construction of a single-frame 4D spatio-temporal graph structure; 2) Latent space normal estimation of 4D spatio-temporal images between adjacent frames; 3) Attitude registration and spatial fusion of point clouds between frames. By establishing a spatial graph network model, the spatio-temporal graph and dynamic graph construction mechanisms were integrated to accurately assess the latent space attitude characteristics between different frames.

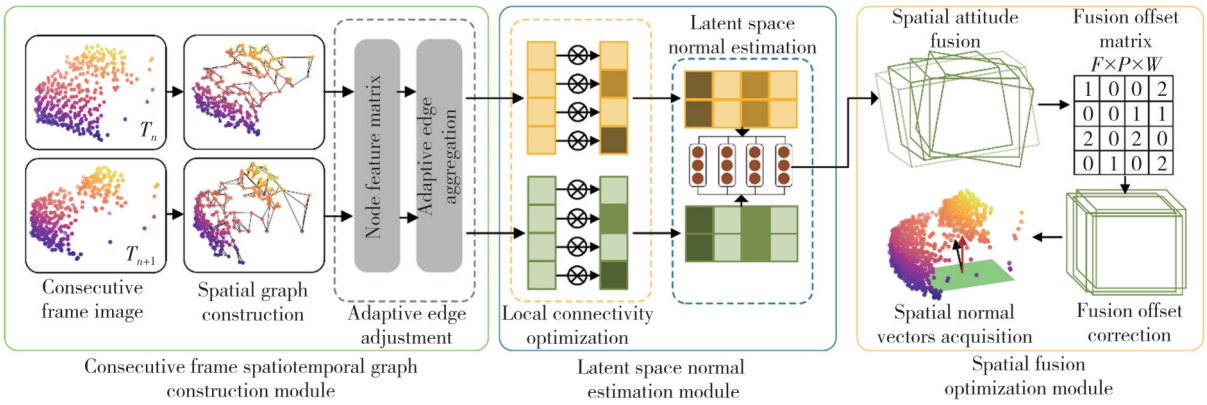


Fig. 1 Flow of consecutive frame latent space normal estimation method

2.1 Graph network construction for single frame point clouds

The construction of a spatial graph based on millimeter-wave 4D point clouds included a node feature matrix and edges between nodes. The node feature matrix converted the point cloud into nodes containing spatial features by combining the spatial position, speed information, and signal strength of the 4D point cloud. Then, the basic connecting edges between nodes were established based on the spatial distance of neighbor nodes.

First, N point clouds are defined in Euclidean space as a set $V=\{v_1,\cdots,v_N\}$, where $v_i=(x_i,s_i,h_i)$, x_i represents the three-dimensional space position feature, s_i represents the velocity feature, and h_i represents the signal intensity characteristics. In order to further compress the features, the number of nodes in each frame is kept within a fixed range. The inner product between adjacent nodes is used to calculate their feature similarity. Then it fuses adjacent nodes with similar features to obtain a fixed number of three-dimensional

space node sets at any time. Specifically, two fully connected layers are used to align feature representations s_i, h_i, and position attributes x_i in the embedding space, and then they are fused to obtain the node's attribute embedding.

$$Z = Concat\left[\sigma(x_i W_1), \sigma(s_i W_2), \sigma(h_i W_3)\right], \quad (1)$$

where W_1, W_2, and W_3 are the learnable weights; $Concat[\cdot]$ represents the concatenation operation and σ is the LeakyReLU activation function. The above steps can be used to complete the comprehensive characterization of node features. Then a three-dimensional spatial graph $G=(V,E)$ can be constructed. Taking V as a node, nodes are connected to other nodes in the radius neighbor region, and E denotes the connection relationship with neighboring nodes. Given a set of edges $\mathcal{E}=\{\mathcal{E}_{ij}|i,j\in V_t, \mathcal{E}_{ij}\notin 0\}$, after obtaining the node characteristics, the edges are established based on the affinity of node attributes, and the correlation between two node attributes is defined as $s(v_i,v_j)$. For $\forall v_i$, each node selects k ($k\leqslant|V|$) sets of neighboring nodes N_{v_i} and connects them through edges $\mathcal{E}_{ij}\in\mathcal{E}$. If the adjacency matrix $\boldsymbol{A}_{i,j}$ is not empty, it means that there is a dependency between $v_{i,t}$ and $v_{j,t}$. The correlation calculation process is

$$s(v_i,v_j)=a_t=l_2(\varphi(v_i)^{\mathrm{T}}\phi(v_j)), \quad (2)$$

where φ and ϕ represent two linear transformation functions, which can be expressed as $\varphi=W_\varphi Z$, $\phi=W_\phi Z$, W_φ and W_ϕ are the learnable parameters in the two transformation functions, respectively, which are multiplied by the two matrices in order to obtain the adjacency matrix $\boldsymbol{A}_t$. Finally, a l_2 normalization is performed for each row in the matrix. This approach allows adaptive selection and establishment of topological dependencies, where edges describe spatial geometric associations within the same frame. After constructing nodes and edges, the established adaptive spatial graph topology is obtained.

2.2 Latent space normal estimation algorithm between consecutive frames

The proposed method payed more attention to the normal estimation of the latent space between adjacent frames. The front and back frame images are defined as an image pair $\boldsymbol{I}=\left\{\left(I_t^i, I_{t-1}^i\right)\right\}_{i=1}^{N}$, corresponding to the current frame and the previous frame, respectively, where N is the total number of captured frames, and t denotes the image collection time. Each pair of images $\left(I_t^i, I_{t-1}^i\right)$ corresponds to a normal estimate of the latent space $Y=\left\{\left(y_t\right)\right\}_{i=1}^{N}$, representing the attitude correlation between two consecutive frames. After establishing the spatial graph structure $\mathcal{G}_t^s$, the dynamic graph is unrolled by optimally connecting the edges, and it is converted into a static graph to better characterize the spatial features to construct the spatio-temporal graph $\mathcal{G}^t$ that can describe the nonlinear differences between the front and back of the 4D point cloud at different frames. Specifically, for the spatial graph $\mathcal{G}_t^s$ before and after adjacent frames, ϕ_{sim} is used to calculate the similarity of two nodes (Z_0,Z_1) in $\mathcal{G}_{t0}^s$ and $\mathcal{G}_{t1}^s$. The calculation process is shown in Eq. (3). After establishing the spatial graph structure $\mathcal{G}_{\mathrm{spatial}}$, the temporal domain features can be characterized by adding spatio-temporal connecting edges. In general, the spatio-temporal edge $\mathcal{E}_{\mathrm{temporal}}$ can be constructed through the feature similarity of two nodes.

$$\mathcal{E}_{\mathrm{temporal}}=\phi_{\mathrm{sim}}(Z_0, Z_1). \quad (3)$$

In other words, for each $v_{i,t0}$ in $\mathcal{G}_{t0}^s$, the most similar node v_j is selected from $\mathcal{G}_{t1}^s$. Visually similar and semantically aligned inter-node associations are established through spatio-temporal edges to model the relationship of the same structure across frames, and the semantic similarity is calculated by

$$\boldsymbol{C}_0=\left[\tanh(\theta_1 Z_0)\right]^{\mathrm{T}},\ \boldsymbol{C}_1=\tanh(\theta_2 Z_1),$$
$$\phi_{\mathrm{sim}}=a_{t,t+1}=ReLU\left(\boldsymbol{C}_0\boldsymbol{C}_1-(\boldsymbol{C}_1)^{\mathrm{T}}(\boldsymbol{C}_0)^{\mathrm{T}}\right), \quad (4)$$

where Z_0, Z_1 represent the node features; θ_1, θ_2 are the learnable parameters; and $\boldsymbol{C}_0$, $\boldsymbol{C}_1$ are the calculated features. In the process of updating the spatio-temporal edge $\mathcal{E}_{\mathrm{temporal}}$, two key points are worthy of attention. Firstly, the shape and location have changed due to the randomness of the spatial point cloud obtained by the millimeter-wave radar. If all $\mathcal{E}_{\mathrm{temporal}}$ connections only exist under the current graph structure, it will not be able to effectively characterize the potential features of the point cloud at different times. Therefore, in the process of updating $\mathcal{E}_{\mathrm{temporal}}$, the node connection re-establishment and dissolution were also considered when a certain area of the 4D point cloud of the current and subsequent frames was in a dense state of point clouds. The node features in this area were highly similar, and the spatio-temporal edges were processed by adding the edges.

After completing the optimization of the adjacent frame graph structure, the change in the state of the front and back frames in the adjacent scene is a crucial clue. In

order to utilize this clue, it is necessary to establish a connection in the temporal dimension and project the graphs constructed at the two time points $\mathcal{G}_t^s$ into the same latent space, to extract deeper correlations among the inter-temporal graphs. For the spatio-temporal graph $\mathcal{G}^{st}=(V^{st},E^{st})$, there is

$$V^{st}=\bigcup_{1\leqslant t\leqslant 2}v_t,\ E^{st}=\mathcal{E}\cup\mathcal{E}_{temporal}. \tag{5}$$

The spatio-temporal graph convolution model captures the dynamic dependencies between nodes in the time dimension. For k layers $(1\leqslant k\leqslant K)$ the spatio-temporal convolution process can be expressed as

$$Z_{st}^{(k+1)}=\sigma\left(\tilde{D}^{-\frac{1}{2}}A_{st}\tilde{D}^{-\frac{1}{2}}Z_{st}^{(k)}W_{st}^{(k)}\right), \tag{6}$$

where $Z_{st}^{(k)}$ is the attribute of the node in layer k; $W_{st}^{(k)}$ is the learned parameter matrix, A_{st} is the normalized adjacency matrix, and after each layer of graph convolution, a layer normalization and LeakyReLU are added. After completing the graph convolution, a fully connected layer with a softmax activation function was used to obtain the final single-frame latent space normal estimate. Finally, the attitude adjustment gap was obtained by fusing the attitude between adjacent frames through the fully connected layer.

2.3 Multi-frame 4D point cloud spatial fusion optimization

The ultimate goal of this method was to enhance point cloud data using multi-frame fusion while ensuring confidence. Therefore, after obtaining normal estimations from the latent space of each frame image, multi-frame fusion needed to be performed with point cloud stitching based on the estimation results. The latent space normal estimation of an image only evaluated the normals between the current and previous frames. For multi-frame images, global discretization analysis in the form of an attitude matrix was necessary to ensure the stability of the overall fusion.

Specifically, a sparse adjacency matrix $\boldsymbol{P}$ in this step was first established, which stored the latent space estimation results and key vertices of each frame's point cloud. In each iterative tuning process, by fine-tuning the attitude of a single frame, the discretization between the current frame and the overall point cloud will change, thereby updating the matrix's weights. As shown in Fig.2, a negative number is taken as the initial value in the relative adjacency matrix, the connection relationship between the vertices of different frames is expressed as a positive number, and the number of edges is represented as a positive number in the relative adjacency matrix. When the distance between two nodes is less than the threshold, no edge is established. By gradually iterating the number of adjacent edges connected between nodes, the values in the matrix approach 0.

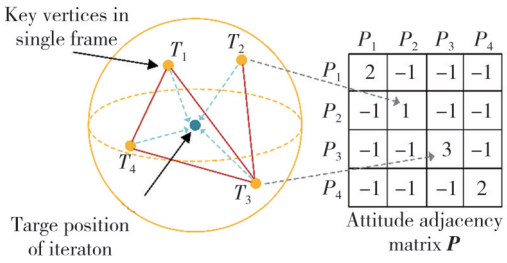


Fig. 2 Establishing a multi-frame 4D point cloud spatial fusion tuning matrix

3 Experiment

3.1 Experimental settings and datasets

Both public and private datasets were used in the study. The public dataset was mainly used for model pre-training. It included the shape-net 3D point cloud dataset[16], which could be transformed into 3D point cloud data from the dense frames collected by a 3D camera, and this dataset enabled the network to form a preliminary understanding of the inter-frame relationships. The private dataset was mainly used for formal training and testing of the model using real-time point clouds collected by a self-built millimeter-wave MIMO radar acquisition system.

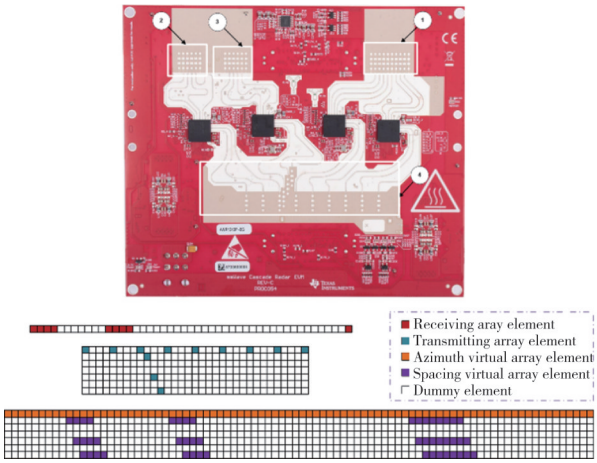


Fig. 3 Four-chip cascade MIMO millimeter wave imaging system and virtual aperture array

As shown in Fig.3, in order to obtain the distance and angular resolution, Texas instruments AWR2243 is used, which employs four millimeter-wave radars in a cascade with the antenna and chip layout, where the upper side antennas are the receiving antennas (1-3) part, forming a uniform sparse array based on the

MIMO mechanism. The lower side antenna was the transmit antenna (4) part, and this layout reduced the effect of coupling. The right side of Fig. 3 depicts the transmit antenna, receive antenna combination, and virtual antenna array offsets. A total of 192 virtual antenna apertures were created by cascading, and a uniform linear array with a maximum of 86 available virtual array elements in the azimuthal direction in the MIMO mode achieved a beamwidth resolution of 2° in azimuth and pitch. The chirp parameters used in the experiment are shown in Table 1.

Table 1 Radar sensor parameter settings

Parameter	Specific value
Starting frequency/GHz	77
FM slope/μs	98
Leisure time/μs	250
ADC start time/μs	10
ADC sampling number	64
Chirp number	16

3.2 Performance evaluation and analysis

To evaluate the method in this paper's normal estimation ability of three-dimensional latent space and the accuracy of spatial attitude correction, a water ripple plane point cloud simulation environment is built, and the model performance is evaluated by identifying the normals of the water ripple plane, as shown in Fig.4.

The black arrow extending out of the point cloud represents the normal plane, and the red arrow is the estimated normal vector. The coordinates in brackets are the calculated vector differences in the three directions (X, Y, Z), exactly the same as the predefined values. It was proved that the method in this paper could complete accurate latent space normal estimation in an ideal environment.

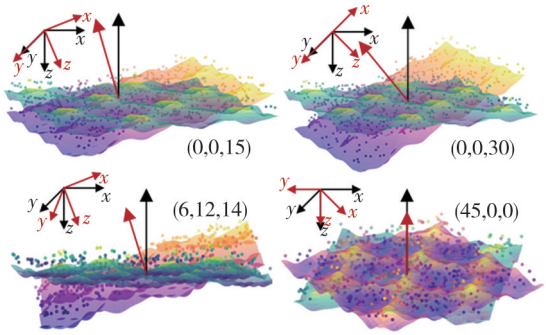


Fig. 4 Latent space normal estimation results in waterline plane point cloud simulation environmen

The construction of point cloud spatial graphs under different scenes was used as an input to the latent space estimation, which directly affected the final evaluation of the model and the results of point cloud fusion. Fig. 5 demonstrates this part of the experiment by performing spatial graph construction for millimeter-wave 4D in various scenarios and adaptively optimizing edges for similar parts. This part used static acquisition to detect millimeter-wave radar point clouds of static scenes. In the fusion part, 12 frames were used to perform densely fuse points cloud fusion of the final scene.

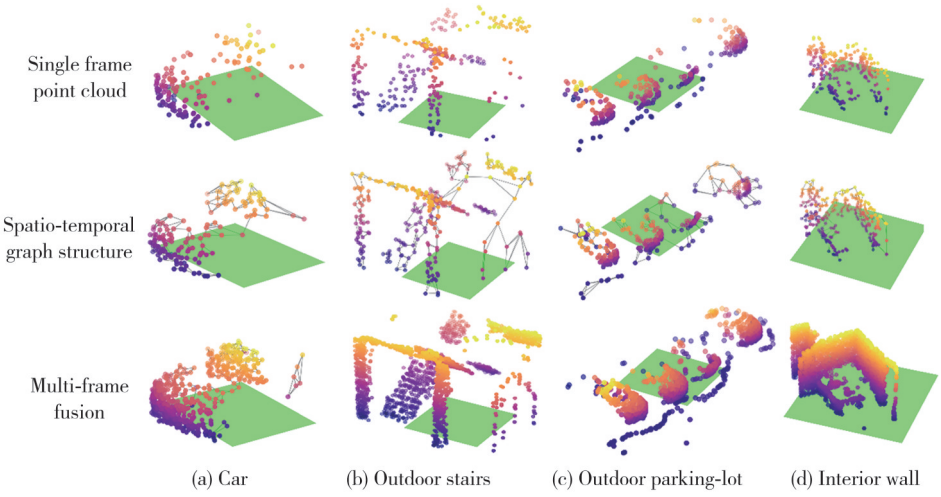
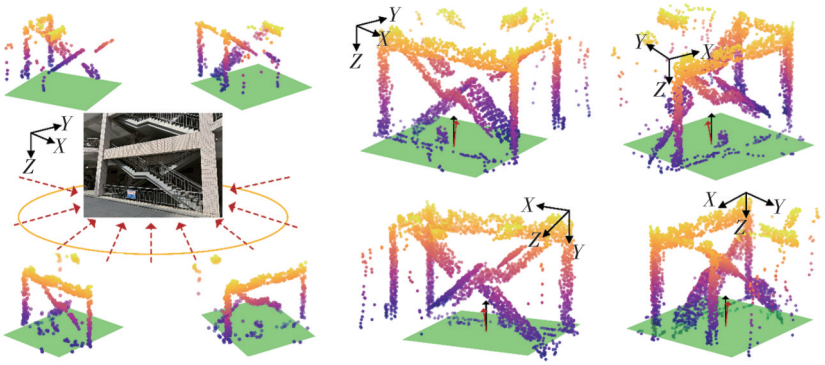


Fig. 5 Spatio-temporal graph construction and multi-frame fusion results for multiple scenario

The proposed method was ultimately designed to achieve sparse point cloud splicing and fusion in motion scenes. Therefore, in this part of the experiment, by splicing 4D point clouds of consecutive frames in a moving scene, the impact of discontinuities caused by lateral displacement and longitudinal vibration of the scene on the performance of latent space normal estimation was evaluated.

As shown in Fig.6, this part of the experiment carries out consecutive frame acquisition of circular trajectory millimeter-wave radar on the complex local space in Fig.6(a), and a total of 12 frames are acquired with 30° intervals by estimating the consecutive latent space normal between each frame and performing point cloud

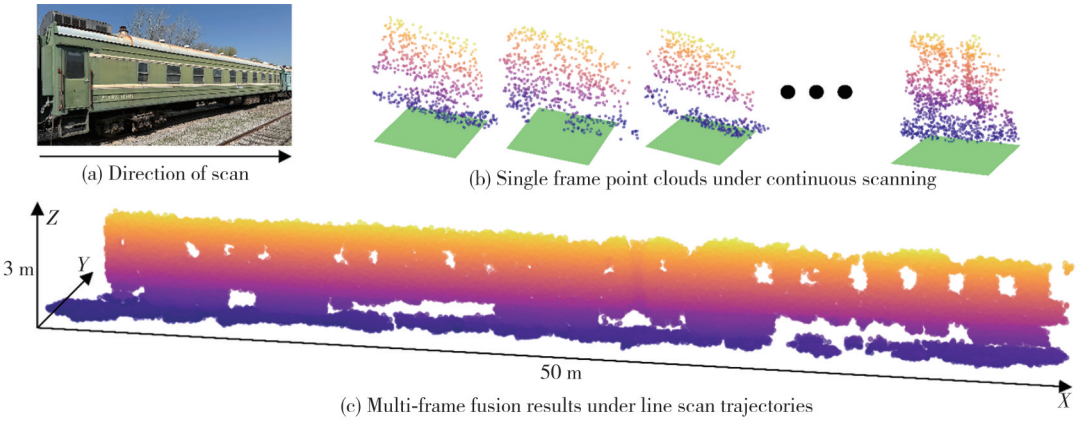
fusion. Fig. 6(b) shows the final point cloud fusion reconstruction results from four angles. The black arrow on the plane coordinate system is the normal vector of the coordinate system, and the red arrow is the comprehensive attitude adjustment after a single frame normal estimation.



(a) Consective frame acquisition of circular trajectory (b) Results after multi-frame fusion of 4D point clouds

Fig. 6 Complex local space circular trajectory scanning multiframe point cloud fusion reconstruction results

For the straight-line scanning task of the automatic driving class, long-distance straight line scanning of train carriages was carried out, and point cloud splicing and fusion was conducted only according to the results of attitude estimation. As shown in Fig. 7, the handheld millimeter-wave radar is used to collect the point cloud along the direction of the train at the speed of 2 m/s, the collection distance is 50 m, the collection number of frames is 100 frames, and the panoramic view of the carriages is generated by the attitude correction in the end. It can be seen that the shape of the carriage did not change its shape due to the vibration generated by the people walking around.

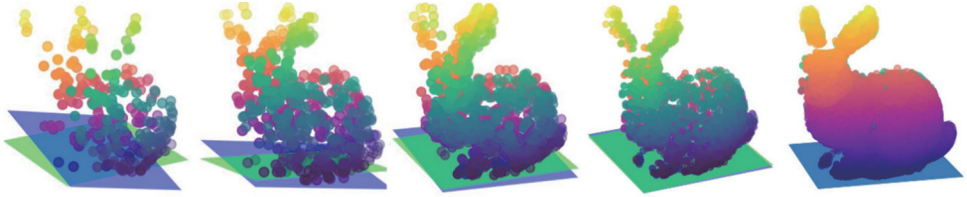


(a) Direction of scan (b) Single frame point clouds under continuous scanning

(c) Multi-frame fusion results under line scan trajectories

Fig. 7 Long-amplitude point cloud fusion reconstruction results under straight line trajectory scanning

Comparative experiments were conducted on the impact of point cloud density on the performance of latent space estimation. As shown in Fig.8, by comparing the potential space estimation results under point clouds of different densities in the same scene (Stanford rabbit), it can be seen that once the point cloud density reaches a specific threshold, further increasing the density does not improve the estimation accuracy of the same scene.



(a) 10–30 (b) 60–100 (c) 100–200 (d) 200– (e) Exact match

Number of point clouds

Fig. 8 Estimated results of point cloud with different densities

When the number of points in the cloud remains constant, the richer the spatial structural features, the higher the estimation accuracy will be. For different scenes, the richer the structural features, the fewer points in the

cloud are required for accurate latent space estimation. Fig.9 further illustrates the relationship between the number of point clouds and registration coincidence degree. When the number of point clouds reached 200, the estimation accuracy basically reached stability.

The generalizability of the public dataset was evaluated using the proposed method. The ShapeNet 3D point cloud dataset[16] contained 16 880 models in 16 classes, of which 13 006 models were used for training, and 3 874 models were used for testing. Table 2 summarizes the quantitative comparison with state-of-the-art methods. The number of input frames indicated how many consecutive input frames could output the normal estimation results, the attitude deviation degree indicated the attitude error between each frame, and the fusion dispersion degree indicated the corresponding point distance of the fused point cloud. Table 2 shows that the model in this paper outperforms all other methods listed in the table for 3D point cloud attitude estimation. The same results as LTSR[7] were obtained for the independent small model class and better than other methods.

In addition, the fusion dispersion of large scene classes was at least 8% lower than other methods. It proved the proposed model could capture latent shape information from 3D points even with limited training samples.

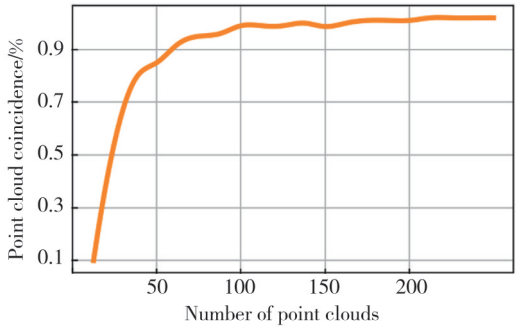


Fig. 9 Relationship between number of point clouds and registration coincidence degree

Table 2 Comparison of decomposition results for each algorithm

Method	Number of input frames	Attitude deviation/(°)	Fusion dispersion
SPLATNet[17]	12	5.6	32
RS-CNN[18]	12	4.2	16
LTSR[7]	12	2.6	9
This paper	12	0.8	3

4 Conclusions

The use of spatio-temporal graph structures was proposed for normal estimation of unstructured 4D point clouds, and ultimately a robust model for surface normal vectors was learned using latent tangent spaces. Additionally, the characterization of latent spatial features was enhanced by employing a feature fusion mechanism incorporating distance weighting and adaptive tuning of connected edges. Furthermore, experiments on point cloud attitude correction and fusion in various scenarios were conducted utilizing a self-built millimeter-wave 4D point cloud acquisition system. The results revealed that the method presented in this paper could estimate highly accurate normals without relying on hand-crafted features or complex networks. Assessments of scanned scenes from public datasets further demonstrated the strong generalization and good portability of the results obtained in this study.

Looking ahead, it aims to explore additional techniques for regularizing the latent representation in normal estimation and other point cloud analysis tasks, with the aim of enhancing the current architecture presented in this paper by integrating advanced convolutional or point cloud networks.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (No.62101512).

Declaration of conflicting interests

The authors have no conflict of interests related to this publication.

References

[1] SUN Y P, HE S F, QU L L. Human identification based on micro-Doppler signal separation and SqueezeNet. Radar Science and Technology, 2023, 21(5): 511-516.

[2] GONG R, WANG L, XU C, et al. A spatial target rotation vector estimation method combining InISAR imaging and micro-Doppler feature extraction. Journal of Electronics and Information, 2021, 43(3): 640-649.

[3] YAAKOB T, ABD RASHID N E, PASYA I. Analysis of pedestrian's Doppler characteristics using MIMO CW radar//2017 IEEE Asia Pacific Microwave Conference (APMC), November 13-16, 2017, Kuala Lumpur, Malaysia. New York: IEEE, 2017: 857-860.

[4] QI F G, LI Z, MA Y Y, et al. Generalization of channel micro-Doppler capacity evaluation for improved finer-grained human activity classification using MIMO UWB radar. IEEE Transactions on Microwave Theory and Techniques, 2021, 69(11): 4748-4761.

[5] KHOMCHUK P, STAINVAS I, BILIK I. Pedestrian motion direction estimation using simulated automotive MIMO radar. IEEE Transactions on Aerospace and Electronic Systems, 2016, 52(3): 1132-1145.

[6] DIAO S, YANG H, XIANG Y, et al. Research on splicing method of point cloud with insufficient features based on spatial reference. Journal of Electronic Imaging, 2021, 30(4): 043008.
[7] CAO J J, ZHU H R, BAI Y P, et al. Latent tangent space representation for normal estimation. IEEE Transactions on Industrial Electronics, 2022, 69(1): 921-929.
[8] JIN F, SENGUPTA A, CAO S Y. mmFall: Fall detection using 4-D mmwave radar and a hybrid variational RNN autoencoder. IEEE Transactions on Automation Science and Engineering, 2022, 19(2): 1245-1257.
[9] XIAO Y Y, CHEN Z G, LIN Z T, et al. Merge-swap optimization framework for supervoxel generation from three-dimensional point clouds. Remote Sensing, 2020, 12(3): 473.
[10] JUNG T W, JEONG C S, KIM I S, et al. Graph convolutional network for 3D object pose estimation in a point cloud. Sensors, 2022, 22(21): 8166.
[11] CUI Y M, LIU X, LIU H M, et al. Geometric attentional dynamic graph convolutional neural networks for point cloud analysis. Neurocomputing, 2021, 432: 300-310.
[12] HONG Y, ZHEN H, CHEN P, et al. 3D-llm: In-jecting the 3d world into large language models. Advances in Neural Information Processing Systems, 2023, 36: 20482-20494.
[13] WANG G M, HU Y Z, WU X R, et al. Residual 3-D scene flow learning with context-aware feature extraction. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 5009609.
[14] ZHANG Z C, NIE J H, YU M J, et al. SharpNet: a deep learning method for normal vector estimation of point cloud with sharp features. Graphical Models, 2022, 124: 101167.
[15] MIKAMO M, FURUKAWA R, OKA S, et al. Active stereo method for 3D endoscopes using deep-layer GCN and graph representation with proximity information//2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), November 1-5, 2021, Mexico. New York: IEEE, 2021: 7551-7555.
[16] YI L, KIM V G, CEYLAN D, et al. A scalable active framework for region annotation in 3D shape collections. ACM Transactions on Graphics, 2016, 35(6): 210.
[17] SU H, JAMPANI V, SUN D Q, et al. SPLATNet: sparse lattice networks for point cloud processing//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE, 2018: 2530-2539.
[18] LIU Y C, FAN B, XIANG S M, et al. Relation-shape convolutional neural network for point cloud analysis//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE, 2019: 8887-8896.

4D毫米波雷达稀疏点云下的连续帧隐空间法线估计

吴泱序[1,2*], 袁新芳[1,2], 陈　平[1,2]

1. 中北大学 信息与通信工程学院, 山西 太原 030051;
2. 中北大学 动态测试技术国家重点实验室, 山西 太原 030051

摘　要: 针对毫米波连续调频雷达在室外运动场景下点云数据稀疏和空间配准精度低的问题, 本文提出了一种轻量级空间配准方法。该方法适用于室外运动场景下毫米波雷达连续多帧之间的点云处理, 能在没有位姿传感器辅助的情况下, 通过时空图神经网络准确估计相邻多帧点云的隐空间法线, 并将每帧雷达点云数据转换到统一的观测坐标系中, 从而实现四维点云的多帧融合与场景配准。实验结果表明, 该方法不仅能准确评估四维点云的空间姿态, 还能有效校正和融合每帧点云的坐标, 在运动及震动过程中实现点云坐标的精准配准。此外, 该算法还能显著提高点云成像的密度, 增强图像的精度和可读性, 同时适用于静态和动态目标的成像, 为毫米波雷达在室外运动场景的应用提供了有力支持。

关键词: 毫米波雷达; 四维点云; 空间配准; 隐空间法线估计; 时空图神经网络

引用格式: WU Yangxu, YUAN Xinfang, CHEN Ping. Consecutive-frame latent space normal estimation under sparse point clouds for 4D millimeter-wave radar. Journal of Measurement Science and Instrumentation, 2024, 15(2): 276-284.