ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-6874-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICASSP49660.2025.10888353

Multi-modal Dynamic Point Cloud Geometric Compression Based on Bidirectional Recurrent Scene Flow*

Fangzhe Nan[1,2] Frederick Li [3] Zhuoyue Wang[4] Gary K.L. Tam [5] Zhaoyi Jiang[2] DongZheng [6] Bailin Yang*[2]

[1]School of Statistics and Mathematics, Zhejiang Gongshang University
[2]School of Computer Science and Technolgy, Zhejiang Gongshang University
[3] University of Durham [4]University of California, Berkeley [5]Swansea University [6] Universal Ubiquitous Technology Co., LTD

***Abstract*—Deep learning methods have recently shown significant promise in compressing the geometric features of point clouds. However, challenges arise when consecutive point clouds contain holes, resulting in incomplete information that complicates motion estimation. To our knowledge, most existing dynamic point cloud compression methods have largely overlooked this critical issue. Moreover, these methods typically employ a multi-scale single-pass approach for motion estimation, performing only one estimation at each scale. This limits accuracy and adversely impacts compression performance. To address these challenges, we propose a dynamic point cloud compression model called M2BR-DPCC (Multi-Modal Multi-Scale Bidirectional Recursion for Dynamic Point Cloud Compression). Our method introduces two key innovations. First, we integrate both point cloud and image data as inputs, leveraging a multi-modal feature representation completion (MFRepC) approach to align information across modalities. This addresses the issue of missing data in point clouds by using complementary information from images. Second, we implement a multi-scale bidirectional recursive (MSBR) motion estimation method. This module iteratively refines motion flows in both forward and backward directions, progressively enhancing point cloud features and improving motion estimation accuracy. Experimental results on widely used datasets, including MVUB and 8iVFB, demonstrate the effectiveness of our approach. Compared to existing methods, M2BR-DPCC achieves superior performance, with an average BD-rate improvement of 95.23% over V-PCC, 12.92% over D-DPCC, and 16.16% over patchDPCC. These results underscore the potential of leveraging multi-modal data and bidirectional refinement for dynamic point cloud compression.**

***Index Terms*—Dynamic point cloud geometric compression, Multi-modal, Multi-scale bidirectional recurrent.**

I. Introduction

Dynamic point cloud geometry compression is a method aimed at efficiently compressing point cloud data that evolves over time. The goal is to achieve low bit-rate compression for continuously moving point cloud sequences while preserving high fidelity. This reduces the bandwidth and storage demands for data transmission and storage, enabling more efficient resource utilization. By compressing dynamic point cloud data effectively, this technique enhances real-time system performance and reliability. It has attracted significant attention from researchers and has found widespread applications in areas such as autonomous driving and robotic navigation.

Traditional dynamic point cloud geometry compression methods, such as those proposed in [1], [2], [3], [4], [5], [6], and [7], can be broadly classified into two categories: 3D-based and 2D projection-based approaches. 3D-based methods perform motion estimation and motion compensation (ME/MC) between frames through block matching techniques, while 2D projection-based methods leverage existing video codecs, focusing primarily on the design of projection algorithms and block arrangements. Both approaches rely on rule-based frameworks with hand-crafted feature extraction and assumption-driven matching rules, which often result in suboptimal coding efficiency. In deep learning-based dynamic point cloud compression [8], [9], a key challenge is embedding motion estimation and compensation into an end-to-end compression network. Some studies [8], [10] explore ME/MC structures for motion estimation and compensation, but these methods are limited to a single-resolution scale, leading to high computational latency. Other works [9], [11] adopt multi-scale motion estimation and compensation networks to improve accuracy, but each layer performs only a single motion estimation, limiting optimization potential. Recently, iterative refinement techniques have emerged, using recurrent schemes to progressively optimize estimated scene flow and enhance motion estimation accuracy. To address the need for both high accuracy and computational efficiency, we propose a multi-scale bidirectional recurrent network in latent space. This network iteratively optimizes the motion scene flow from coarse to fine, enabling effective and efficient motion estimation.

On the other hand, raw point cloud data often suffers from sparsity and incompleteness due to factors like sensor precision, noise, and occlusion, complicating accurate motion estimation [12]. Recovering a complete point cloud from partial data is critical for many real-world applications, yet existing compression models largely overlook this issue. Our key insight is that point cloud data effectively captures the geometric structure of an object's surface, while image textures reveal detailed surface features. By integrating image textures, we can infer the missing geometric details in an incomplete point cloud. This highlights the potential of a multi-modal approach to effectively tackle the challenge of point cloud incompleteness. Although point cloud completion has been extensively studied [13]–[15], most methods rely on unimodal prior information. Some multi-modal approaches, such as Aiello et al. [16], combine image side information with shape priors, and Wang et al. [17] integrate data from cameras and LiDAR for 3D object detection. However, these methods are not designed for compression tasks. Lin et al. [18] proposed a multi-modal compression framework using oc-trees and depth information from paired images of LiDAR point clouds, but it does not address the problem of holes in the point cloud. To the best of our knowledge, no existing methods specifically use images as an auxiliary modality for point cloud compression to resolve incompleteness. Our novel insight is to leverage image texture

Bailin Yang is the corresponding author. This work was supported in part by the National Natural Science Foundation under Grant 62172366, Zhe jiang Province Natural Science Foundation No. LY21F020013, LY22F020013. He is also with Economic Forecasting and Policy Simulation Laboratory, Zhejiang Gongshang University and Collaborative Innovation Center of Statistical Data Engineering, Technology Application, Zhejiang Gongshang University. Gary Tam is supported by the Royal Society grant IEC/NSFC/211159.

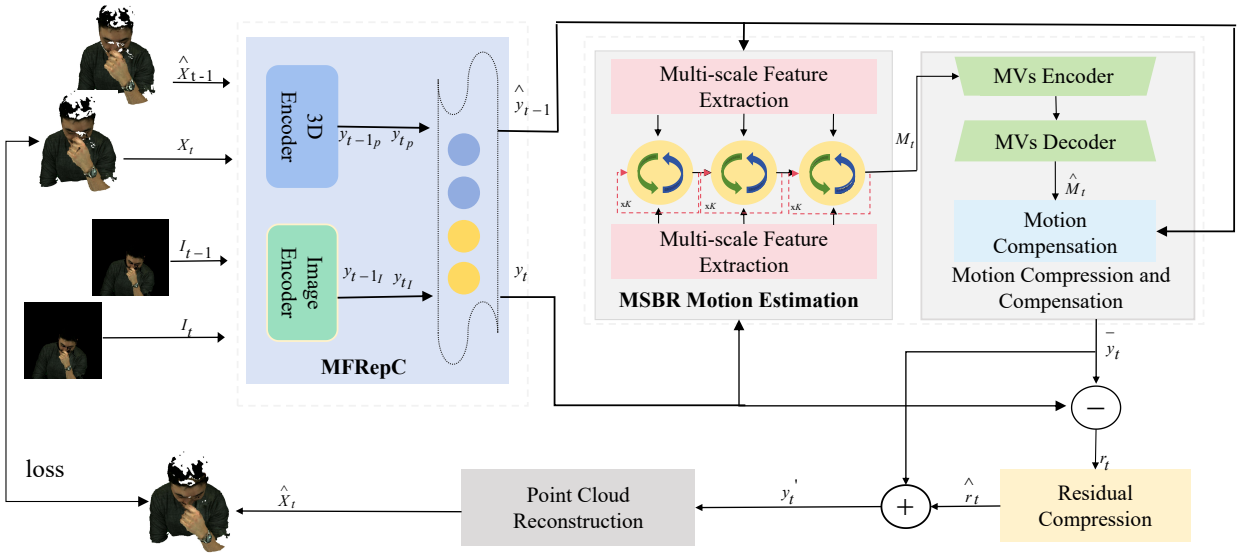

Fig. 1: M2RS-DPCC Network Architecture.

features to fill the geometric holes in point cloud data. We propose the MFRepC model, which integrates information from multiple modalities to improve the completeness and accuracy of compressed point clouds, addressing a critical limitation in current methods.

The main contributions of this work are as follows:

- We propose the M2BR-DPCC model, a dynamic point cloud compression approach that jointly optimizes feature representation completion and motion estimation using multi-modal, multi-scale bidirectional recurrent scene flow.
- We design the Multi-modal Feature Representation Completion (MFRepC) module, which aligns point cloud and image features. MFRepC uses cross-modal information to complete missing geometric details in point clouds, enriching their representation for downstream tasks.
- We introduce a novel Multi-scale Bidirectional Recurrent Scene Flow (MSBR) motion estimation method. MSBR iteratively refines point cloud features across scales, generating optimized motion flows and achieving more accurate inter-frame motion estimation than existing single-pass compression methods.

II. Proposed Method

A. System Overview

The goal of dynamic point cloud compression is to achieve low bit-rate compression of continuous point cloud sequences while maintaining high fidelity. To achieve this, we propose the M2BR-DPCC model, which combines multi-modal representation completion and multi-scale bidirectional recurrent motion estimation, as shown in Figure 1. The input consists of two adjacent point cloud frames, $\hat{X}_{t-1} = \{C_{\hat{X}_{t-1}}, F_{\hat{X}_{t-1}}\}$ and $X_t = \{C_{X_t}, F_{X_t}\}$, along with their corresponding images, I_{t-1} and I_t. Here, $C_{\hat{X}_{t-1}}$ and C_{X_t} represent point cloud coordinates, and $F_{\hat{X}_{t-1}}$ and F_{X_t} are feature matrices indicating voxel occupancy. The architecture first encodes the point cloud and image features. These are fused to produce latent representations $\hat{y}_{t-1}$ and y_t, which are processed through a multi-scale bidirectional cyclic motion estimation module. After motion compensation and compression, the predicted representation $\bar{y}_t$ is obtained, followed by residual compression of r_t. Finally, the point cloud reconstruction module reconstructs $\hat{X}_t$, with remaining steps based on D-DPCC.

B. Multi-modal Representation Completion Learning

To address the limitation in existing models that rely solely on the previous frame's information and fail to effectively fill holes in point clouds, we propose the MFRepC model, which incorporates image information to enhance point cloud data. By leveraging image texture features, this model can infer missing geometric details of the point cloud surface, filling in the gaps in the point cloud data.

The MFRepC model consists of two distinct modules for processing point clouds and images. The point cloud feature extraction module includes two consecutive down-sample blocks designed to capture deeper-level information in the spatial geometry. Each down-sample block contains two sparse convolution layers for point cloud down-sampling and IRN (Interpolation Residual Normalization) blocks for local feature analysis and aggregation. The input point cloud frame X_t is encoded into a sparse tensor $y_{t_p} = \{C^2_{x_t}, F_{y_{t_p}}\}$, where $F_{y_{t_p}}$ represents the feature matrix, and $C^2_{x_t}$ denotes the coordinate matrix with a down-sampling factor of 2. Similarly, the image feature extraction module consists of two down-sample blocks to extract image features. Each block includes two convolutional layers for image down-sampling, followed by ResNet blocks for local feature analysis and aggregation. The input image I_t is encoded into an image feature tensor y_{t_I} after passing through the down-sample blocks.

In the multi-modal representation learning module, our goal is to retain the spatial coordinate matrix of the point cloud while enriching its feature information by integrating image data. This is represented as follows, where F_Q/F indicates feature alignment completion:

$$F_{y_t} = F_Q/F(F_{y_{t_p}}, y_{t_I}). \tag{1}$$

To achieve this, we first adjust the size of the image feature tensor y_{t_I} to match the dimensions of the point cloud feature tensor y_{t_p}. Next, we combine the resized image feature tensor with the point cloud feature tensor to generate a latent spatial feature representation that incorporates both deep visual and geometric information, y_t, i.e., $y_t = (C^2_{x_t}, F_{y_t})$. Specifically, this is described as:

$$F_{y_t} = \begin{cases} F_{y_{t_p}} + resize(y_{t_I}), & I_h > P_h, \\ F_{y_{t_p}} + y_{t_I}, & other. \end{cases} \tag{2}$$

TABLE I: BD-Rate (%) gains compared with existing methods.

M2BR-DPCC						
	VPCC [1]		D-DPCC [8]		jiang's [9]	
Models from MYUB	D1-PSNR	D2-PSNR	D1-PSNR	D2-PSNR	D1-PSNR	D2-PSNR
david	-38.89	-99.9	-15.84	-9.04	-4.76	-11.9
sarah	-32.53	-96.96	-11.71	-6.99	-10.4	-12.9
phil	-64.15	-90.33	-8.61	-5.91	-4.9	-3.3
ricardo	-17.57	-93.73	-15.51	-15.37	-5.8	-8.5
Average	**-38.29**	**-95.23**	**-12.92**	**-9.32**	**-6.5**	**-9.2**
M2BR-DPCC w/o MFRepC						
	GPCC [2]	VPCC [1]	OctAttention [19]	PCGC [20]	D-DPCC [8]	PatchDPCC [21]
Models from 8iVFB	D1-PSNR	D1-PSNR	D1-PSNR	D1-PSNR	D1-PSNR	D1-PSNR
longdress	-82.7	-99.64	-99.96	-36.96	-61.96	-36.96
loot	-51.22	-99.99	-99.63	-56.97	-28.57	-9.64
Average	**-66.96**	**-99.81**	**-99.79**	**-46.96**	**-45.26**	**-23.3**

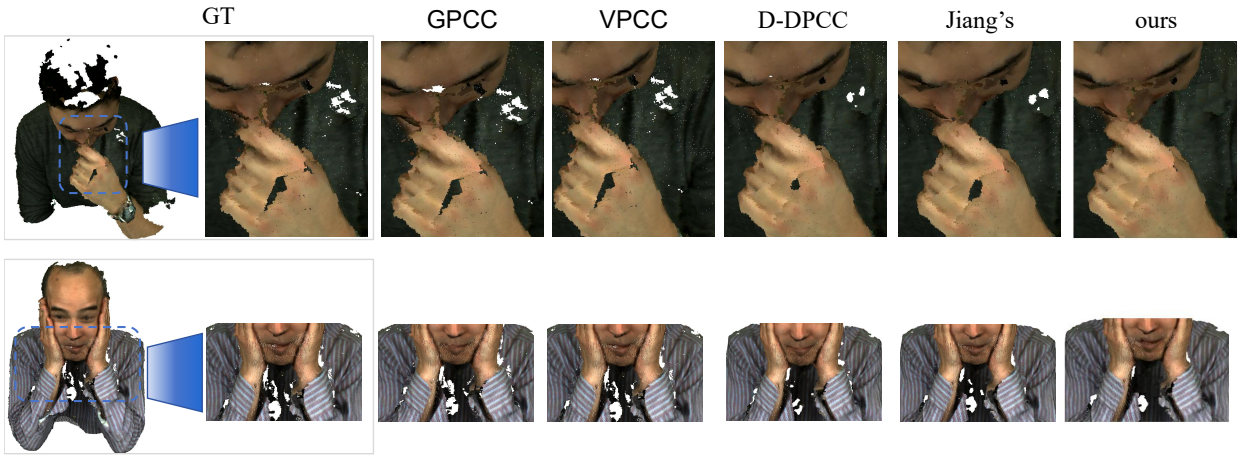

Fig. 2: Visualization results on MVUB (David and Phil) test sequences.

This approach enhances the point cloud features while preserving the original spatial structure, resulting in a richer and more effective feature representation for subsequent deep learning tasks.

C. Multi-scale bidirectional recurrent motion estimation

Existing compression methods often combine motion estimation and compensation (ME/MC) structures with multi-scale networks to achieve accurate motion estimation. However, performing a single estimation at each scale limits their effectiveness. To address this issue, inspired by MSBRN [22], we propose the Multi-scale Bidirectional Recurrent Scene Flow (MSBR) motion estimation module, based on the transformer architecture [23]. The key idea and benefit of our approach is its ability to iteratively refine motion estimation through bidirectional processing across multiple scales, resulting in more accurate and efficient motion flow predictions. This significantly improves the quality of dynamic point cloud compression while reducing computational complexity compared to existing single-pass methods. Unlike MSBRN, which operates in the spatial domain of point clouds and cannot estimate motion vectors (MVs) from latent representations, our framework utilizes latent tensors and sparse convolutions from adjacent frames. After feature fusion, the integrated features y_t and $\hat{y}_{t-1}$ pass through a multi-scale feature extraction module [24]. Bidirectional transformers enhance point features at each scale, optimizing motion prediction. Up-sampling layers propagate features from higher to lower scales, and a Multi-Layer Perceptron (MLP) generates motion vectors M_t.

The Bi-directional transformer structure works as follows: the bidirectional feature aggregator receives the coordinates of points from two frames and their corresponding features, which are obtained from the previous frame through $K-1$ iterations. These frames are denoted as $S^{k-1} = \{(s, f^{k-1})\}$ and $T^{k-1} = \{(t, g^{k-1})\}$. By aggregating features between the two frames, the feature representation of each point in the current iteration is enriched, as described in Equations (3) and (4). The enhanced features are then passed to the hybrid correlation extraction module, which combines in the Euclidean space for local motion and feature-induced grouping for semantic motion, generating candidate correlation features.

$$\boldsymbol{f}^k = \underset{(t_i, g_i^{k-1}) \in N_T^{k-1}\{(s, f^{k-1})\}}{\text{SetConv}\nu} \left([t_i - s, g_i^{k-1}, f^{k-1}] \right). \quad (3)$$

$$g^k = \underset{(s_j, f_j^{k-1}) \in N_S^{k-1}\{(t, g^{k-1})\}}{\text{SetConv}} \left([s_j - t, f_j^{k-1}, g^{k-1}] \right). \quad (4)$$

The SetConv consists of a shared multi-layer perceptron and a max-pooling layer. $N_T(s,)$ denotes the nearest neighboring points of s in T, where i, j represent the indices of points in the neighbor group N, and $'[,]'$ denotes the channel concatenation operator.

D. Loss function

To optimize the model, we use the rate-distortion joint loss function to optimize the loss L:

$$L = R + \lambda D, \tag{5}$$

λ is the lagrange multiplier used to balance distortion and rate. R is the bits per point (bpp) value encoding the current frame. D represents the distortion between X_t and $\hat{X}_t$.

III. Experiments

A. Experiment Settings

Dataset: Common datasets used in dynamic point cloud compression include 8iVFB [25], Owlii [26], and MVUB [27]. For training, we selected four sequences from the Owlii dataset and two sequences from 8iVFB, while the remaining sequences were used for testing.

Training Strategy: Experiments were conducted on an RTX4090. The model was optimized using the Adam optimizer with parameters $\beta = (0.9, 0.999)$ and a learning rate scheduler that reduced the learning rate by a factor of 0.7 every 10 epochs. We used $\lambda = 4, 5, 7, 9$ as training parameters, with a batch size of 1 and an initial learning rate of 0.0008. The model was trained for 100 epochs.

Evaluation Metrics: To assess the effectiveness of our framework, we use bits per point (bpp), which measures the average number of bits per point in each frame. For evaluating the quality of the compressed point cloud sequences, we employ point-to-point error (D1-PSNR↑) and point-to-surface error (D2-PSNR↑).

B. Experimental results

Table I presents the BD-rate values of various methods on the MVUB and 8iVFB datasets, comparing them to our approach. The top half of the table highlights the advantages of multimodal processing, while the bottom half demonstrates that our model still achieves optimal results even without image information. Specifically, compared to D-DPCC, our method shows average BD-rate improvements of 12.92% and 9.32%, respectively. Furthermore, when compared to VPCC and Jiang's method, our method improves BD-rate by 6.47%, 9.15%, and 95.23%, 38.29%, respectively.

Figure 2 provides a visual comparison between our method and other approaches on the MVUB dataset. By incorporating image information, our method effectively fills in the missing parts of the original point cloud, preserving its intrinsic features and addressing small- and alleviating large-hole issues.

C. Ablation Study

Multi-modal Representation Completion Learning: Figure 3 presents a comparison of results on the Sarah sequences. It is evident that, at the same bits per pixel (bpp), our model achieves higher D2-PSNR values than models without MFRepC information. Additionally, models enhanced with MFRepC, such as D-DPCC + MFRepC and Jiang's + MFRepC, also outperform their original counterparts. This demonstrates that our proposed multi-modal feature processing module not only boosts model performance but can also be easily integrated into existing point cloud compression models.

Multiple Scale Bidirectional Recurrent Scene Flow (MSBR) Motion Estimation: Figure 4 visualizes motion vectors (MVs) to showcase the effectiveness of our motion estimation method. The first three columns display the last frame, current frame, and overlap results, with white areas indicating significant motion. The last three columns show the motion vectors from single-scale, multi-scale, and our MSBR method, with red indicating areas of motion change. A closer alignment of the red regions with the white areas indicates better performance. It is clear that the sixth column (MSBR) exhibits the highest similarity between the red and white regions, demonstrating that our motion estimation method is the most accurate.

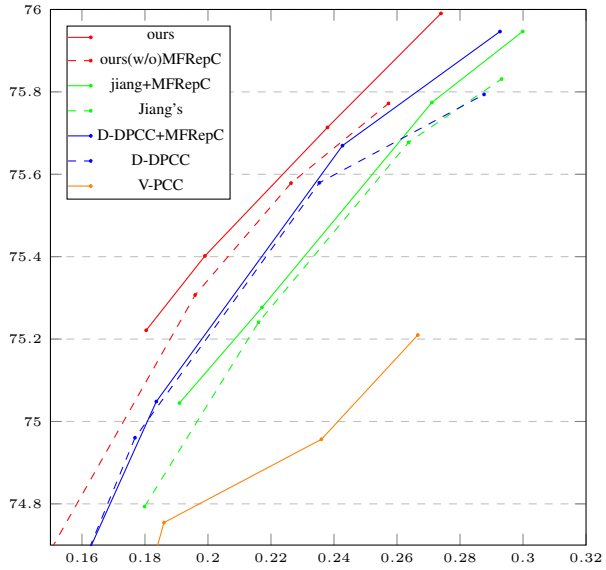

Fig. 3: Performance of different models with and without MFRepC modules on Sarah.

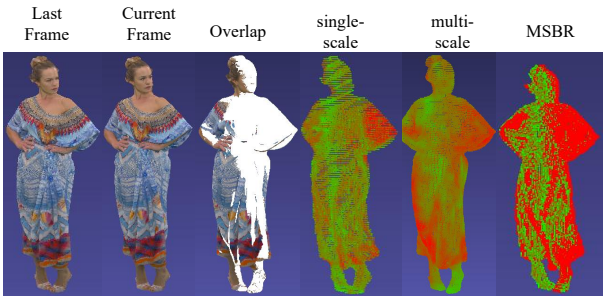

Fig. 4: Visualization of motion vectors (MVs).

IV. Conclusion

This paper addresses the challenges of incomplete point cloud data and inaccurate motion estimation in dynamic point cloud compression. Existing methods struggle with filling missing geometric details and performing precise motion estimation, limiting the quality of compressed point clouds. To overcome these issues, we propose a novel compression model that integrates two key modules: multi-modal representation completion (MFPepC) and multi-scale bidirectional recurrent scene flow (MSBR). The MFPepC module utilizes complementary image and point cloud data to complete missing geometric information, while the MSBR module iteratively refines motion estimation across multiple scales for more accurate compression. Our experiments show that the proposed model outperforms state-of-the-art methods, achieving significant BD-rate improvements and effectively addressing both point cloud incompleteness and motion estimation challenges. Our future research will focus on how to use multi-modal information to deal with the compression task of multi-domain data sets.

REFERENCES

[1] D. Graziosi, O. Nakagami, S. Kuma, A. Zaghetto, T. Suzuki, and A. Tabatabai, "An overview of ongoing point cloud compression standardization activities: Video-based (v-pcc) and geometry-based (g-pcc)," *APSIPA Transactions on Signal and Information Processing*, vol. 9, p. e13, 2020.

[2] S. Schwarz, M. Preda, V. Baroncini, M. Budagavi, P. Cesar, P. A. Chou, R. A. Cohen, M. Krivokuća, S. Lasserre, Z. Li *et al.*, "Emerging mpeg standards for point cloud compression," *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 9, no. 1, pp. 133–148, 2018.

[3] R. Mekuria, K. Blom, and P. Cesar, "Design, implementation, and evaluation of a point cloud codec for tele-immersive video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 4, pp. 828–842, 2016.

[4] D. Thanou, P. A. Chou, and P. Frossard, "Graph-based compression of dynamic 3d point cloud sequences," *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1765–1778, 2016.

[5] W. Zhu, Z. Ma, Y. Xu, L. Li, and Z. Li, "View-dependent dynamic point cloud compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 2, pp. 765–781, 2020.

[6] E. Ramalho, E. Peixoto, and E. Medeiros, "Silhouette 4d with context selection: Lossless geometry compression of dynamic point clouds," *IEEE Signal Processing Letters*, vol. 28, pp. 1660–1664, 2021.

[7] Z. Que, G. Lu, and D. Xu, "Voxelcontext-net: An octree based framework for point cloud compression. 2021 ieee," in *CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 6038–6047.

[8] T. Fan, L. Gao, Y. Xu, Z. Li, and D. Wang, "D-dpcc: Deep dynamic point cloud compression via 3d motion prediction," *arXiv preprint arXiv:2205.01135*, 2022.

[9] Z. Jiang, G. Wang, G. K. Tam, C. Song, F. W. Li, and B. Yang, "An end-to-end dynamic point cloud geometry compression in latent space," *Displays*, vol. 80, p. 102528, 2023.

[10] A. Akhtar, Z. Li, and G. Van der Auwera, "Inter-frame compression for dynamic point cloud geometry coding," *IEEE Transactions on Image Processing*, 2024.

[11] S. Xia, T. Fan, Y. Xu, J.-N. Hwang, and Z. Li, "Learning dynamic point cloud compression via hierarchical inter-frame block matching," in *Proceedings of the 31st ACM International Conference on Multimedia*, 2023, pp. 7993–8003.

[12] H. Zhang, C. Wang, S. Tian, B. Lu, L. Zhang, X. Ning, and X. Bai, "Deep learning-based 3d point cloud classification: A systematic survey and outlook," *Displays*, vol. 79, p. 102456, 2023.

[13] L. Pan, "Ecg: Edge-aware point cloud completion with graph convolution," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 4392–4398, 2020.

[14] X. Yu, Y. Rao, Z. Wang, Z. Liu, J. Lu, and J. Zhou, "Pointr: Diverse point cloud completion with geometry-aware transformers," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 12 498–12 507.

[15] L. Pan, X. Chen, Z. Cai, J. Zhang, H. Zhao, S. Yi, and Z. Liu, "Variational relational point completion network for robust 3d classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 9, pp. 11 340–11 351, 2023.

[16] E. Aiello, D. Valsesia, and E. Magli, "Cross-modal learning for image-guided point cloud shape completion," *Advances in Neural Information Processing Systems*, vol. 35, pp. 37 349–37 362, 2022.

[17] C. Wang, C. Ma, M. Zhu, and X. Yang, "Pointaugmenting: Cross-modal augmentation for 3d object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 11 794–11 803.

[18] Y. Lin, T. Xu, Z. Zhu, Y. Li, Z. Wang, and Y. Wang, "Your camera improves your point cloud compression," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2023, pp. 1–5.

[19] C. Fu, G. Li, R. Song, W. Gao, and S. Liu, "Octattention: Octree-based large-scale contexts model for point cloud compression," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 36, no. 1, 2022, pp. 625–633.

[20] J. Wang, D. Ding, Z. Li, and Z. Ma, "Multiscale point cloud geometry compression," in *2021 Data Compression Conference (DCC)*. IEEE, 2021, pp. 73–82.

[21] Z. Pan, M. Xiao, X. Han, D. Yu, G. Zhang, and Y. Liu, "patchdpcc: A patchwise deep compression framework for dynamic point clouds," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 5, 2024, pp. 4406–4414.

[22] W. Cheng and J. H. Ko, "Multi-scale bidirectional recurrent network with hybrid correlation for point cloud based scene flow estimation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 10 041–10 050.

[23] C. Wang, M. Wu, S.-K. Lam, X. Ning, S. Yu, R. Wang, W. Li, and T. Srikanthan, "Gpsformer: A global perception and local structure fitting-based transformer for point cloud understanding," *arXiv preprint arXiv:2407.13519*, 2024.

[24] W. Wu, Z. Qi, and L. Fuxin, "Pointconv: Deep convolutional networks on 3d point clouds," in *Proceedings of the IEEE/CVF Conference on computer vision and pattern recognition*, 2019, pp. 9621–9630.

[25] E. d'Eon, B. Harrison, T. Myers, and P. A. Chou, "8i voxelized full bodies-a voxelized point cloud dataset," *ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document WG11M40059/WG1M74006*, vol. 7, no. 8, p. 11, 2017.

[26] K. Cao, Y. Xu, Y. Lu, and Z. Wen, "Owlii dynamic human mesh sequence dataset," in *ISO/IEC JTC1/SC29/WG11 m42816, 122th MPEG Meeting*, 2018.

[27] C. Loop, Q. Cai, S. O. Escolano, and P. A. Chou, "Microsoft voxelized upper bodies-a voxelized point cloud dataset," *ISO/IEC JTC1/SC29 Joint WG11/WG1 (MPEG/JPEG) input document m38673 M*, vol. 72012, p. 2016, 2016.