Topic One – 3D CAT for Space Division

Implement the Chordal Axis Transform (CAT) in 3D space and use it to divide the space for disjoint 3D mesh models. Show some examples.

Reference

Prasad L. Rectification of the chordal axis transform and a new criterion for shape decomposition[C]//International Conference on Discrete Geometry for Computer Imagery. Springer, Berlin, Heidelberg, 2005: 263-275.

Ma Y, Chen Z, Hu W, et al. Packing irregular objects in 3D space via hybrid optimization[C]//Computer Graphics Forum. 2018, 37(5): 49-59.

Topic Two - Medial Axis Transform and Its Application for 3D Vision and Shape Analysis

Medial Axis Transform (MAT) is an important concept in computational geometry. It has been widely explored for shape approximation, shape recognition, shape retrieval, shape segmentation, etc. Please implement an 3D MAT algorithm and one related application. If you choose to use others' code as your baseline, improvement by your team is required. Use enough experiments to show the superiority of your idea.

Reference

Li P, Wang B, Sun F, et al. Q-mat: Computing medial axis transform by quadratic error minimization[J]. ACM Transactions on Graphics (TOG), 2015, 35(1): 1-16.

Lin C, Li C, Liu Y, et al. Point2Skeleton: Learning Skeletal Representations from Point Clouds[J]. CVPR, 2021.

Lin C, Liu L, Li C, et al. Seg-mat: 3d shape segmentation using medial axis transform[J]. IEEE Transactions on Visualization and Computer Graphics, 2020.

Yang B, Yao J, Guo X. DMAT: Deformable medial axis transform for animated mesh approximation[C]//Computer Graphics Forum. 2018, 37(7): 301-311.

Topic Three – 3D Object Modeling from Single Image

Single camera is easy to acquire due to the popularity smart phone. 3D object modeling from single image become hotter and hotter because it has wide applications in people's daily life. Please implement an algorithm to model 3D objects or humans from single image. If you choose to use others' code as your baseline, improvement by your team is required. Use enough experiments to show the superiority of your idea.

Reference

Pavlakos G, Choutas V, Ghorbani N, et al. Expressive body capture: 3d hands, face, and body from a single image[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 10975-10985.

Li C, Pan H, Liu Y, et al. Bendsketch: Modeling freeform surfaces through 2d sketching[J]. ACM Transactions on

Graphics (TOG), 2017, 36(4): 1-14.
Li C, Pan H, Liu Y, et al. Robust flow-guided neural prediction for sketch-based freeform surface modeling[J]. ACM Transactions on Graphics (TOG), 2018, 37(6): 1-12.

Topic Four – 3D Object Perception from Point Cloud

Point cloud provides exact depth information of objects in 3D space, which is helpful for 3D object detection or segmentation. Please implement an algorithm of 3D perception for indoor or outdoor scenarios. If you choose to use others' code as your baseline, improvement by your team is required. Use enough experiments to show the superiority of your idea.

Reference

Zhou H, Zhu X, Song X, et al. Cylinder3d: An effective 3d framework for driving-scene lidar semantic segmentation[J]. arXiv preprint arXiv:2008.01550, 2020.
Yin T, Zhou X, Krahenbuhl P. Center-based 3d object detection and tracking[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021: 11784-11793.
Liu, Z., Tang, H., Amini, A., Yang, X., Mao, H., Rus, D., & Han, S. (2022). BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation. 2023 IEEE International Conference on Robotics and Automation (ICRA), 2774-2781.
Hu, H., Wang, F., Su, J., Wang, Y., Hu, L., Fang, W., Xu, J., & Zhang, Z. (2023). EA-LSS: Edge-aware Lift-splat-shot Framework for 3D BEV Object Detection.
Shi, S., Guo, C., Jiang, L., Wang, Z., Shi, J., Wang, X., & Li, H. (2019). PV-RCNN: Point-Voxel Feature Set Abstraction for 3D Object Detection. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 10526-10535.
Schult, J., Engelmann, F., Hermans, A., Litany, O., Tang, S., & Leibe, B. (2022). Mask3D: Mask Transformer for 3D Semantic Instance Segmentation. 2023 IEEE International Conference on Robotics and Automation (ICRA), 8216-8223.
Kolodiazhnyi, M., Vorontsova, A., Konushin, A., & Rukhovich, D.D. (2023). OneFormer3D: One Transformer for Unified Point Cloud Segmentation. ArXiv, abs/2311.14405.

Topic Five – Point Cloud Completion

Point cloud captured by view-dependent sensors are not complete by sparse views. Partial point cloud will affect the algorithm performance on understanding. Thus, point cloud completion is important for downstream tasks, like reconstruction, perception, etc. Please implement an algorithm of point cloud completion. If you choose to use others' code as your baseline, improvement by your team is required. Use enough experiments to show the superiority of your idea.

Reference

Ren, Y., Cong, P., Zhu, X., & Ma, Y. (2022). Self-supervised Point Cloud Completion on Real Traffic Scenes via Scene-concerned Bottom-up Mechanism. ICME 2022.
Wen X, Li T, Han Z, et al. Point cloud completion by skip-attention network with hierarchical

folding[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 1939-1948.
Wang, X., Ang Jr, M. H., & Lee, G. H. (2020). Cascaded refinement network for point cloud completion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 790-799).
Huang, Z., Yu, Y., Xu, J., Ni, F., & Le, X. (2020). Pf-net: Point fractal network for 3d point cloud completion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 7662-7670).

Topic Six – Advanced Scene Rendering with 3D Gaussian Splatting

3D Gaussian Splatting (3DGS) is a cutting-edge technique for rendering complex 3D scenes with high efficiency and detail. The method takes multi-view calibrated images as input and produces high-fidelity visual outputs by modeling scene geometry with learnable 3d Gaussians. This technique has found diverse applications in areas such as motion capture, relighting, autonomous driving, SLAM, and AIGC, just name a few. You can select from 3DGS-related research tasks according to your own research interests. If you choose to use others' code as your baseline, improvement by your team is required. Use enough experiments to show the superiority of your idea.

Reference

Kerbl, Bernhard, et al. "3D Gaussian Splatting for Real-Time Radiance Field Rendering." ACM Trans. Graph. 42.4 (2023): 139-1. https://github.com/MrNeRF/awesome-3D-gaussian-splatting

Topic Seven – 3D Perception from Multimodal Visual Data

Considering that the image contains rich appearance features and the point cloud possesses accurate location and geometry features, many works explore effective fusion ways to make these two sensors complement each other for more precise 3D perception. Please implement a 3D perception algorithm with multimodal visual data. If you choose to use others' code as your baseline, improvement by your team is required. Use enough experiments to show the superiority of your idea.

Reference

Vora, S., Lang, A.H., Helou, B., & Beijbom, O. (2019). PointPainting: Sequential Fusion for 3D Object Detection. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4603-4611.
Wang, C., Ma, C., Zhu, M., & Yang, X. (2021). PointAugmenting: Cross-Modal Augmentation for 3D Object Detection. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 11789-11798.
Zhuang, Z., Li, R., Li, Y., Jia, K., Wang, Q., & Tan, M. (2021). Perception-Aware Multi-Sensor Fusion for 3D LiDAR Semantic Segmentation. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 16260-16270.
Yan, X., Gao, J., Zheng, C., Zheng, C., Zhang, R., Cui, S., & Li, Z. (2022). 2DPASS: 2D Priors Assisted Semantic Segmentation on LiDAR Point Clouds. European Conference on Computer Vision.
Chitta, K., Prakash, A., Jaeger, B., Yu, Z., Renz, K., & Geiger, A. (2022). TransFuser: Imitation With Transformer-Based Sensor Fusion for Autonomous Driving. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45,

12878-12895.
Li, Y., Yu, A.W., Meng, T., Caine, B., Ngiam, J., Peng, D., Shen, J., Wu, B., Lu, Y., Zhou, D., Le, Q.V., Yuille, A.L., & Tan, M. (2022). DeepFusion: Lidar-Camera Deep Fusion for Multi-Modal 3D Object Detection. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 17161-17170.

Topic Eight – 3D Motion Generation

The generation of 3D motion, particularly in the realms of human, hand, or object movements, has been extensively explored. The aim is to produce realistic and plausible 3D motions, whether unconditionally or conditioned on factors such as text, audio, scenes, trajectory, and so on. In this project, our goal is to implement an algorithm capable of generating 3D human/hand/object motion. Either unconditional generation or conditional generation are acceptable. If you choose to use others' code as your baseline, improvement by your team is required. Use enough experiments to show the superiority of your idea.

Reference

Zhu, Wentao, et al. "Human motion generation: A survey." IEEE Transactions on Pattern Analysis and Machine Intelligence (2023).
Tevet, Guy, et al. "Human motion diffusion model." arXiv preprint arXiv:2209.14916 (2022).
Ghosh, Anindita, et al. "IMoS: Intent - Driven Full - Body Motion Synthesis for Human - Object Interactions." Computer Graphics Forum. Vol. 42. No. 2. 2023.
Karunratanakul, Korrawe, et al. "Guided motion diffusion for controllable human motion synthesis." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.
Zhang, He, et al. "Manipnet: neural manipulation synthesis with a hand-object spatial representation." ACM Transactions on Graphics (ToG) 40.4 (2021): 1-14.
Zheng, Juntian, et al. "CAMS: CAnonicalized Manipulation Spaces for Category-Level Functional Hand-Object Manipulation Synthesis." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023.

Top Nine – 3D Motion Prediction

3D motion prediction involves forecasting the future movements or trajectories of human or objects a three-dimensional space. The input for a 3D motion prediction task typically involves data that describes the current state and their historical motion information, or adding additional modal information guidance. This is a common task in various fields, including computer vision, robotics, autonomous systems, and animation. Input of any modality is acceptable, including images, videos, point clouds, etc. If you choose to use others' code as your baseline, improvement by your team is required. Use enough experiments to show the superiority of your idea.

Reference

Zheng Y, Yang Y, Mo K, et al. Gimo: Gaze-informed human motion prediction in context[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 676-694.

Li M, Chen S, Zhang Z, et al. Skeleton-parted graph scattering networks for 3d human motion prediction[C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 18-36.
Xu C, Tan R T, Tan Y, et al. Auxiliary Tasks Benefit 3D Skeleton-based Human Motion Prediction[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 9509-9520.
Xu C, Tan R T, Tan Y, et al. EqMotion: Equivariant Multi-agent Motion Prediction with Invariant Interaction Reasoning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 1410-1420.

Topic Ten – Vision-centric 3D Perception

Vision-centric 3D perception has gained significant traction in recent years due to its ease of deployment, cost-effectiveness, and the preservation of intricate visual attributes. According to the view transformation paradigm, these methods can be categorized into three distinct types. LSS-based approaches explicitly lift multi-view image features into 3D space through depth prediction. Another category of works implicitly derives depth information by querying from 3D to 2D. Notably, projection-free methods have recently demonstrated exceptional performance. Please implement an algorithm for vision-centric 3D perception. If you choose to use others' code as your baseline, improvement by your team is required. Use enough experiments to show the superiority of your idea.

Reference

Liu, Z., Tang, H., Amini, A., Yang, X., Mao, H., Rus, D., & Han, S. (2022). BEVFusion: Multi-Task Multi-Sensor Fusion with Unified Bird's-Eye View Representation. 2023 IEEE International Conference on Robotics and Automation (ICRA), 2774-2781.
Philion, J., & Fidler, S. (2020). Lift, Splat, Shoot: Encoding Images From Arbitrary Camera Rigs by Implicitly Unprojecting to 3D. European Conference on Computer Vision.
Li, Z., Wang, W., Li, H., Xie, E., Sima, C., Lu, T., Yu, Q., & Dai, J. (2022). BEVFormer: Learning Bird's-Eye-View Representation from Multi-Camera Images via Spatiotemporal Transformers. ArXiv, abs/2203.17270.
Jiang, Y., Zhang, L., Miao, Z., Zhu, X., Gao, J., Hu, W., & Jiang, Y. (2022). PolarFormer: Multi-camera 3D Object Detection with Polar Transformers. AAAI Conference on Artificial Intelligence.
Liu, Y., Wang, T., Zhang, X., & Sun, J. (2022). PETR: Position Embedding Transformation for Multi-View 3D Object Detection. ArXiv, abs/2203.05625.
Wang, S., Liu, Y., Wang, T., Li, Y., & Zhang, X. (2023). Exploring Object-Centric Temporal Modeling for Efficient Multi-View 3D Object Detection. ArXiv, abs/2303.11926.