Prsnt: Seminar - UWin | Point Cloud Refinement with 3D Gaussian Splatting

MASc Seminar

Zichen Wang

wang361@uwindsor.ca

University of Windsor

2024-03-18

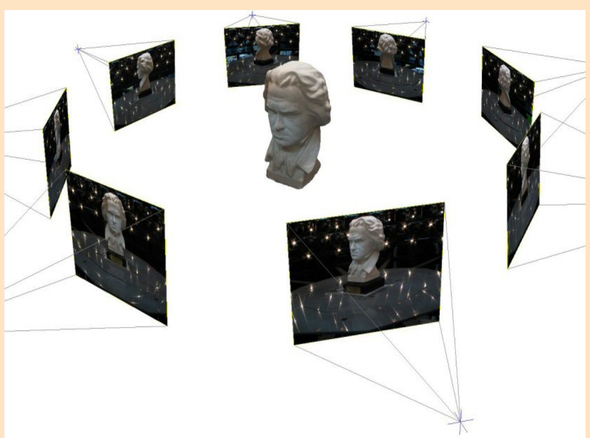
Table of contents

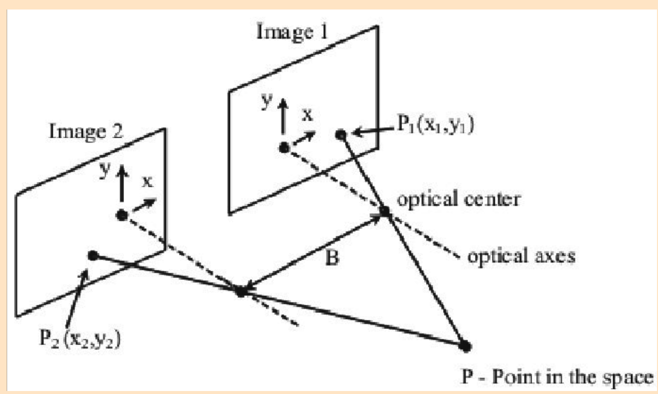
Introduction

- **Point cloud data** will be widely used in the future.
  - Simple and powerful surface representation without topology limitation. [1]
  - Having depths ready make things easy: 3D Apps. [2]
  - Easy to obtain: 3D scanner or photogrammetry. [3]
- **Raw pcd**: noisy, redundant, inaccurate
  - Refinement: accurate geometry, complete
- **Idea**: Consolidating k-NN to representative PC, and then leveraging 3DGS to optimize the point cloud.

Related Works

1. Multi-View Stereo (MVS) methods output **depth map** for each input image.
2. Each depth map produces a **point cloud** individually.
    - Before combination: Apply **constriants** to filter multiple point clouds and mask false values on depth maps.
    - fusible [4] used **reprojection** and **normal** directions to identify reliable dpeth values.
    - The **filtered** averaged depth map will produce a unified point cloud.
3. Evaluate the PC’s **Accuracy** and **Completeness** to verify a MVS method: MVSNet, CasMVSNet, TransMVSNet, …





MVSNet_GT_depth.jpg

4. Consolidate point cloud with KNN

- EC-Net [5] used k-NN to conduct edge-aware convolutions.
- PR-GCN [6] applied k-NN to build a local graph to perform graph convolutions.
- Similarly, use k-NN to determine the 3D positions jointly.

5. 3DGS for optimizting

- 3DGS [7] , a SOTA for novel view synthesis: A efficient point cloud rasterization technique.
- Real-time rendering and fast optimization for points' attributes: colors, opacities, positions, etc.

Components Overview

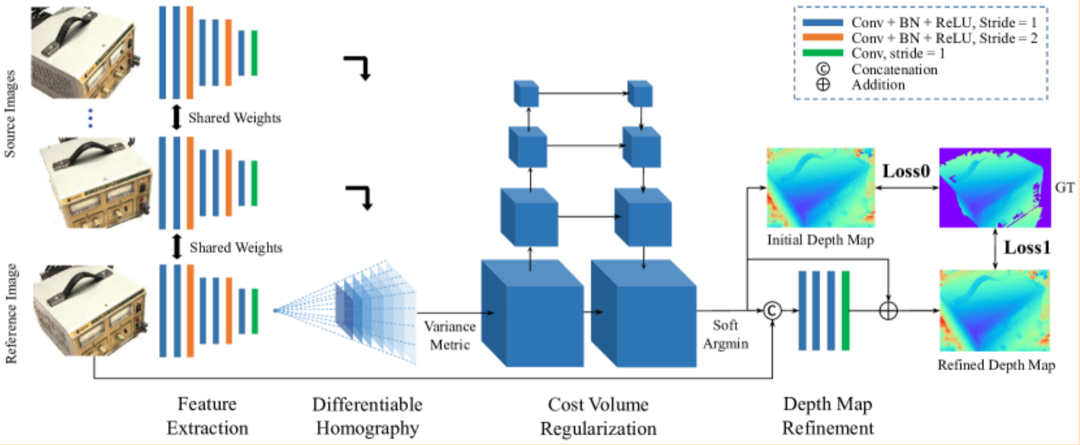
1. Pretrained **CasMVSNet** ➡ Multi-view depth maps
2. **fusible**(MVSNet-pytorch) ➡ Averaged depth map; Unified PC
3. **KNN** (CUDA) ➡ Neighbors; Consolidation
4. **3DGS** ➡ Optimization

PC_01_pipeline.jpg

1. Depth Map Acquisition

- Multi-stereo: multiple **stereo image pairs** based on homography, mapping pixels to another view.
- MVSNet: Depth estimation as a classification problem: 192 predefined hypothetical depth values.
- Cost volume of a view is built on 192 hypothetical depth values. All cost volume are merged to a variance volume.

MVSNet_homographies.jpg



- 3D UNet: 32-channel image feature ➡ 1-channel probability.
- Probability volume: the distribution over 192 depths for each pixel.
- Predicted depth: expectation of 192 hypothetical depths
- A pre-trained NN model generates depth maps:

  CasMVSNet_depthMap_fuse.jpg

2. Restore Point Cloud from Depth Maps

- Every 2-view pair performs once the point cloud filtering:
  1. Photometric constriant: Depth confidence probability > 0.8
  2. Geometric constraint: Reprojection

MVSNet_depthMap_fusion.jpg

- Pseudocode:

Algorithm 1 Depth map fusion

```
for view=1 to 49 do
    Read reference image
    for src=1 to 10 do
        Warp src depth map to be seen
        from the ref view
        Reproject the warped dMap_src
        onto dMap_ref
        if Pixel diff < 1 pix and depth error
        < 0.01 then
            mask_geo = 1
        end if
    end for
    Average reprojected ref depth maps
    and images
    Unproject the mean ref depth map to
    point cloud in world space
end for
Output 49 point clouds
```

CasMVSNet_pl evaluation on scan9:

scan9_CasMVSNet-pl_2024-02-23.jpg

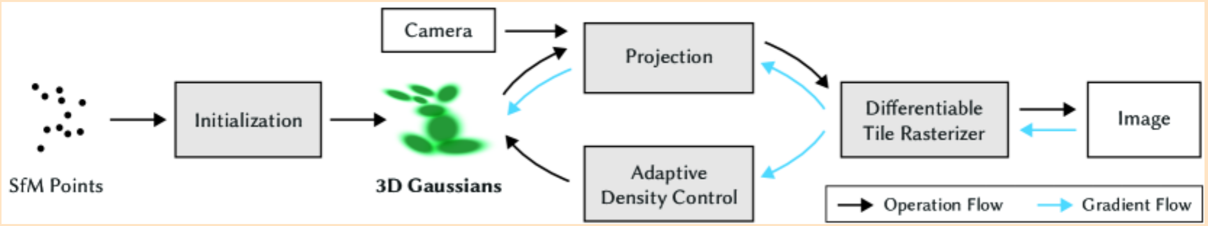
Point cloud quality:

Acc.= 0.36399536,

Comp. = 0.36997940

3. 3D Gaussian Splatting

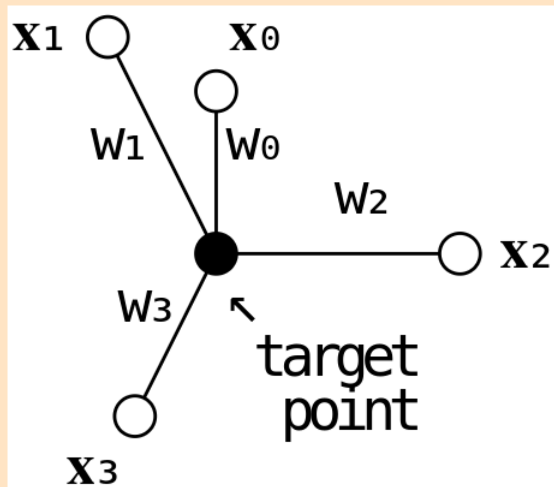
- Scene Representation:
  - Each point is a 3D Gaussian distribution.
  - Adaptive density control: grow and prune point cloud
  - Tile-based Rasterization: parallel programming
- Optimization: the MSE between rendered and input images.



Splat_contributions.jpg

4. Refinement with 3DGS

- Fuse each point with its 3 nearest neighbors through learnable weights
- X,Y,Z direction: expectation of 4 points:



- 3DGS renders the fused point cloud to images, and the fused point cloud gets optimized based on the image loss.

Method Features

Advantages are attributed to 3DGS

1. Fast optimization due to 3DGS.
2. The fused point cloud can be the initialization of 3DGS for various downstream tasks:

   dynamic scene, mesh reconstruction, compression.
3. Investigate the efficacy of consolidating the point cloud for accuracy and completeness.

Current Progress

- [x] 3DGS code understanding
- [x] CasMVSNet for depth maps
- [x] Depth map fusion based on CasMVSNet_pl
- [x] simple-KNN (CUDA) for fetching neighbor points
- [ ] Adapt data format to input into 3DGS for optimizing
- [ ] Evaluate the accuracy and completeness of the optimized point cloud

Footnotes

1. Guennebaud, G., Barthe, L., & Paulin, M. (2004, June). Real-Time Point Cloud Refinement. In PBG (pp. 41-48).

2.

Nvidia Point cloud library: https://developer.download.nvidia.com/GTC/PDF/GTC2012/PresentationPDF/S0088-GTC2012-Point-Cloud-CUDA.pdf

3. Wikipedia: https://en.wikipedia.org/wiki/Point_cloud#cite_note-3DVAE-3

4.

Galliani, S. et al. “Massively Parallel Multiview Stereopsis by Surface Normal Diffusion.” 2015 IEEE International Conference on Computer Vision (ICCV) (2015): 873-881.

5.

Yu, L., Li, X., Fu, C. W., Cohen-Or, D., & Heng, P. A. (2018). Ec-net: an edge-aware point set consolidation network. In Proceedings of the European conference on computer vision (ECCV) (pp. 386-402).

6.

Zhou, G., Wang, H., Chen, J., & Huang, D. (2021). PR-GCN: A Deep Graph Convolutional Network with Point Refinement for 6D Pose Estimation. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 2773-2782.

7.

Kerbl, B., Kopanas, G., Leimkühler, T., & Drettakis, G. (2023). 3d gaussian splatting for real-time radiance field rendering. ACM Transactions on Graphics, 42(4), 1-14.