Research Seminar
Eötvös Loránd University, Budapest,Hungary, 13 May 2025

Exploring Opportunities with 3D Point Clouds

Weisi Lin
Email: wslin@ntu.edu.sg
College of Computing and Data Science
Nanyang Technological University, Singapore

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Two separate worlds for long

- Computer vision: 2D visual data (from physical world)→ 3D understanding
- Computer graphics: 3D models (from virtual world)→ 2D rendering/visualization

3D point clouds (PCs): a bridge between physical and virtual worlds

To facilitate

- mixed reality
- multimedia interaction
- digital transformation
- smart cities

3D Digital Transformation

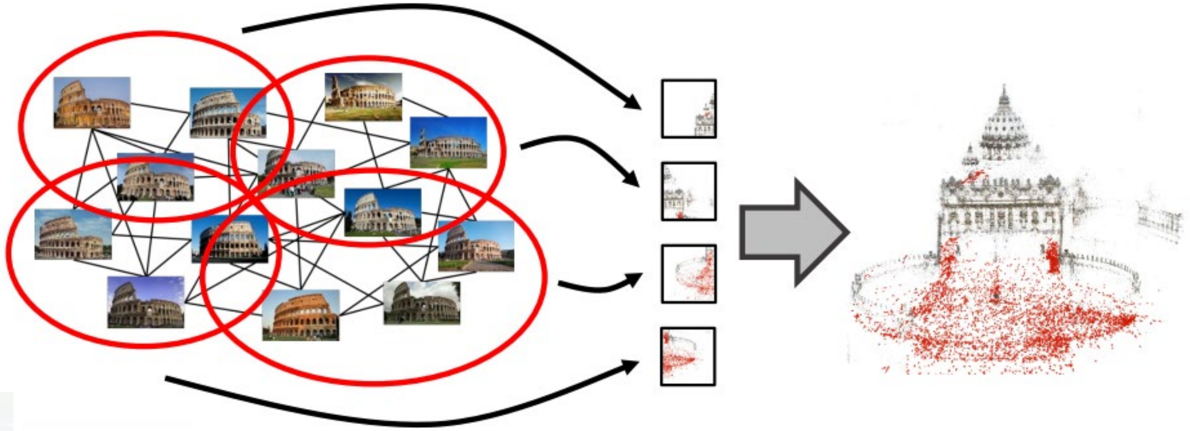
- **Digital Twin** for everything

Point Cloud (PC) Acquisition

1) RGBD sensors: Kinect, Laser scanner, Lidar, …

2) Structure from Motion (SfM) Photogrammetry

Photographs → Matching points → A point cloud



Lin & Lee, "Visual Saliency and Quality Evaluation for 3D Point Clouds and Meshes: An Overview", APSIPA Trans. Signal and Info Processing, 11(1), e28, 2022.

- 3D PCs captured

Laser Scanning Equipment
by LiDAR

Handheld Device

Different scales of 3D data acquisition

- Scanned Raw PCs

Missing Parts

Noisy Parts

…bring difficulties for processing and utilities

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

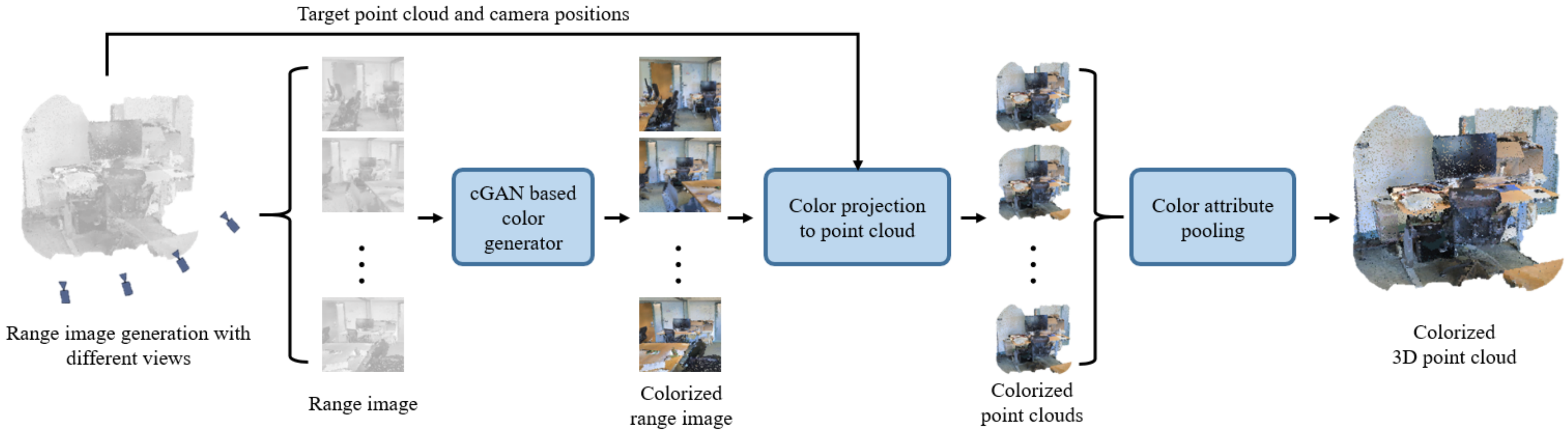
3D reconstruction from a single image

- *K. Chan, et al, "R-Cyclic Diffuser: Reductive and Cyclic Latent Diffusion for 3D Clothed Human Digitalization", CVPR 2024.*
- *K. Chan, et al, "Fine Structure-Aware Sampling: A New Sampling Training Scheme for Pixel-Aligned Implicit Models in Single-View Human Reconstruction", AAAI 2024.*
- K. Chan, *et al*, "S-PIFu: Integrating Parametric Human Models with PIFu for Single-view Clothed Human Reconstruction", NeurIPS 2022.
- K. Chan, *et al*, "IntegratedPIFu: Integrated Pixel Aligned Implicit Function for Single-view Human Reconstruction", ECCV, 2022.

AI (deep learning) to fill missing Point Cloud Data

Coloring of 3D Point Clouds

1. Range image generation from a point cloud
2. Color attribute generation using conditional GAN (cGAN)
3. Color attribute projection and merging



Hou, et al, "Range Image Based Point Cloud Colorization Using Conditional Generative Model", IEEE ICIP, 2019

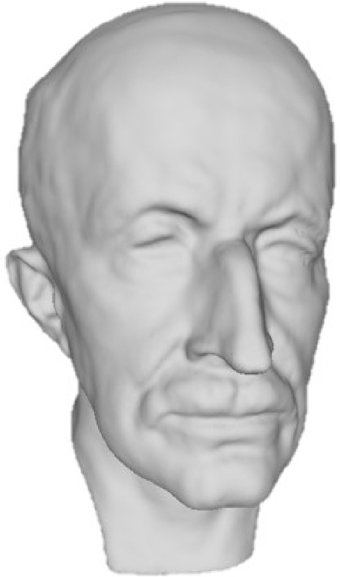
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

3D PCs:

Point clouds data for Bremen city

Large (city) scale

Lucy Model



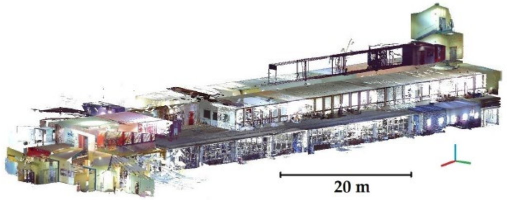
Max Planck Model

Single object
(standalone or to form bigger PCs—like BIM)

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

- A point cloud: a set of points used to represent a 3D object/scene
  - Geometry info: coordinates *(x, y, z)*
  - Possible attribute info: color, normal, SIFT, etc.
- Examples of application scenarios


VR/AR/XR



Building information modelling

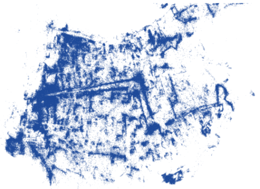
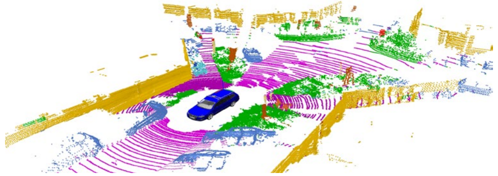


Image-based localization


Autonomous driving

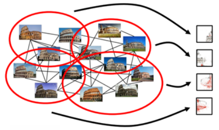
Also: robots, gaming/entertainment, social media, industrial metaverse, BIM, urban surveillance/planning, digital art, cultural heritage preservation, future training/education, crime investigation, discovery in medical/biological/material sciences

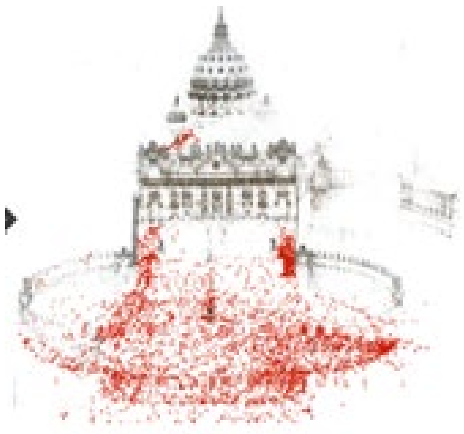
- Challenges & opportunities
  - large-scale, high-resolution point clouds: millions of “nonstructured” points
  - dynamic point clouds at a high frame rate: tens of GB
  - urgent calls for technical innovation in storage, transmission, processing, etc.

Three types of targeted ultimate users for 3D PCs:

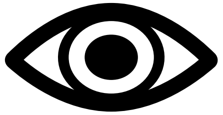
RGBD sensors

Photogrammetry

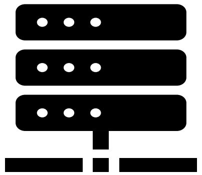




Type I: Humans (HVS)



Type II: Machine Intelligence (MI)



Type III: HVS + MI

The rest of this talk: to present related recent research attempts… to address the major challenges

- PC representation
  - Compression
  - Mesh reconstruction
- PC saliency determination
- Image-based localization, for practical applications
  - PC simplification
  - PC quality assessment
- Moving forward

The rest of this talk: to present related recent research attempts... to address the major challenges

- PC representation
  - ❖Compression
  - ❖Mesh reconstruction
- PC saliency determination
- Image-based localization, for practical applications
  - ❖PC simplification
  - ❖PC quality assessment
- Moving forward

MPEG PC compression (PCC) standardization

- 3 PC categories
  - Category 1: static objects & scenes
  - Category 2: dynamic objects
  - Category 3: dynamically-acquired LiDAR sequences --"one-sided" 3D data

Complete 3D data

- 2 existing coding schemes
  - **TMC13** (geometry-based PCC; **G-PCC**)
    - Proposed for Categories 1 and 3 (but may be used for all PCs)
    - geometry info:
      - use an **octree** structure & 1 bit to tell the occupancy of each node
      - decode first before decoding attribute information
    - 2 attribute/color coding options: Region Adaptive Hierarchical Transform (RAHT); Level of Details (LoD) based
  - **TMC2** (video-based PCC; **V-PCC**)
    - Proposed for Category 2 (but may be used for all PCs)
    - **Project** 3D points with different viewpoints → compress by **existing video codec**
    - 3 major modules: **patch generation**; **patch projection & packing**; **image padding**
    - good for **uniformly-distributed PCs**; **not suitable for large-scale PC compression**

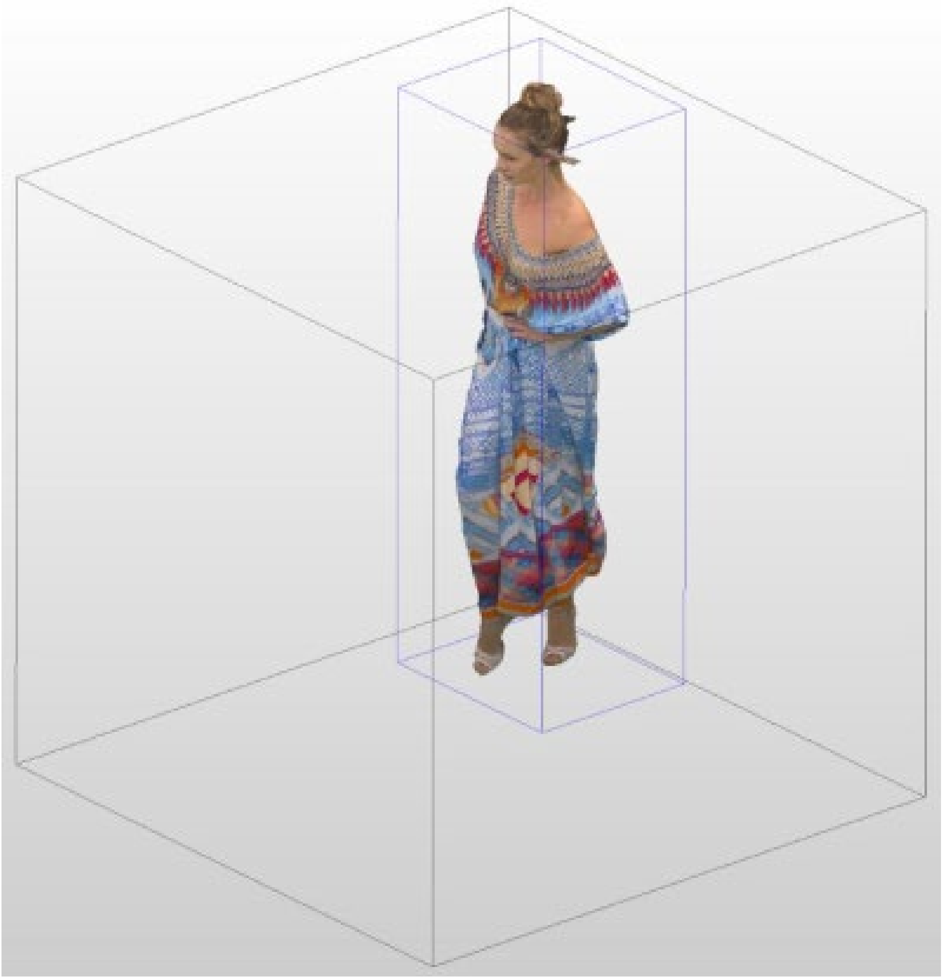
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Lin & Lee, "Visual Saliency and Quality Evaluation for 3D Point Clouds and Meshes: An Overview", APSIPA Trans. Signal and Info Processing, 11(1), e28, 2022.

Dynamic objects :

(Category-2)

- not to compress 3D PCs directly
- project 3D data onto 2D frames
  - ✓ cubic (or cylinder) projection
- using 2D Motion Compensation (MC) & Motion Estimation (ME)
- making full use of existing 2D image/video coding infrastructure



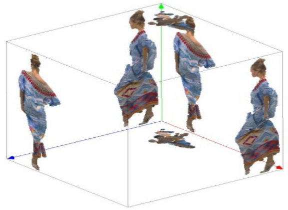
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Patch projection-based V-PCC

A ***patch***: a set of 2D pixels projected from 3D points adjacent to each other and having similar normal vectors.

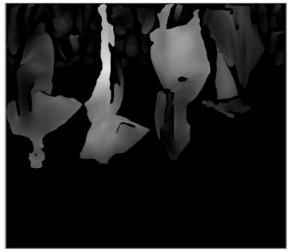
-- Patch generation

a PC → patches



-- Patch Cubic-Projection & Packing

patches → 2D frames

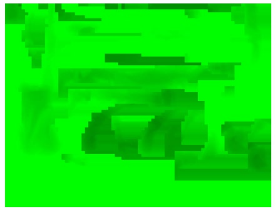


Geometry

Attribute

-- Image Padding

Fill empty space between patches



Geometry frame

Attribute frame

To make generated frames more suitable for existing video coding infrastructure

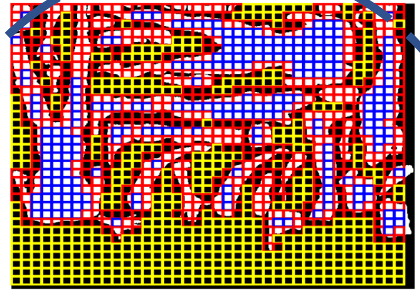
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

projection-based V-PCC: high computational complexity

- Patches: contain a large number of empty pixels→ generated videos are with **high dimensions**
- A PC: decomposed into **3 videos** (joining all projections)

(a) Attribute Frame

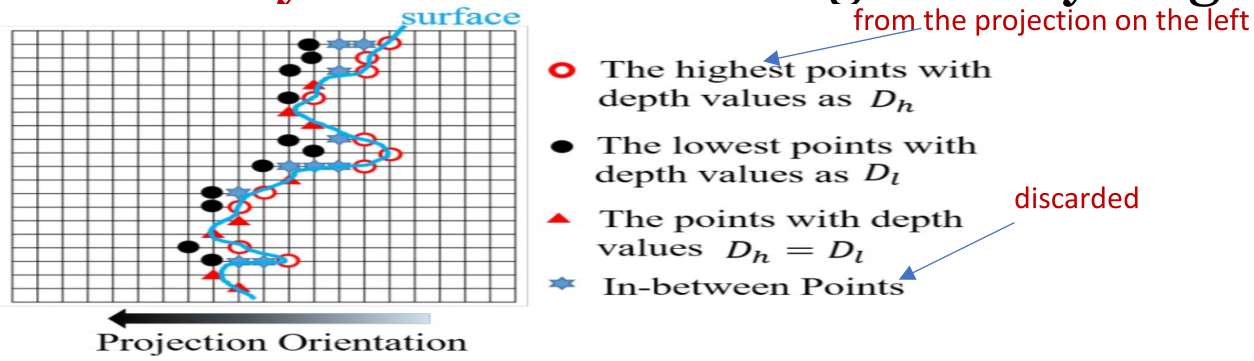
(b) Geometry Frame



(c) Occupancy Map

blue, red & yellow boxes: **occupied, boundary & unoccupied** blocks (size 32x32)

- Provision of far & near layers for attribute & geometry → generated videos: higher **frame**



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

rate-distortion optimization (RDO):

$$m^* = \arg\min_{m_i \in \mathbf{M}} \{J(m_i) = D(m_i) + \lambda R(m_i)\},$$

$\mathbf{M} = \{\mathbf{M}_{Inter}, \mathbf{M}_{Intra}\}$, as all prediction modes, including

$\mathbf{M}_{Intra} = \{\ Intra_{2N\times 2N},\ Intra_{N\times N}\ \}.$ and $\mathbf{M}_{Inter} = \{\ Inter_{skip/merge},\ Inter_{2N\times 2N},$
$\mathbf{M}_{asym},\ \mathbf{M}_{sym}\ \}.$

asymmetric modes

symmetric modes

the modes can be also divided, according to partition size:

$\mathbf{M}_{2N\times 2N} = \{\ Inter_{skip/merge},$
$Inter_{2N\times 2N}\ , Intra_{2N\times 2N}\ \}.$
the modes with the partition size $2N\times 2N$

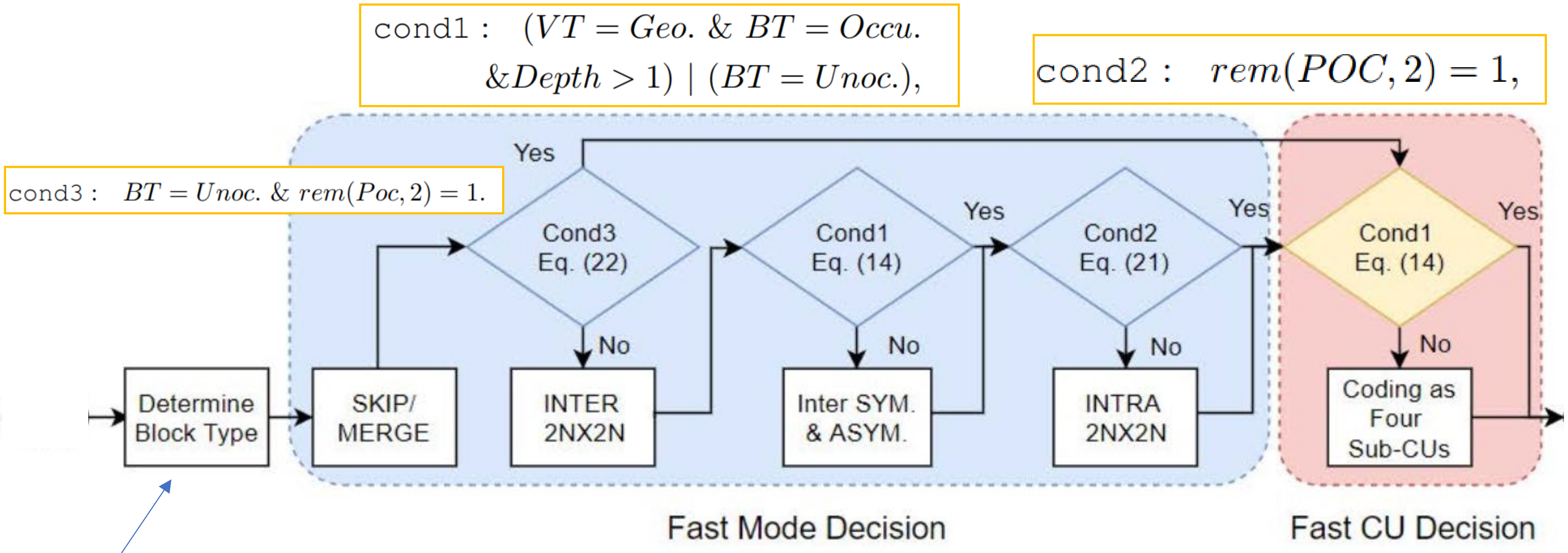
and $\mathbf{M}_{non-2N\times 2N} = \{Intra_{N\times N}, \mathbf{M}_{asym},\ \mathbf{M}_{sym}\ \}.$

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Occupancy-map guided fast coding mode decision

Both theoretical analysis & experimental demonstration:

3 simple & effective occupancy map based rules (also using far/near info)



To iterate if sub-CUs are selected

$$BT = \begin{cases} Unoc., & \sum_{o_i \in \mathbf{O}} = 0; \\ Occu., & \sum_{o_i \in \mathbf{O}} = 4N^2; \\ Boun., & otherwise, \end{cases}$$

to avoid substantial RD calculation for CU and PU decisions

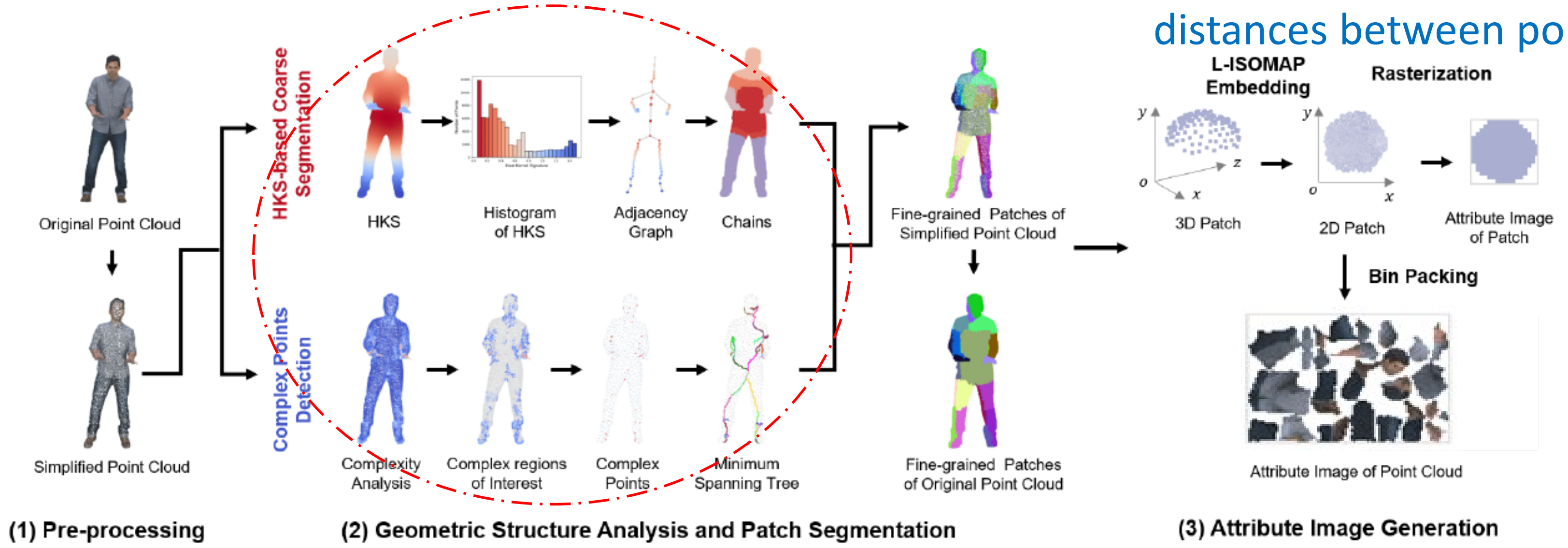
NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Xiong,et al, “Occupancy Map Guided Fast Video based Dynamic Point Cloud Coding”, IEEE T-CSVT,2021

Can we avoid overhead caused by many projections of a PC?

segmentation into right patches
(with simplified PC for efficiency)

3-D to 2-D (nonlinear dimensionality reduction) while preserving geodesic distances between points



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Zhao, et al, "Fine-Grained Patch Segmentation and Rasterization for 3D Point Cloud Attribute Compression", IEEE Trans. on Circuits and Systems for Video Technology, 31(12): 4590-4602, 2021.

More work in PC coding:

Xiong, et al, "Efficient Geometry Surface Coding in V-PCC", *IEEE Transactions on Multimedia*, 2023.

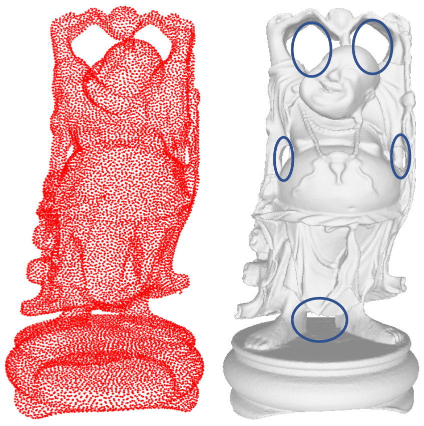
Xiong, et al, "Occupancy Map Guided Fast Video based Dynamic Point Cloud Coding", IEEE Transactions on Circuits and Systems for Video Technology, 32(2): 813-825, 2022.

- **To reconstruct 3D meshes from a 3D point cloud**

Limitations of 3D Point Clouds:

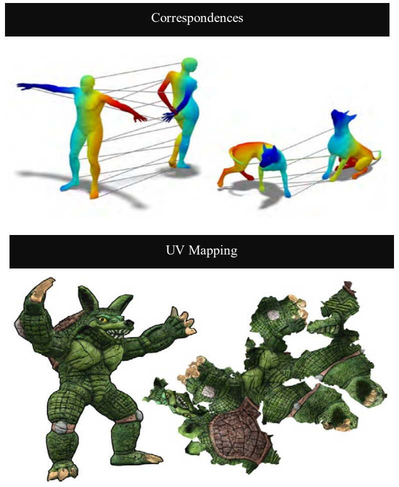
- no info for ***topological structure***
- not support ***shape analysis*** algorithms
- not support high-precision ***rendering*** tasks

Topological Structure

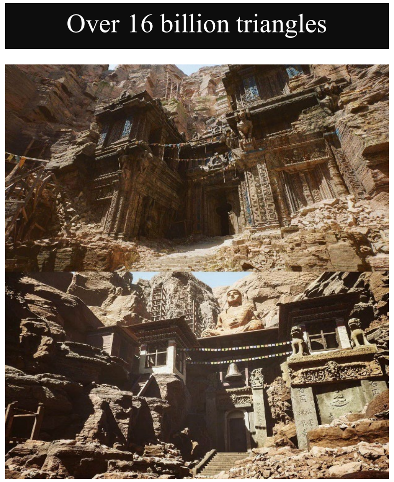


× √

Shape analysis



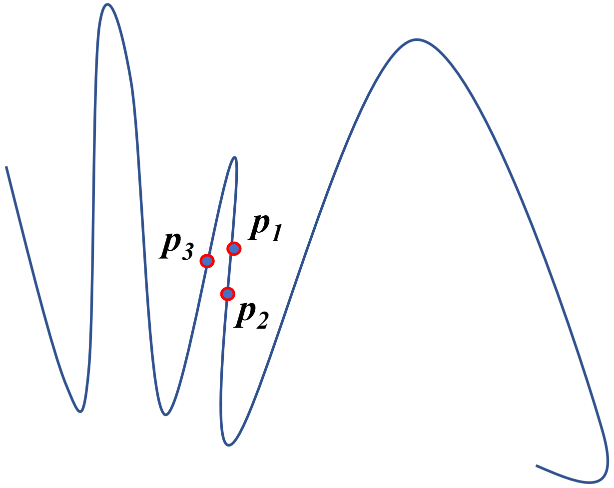
Rendering



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

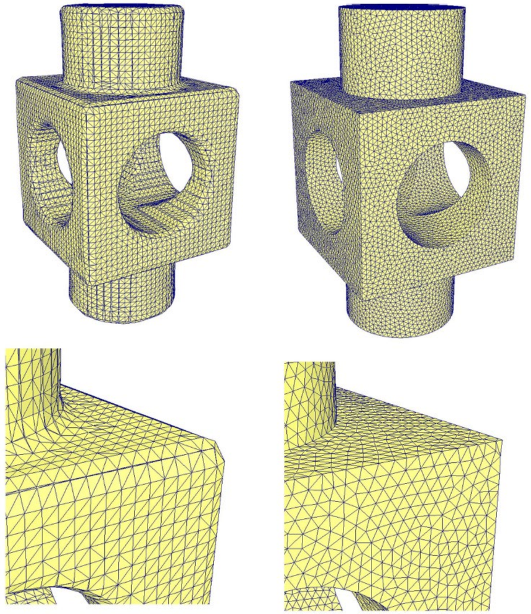
3 key requirements for mesh reconstruction

Structure search



searching the real neighbor point of p_1

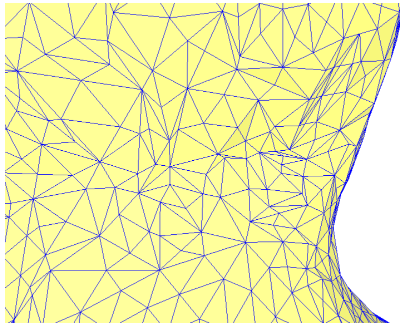
Geometric Feature keeping



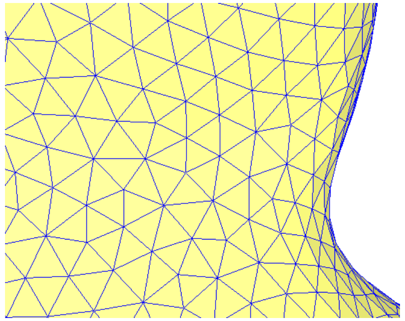
×

√

Mesh quality [PASCAL99]



× **Anisotropic**

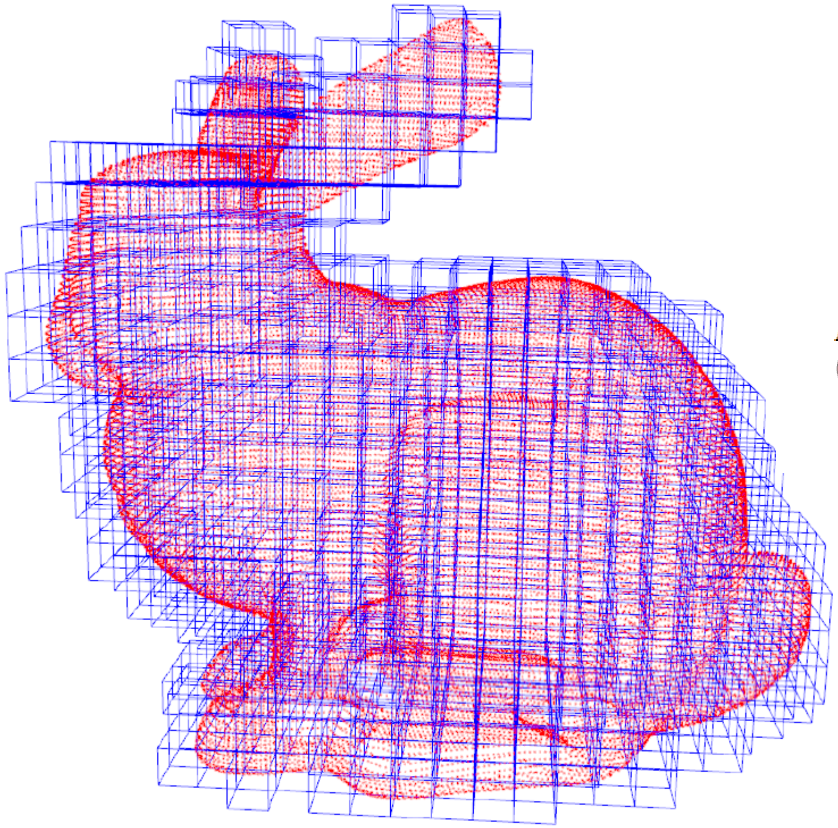


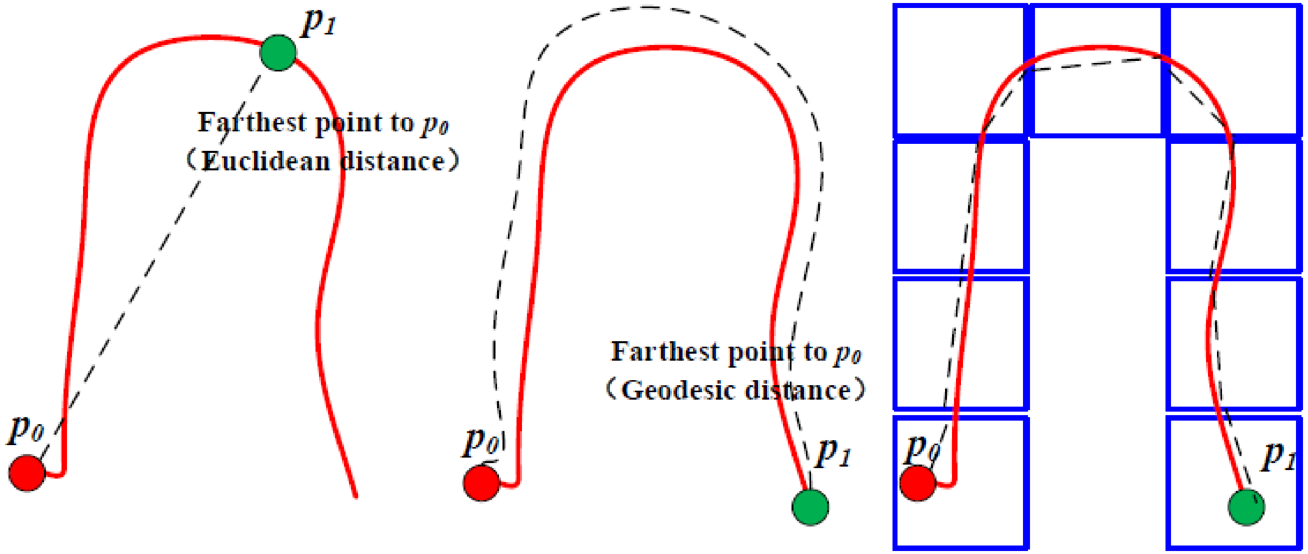
√ **Isotropic**—with equilateral triangles (as close as possible)

Lv, et al, “Voxel Structure-based Mesh Reconstruction from a 3D Point Cloud”, *IEEE Transactions on Multimedia*, 24 (2021): 1815-1829.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Voxel Structure



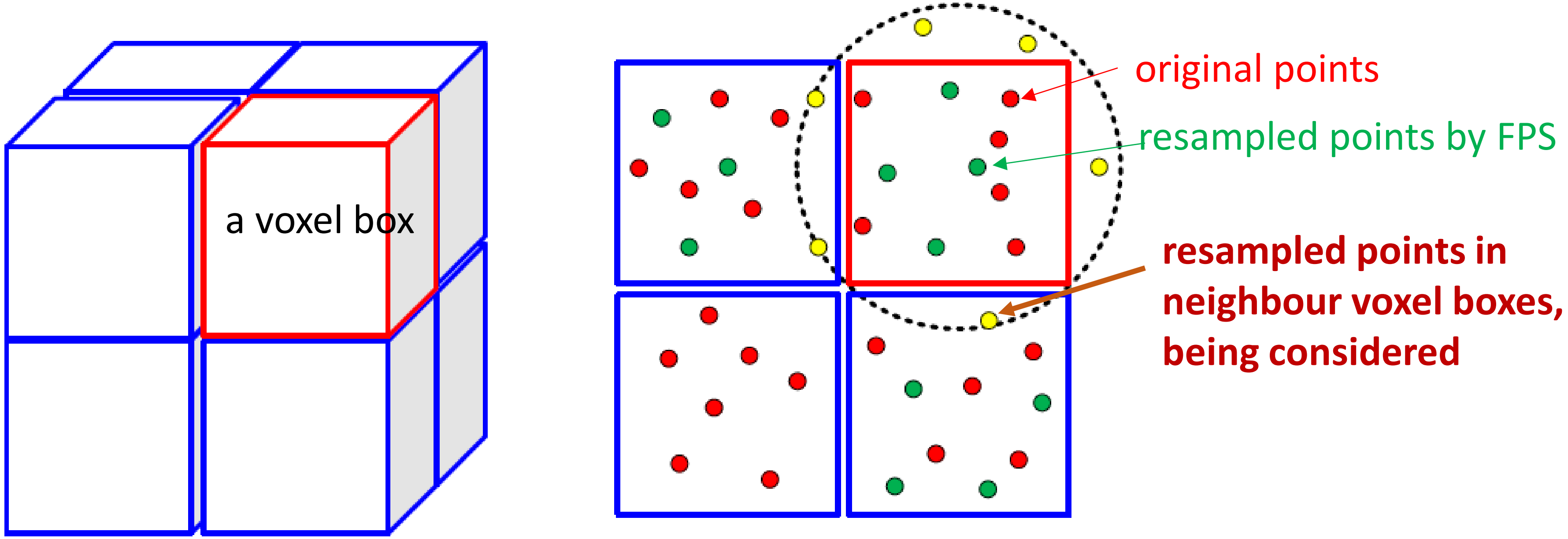


voxel structure: **intrinsic control** (voxel-based **geodesic** distance) between different points.

Based on the voxel structure, more accurate neighbor structure can be achieved from a PC.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Resampling – preparing data for mesh reconstruction ***in parallel (for efficiency)***



Based on ***voxel structure***, points are ***resampled*** by FPS (Farthest Point Sampling).

Lv, et al, "Voxel Structure-based Mesh Reconstruction from a 3D Point Cloud", *IEEE Transactions on Multimedia*, 24 (2021): 1815-1829.

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

More details of related research:

Lv, et al, “Intrinsic and Isotropic Resampling for 3D Point Clouds”, IEEE Trans on Pattern Analysis and Machine Intelligence, 45(3): 3274-3291, 2023.

Lv, et al, "KSS-ICP: Point Cloud Registration based on Kendall Shape Space", IEEE Transactions on Image Processing, 32: 1681-1693, 2023.

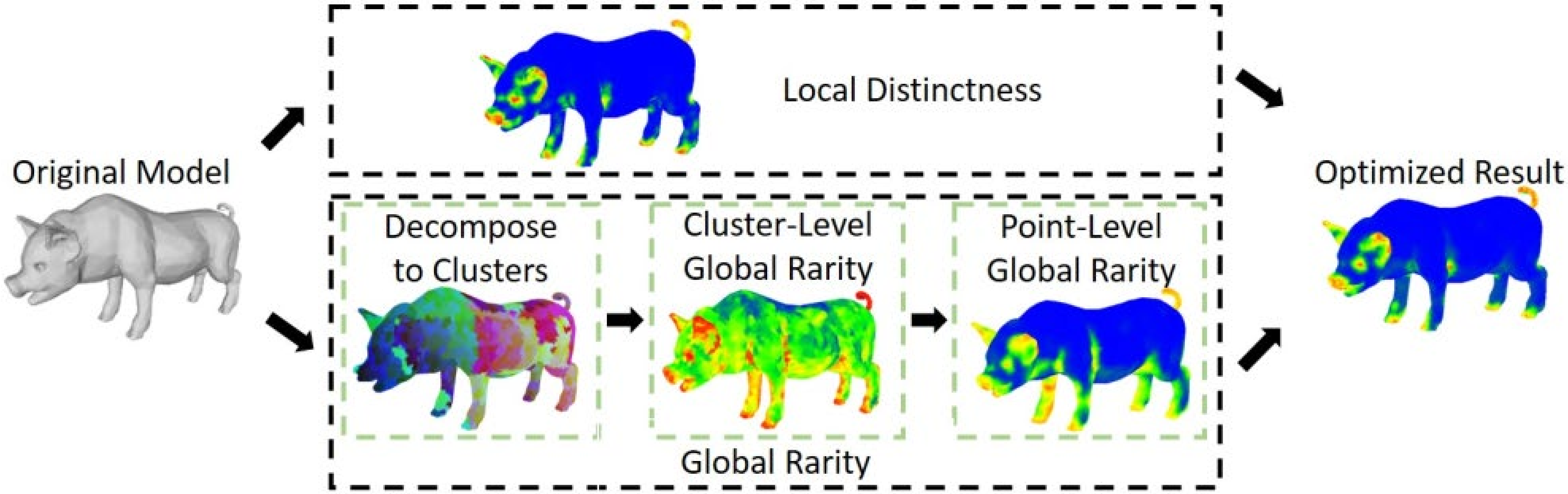
Lv, et al, "Approximate Intrinsic Voxel Structure for Point Cloud Simplification", IEEE Transactions on Image Processing, 30(9): 7241 – 7255, 2021

The rest of this talk: to present related recent research attempts… to address the major challenges

- PC representation
  - Compression
  - Mesh reconstruction
- **PC saliency determination**
- Image-based localization, for practical applications
  - PC simplification
  - PC quality assessment
- Moving forward

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

local distinctness & **global rarity**: both important; like in cases of 2D images



Main contributions:

- Consider local geometric features of each point for cluster-level global rarity refinement
- Propose an adaptive optimization framework.

X. Ding, et al., “Point Cloud Saliency Detection by Local and Global Feature Fusion”, IEEE Transactions on Image Processing, 2019

Local Distinctness

Characterize local geometric features

- **Zernike coefficients** (performance not stable across different models)
- **SHOT descriptor** (10x in size compared with FPFH; slow processing speed)
- **Fast Point Feature Histograms (FPFH) descriptor**

Measure the difference/dissimilarity between two points (p_i, p_j)

$$\chi^2(p_i, p_j) = \sum_{n=1}^{N} \frac{(FPFH_n(p_i) - FPFH_n(p_j))^2}{FPFH_n(p_i) + FPFH_n(p_j)}$$

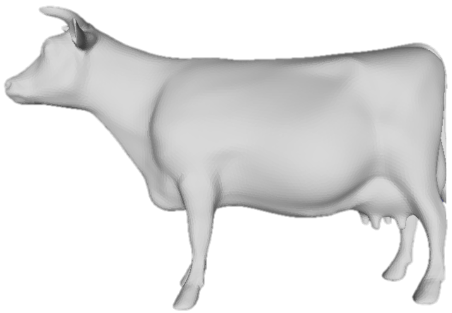
local distinctness (within a neighborhood R)

$$D(p_i) = 1 - \exp\left(-\frac{1}{R}\sum_{j=1}^{R} \frac{\chi^2(p_i, p_j)}{1 + ||p_i - p_j||}\right)$$

Global Rarity

PC segmentation

- Voxel cloud connectivity segmentation [24]


Original model


segmented model

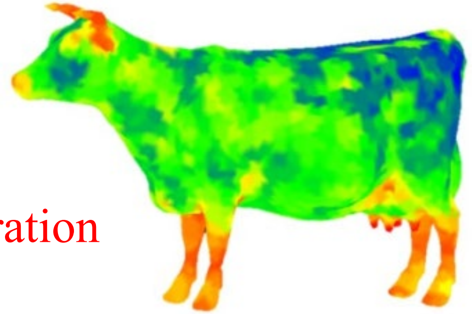
Cluster-level global rarity calculation (similar to local distinctness calculation):

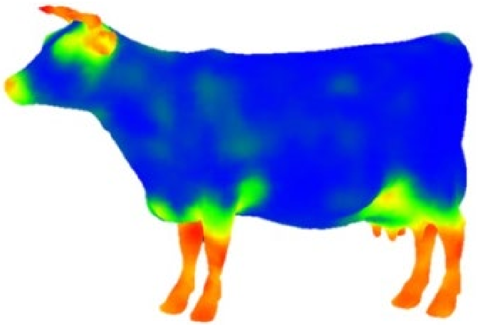
$$G(c_i) = 1 - \exp\left(-\frac{1}{N}\sum_{j=1}^{N}\frac{\chi^2(c_i, c_j)}{1 + ||c_i - c_j||}\right)$$

Point-level global rarity refinement (details in the next slide)

Why refinement is needed?

- Take detailed information of the model into consideration


Cluster-level result


Point-level result

mouth & horn preserved

body liminated

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Random Walk Ranking: to distribute cluster-level global rarity G to point-level

Initialization

1. **Seed points**

   Salient: points nearest to a salient-cluster (i.e., G of a cluster > th1) center,

   Non-salient: points nearest to non-salient-cluster (i.e., G of a cluster < th2) center

2. **Unseeded points:**

   all points for clusters with th2≤G≤th1;

   All other points in Case 1 above

$$th_1 = mean(G) + \frac{1}{2}\Delta$$

$$th_2 = mean(G)$$

$$\Delta = \max(G) - mean(G)$$

a random walker:

- starting at each of the unseeded points
- calculate the probability of first reaching one of the seed points

Optimization Framework

Integrate local & global **saliency cues**

- **Linear combination** (not ideal)
- **Adaptive integration to minimize**

saliency value to be decided:
s=1 if salient; s=0 if not

No. of points in the PC

weight matrix to smooth the result

$$W_{ij} = \exp(-\frac{||p_i - p_j||^2}{2\sigma_2^2})$$

Min: $$L = \sum_{k=1}^{2}\sum_{i=1}^{N}(\frac{s_i^2}{Z_k(p_i)} + Z_k(p_i)(1-s_i)^2) + \sum_{j\in R} W_{ij}(s_i - s_j)^2$$

Neighbours of point p_i

k=1, Z_1 : local distinctnesss;
k=2, Z_2 : point-level global rarity refinement

The framework enables

- Higher local distinctness & higher global rarity to obtain higher S_i
- Lower local distinctness & lower global rarity to obtain lower S_i

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

The rest of this talk: to present related recent research attempts… to address the major challenges

- PC representation
  - ❖Compression
  - ❖Mesh reconstruction
- PC saliency determination
- Image-based localization, for practical applications
  - ❖PC simplification
  - ❖PC quality assessment
- Moving forward

Structure-from-Motion (SfM) Point Clouds (PCs)

Image Collections

PC

Exemplary Views

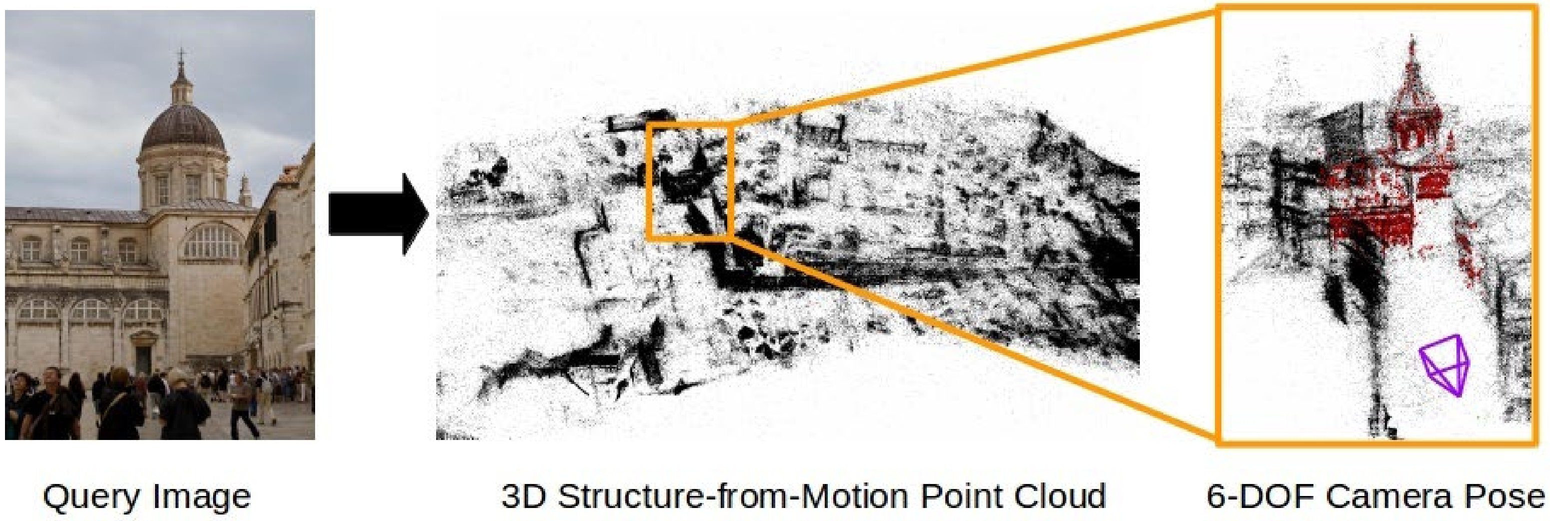
Position: X,Y,Z; ***Color***: R,G,B
Visibility:
Image 1 -> feature descriptor
Image 2 -> feature descriptor
……

Information a 3D Point

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

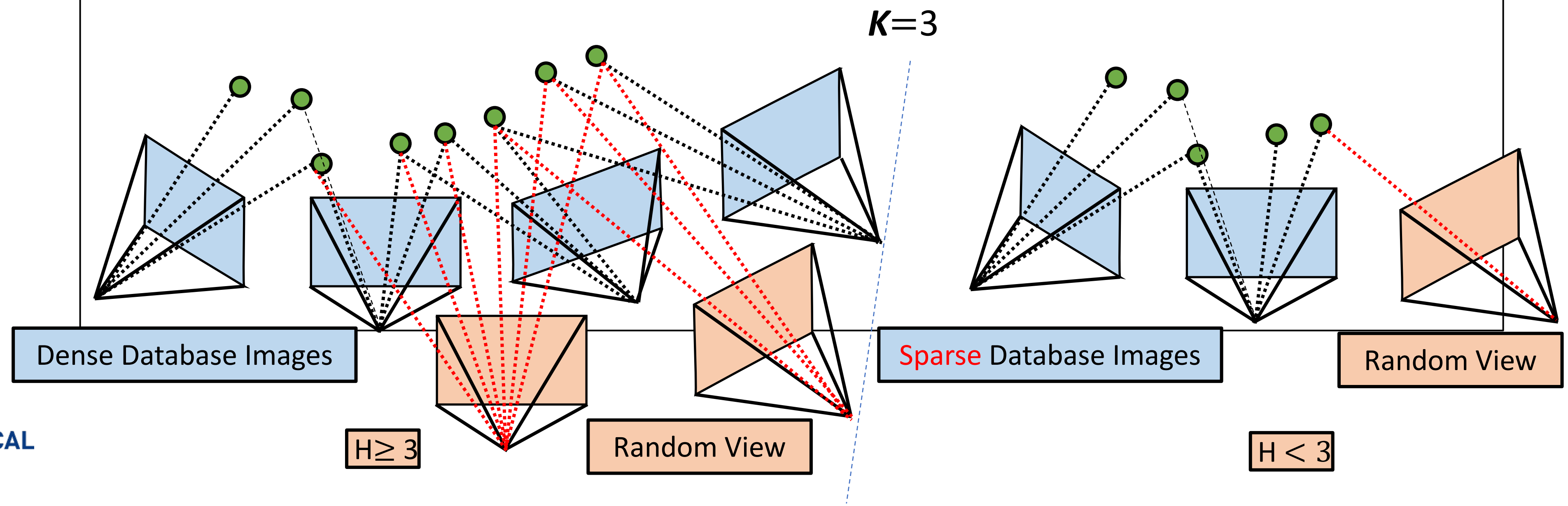
What is Image-based Localization?

- Given a 3D SfM PC, compute the 6-DOF camera pose for a query image



Further Analysis

- A Key Factor for Image-based Localization
  - A random view from underlying scene should be able to establish sufficient 2D-3D correspondences (denoted as H)
- Assumption of K-Cover based Methods
  - Discrete-database images well describe continuous 3D geometric space
  - Each database image observes at least ***K*** (H>=K) points -> A random view is likely to observe at least ***K*** points, i.e., the corresponding query image is likely to establish at least ***K*** 2D-3D correspondences (also H>=K)
  - The assumptions do not always hold (i.e., H may be lower than K)



NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Visibility (as a measure of point saliency) Probability

- Straightforward definition for a single point P_j

$$\phi(P_j) = \frac{d(P_j)}{m}$$

$\longrightarrow$ No. of database images observing point P_j

$\longrightarrow$ No. of all database images

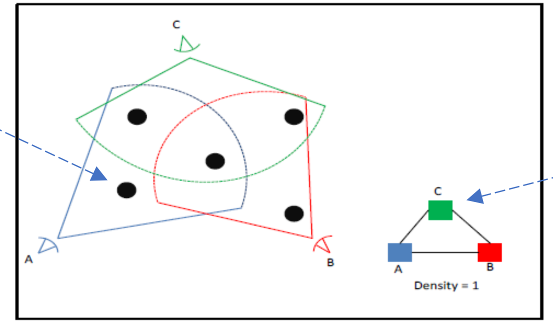
$$= \frac{\sum_i G_{ij}}{m}$$

- Resultant Visibility Probability $V(P_j)$ in a PC
  - $\phi(P_j)$ is an approximation of $V(P_j)$ when database images are densely distributed
  - Need to evaluate the distribution of database images in the PC (next slide)

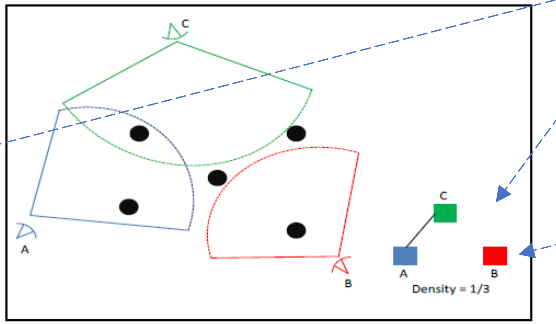
Density Estimation on Graph

Density of database images can be computed on image overlap graph $\mathcal{O}$ (constructed from G)

3D point



(a) **Dense Case**



(b) **Sparse Case**

(database images are not dense, or not from many angles)

Do not see a common 3D point, with A and C

Graph Density (D) Estimation:

$$D = \frac{2e}{m(m-1)}$$

e: number of edges in $\mathcal{O}$
m: number of nodes in $\mathcal{O}$

Computing $V(P_j)$:

$$V(P_j) = f(D)\phi(P_j)$$

$f(D)$: a weighting function based on D
e.g.,
$$f(D) = (1 - 6.72e^{-213D})$$

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Model (K,H)

Poisson Binomial Distribution, based on 3D point's visibility probability $V(P_j)$:

$$X_j \sim Bernoulli(V(P_j)), j = 1, ..., \tau$$

τ=n for original PC

Probability for P_j being visible: $\boxed{\Pr(X_j = 1) = V(P_j)}$

Probability for P_j being not visible: $\boxed{\Pr(X_j = 0) = 1 - V(P_j)}$

Total no. of points visible with a view: $X = \sum_{j=1}^{\tau} X_j.$

Approximation using Central Limit Theorem (check the paper for more details):

Cumulative Distribution Function
for γ points visible with a view: $Pr(X > \gamma) \approx 1 - \Phi(\frac{\gamma + 0.5 - \mu}{\sigma})$

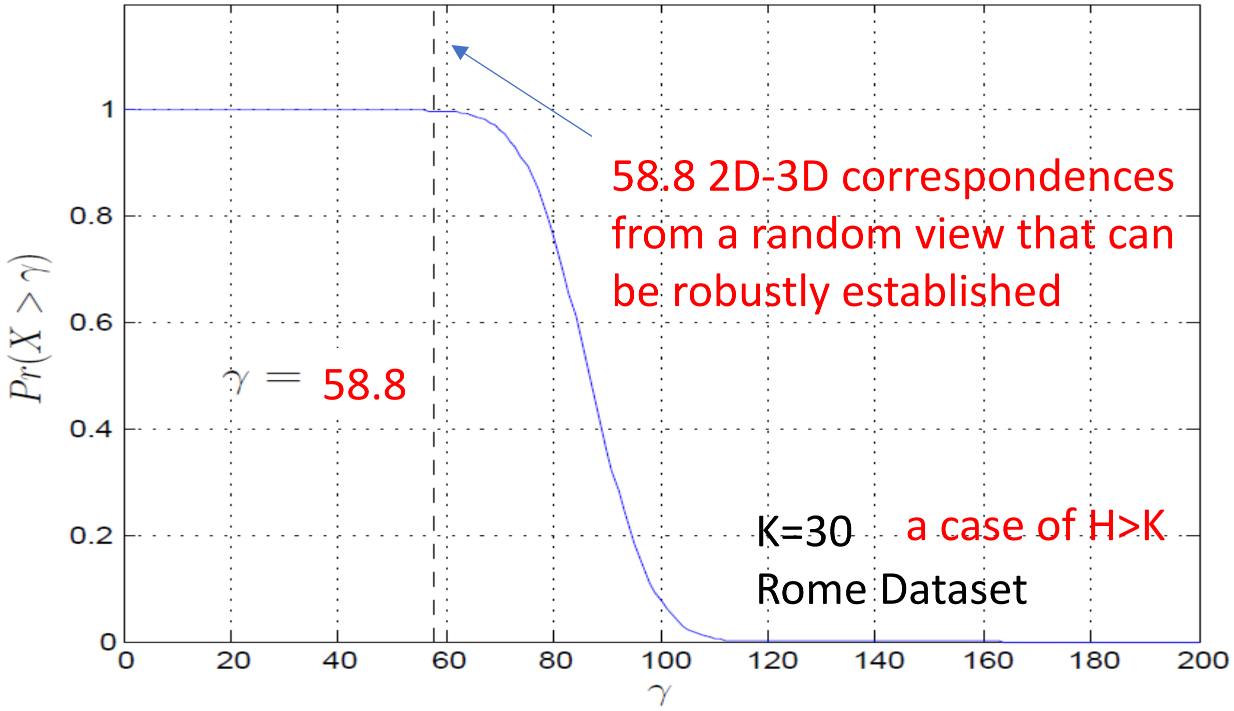
Φ: Normal distribution

$$\mu = E(X) = \sum_{j=1}^{\tau} V(P_j)$$

$\sigma = (\mu)^{1/2}$, when $V(P_j) \ll 1$ (i.e., for large databases)

$$H = \underset{\gamma}{\operatorname{argmax}} \overbrace{\left(\underline{Pr(X > \gamma)} = 1 \right)}^{\textit{seek the largest } \gamma \textit{ possible}}$$

a PC is robust to establish γ 2D-3D correspondences



For each database image/view

For each 3D point

Database	K	H	P: ave. No. of observed 3D points per database image
Dubrovnik	25	28.6	73.7
	30	35.6	87.4
Rome	20	36.6	61.9
	30	58.8	88.8
Aachen	60	8.2	96.4
	320	78.9	461.7

from Internet images with denser visual overlap among database images

from images captured by a single user (walking around the downtown of Aachen with a mobile-phone: database images more sparsely located; a higher K needed, and H<<K

P>max(H,K): because database images make the PC

H: the minimum/required no. of 2D-3D correspondences from any view for a required R

or

the largest no. of 2D-3D correspondences from any view for a K for simplification

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Cheng, et al, A Data-Driven Point Cloud Simplification Framework for City-Scale Image-Based Localization. IEEE **TIP**, 2017.

Evaluating AEWKC -Cheng, et al,IEEE **TIP**'17

localization performance

- Especially when simplifying to < 1% original size

The proposed methods

(A) Dubrovnik dataset

K	#points	%points	KC(%)	PKC(%)	WKC(%)	AEWKC(%)
12	5808	0.31	40.43	40.53	59.82	**67.65**
15	7571	0.40	51.79	52.68	71.15	**76.10**
18	9391	0.50	61.98	62.56	78.06	**80.44**
20	10615	0.56	68.99	67.44	80.71	**82.96**
22	11894	0.63	72.25	73.75	83.22	**84.20**
25	13877	0.74	78.78	79.37	**86.25**	86.20
28	15895	0.84	82.50	83.25	88.43	**89.06**
30	17147	0.91	84.61	84.50	89.50	**89.83**
32	18700	1.00	86.54	86.74	**90.00**	89.75
35	20807	1.10	88.01	88.19	**91.05**	90.68

(B) Rome dataset

K	#points	%points	KC(%)	PKC(%)	WKC(%)	AEWKC(%)
6	5109	0.13	45.61	46.05	61.78	**67.66**
8	7053	0.17	63.22	64.43	72.82	**74.69**
10	9117	0.22	72.03	72.99	78.27	**79.45**
12	11216	0.28	77.93	78.66	82.23	**82.75**
15	14598	0.36	83.83	83.90	83.98	**86.78**
18	18071	0.44	86.84	86.66	87.66	**88.25**
20	20426	0.50	88.28	88.25	89.12	**89.24**
22	22828	0.56	89.04	88.92	**90.35**	90.05
25	26560	0.65	90.65	90.30	91.15	**91.33**
28	30367	0.75	91.75	91.52	91.82	**91.99**

(C) Aachen dataset

K	#points	%points	KC(%)	PKC(%)	WKC(%)	AEWKC(%)
50	19487	1.26	46.07	46.31	55.28	**60.43**
60	24036	1.56	53.38	54.90	59.62	**63.68**
70	28693	1.86	58.53	56.36	62.87	**65.58**
80	33445	2.17	63.14	63.68	65.85	**67.20**
90	38290	2.49	64.22	64.66	**68.83**	67.15
100	43186	2.80	67.20	64.55	68.29	**69.10**
110	48141	3.12	67.20	66.80	71.00	**71.17**
120	53220	3.45	68.02	68.56	**72.08**	**72.08**
130	58394	3.79	70.00	70.00	72.95	**73.14**
140	63613	4.13	71.05	71.07	73.71	**74.07**

KC: Li, *et.al.*, ECCV 10
PKC: Cao, *et.al.*, CVPR 14
WKC: Cheng, *et.al.*, ICMEW 15
AEWKC: our method

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

More details of related research:

Cheng, et al, A Two-stage Outlier Filtering Framework for City-Scale Localization using 3D SfM Point Clouds.. IEEE Transaction on Image Processing (TIP), 2019.

Cheng, et al, Cascaded Parallel Filtering for Memory-efficient Image-based Localization. International Conference on Computer Vision (ICCV) 2019

W. Lin, S. Lee, "Visual Saliency and Quality Evaluation for 3D Point Clouds and Meshes: An Overview", APSIPA Trans. on Signal and Information Processing, 11(1): e28, 2022. http://dx.doi.org/10.1561/116.00000125

Chen, et al, “No-Reference Point Cloud Quality Assessment via Graph Convolutional Network”, IEEE Trans. on Multimedia, 27: 2489 – 2502, 2024.

Chen, et al,, “Dynamic Hypergraph Convolutional Network for No-Reference Point Cloud Quality Assessment”, IEEE Trans. on Circuits and Systems for Video Technology, 34(10):10479 – 10493, 2024.

Lee,et al, “3D-PSSIM: Projective Structural Similarity for 3D Mesh Quality Assessment Robust to Topological Irregularities”, IEEE Trans on Pattern Analysis and Machine Intelligence, 46(12): 9595 – 9611, 2024.

The rest of this talk: to present related recent research attempts… to address the major challenges

- PC representation
  - Compression
  - Mesh reconstruction
- PC saliency determination
- Image-based localization, for practical applications
  - PC simplification
  - PC quality assessment
- Moving forward

Summary

- Introduction to 3D point clouds (PCs)
  Increasingly important area; a lot to be done
- PC representation
  ❖Intrinsic measure
  ❖Resampling/simplification
  ❖Mesh construction
  ❖Compression & other representations
- PC saliency & quality
  Initial models for human and machine uses, respectively
- Exemplary utilities of PCs
  image-based localization with PCs, PC registration, and remeshing

further discussion

- Information Saliency & Sufficiency/Quality of 3D PCs
  - For humans

    According to visualization result: esp for small objects, or relatively simple scene
  - For machines: utility-based definitions, esp for city-scale PCs

    e.g., visibility, H measure, etc, toward image-based localization

To further exploit (cont’d):

- 3D PC compression
  - ❑PC-dedicated coders
  - ❑emerging deep learning-based coding
  - ❑lightweight learning model compression for green computing
- PC denoising, SR, edge detection, inpainting, …
- More utilities for PCs: recognition, classification, autonomous driving & flying, robots, building Information Models (BIM), VR/AR/MR, smart manufacturing, urban surveillance and planning, cultural heritage preservation, and other tasks in smart cities

Lin & Lee, “Visual Saliency and Quality Evaluation for 3D Point Clouds and Meshes: An Overview”, APSIPA Trans. Signal and Info Processing, 11(1), e28, 2022.

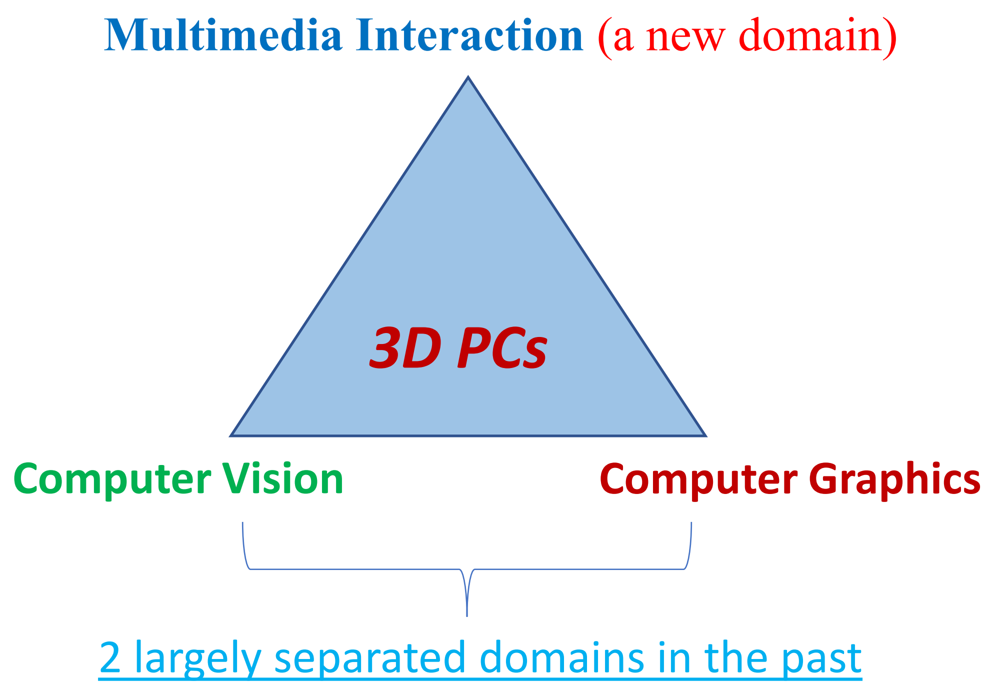
To further exploit (cont'd):

- PC saliency & quality modeling
  - ❑For human tasks
  - ❑for machine tasks
  - ❑possibilities for formulating for both human & machine uses
  - ❑Just-noticeable difference (JND) for PCs
  - ❑Utility-oriented models
  - ❑To consider technical quality, aesthetic quality, visual discomfort & Uncanny Valley together
- Medical and biological scenarios
- To explore possible advantages enabled by connected computer vision and graphics …

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

Lin & Ghinea, "Progress and Opportunities in Modelling Just-Noticeable Difference (JND) for Multimedia", IEEE TMM, 2022.
Lin & Lee, "Visual Saliency and Quality Evaluation for 3D Point Clouds and Meshes: An Overview", APSIPA Trans. Signal and Info Processing, 11(1), e28, 2022.

More opportunities?



PC representation (as presented earlier)

The j^{th} point in a PC, Ω, with K points totally, can be expressed as a vector:

$$\omega_j = [g_j, c_j, a_j], j = 1, 2, ..., K \quad (1)$$

where $g_j = [x_j, y_j, z_j]$ and $c_j = [r_j, g_j, b_j]$ being the geometry information (point position) and attribute information (RGB values), respectively; a_j represents auxiliary descriptor(s), such as the derived surface normal, and in a SfM PC, it can represent the SIFTs and database image associated to the point; and

$$\Omega = \{\omega_j, j = 1, 2, ..., K\}.$$

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE

PC representation extended to true multimedia

Take olfaction as an example to add to the PC representation for an arbitrary point:

$$\widetilde{\omega}_j = [\omega_j, o_j], j = 1, 2, \ldots, K,$$

where ω_j is defined in Equation (1), and $o_j = [e_j^1, e_j^2, \ldots, e_j^{L'}, f_j, d_j, i_j]$ denotes the olfactory descriptor.

Cross-modal effects must be considered (e.g., between visual and olfaction and among visual, olfaction, and taste):

$$\boldsymbol{S} = \sum_{\iota=1}^{n} S(\iota) - \sum_{\iota=1}^{n} \sum_{\rho=\iota+1}^{n} \varepsilon(\iota, \rho) \Psi(S(\iota), S(\rho))$$

Lin & Ghinea, "Progress and Opportunities in Modelling Just-Noticeable Difference (JND) for Multimedia", IEEE TMM, 2022.

Lin & Lee, "Visual Saliency and Quality Evaluation for 3D Point Clouds and Meshes: An Overview", APSIPA Trans. Signal and Info Processing, 11(1), e28, 2022.

Some of the speaker’s papers closely related to this presentation

- C. Lv, W. Lin, J. Zheng, “Adaptively Isotropic Remeshing based on Curvature Smoothed Field”, IEEE Transactions on Visualization and Computer Graphics, accepted.
- C. Lv, W. Lin, B. Zhao, “Intrinsic and Isotropic Resampling for 3D Point Clouds”, IEEE Trans on Pattern Analysis and Machine Intelligence, accepted
- C. Lv, W. Lin, B. Zhao, "KSS-ICP: Point Cloud Registration based on Kendall Shape Space", IEEE Transactions on Image Processing, 32: 1681-1693, 2023
- C. Lv, W. Lin, B. Zhao,“Voxel Structure-based Mesh Reconstruction from a 3D Point Cloud”, *IEEE Transactions on Multimedia*, accepted
- W. Lin, S. Lee, “Visual Saliency and Quality Evaluation for 3D Point Clouds and Meshes: An Overview”, APSIPA Trans. on Signal and Information Processing, 11(1), e28, 2022.
- C. Lv, W. Lin, B. Zhao, "Approximate Intrinsic Voxel Structure for Point Cloud Simplification", IEEE Transactions on Image Processing, 30(9): 7241 – 7255, 2021
- J. Xiong, H. Gao, M. Wang, H. Li, W. Lin, “Occupancy Map Guided Fast Video based Dynamic Point Cloud Coding”, IEEE Transactions on Circuits and Systems for Video Technology, accepted.
- W. Lin, G. Ghinea, “Progress and Opportunities in Modelling Just-Noticeable Difference (JND) for Multimedia”, IEEE TMM, 2022.
- B. Zhao, W. Lin, C. Lv, “Fine-Grained Patch Segmentation and Rasterization for 3D Point Cloud Attribute Compression”, IEEE Trans. on Circuits and Systems for Video Technology, 31(12): 4590-4602, 2021.
- W. Cheng, W. Lin, K. Chen, X. Zhang, “Cascaded Parallel Filtering for Memory Efficient Image-based Localization”, International Conference on Computer Vision (ICCV), 2019
- X. Ding, W. Lin, Z. Chen, X. Zhang, “Point Cloud Saliency Detection by Local and Global Feature Fusion”, IEEE Transactions on Image Processing, 28(11): 5379–5393, 2019
- W. Cheng, K. Chen, W. Lin, M. Goesele, X. Zhang, Y. Zhang, "A Two-stage Outlier Filtering Framework for City-Scale Localization using 3D SfM Point Clouds", IEEE Transactions on Image Processing, 28(10): 4857 - 4869, 2019
- J. Hou, B. Zhao, N. Ansari, W. Lin, “Range Image Based Point Cloud Colorization Using Conditional Generative Model”, *IEEE International Conference on Image Processing (ICIP),* 2019
- W. Cheng, W. Lin, X. Zhang, M. Goesele, M-T Sun, “A Data-driven Point Cloud Simplification Framework for City-scale Image-based Localization”, IEEE Transactions on Image Processing, 26(1): 262-275, 2017
- S. M. Prakhya, B. Liu, W. Lin, V. Jakhetiya, S. C. Guntuku, “B-SHOT: A Binary 3D Feature Descriptor for Fast Keypoint Matching on 3D Point Clouds”, Autonomous Robots, 41(7):1501–1520, 2017.
- S. M Prakhya, J. Lin, V. Chandrasekhar, W. Lin, B. Liu, "3DHoPD: A Fast Low Dimensional 3D Descriptor", IEEE Robotics and Automation Letters, 2(3): 1472-1479, 2017.
- S. M. Prakhya, W. Lin, V. Chandrasekhar, B. Liu, J. Lin, “Low Bit-rate 3D Feature Descriptors for Depth Data from Kinect-style Sensors”, Signal Processing: Image Communication, 51: 40–49, 2017
- S. M. Prakhya, B. Liu, W. Lin, “Detecting Keypoint Sets on 3D Point Clouds via Histogram of Normal Orientations”, Pattern Recognition Letters, 83 (Part 1): 42–48, 2016.
- S. M. Prakhya, B. Liu, W. Lin, “B-SHOT: A binary feature descriptor for fast and efficient keypoint matching on 3D point clouds”, IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2015

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE