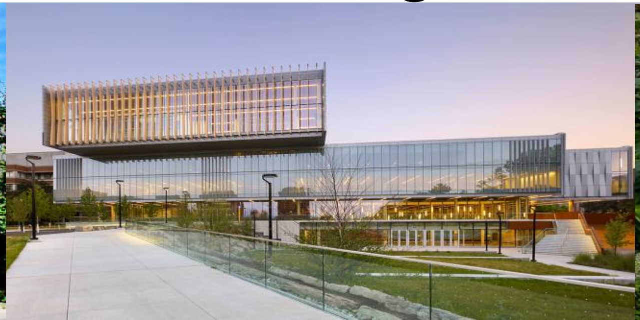
3D Point Cloud Enhancement using Graph-Modelled Multiview Depth Measurements

Xue Zhang, Gene Cheung,
Jiahao Pang, and Dong Tian

Session SS-04: Learning-Based Processing of 3D Visual Data --26 October, 2020



Outline

- Background
- Contributions
- Problem Formulation
- Graph Construction
- Experiments
- Conclusion

Gene Cheung (genec@yorku.ca)

Outline

- Background
- Contributions
- Problem Formulation
- Graph Construction
- Experiments
- Conclusion

Background

➢ 3D point cloud (PC) generation by deploying multiple depth sensors

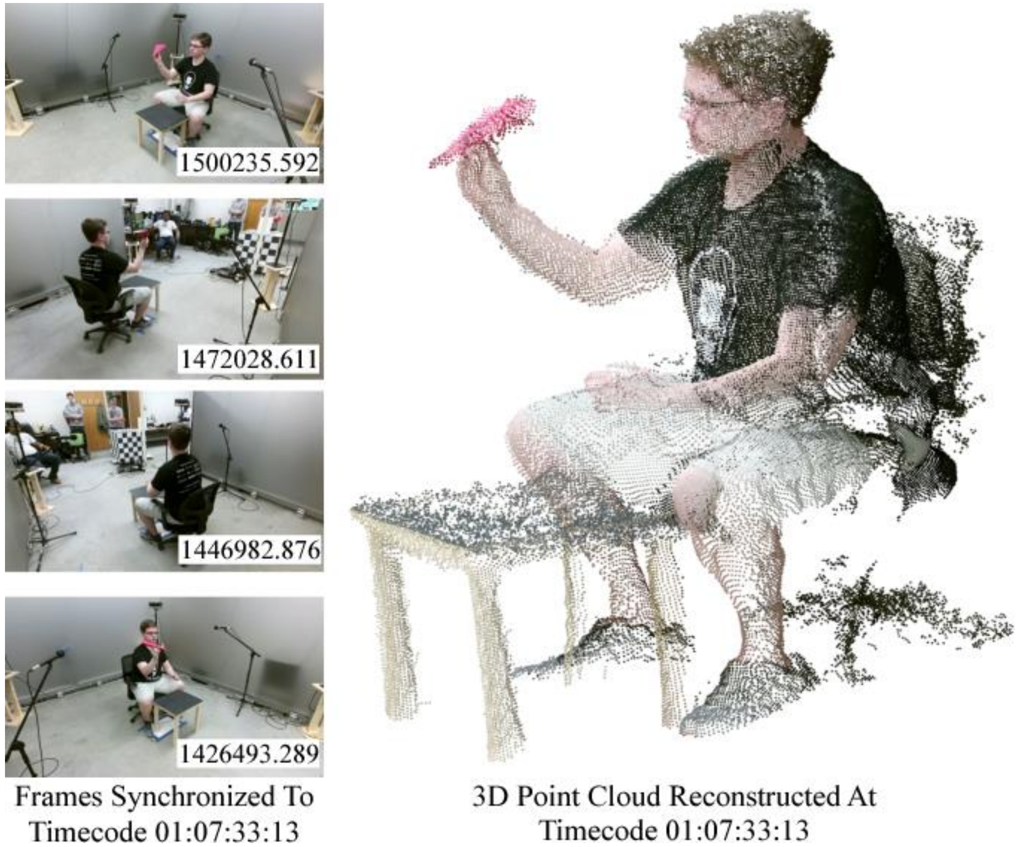


Fig.1 Synthesizing a 3D point cloud using multiple Kinects [1].

Fig.2 Asus Xtions mounted on drones [2].

- **Advantage**: sensors are lightweight, low-powered and inexpensive.
- **Challenge**: acquired depth measurements suffer from both imprecision and additive noise, resulting in a corrupted synthesized PC.

[1] Y., Jiang, D. Russell, T. Godisart, N. K. Banerjee, and S. Banerjee, "Hardware Synchronization of Multiple Kinects and Microphones for 3D Audiovisual Spatiotemporal Data Capture." *2018 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2018.

[2] L. Xu, Y. Liu, W. Cheng, K. Guo, G. Zhou, Q. Dai, and L. Fang, "Flycap: Markerless motion capture using multiple autonomous flying cameras." *IEEE transactions on visualization and computer graphics* 24.8 (2017): 2284-2297.

Background (cont'd)

➢ Related works

- **Point cloud denoising**
  - **previous works**: 1) low-rank prior [3],
    2) low-dimensional manifold model (LDMM) [4],
    3) surface smoothness priors expressed as graph total variation (GTV) [5],
    graph Laplacian regularizer (GLR) [4],
    feature graph Laplacian regularizer (GFLR) [6]
  - **con**: enhance a PC *a posteriori*

    (Denoise raw RGB measurements *before* demosaicking, >15dB PSNR gain [7])

[3] K. Sarkar, F. Bernard, K. Varanasi, C. Theobalt and D. Stricker, "Structured low-rank matrix factorization for point-cloud denoising," in *International Conference on 3D Vision (3DV),* 2018, pp. 444–453.
[4] J. Zeng, G. Cheung, M. Ng, J. Pang and C. Yang, "3D point cloud denoising using graph Laplacian regularization of a low dimensional manifold model," *IEEE Transactions on Image Processing*, vol. 29, pp. 3474–3489, 2019.
[5] C. Dinesh, G. Cheung, I.V. Bajić and C. Yang, "Local 3D point cloud denoising via bipartite graph approximation & total variation," in *MMSP*, 2018.
[6] C. Dinesh, G. Cheung, and I.V. Bajić, "Point cloud denoising via feature graph Laplacian regularization." *IEEE Transactions on Image Processing*, vol. 29, pp. 4143-4158, 2020.
[7] A. Punnappurath, and M.S. Brown, "Learning raw image reconstruction-aware deep image compressors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42,no. 4, pp. 1013 – 1019, 2020.

Background (cont'd)

➢ Related works

- **Depth image enhancement**
  - **previous works**: using image-based signal priors.
  - **con**: 1）ignoring the inherent cross-correlation among the views (sub-optimal) [8-10]

    2）quantizing observations per pixel from two views but considering noiseless [11]

[8] W. Hu, G. Cheung, and M. Kazui, "Graph-based dequantization of block-compressed piecewise smooth images," *IEEE Signal Processing Letters*, February 2016, vol. 23, no.2, pp.242–246.

[9] S. Gu, W. Zuo, S. Guo, Y. Chen, C. Chen, and L. Zhang, "Learning dynamic guidance for depth image enhancement," in *Proceedings of the IEEE conference on computer vision and pattern recognit*ion, 2017, pp. 3769–3778.

[10] J. Jeon and S. Lee, "Reconstruction-based pairwise depth dataset for depth image enhancement using CNN," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 422–438.

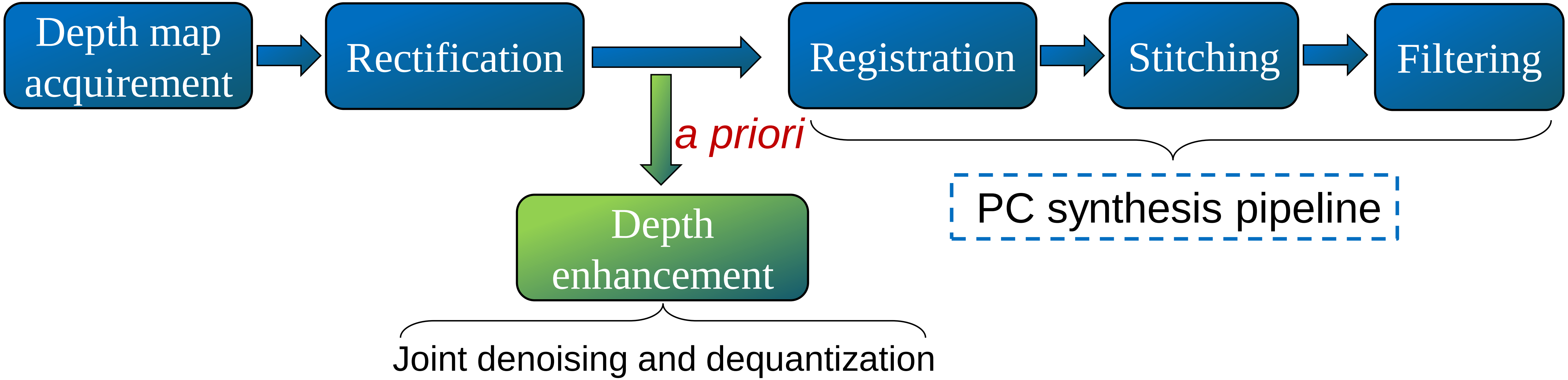
[11] P. Wan, G. Cheung, P. Chou, D. Florencio, C. Zhang, and O. Au, "Precision enhancement of 3D surfaces from com-pressed multiview depth maps," *IEEE Signal Processing Letters*, October 2015, vol. 22, no.10, pp. 1676–1680.

Outline

- Background
- Contributions
- Problem Formulation
- Graph Construction
- Experiments
- Conclusion

Gene Cheung (genec@yorku.ca)

Contributions



- Enhance closer to the actual physical sensing process;
- Construct a sparse graph reflecting inter-pixel similarities via metric learning [12];
- Write a non-linear mapping function exploiting inter-view correlation;
- Formulate a *maximum a posteriori* (MAP) optimization problem after suitable linear approximations;
- Solve by fast gradient method (FGM).

[12] W. Hu, X. Gao, G. Cheung, and Z. Guo, "Feature graph learning for 3D point cloud denoising," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2841–2856, 2020.

Outline

Gene Cheung (genec@yorku.ca)

System Overview

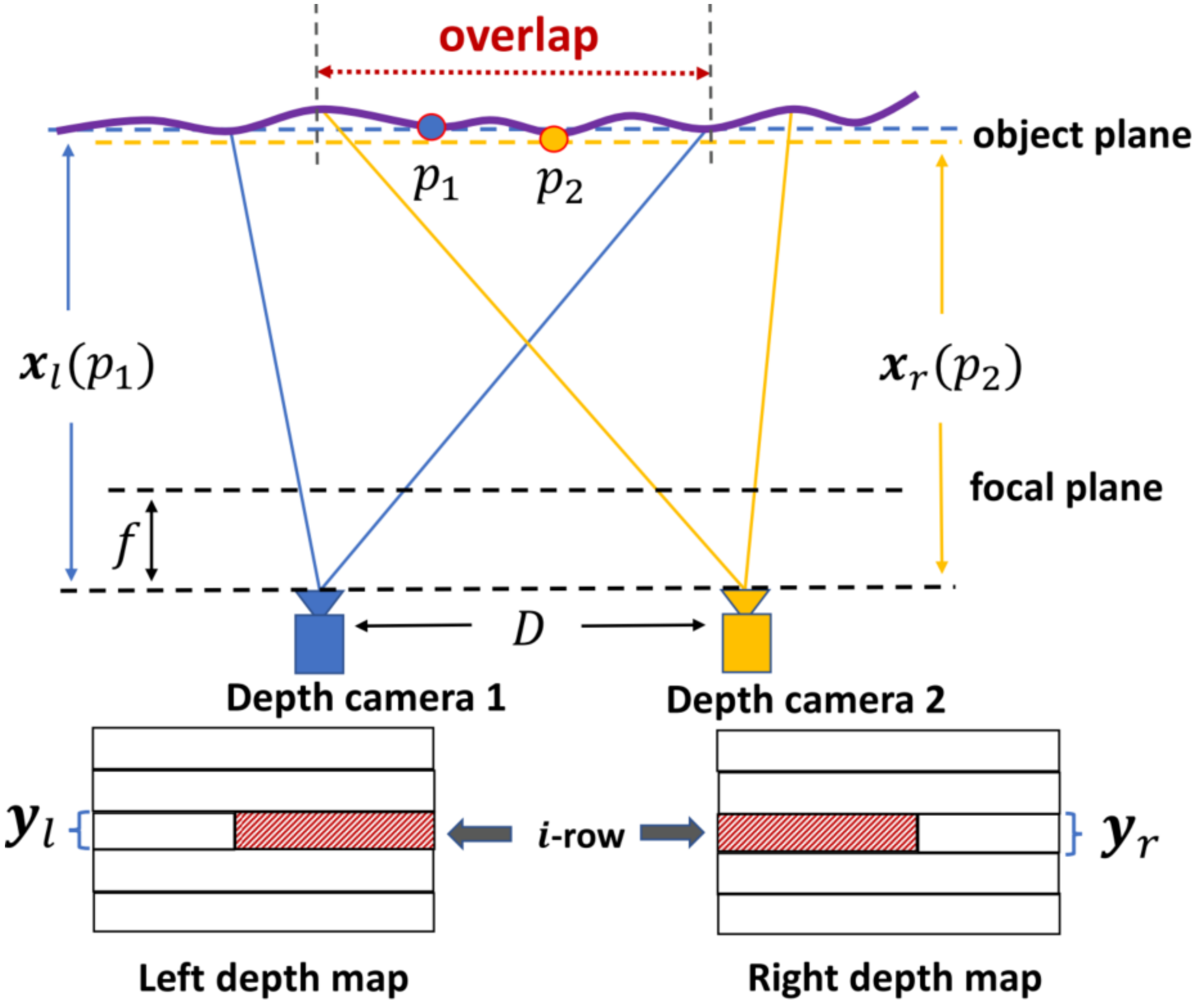


Fig.3 An example of the camera system

- Each pixel is a noise-corrupted and quantized to a *B*-bit representation.
- We assume that the depth map is the "rawest" signal we can acquire from the sensor.

Image Formation Model

true depth measurements

$\mathbf{n}_l \in \mathbb{R}^N$ is a zero mean additive noise.

$$\mathbf{y}_l = \text{round}\left(\frac{\mathbf{x}_l^o + \mathbf{n}_l}{Q}\right) Q$$

$\mathbf{y}_l \in \mathbb{R}^N$ is an observed depth pixel row i in the left view.

$Q \in \mathbb{R}^+$ is a quantization parameter

Goal: optimally reconstruct signal $\widetilde{\mathbf{X}} = \{\widetilde{\mathbf{x}}_l, \widetilde{\mathbf{x}}_r\}$ given observation $\mathbf{Y} = \{\mathbf{y}_l, \mathbf{y}_r\}$.

Gene Cheung (genec@yorku.ca)

View-to-view Mapping

- **1D warping procedure**

j-th pixel in the left view.

$$\underline{s(j, x_{l,j})} = j - \underline{\delta(x_{l,j})}$$

the (non-integer) horizontal position in the right view after projection

disparity $\delta(x_{l,j}) = \frac{fD}{x_{l,j}}$

- **Weighted linear interpolation to estimate the on-grid pixels**

$$\mathbf{x}_r = \underline{\mathbf{W}(\mathbf{x}_l)}\,\mathbf{x}_l = \mathbf{g}(\mathbf{x}_l)$$

→ *Linear approximation*

weights $\omega_{ij} = \frac{1}{\sum_{m=1}^{N} \exp\left(-\frac{(s(m,x_{l,m})-i)^2}{\sigma_s^2}\right)} \exp\left(-\frac{(s(j,x_{l,j})-i)^2}{\sigma_s^2}\right)$

Gene Cheung (genec@yorku.ca)

Approximating the Likelihood Term

- **Likelihood term**

$$\Pr(\mathbf{y}_l|\mathbf{x}_l) = \int_{\mathcal{R}_l} \Pr(\mathbf{n}_l)\, d\mathbf{n}_l \begin{cases} \text{jointly Gaussian distribution } \Pr(\mathbf{n}_l) = \exp\left(-\dfrac{\mathbf{n}_l^\top \mathbf{P}_l \mathbf{n}_l}{\sigma_n^2}\right) \\ \mathcal{R}_l = \left\{ \mathbf{n}_{l,i} \mid \mathbf{y}_{l,i} - \dfrac{Q}{2} \leq \mathbf{x}_{l,i} + \mathbf{n}_{l,i} < \mathbf{y}_{l,i} + \dfrac{Q}{2} \right\} \end{cases}$$

$$\Pr(\mathbf{n}_l) \approx \mathbf{a}^\top \mathbf{n}_l + b$$

$$\Pr(\mathbf{y}_l|\mathbf{x}_l) \approx \int_{\mathcal{R}_l} (\mathbf{a}^\top \mathbf{n}_l + b)\, d\mathbf{n}_l$$
$$= Q^N \left(\mathbf{a}^\top (\mathbf{y}_l - \mathbf{x}_l) + b\right)$$

Affine approximation

For reasonably small Q, this is a good approximation.

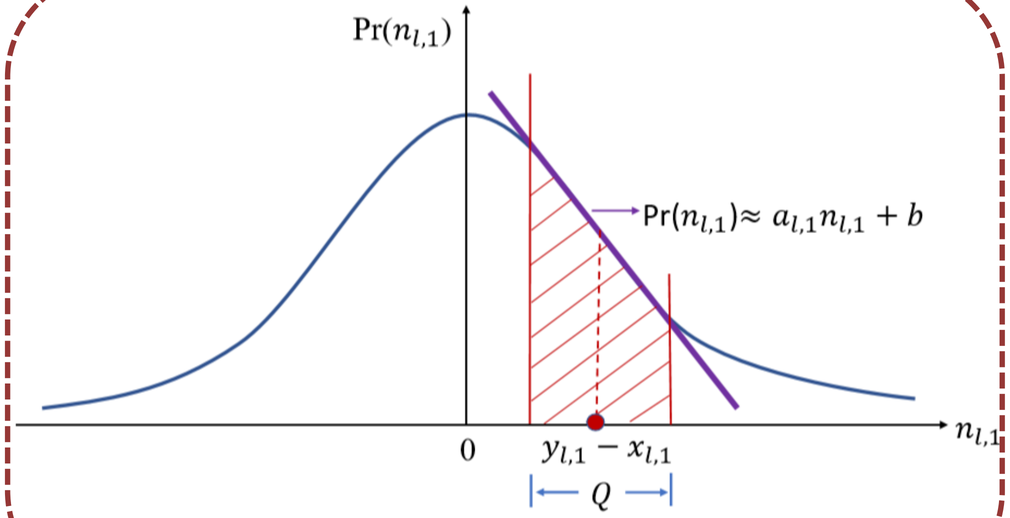


Fig.4 Affine approximation of the Gaussian pdf

Gene Cheung (genec@yorku.ca)

MAP Formulation

- **Signal prior**

$$\Pr(\mathbf{x}_l) = \exp\left(-\frac{\mathbf{x}_l^\top \mathbf{L}_l \mathbf{x}_l}{\sigma^2}\right)$$

A **graph Laplacian matrix** capturing inter-pixel similarities [13], learned from the previous *K* rows via metric learning (to be discussed)

- **MAP estimation**

$$\begin{aligned}\max_{\mathbf{x}_l,\mathbf{x}_r} \; & \Pr(\mathbf{y}_l,\mathbf{y}_r|\mathbf{x}_l,\mathbf{x}_r)\Pr(\mathbf{x}_l,\mathbf{x}_r) \\ = \; & \Pr(\mathbf{y}_l|\mathbf{x}_l)\Pr(\mathbf{y}_r|\mathbf{g}(\mathbf{x}_l))\Pr(\mathbf{x}_l)\Pr(\mathbf{g}(\mathbf{x}_l)) \\ \approx \; & Q^{2N}(\mathbf{a}^\top(\mathbf{y}_l-\mathbf{x}_l)+b)(\mathbf{a}^\top(\mathbf{y}_r-\mathbf{g}(\mathbf{x}_l))+b) \\ & \exp\left(-\frac{\mathbf{x}_l^\top\mathbf{L}_l\mathbf{x}_l}{\sigma^2}\right)\exp\left(-\frac{\mathbf{g}(\mathbf{x}_l)^\top\mathbf{L}_r\mathbf{g}(\mathbf{x}_l)}{\sigma^2}\right)\end{aligned}$$

negative log

$$\begin{aligned}\min_{\mathbf{x}_l} \; & -\ln(\mathbf{a}^\top(\mathbf{y}_l-\mathbf{x}_l)+b) - \ln(\mathbf{a}^\top(\mathbf{y}_r-(\mathbf{H}\mathbf{x}_l+\mathbf{d}))+b) \\ & + \mathbf{x}_l^\top\mathbf{L}_l\mathbf{x}_l + \mathbf{x}_l^\top\mathbf{H}^\top\mathbf{L}_r\mathbf{H}\mathbf{x}_l + 2\mathbf{d}^\top\mathbf{L}_r\mathbf{H}\mathbf{x}_l + \mathbf{d}^\top\mathbf{L}_r\mathbf{d}\end{aligned}$$

Unconstrained convex and differentiable

Solving via *fast gradient method*

[13] Y. Bai, G. Cheung, X. Liu, and W. Gao, "Graph-based blind image deblurring from a single photograph," IEEE Transactions on Image Processing, vol. 28, no. 3, pp. 1404–1418, 2018

Outline

- Background
- Contributions
- Problem Formulation
- Graph Construction
- Experiments
- Conclusion

Gene Cheung (genec@yorku.ca)

Feature Graph Learning

We use *metric learning* in [12] to construct a graph given chosen features considering *K* previous pixel rows.

$$\min_{\mathbf{M}\succ 0} \sum_{k=1}^{K} \left(\tilde{\mathbf{x}}_l^k\right)^\top \mathbf{L}_l^k(\mathbf{M})\tilde{\mathbf{x}}_l^k$$

$$= \sum_{k=1}^{K}\sum_{i,j} w_{ij}^k \left(\tilde{x}_{l,i}^k - \tilde{x}_{l,j}^k\right)^2$$

$\mathbf{M} \in \mathbb{R}^{F\times F}$
symmetric and
PD *metric* matrix

edge weights $w_{ij} = \exp\left(-d_{ij}\right)$

feature distance $d_{ij} = (\mathbf{f}_i - \mathbf{f}_j)^\top \mathbf{M}(\mathbf{f}_i - \mathbf{f}_j)$

feature vector $\mathbf{f}_i \in \mathbb{R}^6$

- *surface normal* $\mathbf{n}_i \in \mathbb{R}^3$
- depth value x_i
- location $\mathbf{l}_i \in \mathbb{R}^2$

[12] W. Hu, X. Gao, G. Cheung, and Z. Guo, “Feature graph learning for 3D point cloud denoising,” *IEEE Transactions on Signal Processing*, vol. 68, pp. 2841– 2856, 2020.

Outline

- Background
- Contributions
- Problem Formulation
- Graph Construction
- Experiments
- Conclusion

Gene Cheung (genec@yorku.ca)

Experimental Setup

- **Five depth image pairs provided in Middlebury datasets** [14]
  - Three generate PCs with around 700K points
  - Two are with round 337K points and 192K points, respectively.
- **Comparison: three PC denoising algorithms**
  - APSS [15], RIMLS [16], MRPCA [17]
- **Evaluation metrics**:
  - point-to-point (C2C) error
  - point-to-plane (C2P) error

[14] D. Scharstein, H. Hirschm¨uller, Y. Kitajima, G. Krathwohl, N. Nevsi´c, X. Wang, and P. Westling, "High-resolution stereo datasets with subpixel-accurate ground truth," in *German conference on pattern recognition*. Springer, 2014, pp. 31–42.
[15] G. Guennebaud and M. Gross, "Algebraic point set surfaces," in *ACM SIGGRAPH 2007 papers*, pp. 23–es. 2007.
[16] A C. ¨Oztireli, G. Guennebaud, and M. Gross, "Feature preserving point set surfaces based on non-linear kernel regression," in *Computer Graphics Forum*. Wiley Online Library, 2009, vol. 28, pp. 493–501.
[17] E. Mattei and A. Castrodad, "Point cloud denoising via moving RPCA," in *Computer Graphics Forum*. Wiley Online Library, 2017, vol. 36, pp. 123–137.

Experimental Results

Table 1. C2C and C2P results of competing methods at three noise levels.

σ_n^2	methods	Adirondack	ArtL	Teddy	Recycle	Playtable
50	APSS	3.63	3.47	2.76	3.92	4.08
		14.45	11.73	7.26	15.79	17.42
	RIMLS	3.47	3.35	2.67	3.72	4.09
		13.26	11.21	7.09	15.11	17.06
	MRPCA	2.91	3.05	2.55	3.17	3.21
		8.86	8.73	6.21	10.21	9.17
	Proposed	**2.11**	**2.26**	**1.56**	**2.45**	**3.09**
		4.88	**6.34**	**2.79**	**7.00**	**8.78**
70	APSS	4.12	3.80	3.09	4.34	4.46
		18.56	13.88	8.96	20.07	18.91
	RIMLS	3.83	3.67	3.00	4.16	4.38
		17.26	13.41	8.73	19.42	18.92
	MRPCA	3.42	3.45	2.89	3.76	3.48
		12.57	11.39	7.98	14.80	11.00
	Proposed	**2.32**	**2.48**	**1.68**	**2.68**	**3.26**
		5.97	**7.64**	**3.32**	**8.47**	**10.43**
90	APSS	4.40	4.28	3.38	4.80	4.91
		21.97	17.07	11.08	25.11	26.18
	RIMLS	4.19	4.13	3.30	4.59	4.83
		21.15	16.52	10.70	24.16	23.46
	MRPCA	3.78	3.91	3.20	4.20	3.95
		16.11	14.07	9.69	19.10	14.52
	Proposed	**2.47**	**2.70**	**1.84**	**2.92**	**3.45**
		6.95	**9.15**	**4.08**	**10.45**	**13.22**

Our method achieves by far the best performance in both metrics and all three noise levels, with C2C reduced by 0.68, 0.92, 1.13; and C2P reduced by 2.68, 4.38, 5.93 on average compared to the second best algorithm, respectively.

Experimental Results (cont'd)

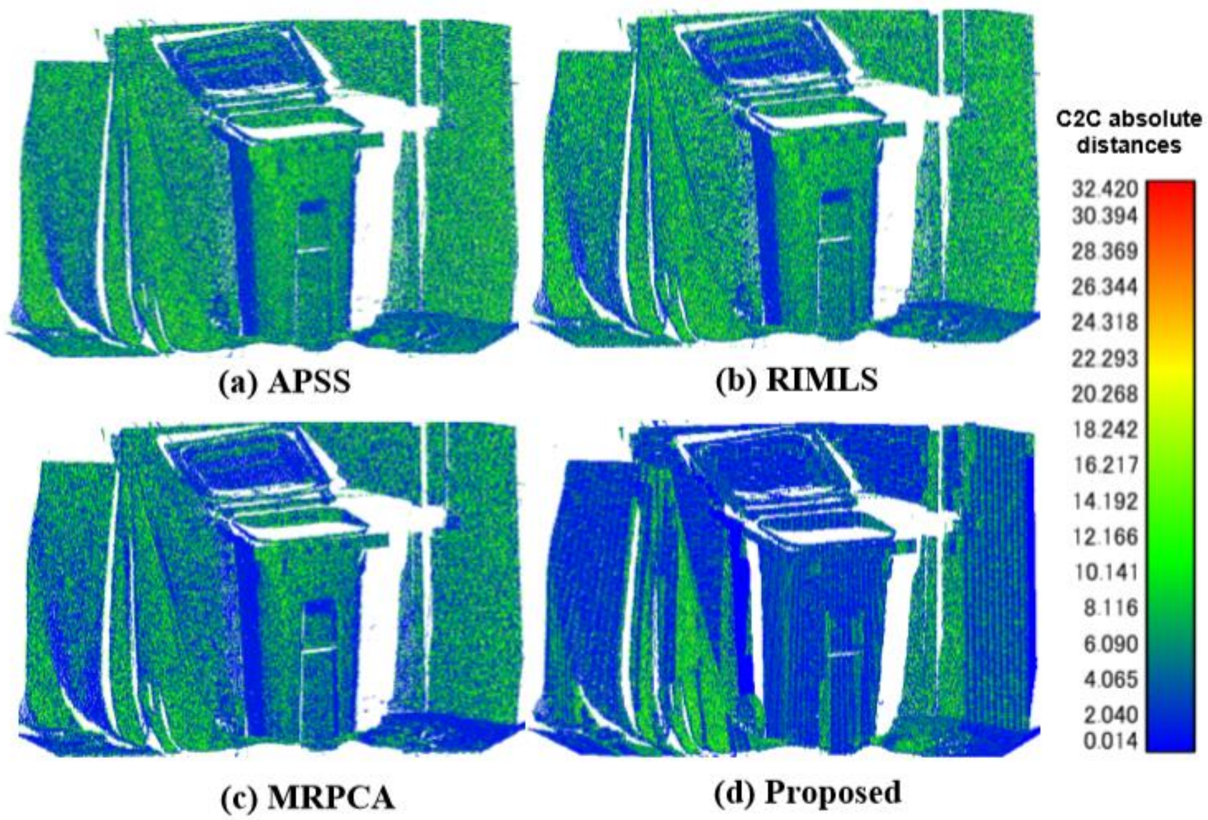


Fig. 3. Comparison of visual results for `Recycle` when $\sigma_n^2 = 50$. From blue to red, C2C absolute errors gradually become larger. More blue points are noticely included in the proposed method.

We observe that our proposed method achieves smaller C2C errors (in blue) compared to the competitors.

Outline

- Background
- Contributions
- Problem Formulation
- Graph Construction
- Experiments
- Conclusion

Gene Cheung (genec@yorku.ca)

Conclusion

- Jointly enhance **multiview** depth images
- Prior to modules (in a typical PC synthesis pipeline) that obscure acquisition noise
- A graph based MAP optimization: an ***image formation model*** accounting for both **additive noise** and **quantization**
- Performance**: outperform** the competitors (denoising PCs after the synthesis pipeline)

Gene Cheung (genec@yorku.ca)

Thank You !

Xue Zhang, Post-doc, York University (xuezhang@yorku.ca)