RD-OPTIMIZED 3D PLANAR MODEL RECONSTRUCTION & ENCODING FOR VIDEO COMPRESSION

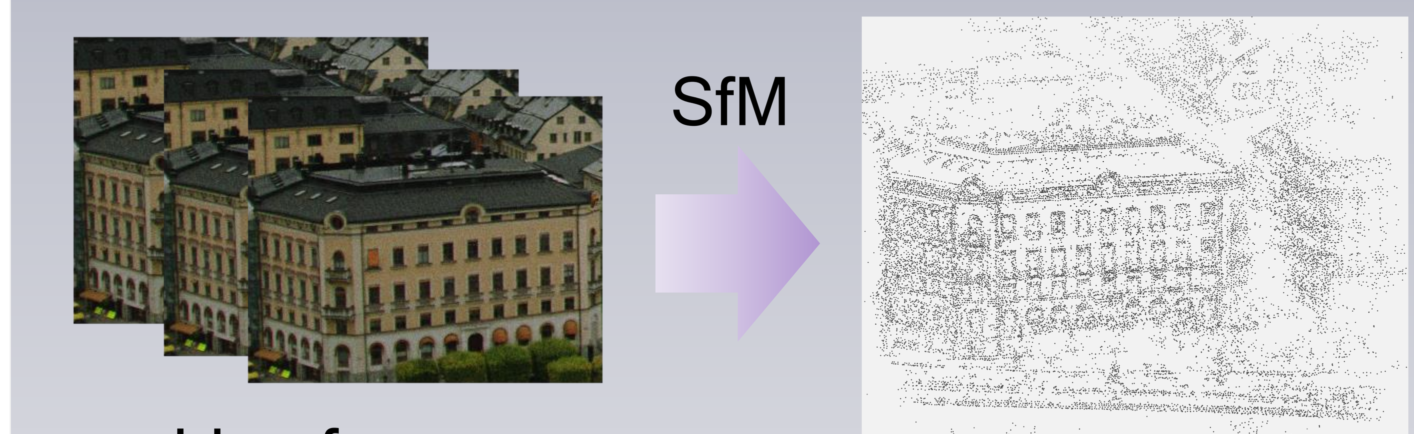
Cheng Yang[1], Gene Cheung[2], Seishi Takamura[3]

[1]National Institute of Informatics, Japan [2]York University, Canada [3]NTT, Japan

cheng@nii.ac.jp, genec@yorku.ca, takamura.seishi@lab.ntt.co.jp

Background

- Motion prediction/transform coding - lack discovery of multi-frame redundancy.
- Structure-from-Motion (SfM) for 3D structure reconstruction.



- Idea: exploit frame redundancy with 3D planar modeling.

Related work

- Model-based video coding [1],[2],[3] - high cost for roughly flat regions.
- Model accuracy v.s. coding cost.
- Object tracking-based coding [4] - lack of appropriate planar grid density setting.

Contribution

1. 3D planar modeling with appropriate planar grid density
2. Don't-care-region (DCR) optimized planar model coding

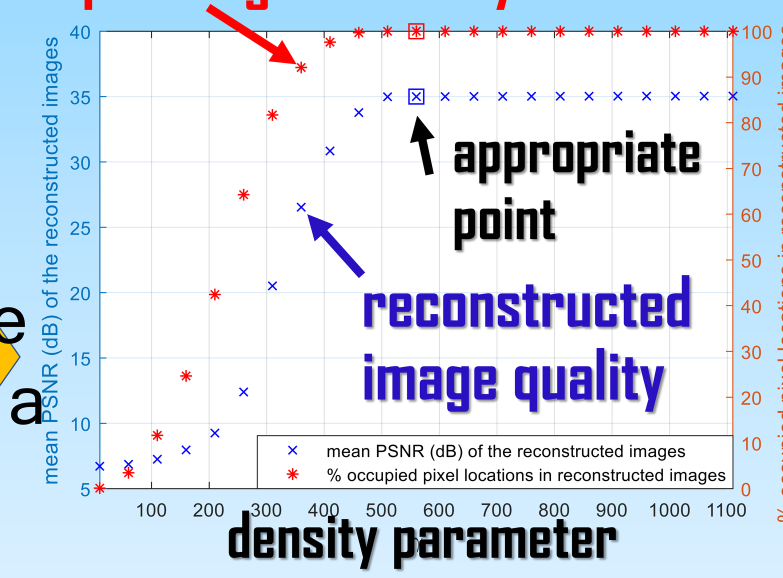
3D planar modelling

video frames → SfM → point cloud set →

RANSAC plane fitting

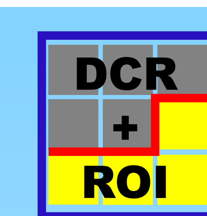
Side information = planar boundary, texture, camera params, frame translation/rotation

Project the frame pixels to a plane



DCR optimization

- Objective: minimize the number of non-zero transform coefficients.



pixels inside region-of-interest (ROI)

DCT, Intra prediction, sampling matrix

$$\min_{\mathbf{x}} \|\mathbf{\Phi}(\mathbf{x}-\mathbf{z})\|_0 \quad \text{s.t.} \quad \mathbf{D}\mathbf{x} = \mathbf{y}$$

convex relaxation

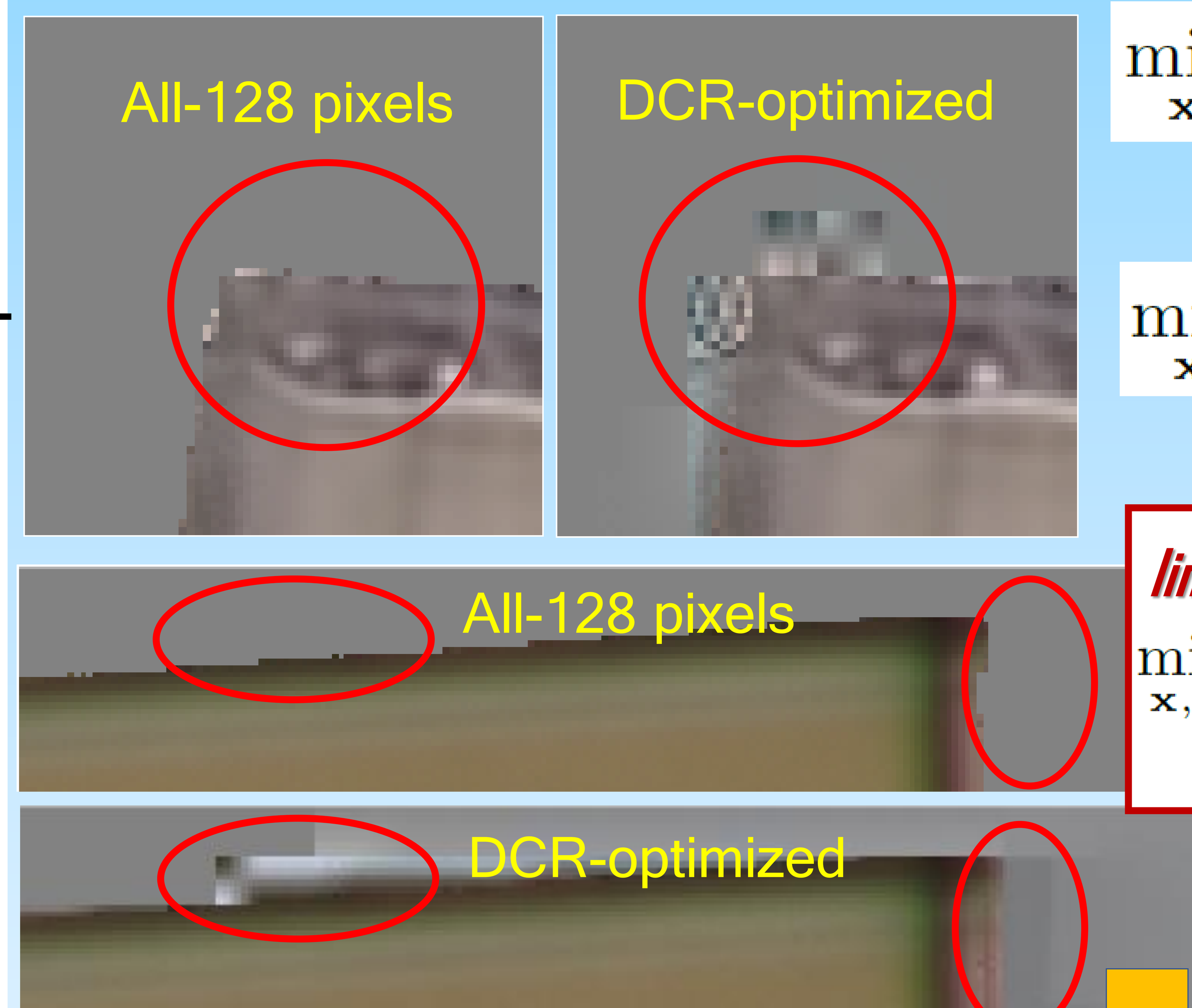
$$\min_{\mathbf{x}} \|\mathbf{\Phi}(\mathbf{x}-\mathbf{z})\|_1 \quad \text{s.t.} \quad \mathbf{D}\mathbf{x} = \mathbf{y}$$

reformulation

linear objective / *linear constraints*

$$\min_{\mathbf{x},\mathbf{u}} \mathbf{1}^\top \mathbf{u} \quad \text{s.t.} \begin{cases} -\mathbf{u} \leq \mathbf{\Phi}(\mathbf{x}-\mathbf{z}) \leq \mathbf{u}, \\ \mathbf{D}\mathbf{x} = \mathbf{y}, \\ \mathbf{u} \geq 0. \end{cases}$$

solved by linear programming (LP)



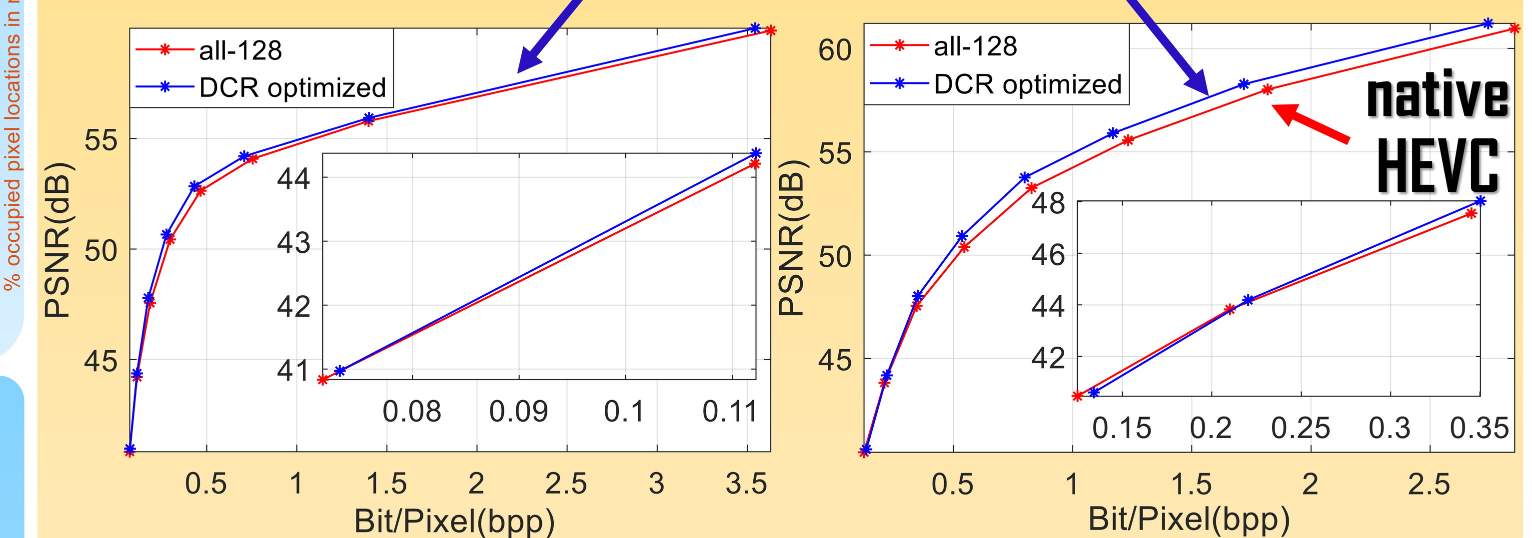
Planar model and video coding

planar boundary - AEC [5]; video frames – HEVC plus planar prediction (project the planar model pixels back to video frames)

Results (2 videos)

- Planar model coding with QP's 2-37

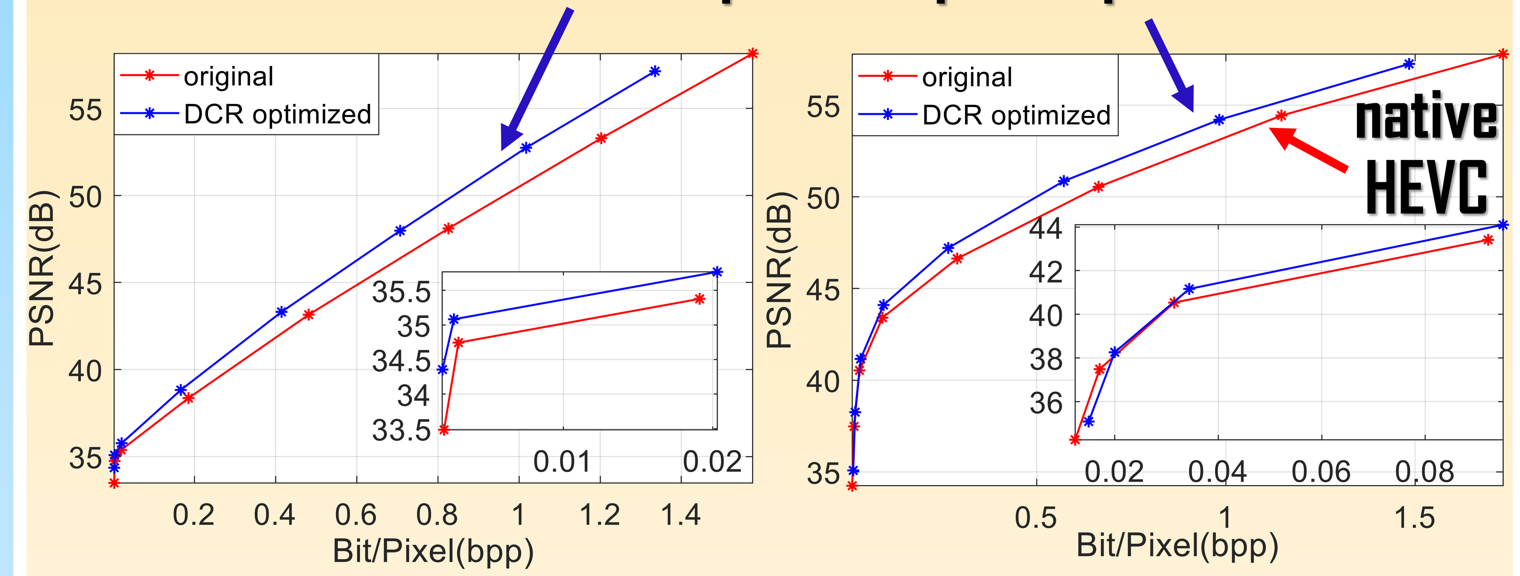
HEVC with DCR-optimization



BD-rate reduction(Y): **-7.45%, -3.65%**

- Video coding with QP's 2-37

HEVC with DCR-optimized planar prediction



BD-rate reduction(Y): **-15.72%, -11.72%**

References

[1] W. Sun et al., "Rate-constrained 3D surface estimation from noise-corrupted multiview depth videos," *IEEE TIP*, vol. 23, no. 7, July 2014.

[2] E. Imre et al., "Rate-distortion efficient piecewise planar 3D scene representation from 2-D images," *IEEE TIP*, vol. 18, no. 3, March 2009.

[3] B. O. Ozkalayci and A. A. Alatan, "3D planar representation of stereo depth images for 3DTV applications," *IEEE TIP*, vol. 23, no. 12, Dec. 2014.

[4] S. Takamura and A. Shimizu, "Efficient video coding using rigid object tracking," in *APSIPA*, Dec. 2017.

[5] I. Daribo et al., "Arbitrarily shaped motion prediction for depth video compression using arithmetic edge coding," *IEEE TIP*, vol. 23, no. 11, Nov 2014.