Depth from Accidental Motion using Geometry Prior

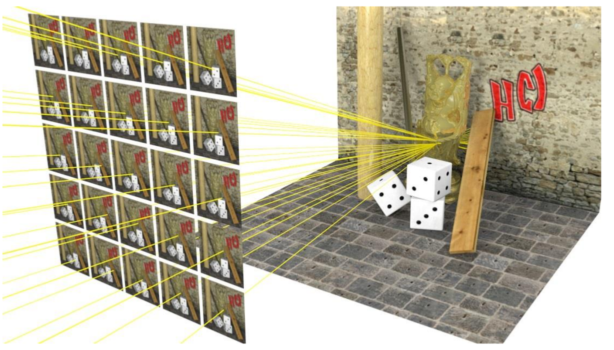
Sunghoon Im, Gyeongmin Choe, Hae-Gon Jeon, In So Kweon

Korea Advanced Institute of Science and Technology

★ Top 10% paper

Motivation

- Cameras for depth estimation
  - Light-field camera
    - Low spatial resolution
    - Short depth range
  - Stereo camera
    - Expensive



- Depth from accidental motion
  - Mobile phone camera (or) DSLR
    - Off-the-shelf Camera
    - Handshaking (Inevitable motion)
    - Narrow-baseline
    - 3sec (100 frames)

User-friendly

- Contribution
  - High quality depth map is estimated from popularized mobile phone
  - Feature extraction — remove features on moving objects and extract very accurate features
  - Sparse 3D reconstruction — set reliable initial camera poses
  - Dense 3D reconstruction — geometry prior assists to remove an artifact on frontal parallel scene

Feature extraction

- Feature extraction
  - Harris corner & KLT tracker
- Remove features on moving objects
  - Compute homography
  - Initial R, T compute by homography decomposition
  - Feature Inlier voting

$$H = K(R^T + \frac{Tn^T}{d})K^{-1}$$

Sparse 3D reconstruction

- Geometric model for small motion
  - Rotation matrix - Small angle approximation
  - Key to estimate pose and depth without any prior depth information
- Bundle adjustment
  - LM optimization
  - Without depth information - Random depth initialization

$$F = \sum_{i=1}^{N_I}\sum_{j=1}^{N_F} ||p_{ij} - \pi(K(R_iP_j + T_i))||^2$$

Dense 3D reconstruction

- Objective function

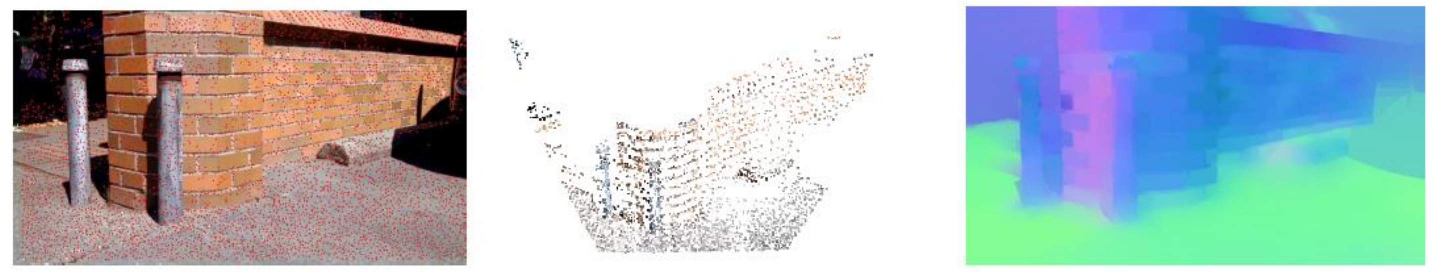
$$E(D) = E_c(D) + \lambda E_g(D)$$

- Color consistency
  - Color term — Neighboring pixel with similar color have similar depth [2,3]

$$E_c(D) = \sum_{p \in I_{ref}} \left(D_p - \sum_{q \in N_8(p)} w_{pq}^c D_q \right)^2$$

$$w_{pq}^c = \frac{1}{N_c} \exp\left(\frac{-|I_p^{lab} - I_q^{lab}|}{\gamma_c} \right)$$

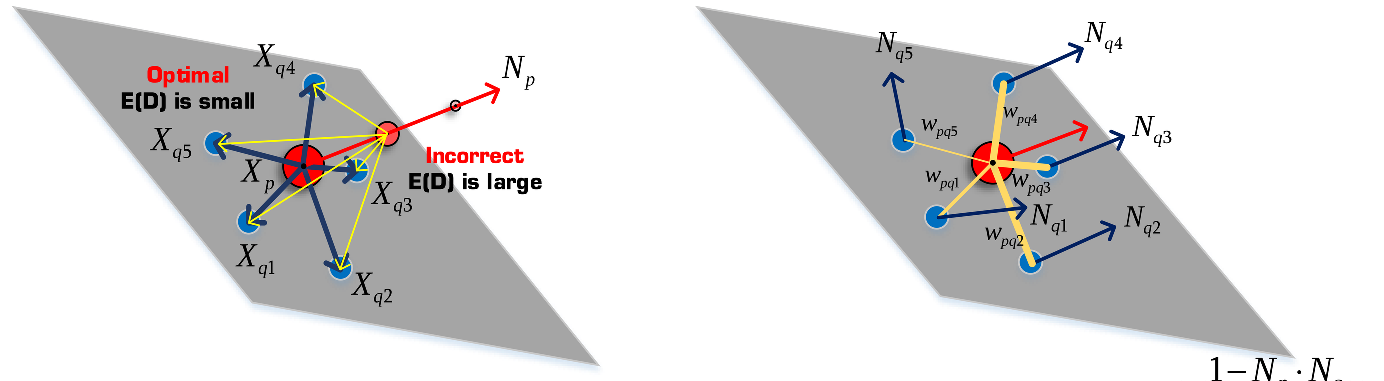
- Normal estimation
  - Using local plane fitting
  - Color-based propagation



Input **Sparse points** **Normal map**

- Geometry consistency
  - Low spatial resolution
  - Normal should be pre-calculated

$$E_g(D) = \sum_{p \in I_{ref}} \sum_{q \in G_w(p)} w_{pq}^g |\mathbf{n}_p \cdot (D_p X_p - D_q X_q))|$$

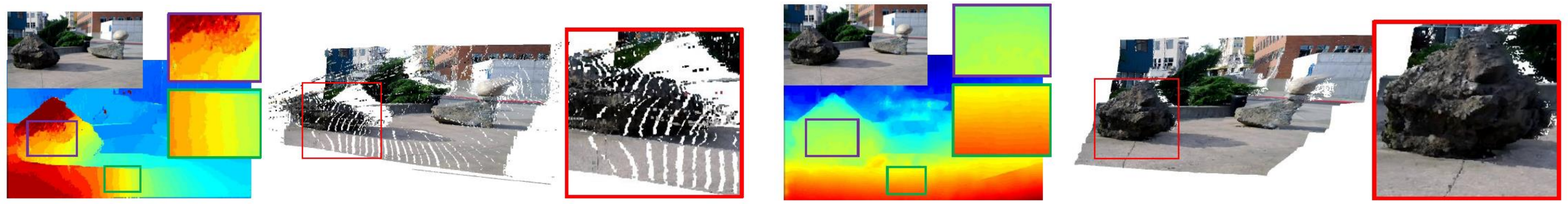


$$(N_p \cdot (X_p - X_q))$$

$$w_{pq} = \exp(-(\frac{1 - N_p \cdot N_q}{\gamma_c}))$$

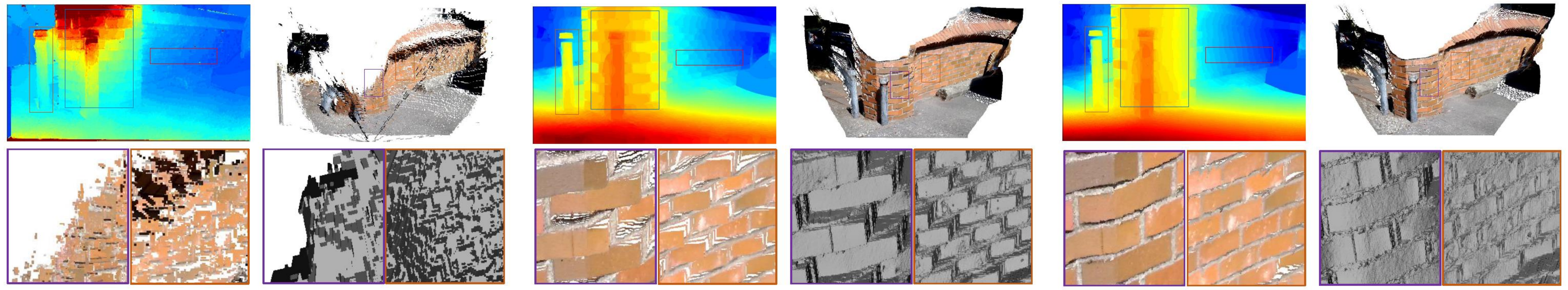
Experiments

- Comparison with state-of-the-art (Captured by Google Galaxy Nexus)



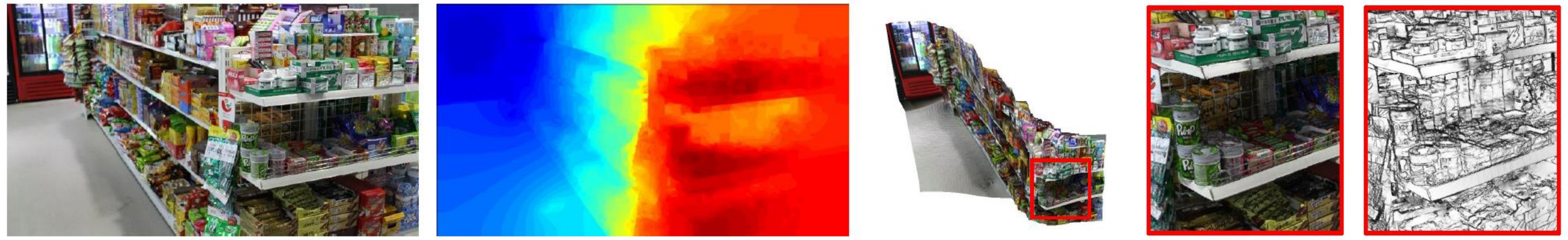
Yu and Gallup [1] **Proposed**

- Comparison with/without Geometry consistency (Captured by Google Galaxy Nexus)



Yu and Gallup [1] **Conventional Propagation [3]** **Proposed Propagation**

- Result of our datasets (Captured by Canon EOS60D)



- Reference

[1] Yu, Fisher, and David Gallup. "3D Reconstruction from Accidental Motion."Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, 2014.

[2] Wang, Liang, and Ruigang Yang. "Global stereo matching leveraged by sparse ground control points." Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011.

[3] Park, Jaesik, et al. "High quality depth map upsampling for 3D-TOF cameras." IEEE International Conference on Computer Vision (ICCV), 2011.

- Acknowledgement

This research is supported by the Study on Imaging Systems for the next generation cameras funded by the Samsung Electronics Co., Ltd (DMC R&D center) (IO130806-00717-02).

- Homepage

https://sites.google.com/site/shimrcv/