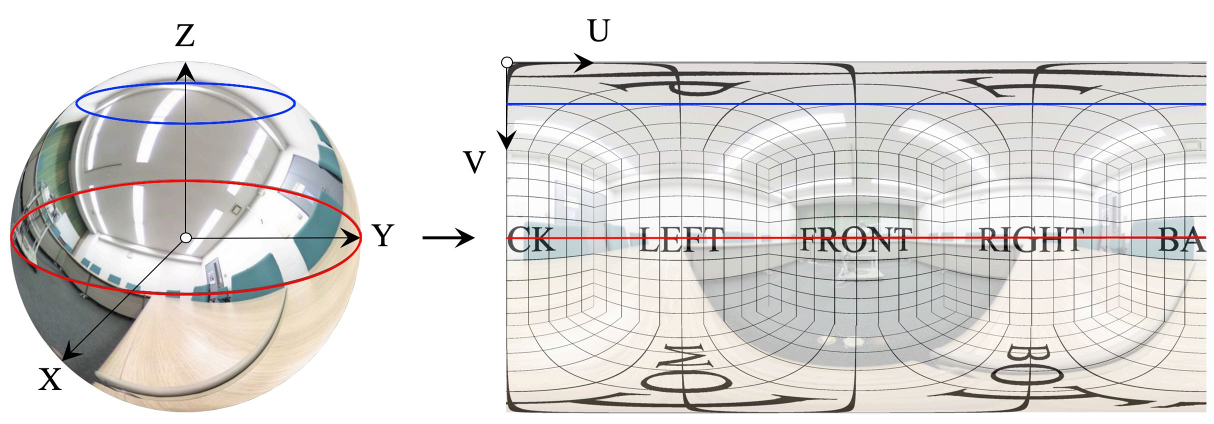
E-CNN: Accurate Spherical Camera Rotation Estimation via Uniformization of Distorted Optical Flow Fields

Dabae Kim, Sarthak Pathak, Alessandro Moro, Ren Komatsu, Atsushi Yamashita, and Hajime Asama
Department of Precision Engineering, The University of Tokyo, Japan
{kimdabae, pathak, moro, komatsu, yamashita, asama}@robot.t.u-tokyo.ac.jp

① Introduction

Spherical Camera (Ricoh Theta S)



Spherical and Equirectangular Images with distortion

- Rotation estimation is an important task for camera-equipped robots (Visual Odometry, Structure from Motion, Visual SLAM)
- **Spherical cameras** are effective for this due to their 360-degree FoV
- **CNNs** estimate robustly against environment variations
- Applying CNNs to spherical images is difficult

→ equirectangular images, which are a distorted planar projection can be used

Distortion remains in equirectangular images → low accuracy

② Motion Estimation by Optical Flow

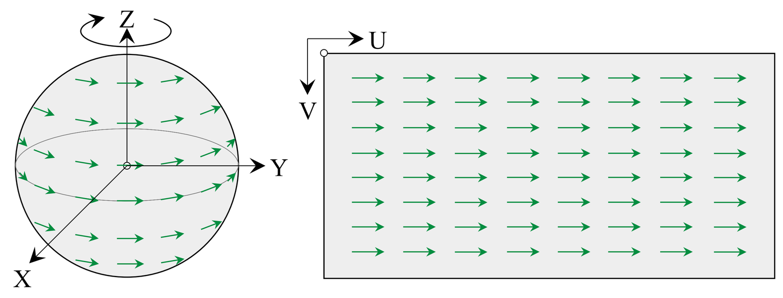
- **Optical Flow**

: Scene-independent vector representation of pixels movement

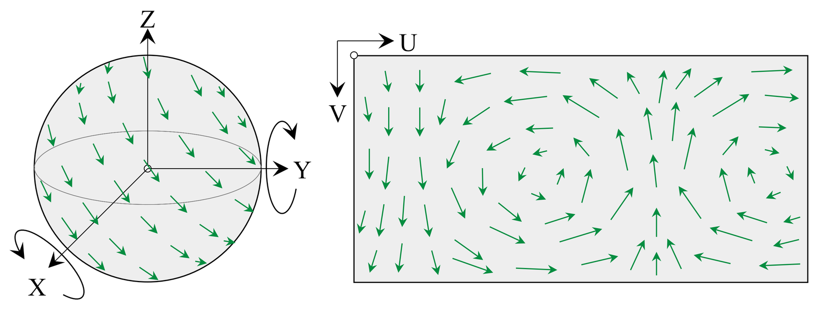
- Different lined/curved patterns appear on equirectangular images even if cameras rotate same quantity around a different axis

→ rotation estimation becomes axis dependent

→ this leads difficulty to learning and interrupt accurate estimation



Lined Optical Flow Patterns



Curved Optical Flow Patterns

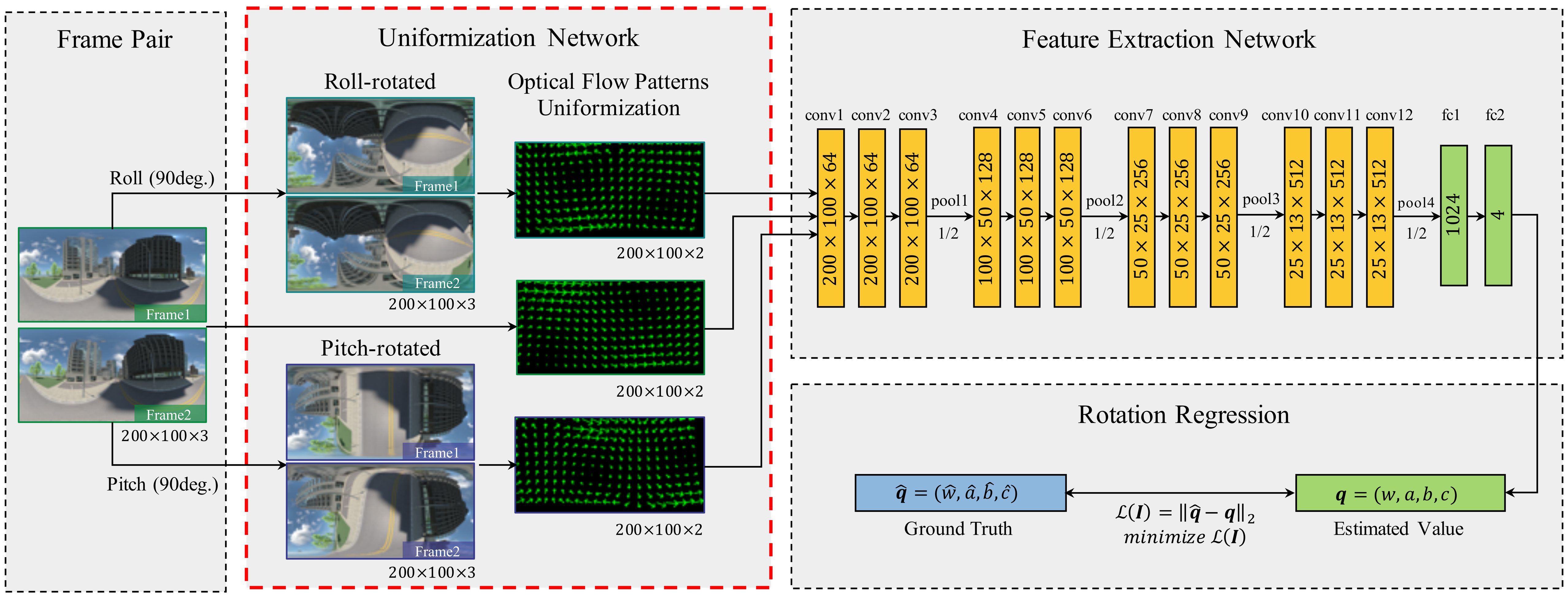
Objective

Accurate spherical camera rotation estimation considering non-uniformity of equirectangular projection

- **E-CNN**: uniformizes non-uniform rotation around all axes

③ Equirectangular-Convolutional Neural Network Structure

- **Uniformization Network**: distorted optical flow patterns uniformization network by spherical image rotation
- **Feature Extraction Network**: feature extraction network by CNN, and rotation regressor



Equirectangular-Convolutional Neural Network (E-CNN) Structure

- **E-CNN process**

(1) takes two image frames as input

(2) rotates them in the directions of roll and pitch by 90 [deg.]

(3) calculates three non-rotated, roll-rotated, and pitch-rotated optical flow

(4) stacks three optical flow as 200×100×6

(5) extracts features by CNN

(6) regresses euclidean distance between ground truth and estimated quaternion

$$\mathcal{L}(\boldsymbol{I}) = \|\hat{\boldsymbol{q}} - \boldsymbol{q}\|_2$$

④ Experimental Verification

- **Datasets**

- different 20 indoor/outdoor real world scenes for training, validation, test datasets
- training data (9,261): 0.5 [deg.] increments in every axis limited to 0~10 [deg.]
- validation (1,000), test data (1,000): randomly chosen limited to 0~10 [deg.]

- **Networks**

- case1(*4lyrs.*), case2(*6lyrs.*), case3(*8lyrs.*) case4(*10lyrs.*), case5(*12lyrs.*)

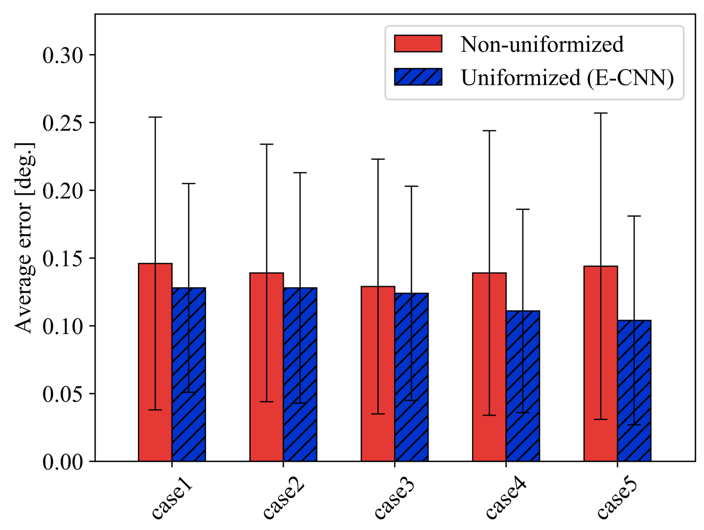
Indoor/Outdoor real world scenes (Ricoh Theta S)

- **Method**

- experiments to verify the effectiveness of uniformization
- same input data size (200×100×6) for a fair comparison

- **Results**

- E-CNN increased the accuracy of estimation by 27.8%
- Higher precision by the lower standard deviation
- E-CNN performed better with increasing network depth whereas the naïve approach did not



Results for Various Network Depths

⑤ Conclusion

- **Conclusion**

- 360 degree FoV property of spherical cameras enables uniformization of distortion
- The effectiveness of our proposed E-CNN was verified in various indoor/outdoor scenes

- **Future works**

- Handling the large angle rotation
- Scale independent rotation estimation
- Simultaneous estimation of rotation and translation