A NOVEL KINECT V2 REGISTRATION METHOD FOR LARGE-DISPLACEMENT ENVIRONMENTS USING CAMERA AND SCENE CONSTRAINTS

Yuan Gao, Sandro Esquivel, Reinhard Koch,

Christian-Albrechts-Universität zu Kiel

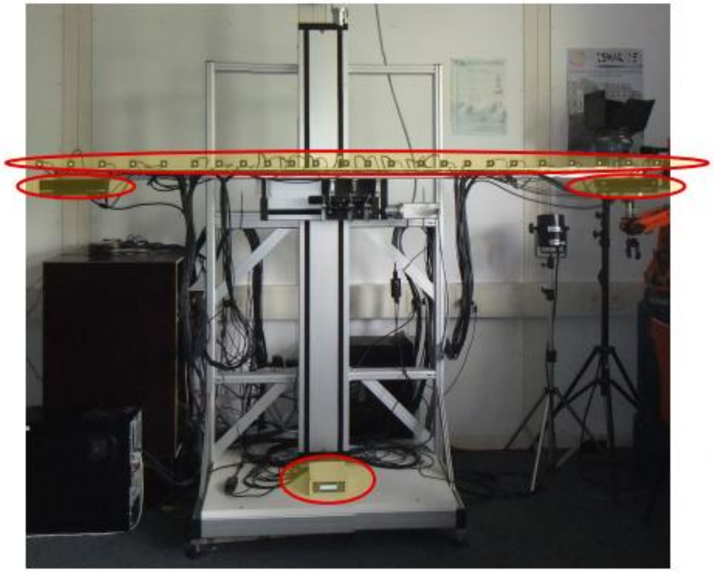
Matthias Ziegler, Frederik Zilly, Joachim Keinert

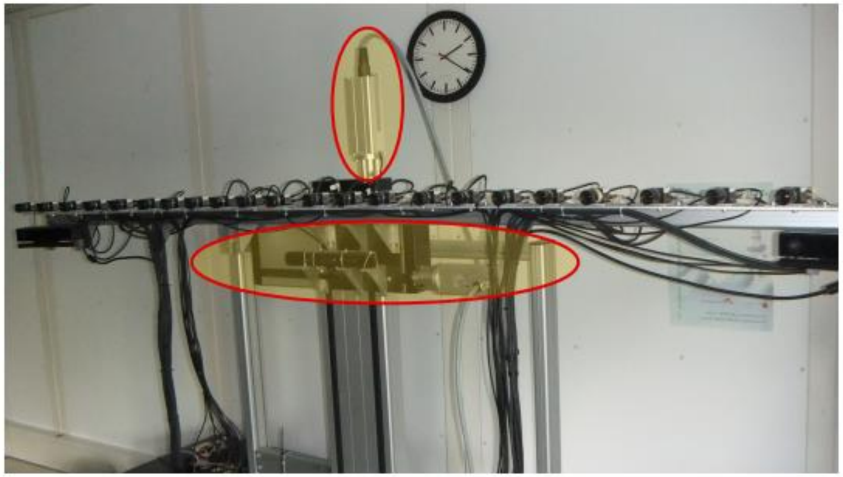
Fraunhofer IIS

Outline

- ❑ **Motivation**
- ❑ Related Work
- ❑ Proposed Method
- ❑ Experiments
- ❑ Conclusion

Motivation





- Multi-camera light field rig
  - 24× IDS USB 3 uEye CP **RGB cameras** at 2 hosts (12 per host)
  - 2× Microsoft **Kinect V2** RGB-D cameras at 1 host
  - 1× Isel iMC-S8 **microstep controller** for 2 linear axes
  - 1× **Hardware trigger** for synchronized uEye camera capture

Motivation

- Kinect V2 cameras
  - **Large displacement**: **2.4** meters
- ToF sensor in a Kinect V2 camera
  - Resolution: **512 × 424** pixels
  - FOV: **71°** × **60°**
- RGB sensor in a Kinect V2 camera
  - Resolution: **1,920 × 1,080** pixels
  - FOV: **84°** × **54°**
- The traditional checkerboard-based calibration method [1] is prone to fail if the checkerboard is not huge enough for being captured.

[1] Zhang, A flexible new technique for camera calibration, TPAMI 2000, vol. 22, no. 11, pp. 1330-1334.

Outline

- ❑Motivation
- ❑**Related Work**
- ❑Proposed Method
- ❑Experiments
- ❑Conclusion

Related Work

- For **non-large-displacement** environment
  - BAICP, ICPR 2014 [2]
  - 3D correspondence-based method, ICMEW 2015 [3]
  - An optical tracking system, 3DUI 2015 [4]
  - OpenPTrack, RAS 2016 [5]
  - Coarse-to-fine framework
    - 3DV 2015 [6]
      - Coarse estimation: Marker (2D)
      - Estimation refinement: Iterative Closest Point (ICP)
    - SCP 2017 [7]
      - Coarse estimation: Wand (1D)
      - Estimation refinement: R-Nearest Neighbor (RNN)

Related Work

[2] Afzal et al., RGB-D multi-view system calibration for full 3D scene reconstruction, ICPR 2014, pp. 2459-2464.

[3] Palasek et al., A flexible calibration method of multiple Kinects for 3D human reconstruction, ICMEW 2015, pp. 1-4.

[4] Beck and Froehlich, Volumetric calibration and registration of multiple RGBD-sensors into a joint coordinate system, 3DUI 2015, pp. 89-96.

[5] Munaro et al., Openptrack: Open source multi-camera calibration and people tracking for RGB-D camera networks, RAS 2016, vol. 75, pp. 525–538.

[6] Kowalski et al., LiveScan3D: A fast and inexpensive 3D data acquisition system for multiple Kinect V2 sensors, 3DV 2015, pp. 318–325.

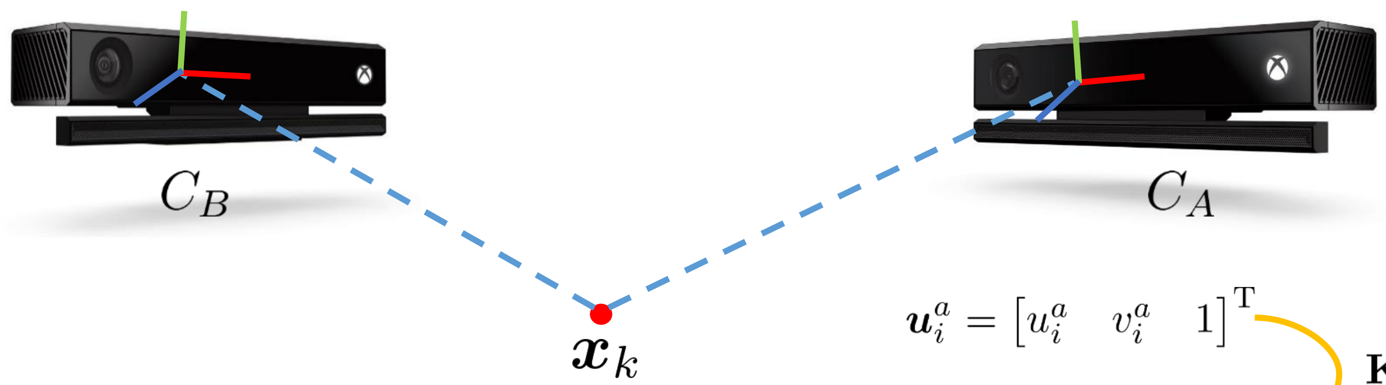
[7] Cordova-Esparza et al., A multiple camera calibration and point cloud fusion tool for Kinect V2, Science of Computer Programming (SCP 2017), vol. 143, pp. 1-8.

Outline

- ❑Motivation
- ❑Related Work
- ❑**Proposed Method**
- ❑Experiments
- ❑Conclusion

Proposed Method

❑ Preliminary



$$\boldsymbol{u}_i^a = \begin{bmatrix} u_i^a & v_i^a & 1 \end{bmatrix}^{\mathrm{T}} \xrightarrow{\mathbf{K}^a,\, z_i^a} \boldsymbol{x}_i^a = \begin{bmatrix} x_i^a & y_i^a & z_i^a & 1 \end{bmatrix}^{\mathrm{T}}$$

$$\boldsymbol{x}_j^b = \begin{bmatrix} \mathbf{R} & \boldsymbol{t} \\ \mathbf{0} & 1 \end{bmatrix} \boldsymbol{x}_i^a = \underbrace{\begin{bmatrix} \mathbf{R}_2 & \boldsymbol{t}_2 \\ \mathbf{0} & 1 \end{bmatrix}}_{\text{Estimation refinement}} \underbrace{\begin{bmatrix} \mathbf{R}_1 & \boldsymbol{t}_1 \\ \mathbf{0} & 1 \end{bmatrix}}_{\text{Coarse estimation}} \boldsymbol{x}_i^a$$

Proposed Method

Stage 1: Coarse estimation

- With calibration objects (**baseline approach**)

$$\min_{\mathbf{R}_1, \boldsymbol{t}_1} \sum_{i=1}^{n} \left\| \boldsymbol{u}_i^b - \hat{\boldsymbol{u}}(\boldsymbol{x}_i^a, \mathbf{K}^b, \mathbf{R}_1, \boldsymbol{t}_1) \right\|^2$$

$$\text{where } \hat{\boldsymbol{u}}(\boldsymbol{x}, \mathbf{K}, \mathbf{R}, \boldsymbol{t}) = \text{proj}(\mathbf{K} \begin{bmatrix} \mathbf{R} & \boldsymbol{t} \end{bmatrix} \boldsymbol{x})$$

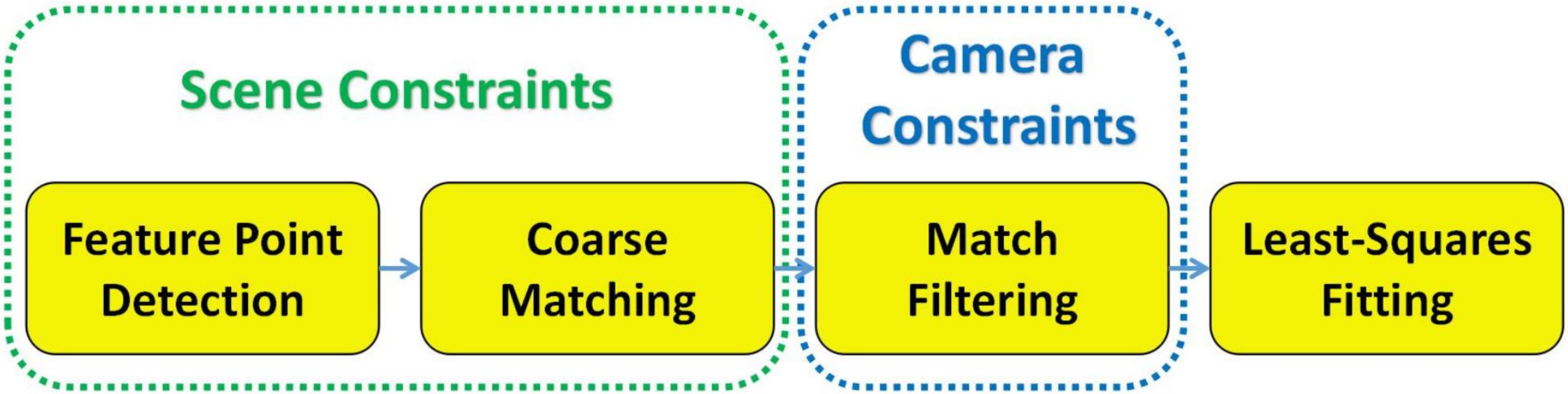
A standard Perspective-*n*-Point (P*n*P) problem

- Levenberg-Marquardt optimization algorithm;
- EPnP [8], RPnP [9], etc.

[8] Lepetit et al., Epnp: An accurate O(*n*) solution to the P*n*P problem, IJCV 2009, vol. 81, no. 2, pp. 155-166.
[9] Li et al., A robust O(*n*) solution to the Perspective-*n*-Point problem, TPAMI 2012, vol. 34, no. 7, pp. 1444-1450.

Proposed Method

- Stage 1: Coarse estimation
  - With camera and scene constraints



SURF

k-Nearest-Neighbors (KNN)
Ratio test

RANSAC

$$\boldsymbol{u}_i^b \sim \mathbf{H}\boldsymbol{u}_i^a$$

$$(\boldsymbol{u}_i^b)^{\mathrm{T}}\mathbf{F}\boldsymbol{u}_i^a = 0$$

$$\min_{\mathbf{R}_1, \boldsymbol{t}_1} \sum_{i=1}^{n} \| \boldsymbol{x}_i^b - \begin{bmatrix} \mathbf{R}_1 & \boldsymbol{t}_1 \\ \mathbf{0} & 1 \end{bmatrix} \boldsymbol{x}_i^a \|^2$$

Proposed Method

Stage 2: Estimation refinement

input : P_A - Point cloud of C_A after the rigid transformation of the coarse estimation stage;
P_B - Point cloud of C_B;
N - Number of point cloud registration iterations;
$\mathbf{R}^a$, $\mathbf{R}^b$ - 3×3 identity matrices;
$\boldsymbol{t}^a$, $\boldsymbol{t}^b$ - 3×1 zero vectors.
output: $\mathbf{R}^a$, $\mathbf{R}^b$, $\boldsymbol{t}^a$, $\boldsymbol{t}^b$.

for $n \leftarrow 1$ **to** N **do**
  $\mathbf{R}, \boldsymbol{t} \leftarrow ICP(P_A, P_B)$;
  for *each point* $\boldsymbol{x}_i^a$ *in* P_A **do**
    $\boldsymbol{x}_i^a \leftarrow \begin{bmatrix} \mathbf{R} & \boldsymbol{t} \\ \mathbf{0} & 1 \end{bmatrix} \boldsymbol{x}_i^a$;
  end
  $\mathbf{R}^a \leftarrow \mathbf{R}\mathbf{R}^a$;
  $\boldsymbol{t}^a \leftarrow \mathbf{R}\boldsymbol{t}^a + \boldsymbol{t}$;
  $\mathbf{R}, \boldsymbol{t} \leftarrow ICP(P_B, P_A)$;
  for *each point* $\boldsymbol{x}_i^b$ *in* P_B **do**
    $\boldsymbol{x}_i^b \leftarrow \begin{bmatrix} \mathbf{R} & \boldsymbol{t} \\ \mathbf{0} & 1 \end{bmatrix} \boldsymbol{x}_i^b$;
  end
  $\mathbf{R}^b \leftarrow \mathbf{R}\mathbf{R}^b$;
  $\boldsymbol{t}^b \leftarrow \mathbf{R}\boldsymbol{t}^b + \boldsymbol{t}$;
end

Algorithm 1: Point cloud registration

Fix P_B
Move P_A

Fix P_A
Move P_B

$$\boldsymbol{x}_k = \begin{bmatrix} \mathbf{R}^a & \boldsymbol{t}^a \\ \mathbf{0} & 1 \end{bmatrix} \boldsymbol{x}_i^a = \begin{bmatrix} \mathbf{R}^b & \boldsymbol{t}^b \\ \mathbf{0} & 1 \end{bmatrix} \boldsymbol{x}_j^b$$

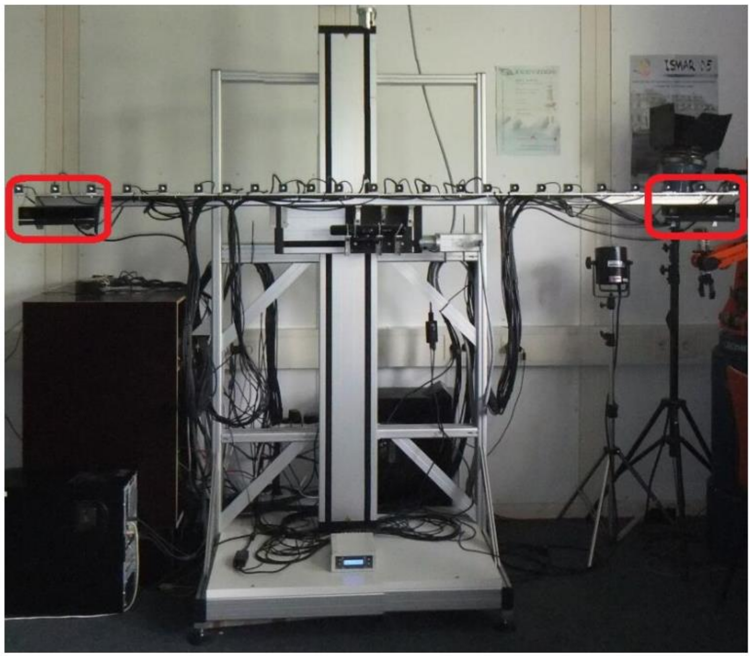
$$\mathbf{R}_2 = (\mathbf{R}^b)^{\mathrm{T}} \mathbf{R}^a$$
$$\boldsymbol{t}_2 = (\mathbf{R}^b)^{\mathrm{T}} (\boldsymbol{t}^a - \boldsymbol{t}^b)$$

Outline

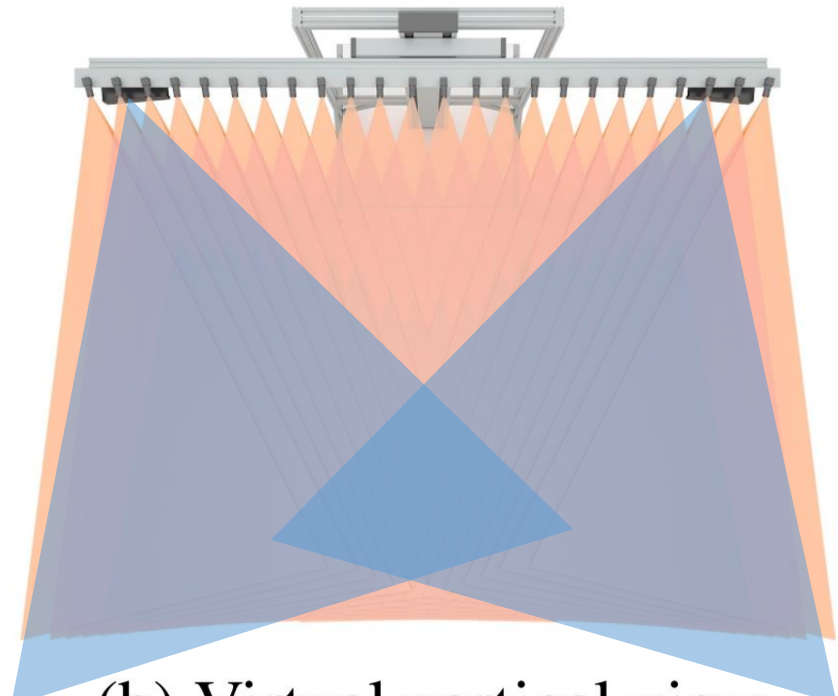
- ❑ Motivation
- ❑ Related Work
- ❑ Proposed Method
- ❑ **Experiments**
- ❑ Conclusion

Experiments

❑ Capture device



(a) Frontal view

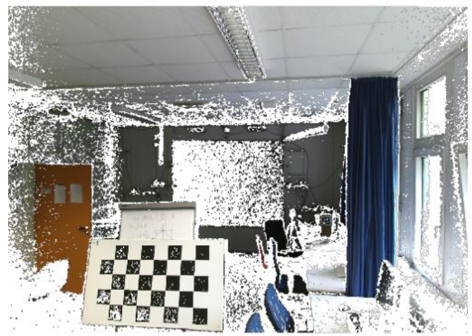


(b) Virtual vertical view

Experiments

❑ Registered color images

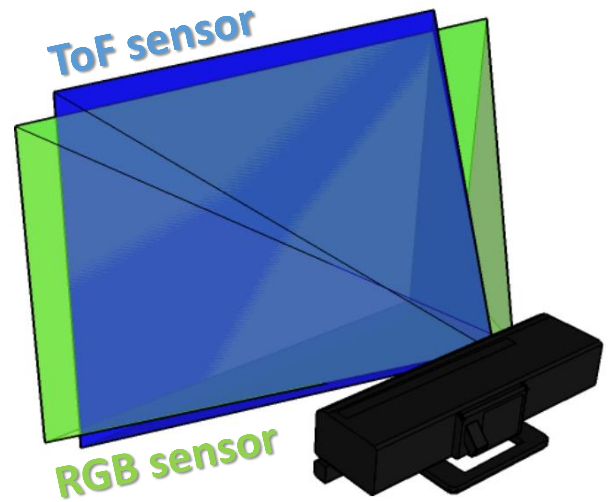
(a) C_A view of the checkerboard



(b) C_B view of the checkerboard

(c) C_A view of the scene

(d) C_B view of the scene



Experiments

Experimental Settings

- Checkerboard-captured data
  - In front of the rig at a distance of **2.8** m;
  - **28** (4 × 7) inner corners;
  - Square size: 124 × 124 mm.
- Scene-captured data
  - Captured room size: 5.5 × 3.0 × **7.8** m (w × h × d)
- Evaluation metric
  - Root-Mean-Square Error (**RMSE**)

$$\sqrt{\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{x}_i^b - \begin{bmatrix}\mathbf{R} & \boldsymbol{t}\\ \mathbf{0} & 1\end{bmatrix}\boldsymbol{x}_i^a\|^2}$$

- Others
  - *libfreenect2* -> **registered** color and depth image

Experiments

Quantitative evaluation

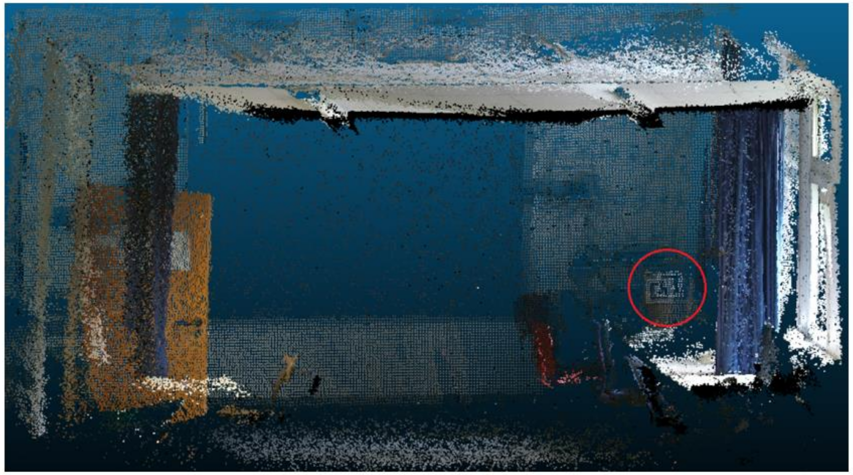
Table I. RMSE of the coarse-to-fine registration methods.

Coarse Estimation	RMSE (mm)	Estimation Refinement	RMSE (mm)
Checkerboard-based	**78.33**	ICP-based point cloud registration	84.11
Homography matrix-based	302.05	ICP-based point cloud registration	44.82
Fundamental matrix-based	295.58	ICP-based point cloud registration	**34.34**

- In the coarse estimation stage,
  - checkerboard-based method achieves much more precise results than the homography and fundamental matrices-based methods.
- In the estimation refinement phase,
  - the precision of the checkerboard-based method **decreases a little bit**;
  - the precision of coarse estimation methods using camera and scene constraints **improves dramatically**.

Experiments

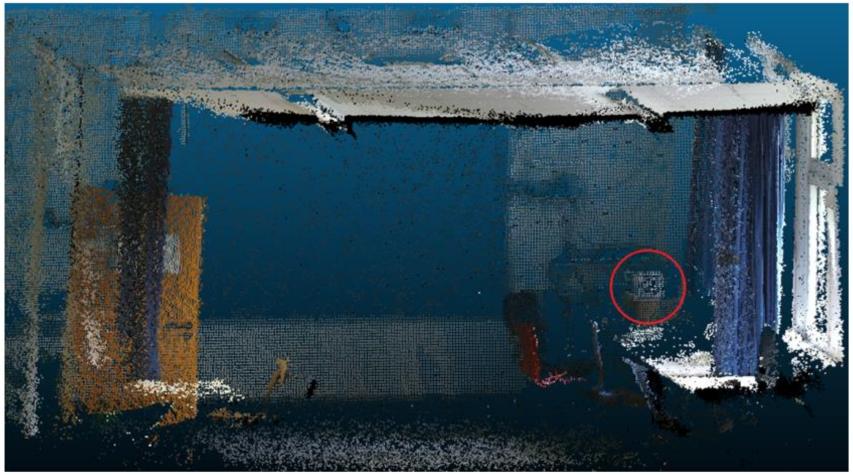
❑Qualitative evaluation



(a) Checkerboard-based coarse estimation with estimation refinement

(b)

Checkerboard-based coarse-to-fine method



(c) Fundamental matrix-based coarse estimation with estimation refinement

(d)

Fundamental matrix-based coarse-to-fine method

Outline

- ❑Motivation
- ❑Related Work
- ❑Proposed Method
- ❑Experiments
- ❑**Conclusion**

Conclusion

- Camera and scene constraints are exploited inside a **coarse-to-fine framework** to solve the Kinect V2 registration problem in the **large-displacement** environment;
- The **fundamental matrix-based** coarse-to-fine registration method outperforms the checkerboard-based coarse-to-fine registration method on a multi-camera rig with a large displacement between two Kinect V2 sensors.

Marie Skłodowska-Curie Actions

DFG Deutsche Forschungsgemeinschaft

intel VISUAL COMPUTING INSTITUTE

Fraunhofer IIS

MAX-PLANCK-GESELLSCHAFT

Thank you!