Department of Computer Science
Kiel University, Germany

Fraunhofer IIS

A Linear Method for Recovering the Depth of Ultra HD Cameras Using a Kinect V2 Sensor

Yuan Gao, Matthias Ziegler, Frederik Zilly, Sandro Esquivel, Reinhard Koch

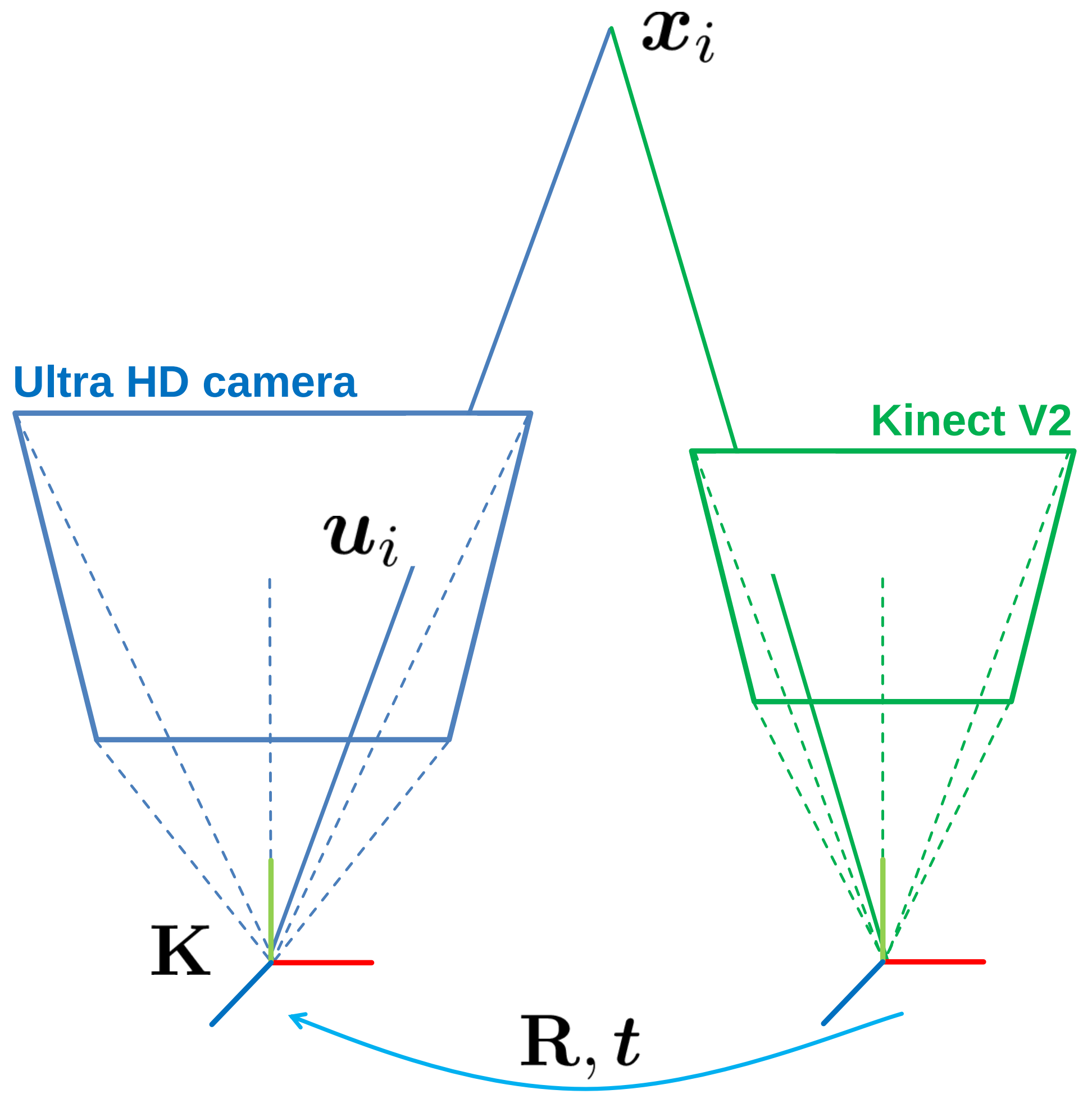
Motivation

- A lot of Ultra HD resolution 3D TVs and high resolution Virtual Reality (VR) headsets have been launched into the market. How to create high-resolution and high-quality contents for these devices is a research hotspot;
- Depth-Image-Based Rendering (DIBR) is a common method for creating free-viewpoint videos. It relies on accurate depth images.

Preliminary

- Prepare a Kinect V2 camera and an Ultra HD camera;
- Intrinsic parameters of these two cameras are known and lens distortions of them are compensated;
- The optical axes of these cameras are approximately parallel;
- Corresponding point pairs are established.

Linear Method



- A 3D point in Kinect V2 camera coordinates is given as:

$$\boldsymbol{x}_i = \begin{bmatrix} x_i & y_i & z_i \end{bmatrix}^{\mathrm{T}}$$

- The corresponding point on the Ultra HD camera image plane is denoted by:

$$\boldsymbol{u}_i = \begin{bmatrix} u_i & v_i & 1 \end{bmatrix}^{\mathrm{T}}$$

- The rigid transformation from the Kinect V2 camera to the UHD camera can be described as:

$$\mathbf{K}(\mathbf{R}\boldsymbol{x}_i + \boldsymbol{t}) = \lambda \boldsymbol{u}_i$$

- To simplify the above equation, an image point on the normalized image plane of the UHD camera is utilized:

$$\boldsymbol{p}_i = \begin{bmatrix} p_i & q_i & 1 \end{bmatrix}^{\mathrm{T}} = \mathbf{K}^{-1}\boldsymbol{u}_i$$

- The rigid transformation formula is simplified as:

$$\mathbf{R}\boldsymbol{x}_i + \boldsymbol{t} = \lambda \boldsymbol{p}_i$$

- The scaling factor can then be derived as:

$$\lambda = \begin{bmatrix} r_{31} & r_{32} & r_{33} \end{bmatrix} \boldsymbol{x}_i + t_3$$

- The rigid transformation formula is finally written as:

$$\begin{bmatrix} \boldsymbol{x}_i^{\mathrm{T}} & \mathbf{0}^{\mathrm{T}} & -p_i\boldsymbol{x}_i^{\mathrm{T}} & 1 & 0 & -p_i \\ \mathbf{0}^{\mathrm{T}} & \boldsymbol{x}_i^{\mathrm{T}} & -q_i\boldsymbol{x}_i^{\mathrm{T}} & 0 & 1 & -q_i \end{bmatrix} \boldsymbol{h} = \mathbf{0}$$

where

$$\boldsymbol{h} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{33} & t_1 & t_2 & t_3 \end{bmatrix}^{\mathrm{T}}$$

- The above homogeneous system of linear equations has a least-squares solution. At least six corresponding point pairs and Singular Value Decomposition (SVD) are used to solve the depth registration problem.

Setup

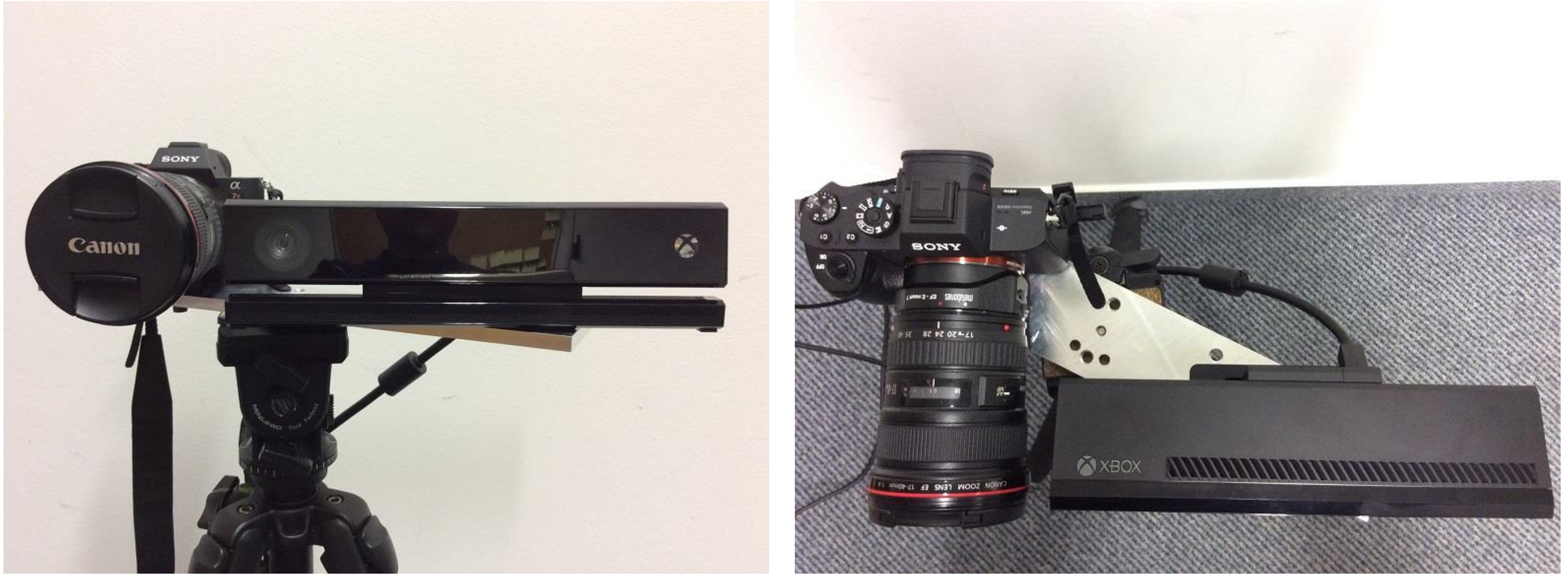


Figure 1. Cameras in our system.

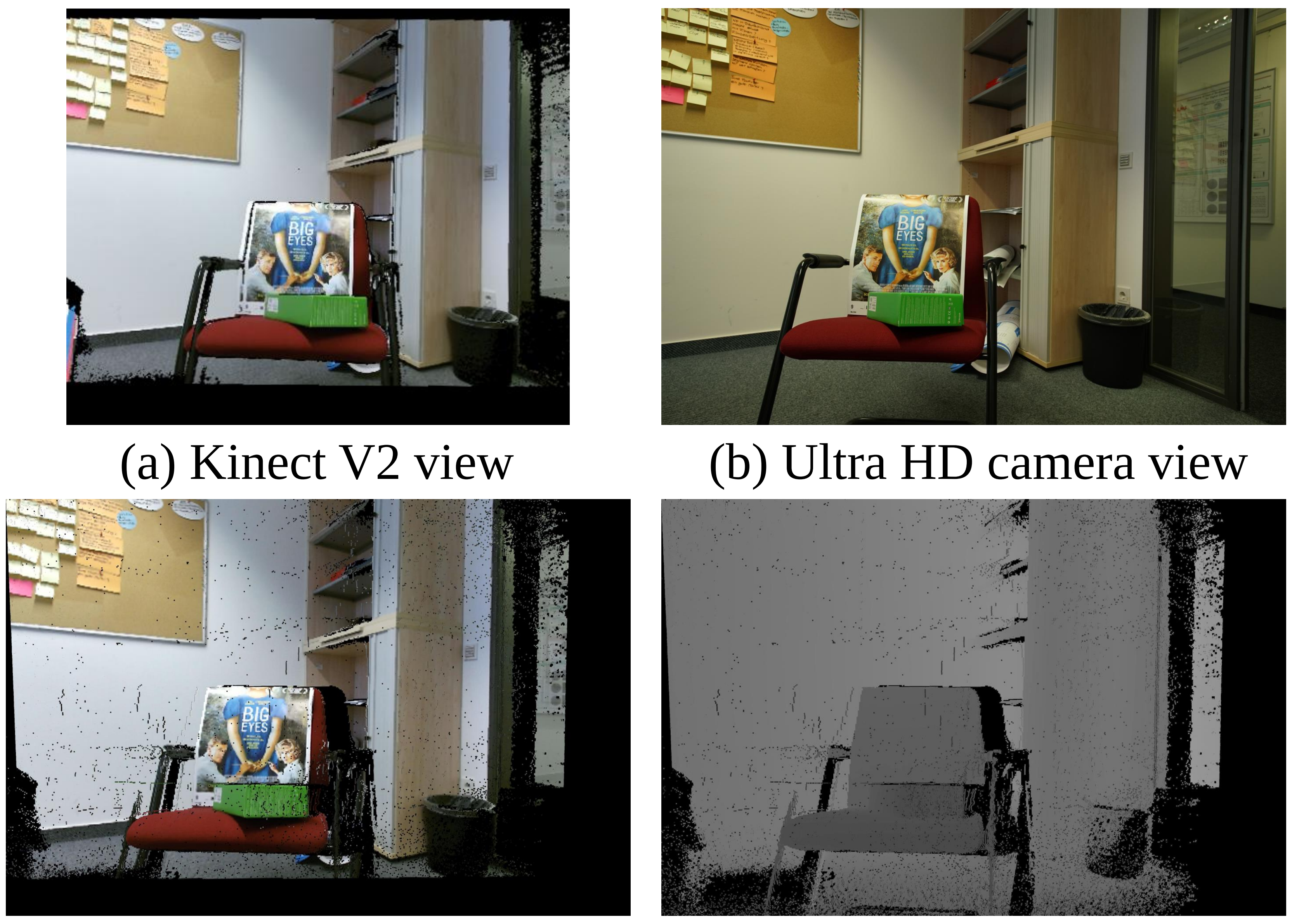
- A Kinect V2 sensor and a Sony α7R II DSLR camera with a Canon lens constitute the system;
- A checkerboard with 54 (9 × 6) corners is placed in front of the system twice with different poses;
- An oversampling strategy in DIBR is employed here, *i.e.* the depth image of the Kinect V2 is oversampled by a factor of 10 via the Nearest-Neighbor method;
- The Root-Mean-Square Error (RMSE) is adopted to evaluate the registration effects of the proposed method.

Experimental Results

Table I. The RMSE and computing time of the proposed registration method and two baseline calibration approaches.

Method	RMSE (pixel)	Time (ms)
Linear Method	0.781	1.2
Non-linear Method (Coarse)	2.045	1.6
Non-linear Method (Refine)	0.993	16.0

- The proposed linear method outperforms the coarse-to-fine non-linear method in precision;
- The linear method is around 14 times faster in speed.



(a) Kinect V2 view (b) Ultra HD camera view

(c) Recovered color (d) Recovered depth

Figure 2. Qualitative evaluation by visualization.

E-mail: *yga@informatik.uni-kiel.de*

IAPR International Conference on Machine Vision Applications, May 8-12, 2017, Nagoya, Japan