Department of Computer Science
Kiel University, Germany

Fraunhofer IIS

Marie Skłodowska-Curie Actions

VISUAL COMPUTING INSTITUTE

DFG Deutsche Forschungsgemeinschaft

max planck institut informatik

A Novel Self-Calibration Method for a Stereo-ToF System Using a Kinect V2 and Two 4K GoPro Cameras

Yuan Gao*, Sandro Esquivel, Reinhard Koch, Joachim Keinert

Motivation

Fig. 1. The multi-camera rig.

- The Kinect V2 is impossible to be fixed due to the changeable tilt, which affects the extrinsic parameters;
- Little literature focuses on the stereo- and ToF-depth fusion problem in 4K resolution.

Depth Correction

Analysis of [1] has shown that typically the Kinect V2 has a depth bias of **-18** mm. Correcting for the bias improves the result.

[1] O. Wasenmüller and D. Stricker. Comparison of Kinect V1 and V2 depth images in terms of accuracy and precision. In ACCVW, pages 34–45, 2016.

Self-Calibration

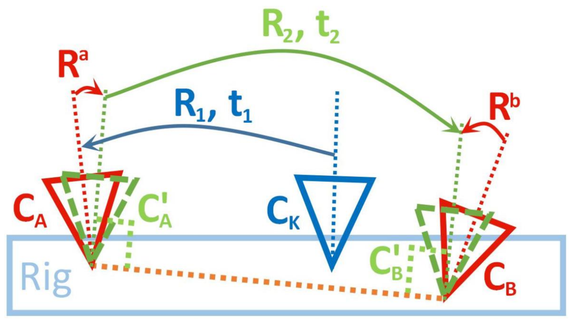


Fig. 2. A virtual vertical view of the rig.

Depth Fusion

In the camera image space of $\mathbb{C}'_{\mathrm{A}}$:

- $\mathbf{D}'_k$ is projected from the Kinect V2 sensor;
- $\mathbf{D}'_s$ is the depth result of using the stereo matching method.

$$\mathbf{D}'_f(i,\ j) = \begin{cases} \mathbf{D}'_k(i,\ j), & \text{if } \mathbf{D}'_k(i,\ j) > 0\ ; \\ \mathbf{D}'_s(i,\ j), & \text{else.} \end{cases}$$

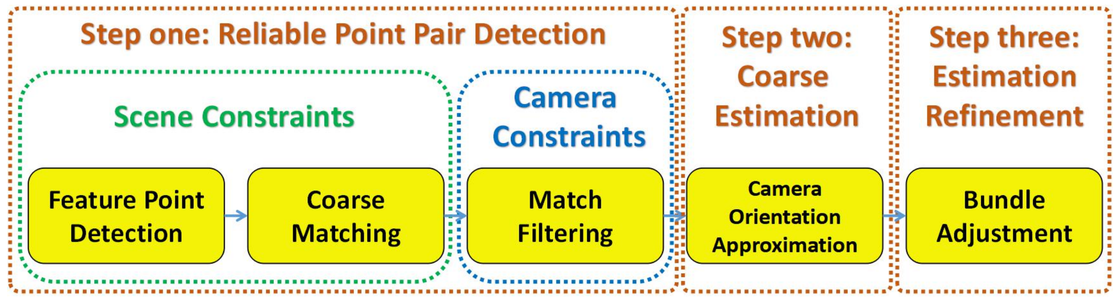


Fig. 3. The flow diagram of the self-calibration process.

Notation

A 2D point of $\mathbb{C}_{\mathrm{A}}$: $\boldsymbol{u}_i^a = [u_i^a \quad v_i^a \quad 1]^{\mathrm{T}}$
A 2D point of $\mathbb{C}_{\mathrm{K}}$: $\boldsymbol{u}_i^k = [u_i^k \quad v_i^k \quad 1]^{\mathrm{T}}$
A 3D point of $\mathbb{C}_{\mathrm{K}}$: $\boldsymbol{x}_i^k = [x_i^k \quad y_i^k \quad z_i^k \quad 1]^{\mathrm{T}}$
Camera intrinsic parameters: $\mathbf{K}_a, \mathbf{K}_k$
A 2D point of $\mathbb{C}_{\mathrm{A}}$ on the normalized image plane: $\boldsymbol{p}_i^a = [p_i^a \quad q_i^a \quad 1]^{\mathrm{T}}$

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

$$\boldsymbol{r} = [\alpha \quad \beta \quad \gamma]^{\mathrm{T}} \quad \min_{\mathbf{R}_1^j, \boldsymbol{t}_1^j, \boldsymbol{x}_i^k} \sum_{j=1}^{2} \sum_{i=1}^{n} \|\boldsymbol{u}_i^j - \hat{\boldsymbol{u}}(\boldsymbol{x}_i^k, \mathbf{K}_j, \mathbf{R}_1^j, \boldsymbol{t}_1^j)\|^2$$

$$\mathbf{R}_1^2 = \begin{bmatrix} 1 & -\gamma & \beta \\ \gamma & 1 & -\alpha \\ -\beta & \alpha & 1 \end{bmatrix}$$

$$\mathbf{A} \begin{bmatrix} \boldsymbol{r} \\ \boldsymbol{t}_1^2 \end{bmatrix} = \begin{bmatrix} p_i^a z_i^k - x_i^k \\ q_i^a z_i^k - y_i^k \end{bmatrix}$$

$$\mathbf{R}_1 = \mathbf{R}_1^2 (\mathbf{R}_1^1)^{\mathrm{T}}$$
$$\boldsymbol{t}_1 = \boldsymbol{t}_1^2 - \mathbf{R}_1 \boldsymbol{t}_1^1$$

$$\mathbf{A} = \begin{bmatrix} -p_i^a y_i^k & (p_i^a x_i^k + z_i^k) & -y_i^k & 1 & 0 & -p_i^a \\ -(q_i^a y_i^k + z_i^k) & q_i^a x_i^k & x_i^k & 0 & 1 & -q_i^a \end{bmatrix}$$

Experiments

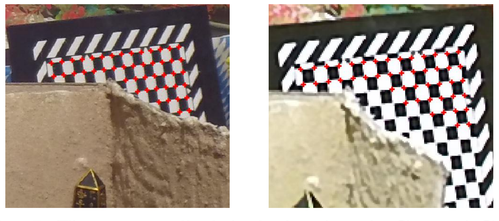


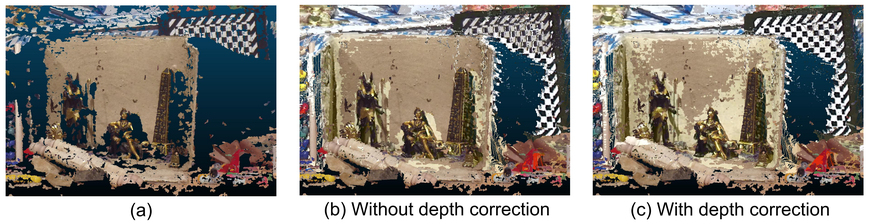
Fig. 5. The manually-labeled points in $\mathbb{C}_{\mathrm{A}}$ and $\mathbb{C}_{\mathrm{K}}$.

Table I. The RMSE results of the self-calibration.

Coarse-to-Fine Framework	LHM	EPnP	RPnP	Ours
	Before depth correction			
Coarse estimation	1.82	3.14	1.45	0.87
Estimation Refinement	-	2.19	-	-
	After depth correction			
Coarse estimation	1.78	2.83	1.54	**0.87**
Estimation refinement	-	2.60	-	-

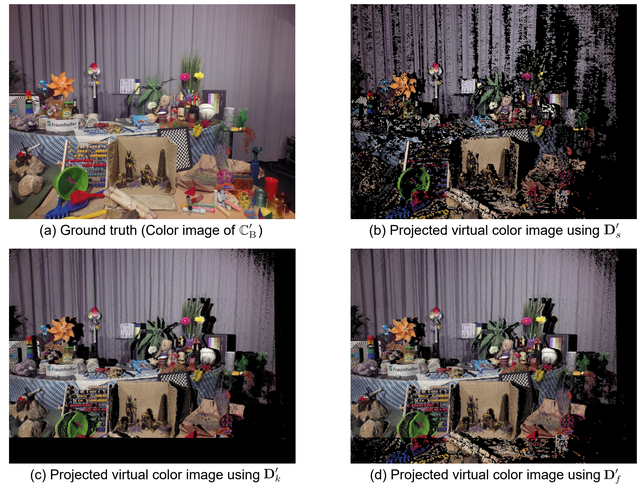
Table II. The performances of different depth sources for the rendering of virtual views using different evaluation metrics.

Depth Source	Before depth correction			After depth correction		
	PSNR	SSIM	NBRP (%)	PSNR	SSIM	NBRP (%)
$\mathbf{D}'_s$	9.29	0.156	61.29	-	-	-
$\mathbf{D}'_k$	11.29	0.235	74.50	11.35	0.245	74.28
$\mathbf{D}'_f$	11.98	0.247	**81.35**	**12.04**	**0.259**	81.15



(a) (b) Without depth correction (c) With depth correction

Fig. 6. The point clouds of a part of the scene in the camera 3D space of $\mathbb{C}'_{\mathrm{A}}$.



(a) Ground truth (Color image of $\mathbb{C}'_{\mathrm{B}}$) (b) Projected virtual color image using $\mathbf{D}'_s$
(c) Projected virtual color image using $\mathbf{D}'_k$ (d) Projected virtual color image using $\mathbf{D}'_f$

Fig. 7. Projected virtual color views on the camera image plane of $\mathbb{C}'_{\mathrm{B}}$.

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

International Conference on 3D Vision, October 10th-12th 2017, Qingdao, China