Data Article

A multi-camera dataset for depth estimation in an indoor scenario

Giulio Marin, Gianluca Agresti, Ludovico Minto, Pietro Zanuttigh[*]

University of Padova, Italy

A B S T R A C T

Time-of-Flight (ToF) sensors and stereo vision systems are two of the most diffused depth acquisition devices for commercial and industrial applications. They share complementary strengths and weaknesses. For this reason, the combination of data acquired from these devices can improve the final depth estimation accuracy. This paper introduces a dataset acquired with a multi-camera system composed by a Microsoft Kinect v2 ToF sensor, an Intel RealSense R200 active stereo sensor and a Stereolabs ZED passive stereo camera system. The acquired scenes include indoor settings with different external lighting conditions. The depth ground truth has been acquired for each scene of the dataset using a line laser. The data can be used for developing fusion and denoising algorithms for depth estimation and test with different lighting conditions. A subset of the data has already been used for the experimental evaluation of the work "Stereo and ToF Data Fusion by Learning from Synthetic Data".

© 2019 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

---

* Corresponding author.
E-mail address: zanuttigh@dei.unipd.it (P. Zanuttigh).

Specifications Table

Subject area	*Computer Science*
More specific subject area	*Computer Vision*
Type of data	*Color, Infra-Red (IR) images and depth maps*
How data was acquired	*Using a Stereolabs ZED stereo vision system, an Intel RealSense R200 active stereo system and a Microsoft Kinect v2 Time-of-Flight camera*
Data format	*Raw data and calibration*
Experimental factors	*Color, IR images and depth maps acquired in an indoor static setting with controlled lighting.*
Experimental features	*Acquisitions with a multi-camera setup made of an active and a passive stereo system and a ToF sensor. Ground truth data acquired with a line laser.*
Data source location	*Palo Alto, CA.*
Data accessibility	*The research data are available at* http://lttm.dei.unipd.it/paper_data/real3ext
Related research article	Agresti, G., Minto, L., Marin, G., & Zanuttigh, P. (2019). Stereo and ToF Data Fusion by Learning from Synthetic Data. *Elsevier Information Fusion.*

Value of the Data
- Multiple acquisitions of the same scene with different sensors are provided: we provide 10 real world static scenes acquired with an active IR stereo camera, a passive stereo camera and a ToF sensor.
- Each scene comes with 4 different lighting conditions allowing to evaluate how the illumination affects depth estimation algorithms.
- Additionally, the ToF amplitude and color data from the depth sensors are also provided allowing to exploit them in depth refinement algorithms.
- Accurate depth ground truth, acquired with a line laser, is provided for all the scenes (only a very few datasets have ToF depth with ground truth, this allows to evaluate the algorithms and machine learning based approaches).
- Accurate calibration information allows to reproject the data to a common reference system, a fundamental step for being able to test data fusion strategies.
- The data can be used by researchers working on both depth estimation and data fusion. Part of the data has been used in previous research works, allowing to perform experimental comparison.
- Suitable for testing fusion and denoising algorithms for depth estimation in different lighting conditions.

1. Data

In this paper we introduce the REAL3EXT dataset, an extended version of the REAL3 dataset used for the experimental evaluation of [1]. The proposed dataset is a collection of static real world acquisitions recorded in an indoor setting with a Microsoft Kinect v2 ToF sensor [2,3,7], a ZED passive stereo system [6] and a RealSense R200 active stereo system [4]. The depth ground truth computed from to the viewpoint of the left camera of the ZED stereo system is also provided for each scene. The three sensors are geometrically calibrated for intrinsic and extrinsic parameters.

The dataset is made of 10 unique indoor scenes under different lighting conditions. The additional acquisitions under various illumination conditions, the two extra scenes and the addition of the RealSense R200 data is what differentiates REAL3EXT from REAL3 [1].

2. Experimental design, materials, and methods

The multi camera acquisition system is arranged as in Fig. 1. The reference system has the ZED camera in the center, underneath the ZED there is the Kinect and above there is the RealSense R200. The three cameras are kept in place by a plastic mount specifically designed to fit them. The depth camera of the Kinect is approximately horizontally aligned with the left camera of the ZED with 40 mm vertical displacement, while the color camera is approximately in between the passive stereo pair. The RealSense R200 is placed approximately 20 mm above the ZED camera, with the two IR and color camera inside the baseline of the passive stereo pair.

The ZED stereo camera from Stereolabs is made of two synchronized color cameras that can acquire images at different resolutions and we provide images at 1080p, that is 1920×1080 [pxl]. The SDK

Fig. 1. Multi-sensors arrangement.

provided with the camera also includes a stereo vision algorithm optimized for real time GPU processing, but we decided to provide only the raw data from the two cameras (and the calibration parameters) that can be then processed with any stereo vision algorithm. This follows the rationale of providing raw data only and leaving the selection of the specific implementation to the user. The baseline between the 2 cameras is 12 [cm] and their horizontal field of view is 110°.

The Kinect v2 can acquire the ToF depth amplitude images at a resolution of 512 × 424 [pxl] and 92° horizontal field of view. Moreover, this device is equipped with a color camera with resolution of 1920 × 1080 [pxl] and horizontal field of view of 85°.

The RealSense R200 is equipped with an IR projector and 2 IR cameras capable of acquiring gray-scale images at a resolution of 640 × 480 [pxl] with a horizontal field of view of 56°. The relative baseline between the cameras is 7 [cm]. The baseline between the IR projector and the right camera is 2 [cm]. The RealSense R200 has also a color camera with resolution 1920 × 1080 [pxl] and a field of view of 69°.
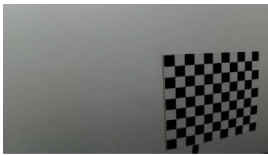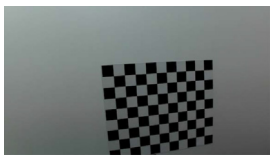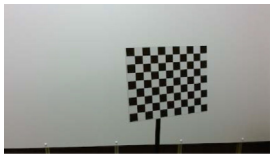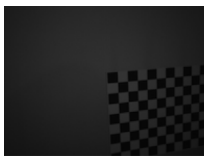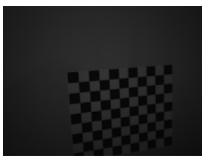

Fig. 2. Images of the same checkerboard acquired from the three sensors during the calibration process.
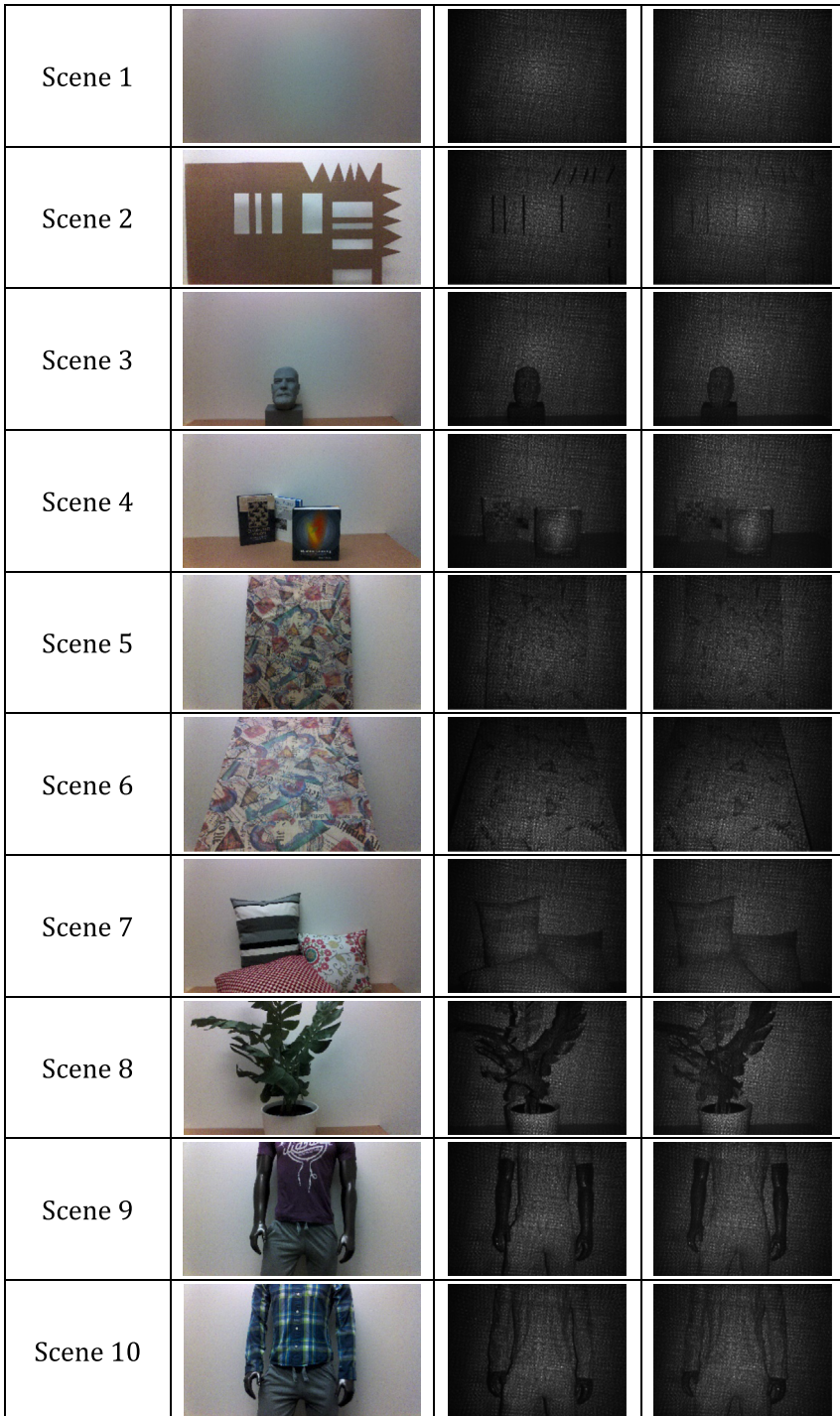
Fig. 3. Color and IR images acquired by the RealSense R200 on the considered 10 scenes.

We provide the geometric calibration for each sensor, which are the intrinsic and extrinsic parameters. We followed the approach of Zhang [5] for camera calibration with a regular black and white checkerboard. Differently from the single camera calibration, a setup including 7 different physical cameras is more complicated to deal with due to different fields of view, resolutions and nature of the imaging system, color, IR, depth or amplitude signal. Different strategies have been proposed for joint calibration of depth and color cameras [9–11].

In this work we used the following procedure. First of all, to avoid misalignment and motion blur we used a tripod for the checkerboard and delayed the acquisition. We also collected 20 images of the same calibration scene for each acquisition, and averaged the images to reduce the noise. To calibrate both the color cameras and the IR cameras at the same time we used a uniform illumination with incandescent light bulbs. Fig. 2 shows some samples of the performed acquisitions. With all the collected image we ran a checkerboard detector obtaining for each camera and for each pose a set of points that are then used in the non-linear optimization to estimate the calibration parameters.

Ground truth depth data come from a non-calibrated line laser used to generate a detailed depth map using the ZED stereo camera. The red illuminator of the laser projects a line in the scene and for each acquisition from the two color cameras we matched the corresponding lit points and compute a disparity value for each pixel. To reduce distortion due to noise and other artifacts, we accumulated multiple acquisitions and stored the median value among the different disparity estimates. The ground truth acquisitions have been carried out without external illumination to increase the contrast of the line. The line laser was kept as close as possible to the cameras to avoid occlusions and illumination artifacts.

The subjects of the 10 scenes in the REAL3EXT dataset try to stress various flaws of the stereo and ToF systems. Critical points are for example lack of texture for the passive stereo system and the presence of Multi-Path interference [12], low reflectivity elements and external illumination for the active sensors [8]. The scenes are composed by flat surfaces with and without textures, plants and objects of various material such as plastic, paper and cotton fabric. These are characterized by various specularity properties as reflective and glossy surfaces and rough materials. The color and IR views acquired by the RealSense R200 on the considered 10 scenes can be found in Fig. 3.
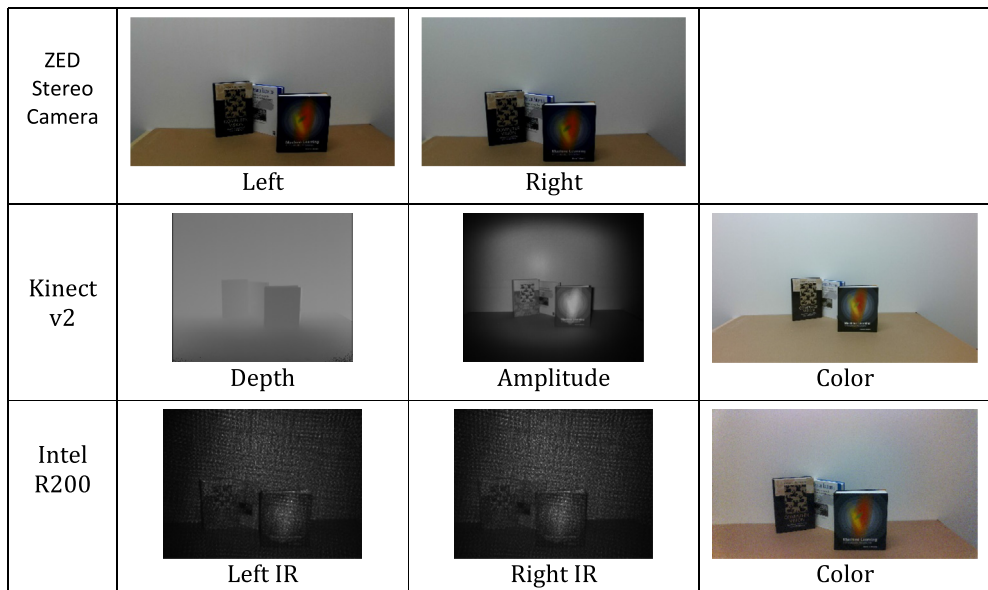

ZED Stereo Camera	Left	Right	
Kinect v2	Depth	Amplitude	Color
Intel R200	Left IR	Right IR	Color

Fig. 4. Data acquired with the considered acquisition system. The left and the right views are provided for the ZED system. The Kinect v2 data are the RGB image, the depth map and the amplitude image related to the ToF acquisition. The RealSense R200 data are the RGB image and the left and the right IR views of its stereo system.
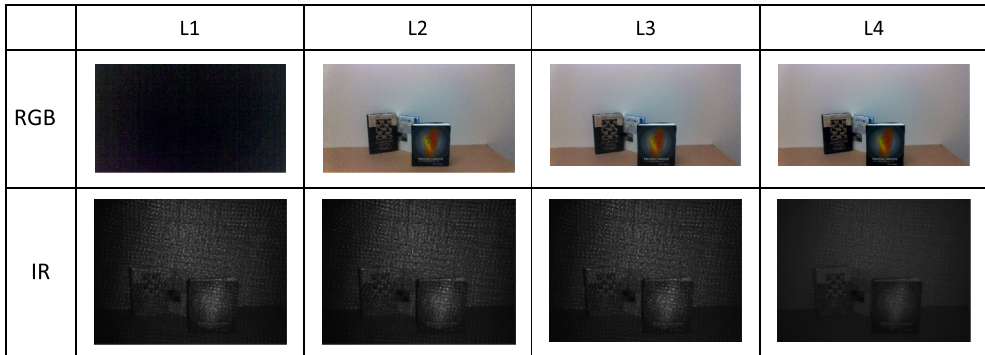
Fig. 5. RGB and IR images acquired respectively with the color and left IR cameras of the RealSense R200 device. These recordings are acquired under 4 different external illumination conditions. L1 stands for "*no external light*"; L2 stands for "*regular lighting*"; L3 stands for "*stronger light*"; L4 stands for "*use of an additional incandescent light source*".

Each scene was recorded under 4 different external lighting conditions, which are the following: with no external light; with regular lighting; with stronger lighting; with an additional incandescent light source.

Each lighting condition can highlight the weakness and strength of the different depth estimation algorithms. We added the acquisitions with the additional incandescent light source since its spectrum, in the IR wavelength, covers the working range of the active depth cameras and this is a known problem for those devices.

Fig. 4 depicts all the data collected for a scene contained in REAL3EXT from all the acquisition systems.

Fig. 5 shows the left IR images and the RGB images acquired with the RealSense R200 under the considered light conditions. From this figure it is possible to notice that the pattern projected by the RealSense R200 is clearly visible from the left IR image in the first 3 external lighting conditions. When the incandescent light source is used in the last column, the pattern starts to fade although there is no clear difference in the visible spectrum (color image).

The dataset is available at http://lttm.dei.unipd.it/paper_data/real3ext.

Acknowledgments

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] G. Agresti, L. Minto, G. Marin, P. Zanuttigh, Stereo and ToF Data Fusion by Learning from Synthetic Data, Elsevier Information Fusion, 2019, p. 49.
[2] P. Zanuttigh, G. Marin, C. Dal Mutto, F. Dominio, L. Minto, G.M. Cortelazzo, Time-of-flight and Structured Light Depth Cameras, Springer, Heidelberg, 2016.
[3] J. Sell, O. Patrick, The xbox one system on a chip and kinect sensor, in: IEEE Micro, vol. 1, 2014, 1-1.
[4] L. Keselman, W.I. Woodfill, J. Grunnet-Jepsen, A. Bhowmik, Intel RealSense stereoscopic depth cameras, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017.
[5] Z. Zhang, A flexible new technique for camera calibration", in: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, 1998.

[6] StereoLabs ZED website. https://www.stereolabs.com/zed/. (Accessed 14 October 2019).

[7] H. Sarbolandi, D. Lefloch, A. Kolb, Kinect range sensing: structured-light versus time-of-flight kinect, in: Computer Vision and Image Understanding, vol. 139, 2015, pp. 1–20.

[8] G. Agresti, L. Minto, G. Marin, P. Zanuttigh, Deep learning for confidence information in stereo and ToF data fusion, in: Proceedings of ICCV Workshops, 2017, pp. 697–705.

[9] C.D. Herrera, J. Kannala, J. Heikkila, Joint depth and color camera calibration with distortion correction, Pattern Analy. Mach. Intell. IEEE Trans. 34 (10) (2012) 2058–2064.

[10] C. Dal Mutto, P. Zanuttigh, G.M. Cortelazzo, A probabilistic approach to tof and stereo data fusion, in: Proceedings of 3DPVT, Paris, France, 2010.

[11] J. Zhu, L. Wang, R. Yang, J. Davis, Fusion of time-of-flight depth and stereo for high accuracy depth maps, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2008.

[12] G. Agresti, H. Schaefer, P. Sartor, P. Zanuttigh, Unsupervised domain adaptation for ToF data denoising with adversarial learning, in: Proceedings of the Int. Conference on Computer Vision and Pattern Recognition (CVPR), 2019.