Application of Intel RealSense Cameras for Depth Image Generation in Robotics

VLADIMIR TADIC, AKOS ODRY, ISTVAN KECSKES, ERVIN BURKUS, ZOLTAN KIRALY, PETER ODRY
Institute of Informatics
University of Dunaujvaros
Tancsics Mihaly u. 1/A Pf.: 152, 2401. Dunaujvaros
HUNGARY
t_laci@stcable.net, odrya@uniduna.hu, kecskes.istvan@gmail.com, ervinbur@appl-dsp.com, kiru@uniduna.hu, podry@uniduna.hu

Abstract: - This paper presents the applications of depth cameras in robotics. The aim is to test the capabilities of depth cameras in order to better detect objects in images based on depth information. In the paper, the Intel RealSense depth cameras are introduced briefly and their working principle and characteristics are explained. The use of depth cameras in the example of painting robots is shown in brief. The utilization of the RealSense depth camera is a very important step in robotic applications, since it is the initial step in a series of robotic operations, where the goal is to detect and extract an obstacle on a wall that is not intended for painting. A series of experiments confirmed that camera D415 provides much more precise and accurate depth information than camera D435.

Key-Words: - Depth image; measuring depth; RealSense cameras; image processing; obstacle detection.

1 Introduction

Recently, computer vision applications based on 2D image processing have been largely limited due to a lack of information on the third dimension, i.e. depth. Unlike 2D vision, 3D vision enables computers and other machines to differentiate accurately and with great precision various shapes, shape dimensions, distances, as well as control in the real, three-dimensional world [1], [2], [3], [4]. For this reason, 3D optical systems have been successfully applied in various areas such as robotics, car manufacturing, mechanics, biomedicine etc.

Not so long ago, 3D cameras were very expensive and their use was complicated and impractical. Today, thanks to technological advancements, the price of 3D cameras that are able to measure image depth has become significantly lower and their use has become much easier [5], [6]. The original version of the Kinect camera was first introduced in 2009 by the company PrimeSense. Equipped with an infrared lens, these cameras were used to detect infrared points projected onto a scene based on which depth information was evaluated along axis Z. In the first version, the resolution of the depth camera was 320x240 pixels with 2048 levels of depth. Later, other companies also started to produce cheaper cameras similar to the Kinect camera that were mainly intended for home use, primarily in video games [1].
In 2013, Apple acquired PrimeSense, after which the development and technology of producing the Kinect camera evolved in the sense that newer cameras became much more accurate and precise in terms of determining the depth information of images. In recent years, some of the depth cameras have also become an integral part of smartphones [1].

In the meantime, Intel also started developing its own family of depth cameras, Intel RealSense cameras. The first version of these cameras was developed in cooperation with Microsoft for Windows Hello as part of the face recognition system when logging on to the OS Windows10 [1], [7].

Although RealSense cameras only recently appeared on the market, they found their application very quickly in various areas such as face recognition, recognition and tracking human movements, in interactive children’s toys, in access control, in robotics and medicine [7], [8], [9], [10], [11], [12].

In the next section, the D400 series of the Intel RealSense cameras will briefly be described, which appeared on the market in 2018 and their application in robotics will be summarized on the example of digital image processing and finding and detecting obstacles/objects in complex images [13], [14].

2 Intel RealSense Cameras

This section will briefly introduce the technology and some of the more important characteristics of RealSense cameras, alongside their working principle.

2.1 The Technology of RealSense Cameras

RealSense technology basically consists of a processor for image processing, a module for forming depth images, a module for tracking movements and depth cameras. These cameras rely on deep scanning technology, which enables computers to see objects in the same way as humans do. The full hardware is also accompanied by an appropriate open-source SDK (Software Development Kit) software platform called librealsense [7]. This software platform provides simple software support for cameras of the D400 series, which allows users to use these cameras. The software platform supports ROS (Robot Operating System), C/C++, Matlab and Python systems and programming languages, which can be used to develop appropriate applications. Intel also offers two applications that can be used to set up and use the cameras [8].

As mentioned, Intel introduced two new RealSense cameras at the beginning of 2018: the models D415 and D435. By introducing these cameras, Intel became a leading manufacturer of depth cameras on the market in terms of balance of price and quality.

These cameras primarily differ in the field of view (FOV) expressed in angles and the type of shutter that adjusts the exposure. RealSense D435 has a wider field of view, (HxVxD – Horizontal x Vertical x Depth): 91ºx65ºx100º that minimizes blind, black spots in the depth image, after which a depth image pleasant for the eye is obtained. For this reason, this camera is suitable for applications in robotics, in cases when no great accuracy and precision are required, but a general visual experience is more important. A frequent use of this camera is in drones and in car manufacturing. Furthermore, this camera has a so-called global shutter that offers a better performance while recording fast movements, in situations when lighting is unsatisfactory, and also reduces the effect of blurring the image. The RealSense D415 camera has a narrower field of view, (HxVxD): 69ºx 42ºx 77º, which results in a higher density of pixels, increasing thus the resolution of the depth image. In this way, if accuracy is the main requirement in applications, e.g. in detecting objects and obstacles in robotics, the D415 camera gives much better results, especially when the distance is around or less than 1m. The D415 camera has a rolling shutter, which makes the performance of this camera better when there are no sudden movements when recording, but rather the image is static [7], [11].

(a) (b)

Fig. 1. (a) RealSense D435 and (b) RealSense D415 cameras [7]

Thanks to Intel RealSense cameras, the user interface of the applications can be improved to the highest levels, which has been unthinkable until recently and which enables the discovery of depth in images as well as tracking human movements [11].

2.2 The Working Principle of the RealSense Cameras

This section will briefly outline the working principle of the cameras. These cameras have three lenses: a conventional RGB (Red Green Blue) camera, an infrared (IR) camera and an infrared laser projector. All three lenses jointly make it possible to assess the depth by detecting the infrared light that has been reflected from the object/body that is in front of it. The resulting visual data, combined with the RealSense software, create a depth estimate, i.e. provide a depth image. After some additional processing, an image obtained in this way can be used, for instance, for tracking movements, by creating an interface that gives the impression of touch which reacts to movements by hand, the head, or any other body part, but also to facial expressions. Of course, since the RealSense camera also has a classical RGB lens, as well as an infrared lens, it is possible to take pictures in color and in conditions of poor lighting [7], [11].

RealSense cameras use stereovision to calculate depth. The implementation of stereovision consists of a left-side and a right-side camera, i.e. sensors, and of an infrared projector. The infrared projector projects invisible IR rays that enhance the accuracy of the depth information in scenes with weak textures. The left-side and right-side cameras record the scene and send data about the image to the processor that, based on the received data, calculates the depth values for each pixel of the image, thus correlating the points obtained with the left-side camera to the image obtained with the right-side camera. The depth values of each pixel processed in this way result in the depth frame, i.e. the depth image. By joining consecutive depth images, a depth video stream is obtained.

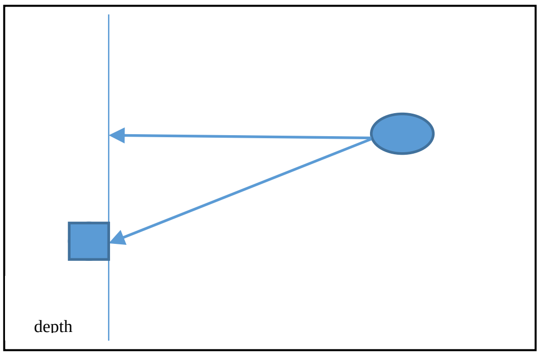


Fig. 2. Diagram showing the method of measuring depth in relation to distance.

As shown in Fig. 2, the value of the depth pixel describing the depth of an object is determined in relation to a parallel plane of the camera doing the recording, and not in relation to the actual distance of the object from the camera.

An important role in the operation of the camera is also reserved for the RealSense D4 processor for image processing, which is capable of processing 36 million depth points in a second. Thanks to this performance, these cameras are built into a high number of devices that require high-speed data processing [7], [11].

3 Results of the Experiment

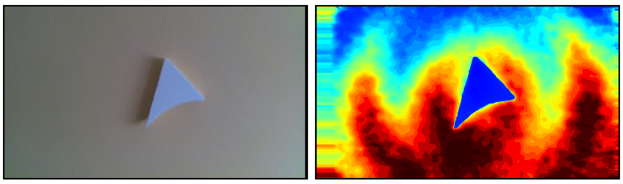
This section will describe the preliminary results of experiments conducted with RealSense cameras D415 and D435. The experiments were part of a project dealing with robotics, or, more precisely, robotic vision [15], [16], [17], [18], [19], [20], [21]. The aim of the project was to construct a robot that would automatically, based on information obtained from the cameras, paint the facades of buildings with a special paint that would simultaneously put on a coat of considerable level of insulation to the object.

The task associated with the RealSense depth cameras can briefly be described as the following steps:

1. Recording images of part of the building/wall
2. Based on the depth image, the obstacles on the wall that must not be painted are to be detected and isolated with algorithms of digital image processing
3. Forward the information about the coordinates of the obstacle to the robot
4. Painting part of the building/wall by bypassing the detected obstacles.

Since the process of developing the afore-mentioned robot is still in its initial phase, the experiments have been conducted with artificially created obstacles with both RealSense cameras. The first task at hand was to determine via the experiments which camera behaves in what way under particular recording conditions and how it is possible to obtain the most accurate depth image of the obstacle in a complex image. This is a crucial step, since the success of detecting the obstacles in a complex image greatly depends on the obtained depth image. The reason for using the depth image lies in the fact that almost all obstacles on the walls protrude from the wall along axis Z, i.e. in their depth. In this way, based on the distance, i.e. depth, it is possible to determine the objects that do not lie on the surface of the wall and which could form some sort of obstacle that is not to be painted. Of course, this method of detecting objects is possible to be used in any robot for detecting and bypassing obstacles, as well as for locating certain objects in the field of movement of the robot. In the described experiments, a depth camera and an RGB camera were used. The infrared camera has no significance in the present phase of the project, as the robot has been designed to operate during daytime or in light, and not in the dark.

The first measurement was performed at a distance of 1m from the wall, since the aim was for the camera to capture as wide a surface on the wall as possible that would potentially need to be painted. It is possible that in the later phases of developing the system, a higher number of cameras will be required for recording to cover the necessary surface. The RealSense software has the ability to control multiple cameras simultaneously; it is thus possible that a future research project will be dealing with this issue.

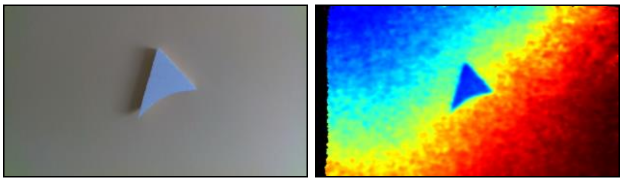


(a) (b)

Fig. 3. The image of camera D415, default setting: (a) RGB image and (b) depth image

Fig. 3 shows images captured with camera D415. The image shows an obstacle of an undefined shape hanging on the wall, which was the aim of the experiment, to explore the capabilities of the camera. It is important to note that the resolution of the RGB and depth cameras was set to the very maximum, i.e. to 1280x720 and the so-called default mode of recording was used, which offers the best visual experience according to the manufacturer's specifications. It can be seen that the object of interest is very clearly discernible in the depth image.

Fig. 4 shows an identical content and recording was performed under identical conditions, but in this case, by using camera D435.

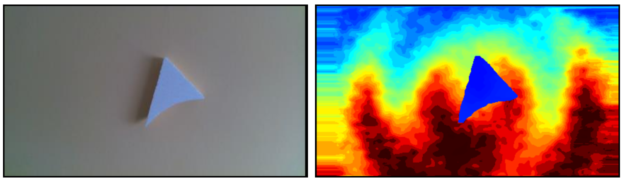


(a) (b)

Fig. 4. The image of camera D435, default setting: (a) RGB image and (b) depth image

As can be seen in Fig. 4, there is almost no difference between the RGB images captured with cameras D415 and D435. Still, the most important result of this pair of images is the drastic difference in the quality of the depth images captured with camera D435 compared to the depth image captured with camera D415. Due to the wider field of view, the pixel density is lower in case of the image captured with camera D435, and hence the depth image is also more blurred, the contours of the obstacle are less pronounced and the obstacle itself is of smaller dimensions compared to the background of the image. Also, there is a strong noise in the image and a higher number of black spots on the left-side part of the image, which is most probably due to a shadow that can be seen on the left side of the scene.

This result shows that it is necessary to use a different setting that would increase the accuracy of determining the obstacle on the wall, primarily in terms of more accurately determining the contours of the obstacle itself that is present. This is necessary for the robot to be able to fully know how far it can or cannot spray the paint. For the above reason, the next experiment was conducted with different settings, where the built-in setting for high-accuracy was used for determining a high-accuracy depth.

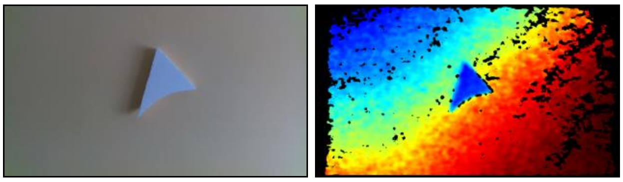


(a) (b)

Fig. 5. The image of camera D415, high-accuracy setting: (a) RGB image and (b) depth image

Fig. 5 shows the same obstacle that was also captured from a distance of 1m under the same conditions of lighting. The resolution of the RGB and depth cameras was also identical as before; only the setting for forming the depth image was changed to high accuracy. In this case, camera D415 was used. As can be seen, the RGB image is almost of the same quality, and the depth image is also of excellent quality where the contours of the obstacle are clearly visible. If Fig. 5 (b) is examined more closely, a darker line can be seen along the circumference of the obstacle, with the help of which it can be differentiated from the wall with great precision. This is a highly important result, as in robotics, it is crucial to clearly separate a possible obstacle so that the robot is able to do the work for which it has been designed.

In the following example, in Fig. 6, the same scene was captured, under identical conditions, only by using camera D435.



(a) (b)

Fig. 6. The image of camera D435, high-accuracy setting: (a) RGB image and (b) depth image

It can be seen that the RGB image has practically remained unchanged, but in case of the depth image, the difference is drastic. This is because in this case, as well, due to a wide field of view of camera D435, the pixel density is lower, and so is the resolution. There is also a strong noise, stronger than when recording with the default setting, but the most noticeable difference is a high number of blind-black spots in the image. The reason for this is the changed setting of the recording, since camera D435 is not designed for determining high-accuracy depth, but for making a depth image that gives a good visual experience. This camera is primarily used at longer distances up to about 30m (in some cases, even at longer distances), whereas camera D415 is used at shorter distances up to about 5-6m, but the recommended distance cited in the literature [8] is up to 1m. In this case, somewhat more pronounced contours of the circumference of the obstacle can be discerned in the depth image compared to the image captured with the default setting that is shown in Fig. 4 (b). The reason for this is in using the setting for high-accuracy depth, which can be achieved by changing the parameters of the in-built filters in the RealSense software. The description of the parameters of the above-mentioned filters, alongside the description of the filters themselves, is an enormous task and goes beyond the scope of this paper, but a full description can be found in the technical specifications in the literature [8].

It should be noted that various settings can be achieved by changing a high number of parameters that control the operation of the RealSense cameras. Some of these settings are included by the manufacturer of the devices, but users can also create their own settings according to their needs. Still, in most of the cases, some of the built-in settings give the best results [8].

Based on the described experiments, it can be concluded that camera RealSense D415 gives a much more accurate and precise depth image and is more suitable for applications in robotics. In further research, the emphasis will be on testing the capabilities of camera RealSense D415.

4 Conclusion

In the present paper, the working principle and characteristics of Intel's RealSense depth cameras are briefly described. Appropriate experiments were conducted with two cameras, D415 and D435, and an evaluation of the experiments was performed. The aim of the experiments was to show the possibilities of use of depth cameras in robotics with the aim of developing systems of robotic vision. The use of a depth camera and forming a depth image represents the initial and most important step in developing such a system with the help of digital image processing algorithms, after which the robot would have the capability of automatically detecting obstacles in its field of movement. The main conclusion of the present paper is that camera D415 gives a much higher-quality depth image in which objects and obstacles are more clearly discernible based on the information on the depth of an image compared to camera D435.

5 Acknowledgements

This research has been funded by project EFOP-3.6.1-16-2016-00003, co-financed by the European Union.

References:

[1] Monica Carfagni, Rocco Furferi, Lapo Governi, Chiara Santarelli, Michaela Servi, "Metrological and Critical Characterization of the Intel D415 Stereo Depth Camera", Sensors, 2019, 19, 489; doi:10.3390/s19030489

[2] Jia Hu, Yifeng Niu, Zhichao Wang, "Obstacle Avoidance Methods for Rotor UAVs Using RealSense", 2017 Chinese Automation Congress (CAC), DOI: 10.1109/CAC.2017.8244068

[3] Silvio Giancola, Matteo Valenti, Remo Sala, "A Survey on 3D Cameras: Metrological Comparison of Time-of-Flight, Structured-Light and Active Stereoscopy Technologies", SpringerBriefs in Computer Science, Springer, ISSN 2191-5768, 2018, doi.org/10.1007/978-3-319-91761-0

[4] Leonid Keselman, John Iselin Woodfill, Anders Grunnet-Jepsen, Achintya Bhowmik, "Intel RealSense Stereoscopic Depth Cameras", arXiv:1705.05548

[5] Reginald L. Lagendijk, Ruggero E.H. Franich, Emile A. Hendriks, "Stereoscopic Image Processing", The work was supported in part by the European Union under the RACE-II project DISTIMA and the ACTS project PANORAMA.

[6] Francesco Luke Siena, Bill Byrom, Paul Watts, Philip Breedon, "Utilising the Intel RealSense Camera for Measuring Health Outcomes in Clinical Research", Journal of Medical Systems (2018) 42: 53, doi.org/10.1007/s10916-018-0905-x

[7] "Intel RealSense D400 Series Product Family Datasheet", New Technologies Group, Intel Corporation, 2019, Document Number: 337029-005

[8] Anders Grunnet-Jepsen, Dave Tong, "Depth Post-Processing for Intel® RealSense™ D400 Depth Cameras", New Technologies Group, Intel Corporation, 2018, Rev 1.0.2

[9] "Evaluating Intel's RealSense SDK 2.0 for 3D Computer Vision Using the RealSense D415/D435 Depth Cameras", 2018, Berkeley Design Technology, Inc.

[10] "Intel® RealSense™ Camera Depth Testing Methodology", New Technologies Group, Intel Corporation, 2018, Revision 1.0

[11] Anders Grunnet-Jepsen, John N. Sweetser, John Woodfill, " Best-Known-Methods for Tuning Intel® RealSense™ D400 Depth Cameras for Best Performance", New Technologies Group, Intel Corporation, Rev 1.9

[12] Anders Grunnet-Jepsen, Paul Winer, Aki Takagi, John Sweetser, Kevin Zhao, Tri Khuong, Dan Nie, John Woodfill, "Using the Intel® RealSenseTM Depth cameras D4xx in Multi-Camera Configurations", New Technologies Group, Intel Corporation, Rev 1.1

[13] "Intel RealSense Depth Module D400 Series Custom Calibration", New Technologies Group, Intel Corporation, 2019, Revision 1.5.0

[14] Anders Grunnet-Jepsen, John N. Sweetser, "Intel RealSens Depth Cameras for Mobile Phones", New Technologies Group, Intel Corporation, 2019

[15] Philip Krejov, Anders Grunnet-Jepsen, "Intel RealSense Depth Camera over Ethernet", New Technologies Group, Intel Corporation, 2019

[16] Joao Cunha and Eurico Pedrosa and Cristovao Cruz and Antonio J.R. Neves and Nuno Lau, "Using a Depth Camera for Indoor Robot Localization and Navigation", Conference: RGB-D: Advanced Reasoning with Depth Cameras Workshop, Robotics Science and Systems onference (RSS), At LA, USA, 2011

[17] Hani Javan Hemmat, Egor Bondarev, Peter H.N. de With, "Real-time planar segmentation of depth images: from 3D edges to segmented planes", Eindhoven University of Technology, Department of Electrical Engineering, Eindhoven, The Netherlands

[18] Fabrizio Flacco, Torsten Kroger, Alessandro De Luca, Oussama Khatib, "A Depth Space Approach to Human-Robot Collision Avoidance", 2012 IEEE International Conference on Robotics and Automation RiverCentre, Saint Paul, Minnesota, USA May 14-18, 2012

[19] Ashutosh Saxena, Sung H. Chung, Andrew Y. Ng, "3-D Depth Reconstruction from a Single Still Image", International Journal of Computer Vision, 2008, Volume 76, Issue 1, pp 53–69

[20] Vladimiros Sterzentsenko, Antonis Karakottas, Alexandros Papachristou, Nikolaos Zioulis, Alexandros Doumanoglou, Dimitrios Zarpalas, Petros Daras, "A low-cost, flexible and portable volumetric capturing system", Website: vcl.iti.gr

[21] Nicole Carey, Radhika Nagpal, Justin Werfel, "Fast, accurate, small-scale 3D scene capture using a low-cost depth sensor", 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), DOI: 10.1109/WACV.2017.146