Research Article

INTERNATIONAL JOURNAL OF ADVANCED ROBOTIC SYSTEMS

International Journal of Advanced Robotic Systems
January-February 2018: 1–8

DOI: 10.1177/1729881418755780
journals.sagepub.com/home/arx
SAGE

Kinect v2 infrared images correction

Zdenko Bobovský, Václav Krys and Vladimír Mostýn

Abstract

This article presents a novel correction filter for infrared images captured by Kinect v2 sensor. Intended application areas are described and the basic concept of proposed sensory subsystem as well. Preliminary tests of the sensor Kinect v2 in real conditions gave promising results, therefore in-depth analysis of its applicability was performed. In the framework of the analysis, a relation between infrared value and a distance of captured surface was evaluated for different colors of the surface. Based on that relation and on additional information about the depth of a pixel, a correction filter was created. The filter allows improving infrared image in order to increase the success rate probability to be able to detect specific features and key points by algorithms more easily. Outputs from the filter on real conditions data sets are also presented in the article.

Keywords

Vision system, service robotics, Kinect v2, cargo boxes, coal mine

Date received: 30 August 2017; accepted: 26 December 2017

Topic: Special Issue—Mobile Robots
Topic Editor: Antonio Fernandez-Caballero
Associate Editor: Michal Kelemen

Introduction

In the context of the intense development of robotic applications based on vision systems in industrial and service robotics, and especially with the development of self-driving cars, the development of hardware (HW) tools with sufficient computational power to process large amounts of data in real time is notable in last years. Lowering prices for these HW tools and sensors make it possible to use them in wider application areas. Another trend is the breaking of the boundaries between industrial and service robotics, which results in mounting of manipulators on mobile platforms and their wider application in production and logistics operations where these systems are collaborating with people. For collaborative tasks, the robotic system needs to be able to detect objects in its dynamically changing environment and respond appropriately to these changes in real time.

Two different application areas met in our research projects in our department at the same time, where it was appropriate to verify the usability of the Kinect v2 sensor with regard to its purchase price. The first area is the recognition of boxes in a data set of an unloaded shipping container. Second, the creation of a three-dimensional (3-D) map of the coal mine in locations without risk of explosive environment. Therefore, the Kinect v2 sensor was tested for these application areas using the most common tools (HW and software (SW)) of the sensor subsystem.

The article describes a method of correcting data obtained from the Kinect v2 sensor to increase the success rate of algorithms for detecting objects in the acquired image as well as for joining partial images into the overall scene preview and 3-D map. The obtained results are evaluated from the point of view of the applicability of the Kinect v2 universal sensory subsystem for detecting objects and their positions for gripping boxes in the automated unloading of

Department of Robotics, Faculty of Mechanical Engineering, VŠB-Technical University of Ostrava, Ostrava, Czech Republic

Corresponding author:
Zdenko Bobovský, Department of Robotics, Faculty of Mechanical Engineering, VŠB-Technical University of Ostrava, 17. listopadu 2172/15, Ostrava 70833, Czech Republic.
Email: zdenko.bobovsky@vsb.cz

Figure 1. Cardboard boxes inside a shipping container.

shipping containers and for creating a 3-D map of the coal mine by a mobile robotic system. It is the first step in processing the data obtained from the sensor. The quality of the acquired scene images significantly influences the difficulty and success rate of further processing of the sensed environment image and further works with it.

Current robotic applications of the Kinect v2 sensor are mainly based on refinement and correction of the obtained point cloud using PCL or on algorithms available for ROS.[2] The current solutions[3,4,5] described are based on improvements of RGB images gained by Kinect v2.

The remainder of this article is structured as follows. "Intended application areas" section describes in greater detail the intended application areas. "Synthesis of the correction filter" section details the data collection process, data processing, and the way of a correction filter creation. Finally, "Applications of the filter in real environments" section shows application of the filter on data obtained in real environment conditions.

Intended application areas

Systems for automated unloading of boxes from containers are not as widespread as could be expected today. In almost all cases, unloading of shipping containers in logistics centers is now manually handled. It is a very demanding physical activity, which is problematic in terms of ergonomics and the total weight of manipulated boxes per worker. From this point of view, it is an area with a high potential for automation. The problem is the return on investment and reliability of such a system. The reliability significantly affects the condition of boxes in a container, which is often very poor after a long journey. Therefore, predicted geometric shape of boxes cannot be expected. Observe the same placing and orientation of boxes in all box walls cannot be guaranteed as well. The task of detection boxes for the automatic unloading is a complex task to find objects in a dynamic environment with low or very lowlight intensity (see Figure 1). The closest solution to our sensor subsystem is the robotic truck unloader (RTU) concept,[1] where the cheap RGB-D sensor ASUS Xtion Pro is used to determine a position of the box being removed. Another developed system for unloading containers is a Parcel Robot,[6] whose sensory subsystem is based on a tilted SICK laser scanner to get basic information about the state of the environment and to move an end point of a manipulator to the position of next box removal. The end effector of the manipulator is equipped with precise distance sensors for data refinement and detection of gaps at the location of box removal. Another existing concept of the system for automatic unloading of containers is TEUN.[7] It is based on an industrial robot on a mobile platform.

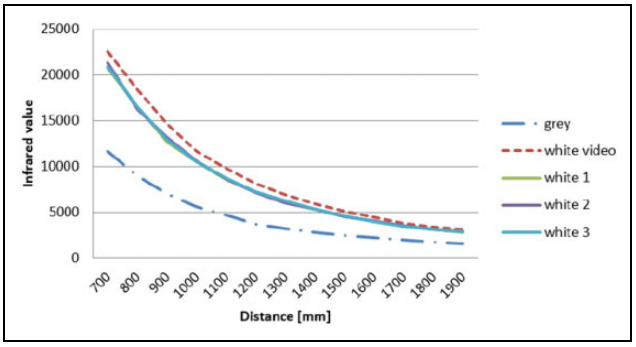
Figure 2. Coal mine tunnel.

Automated creation of 3-D maps of mine tunnels is also a big challenge. An example of such tunnel is shown in Figure 2. Currently described systems, which were used for the creation of 3-D maps of mine tunnels, utilized laser scanners or lidars, followed by post-processing of data from gained point clouds.[8,9]

A common feature of the above-mentioned applications is low or almost no light intensity environment. This led us to the idea of verifying the usability of the Kinect v2 sensor for both of these application areas with the biggest possible amount of common HW and SW tools. The idea is to use information from depth map and infrared image for increasing the success rate of objects and key points detection by algorithms.

Synthesis of the correction filter

The basic requirement of the proposed sensor subsystem is to obtain a depth map and a black-and-white image of the lowlight intensity environment that would be suitable for algorithms that allow detecting objects and key points in it.



Figrue 3. Course of IR value based on a distance and on captured surface color. IR: infrared.

With regard to the low cost of the sensor and the fact that it meets the basic requirements specified for both applications described above, the Kinect v2 sensor applicability was tested. The gained infrared image is modified by the filter based on the information obtained from the depth map. This approach eliminates the need for external lighting in the system.

We had two Kinect v2 sensors available. These were subjected to preliminary tests to create a correction filter, and then data sets were taken under real environmental conditions. In the framework of the preliminary tests, dependencies of infrared values on the distance of the scanned surface from the sensor were determined.

Measurements of preliminary tests were made on different flat surfaces in order to obtain representative data sets. Most of these measurements were done on matt white surfaces, but data sets obtained on bright white, matt grey, and matt black surfaces were evaluated as well. Measurements were made in the whole declared sensor measuring range (700–8000 mm) at the 100 mm measuring step and without artificial light. These tests evaluated an area of ± 10 pixels from the center of the frame for the given distance. Initial measurements were made immediately after the sensor was switched on, but it was verified that the accuracy of the point depth data is significantly affected by the sensor's operating temperature as described in the study by Wasenmüller and Stricker.[10] Subsequent measurements were made after 30 min warm-up of the sensor.

Related SW applications were written in the .NET Framework 4.5 using the Kinect for Windows software development kit (SDK) v2.0 and the C# language. Data from Kinect SDK for depth and infra were collected in raw format. The output image data set for the given distance was created as average values from several subsequent image data sets of the recorded surface. A comparison of measurement deviations with a total deviation based on a change in the number of input subsequent frames was performed before preliminary tests. The difference of deviations for data sets from 10 to 50 subsequent images was 0.5–1%. That is why the average values from 10 subsequent images were used for preliminary tests on flat surfaces.

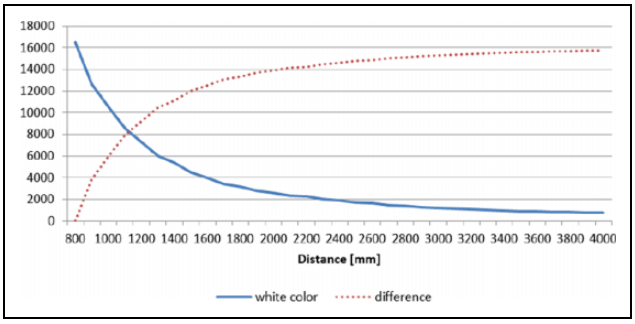


Figure 4. Infra values for white color and their deviations.

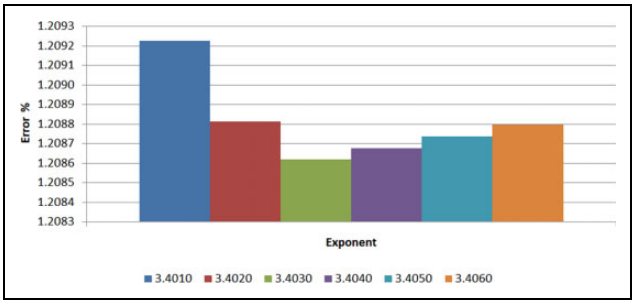


Figure 5. Percentage error at different exponent.

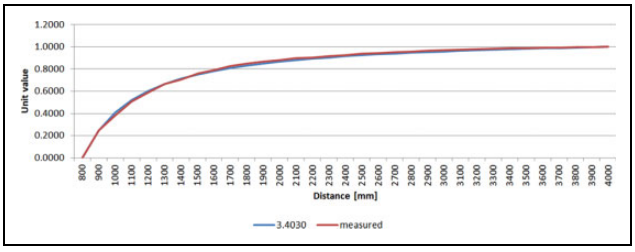


Figure 6. The measured and the calculated characteristic.

Figure 3 shows the difference between performed preliminary tests for different colors of flat surfaces except black color. The maximum infrared value for the black color that we were able to measure was 2395 at a distance of 500 mm from the surface and at a distance of 800 mm, it was 927. Infrared value for the white color was 16,207 and 18,348 at the same distances, respectively.

An infrared characteristic for white color was selected as determinative one, because of highest values for the given distances. Figure 4 shows the characteristic (solid line) in more detail up to a distance of 4 m with the infrared value of 749. Above this distance is a course of the characteristic nearly linear up to a distance of 8000 mm with infrared value of 438. Dotted curve in Figure 4 shows a difference to a required value. The difference is a deviation of the infrared image which has to be corrected.

For the correction, a relation that roughly corresponds to the characteristic "difference" in Figure 4 was used.

The gamma that is used for gamma correction of the infra pixel for a given depth is calculated based on the depth value

Figure 7. Infrared correction of a room image. (a) Default image, (b) gamma correction 0.5, and (c) correction by depth.

Figure 8. Infrared correction of the data set from a coal mine. (a) Default image, (b) gamma correction 0.5, and (c) correction by depth.

$$\text{gamma} = 2 - \left(\frac{1}{s^{\exp}} + \frac{4}{s}\right) * n \quad (1)$$

where s is the coefficient of a distance

$$s = \frac{\text{depth}}{\text{depth}_{\text{nominal}}} \quad (2)$$

where depth is the depth of the given pixel; $\text{depth}_{\text{nominal}}$ the boundary depth from which it is calculated; exp the exponent—determines the curvature of the curve; and n the multiplier—determines max value of the curve.

The exponent was set as 3.41. At this value, the minimum total deviation between the calculated characteristic and the measured one was obtained (see Figure 5).

The maximal deviation between the calculated and the measured characteristic is within 1.21%. Both characteristic are shown in Figure 6.

Applications of the filter in real environments

Results of the correction filter applications are provided in this section. To verify the functionality of the filter, first data sets were recorded in the laboratory of our department. Outputs are shown in Figure 7. Central distance of the image is 4000 mm.

Data sets from coal mine (mine museum) were used for next verification step, because the original obtained image was not the best one from all intended environments. Results are presented in Figure 8.

Afterward, the correction filter was applied to data from shipping container environment. Obtained results are shown in Figure 9.

The correction filter was applied to data gained in logistic center. Deformed and randomly oriented cardboard boxes are stored before their unpacking. Application of the filter for this environment is presented in Figure 10.

Canny edge detection

The Canny filter[11] for edge detection was applied to the corrected images. The filter settings are mask size 7 and sigma 1. The images are displayed with both bold and thin edges, regardless of the threshold. Due to the depth of the mine tunnel image, the difference between the resulting frames (gamma vs. depth correction) after the Canny filter application is very significant (see Figure 11).

Figure 12 presents output from Canny filtered images of cardboard boxes. Results for boxes close enough to the

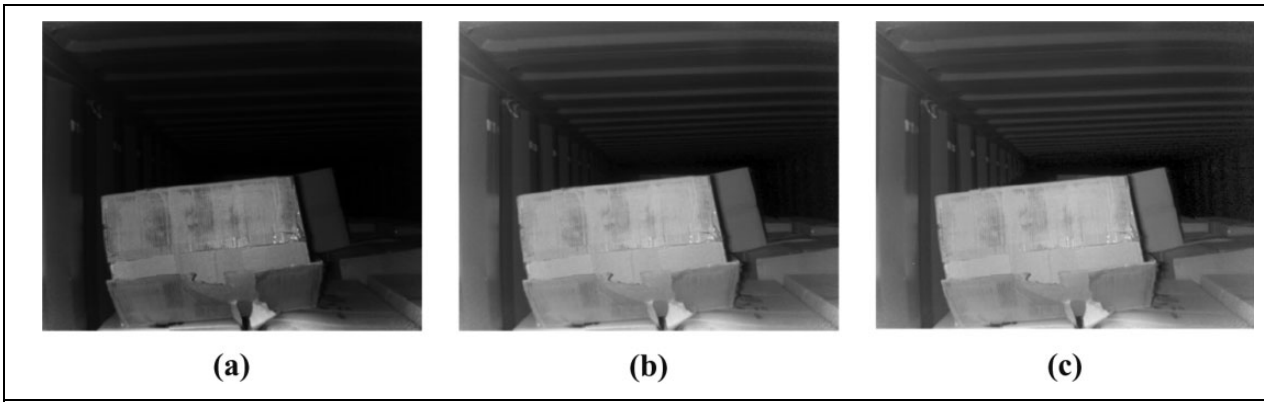


Figure 9. Infrared correction of the data set from a shipping container. (a) Default image, (b) gamma correction 0.5, and (c) correction by depth.

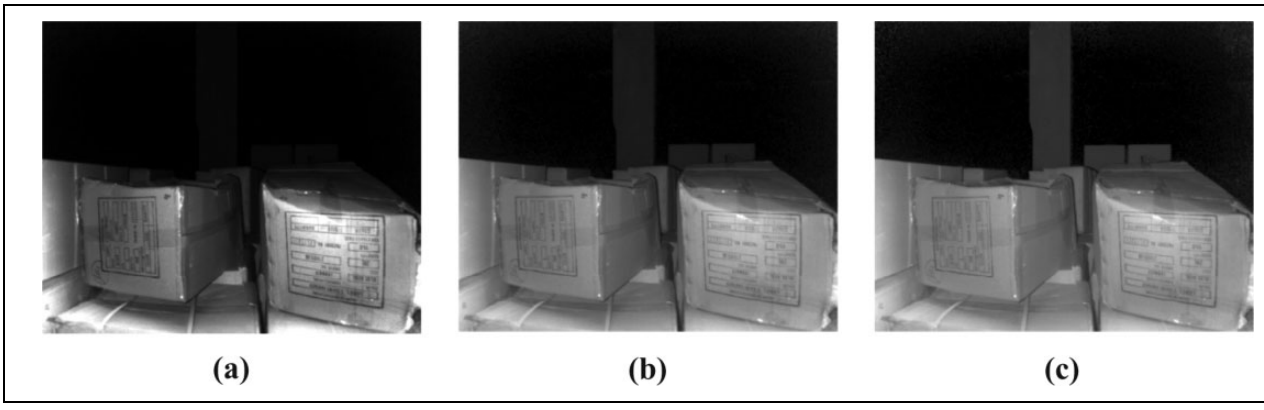


Figure 10. Infrared correction of the data set of cardboard boxes. (a) Default image, (b) gamma correction 0.5, and (c) correction by depth.

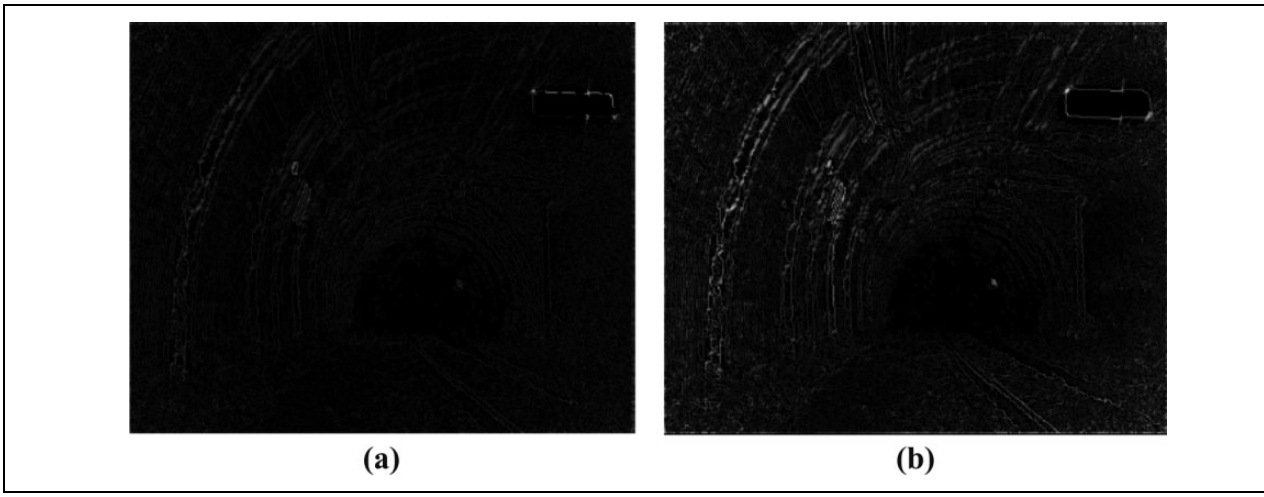


Figure 11. Comparison of outputs of Canny filter applied to images from a coal mine. (a) IR image corrected by gamma 0.5 and (b) correction by depth (our filter). IR: infrared.

sensor are comparable and the same edges of boxes are detected, but at longer distances, images corrected with depth filter give better results.

SIFT, SURF, and Harris detector

MATLAB, specifically SIFT, SURF and Harris detectors, was used to verify the detection of features in the corrected

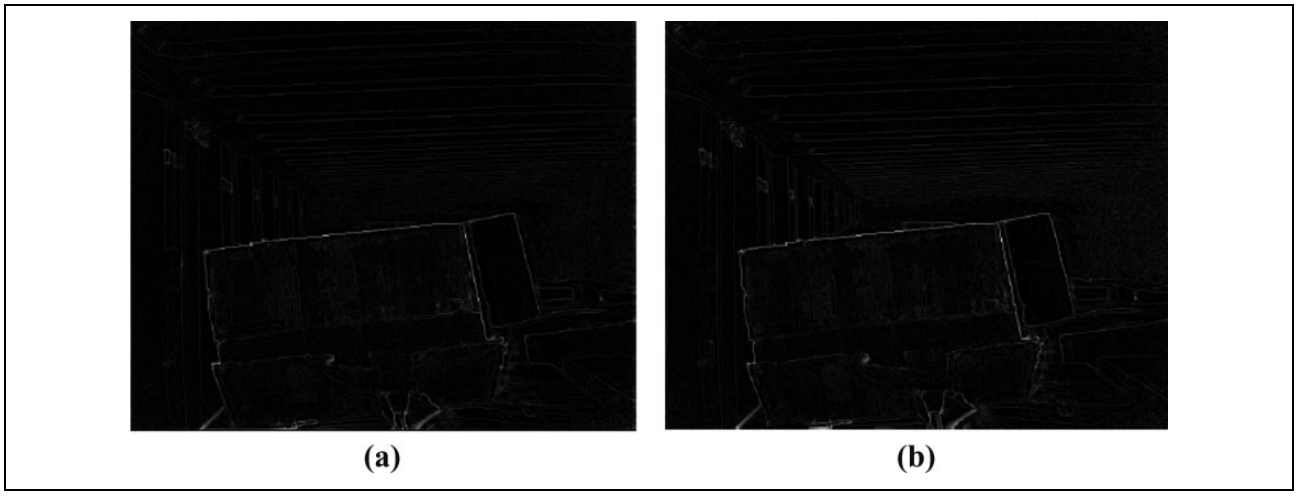


Figure 12. Canny filtered images of cardboard boxes. (a) IR image corrected by gamma 0.5 and (b) correction by depth (our filter). IR: infrared.

images. The SIFT detector is an optional MATLAB extension,[12] while the SURF and Harris detectors are already in as vision tools. The results are shown in Table 1. It is obvious that images after our correction (correction by depth) are best for features detection. In some cases, the number of detected points is higher up to 300%.

Images matching based on matching of identified features

To verify the applicability of corrected images, an image matching was performed using sequentially adjacent images for data obtained from scanning boxes and scanning in a coal mine. The matching features were obtained using the SIFT, SURF, and Harris detector.

In the case of cardboard box scanning, adjacent images were obtained by moving the sensor sideways. The resulting matched images show differences in the number of detected feature pairs. The number of detected feature pairs in our correction (correction by depth) is higher for all detectors applied (see Table 2).

Images from the coal mine were obtained by placing the sensor on a mobile robot that was moving forward through the mine tunnel. Even in this case, it is obvious that the images adjusted by our correction allow us to detect a higher number of feature pairs (see Table 3). For the Harris detector, the number of detected features was not sufficient to correctly match the images.

Based on above shown figures, it is clear that images improved by the depth correction enable better followed-up detection of edges, which in turn makes it easy to detect features and key points.

Table 1. Comparison of feature detection.

Figure	Correction	Features detected		
		SIFT	SURF	Harris
	Default	0	0	6
	Gamma	39	9	121
	Depth	67	20	36
	Default	19	7	6
	Gamma	80	12	13
	Depth	418	21	31
	Default	151	47	51
	Gamma	170	57	124
	Depth	196	64	305
	Default	269	25	25
	Gamma	244	22	49
	Depth	299	38	362

Future works

Creation and verification of the depth filter for infrared images is the first step in the long chain of robotic system synthesis for the intended above-mentioned applications. To automate the container unloading process, we need to trace and verify or create our own algorithm for

Table 2. Matching of neighborhood figures of carbon boxes.

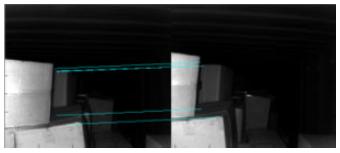
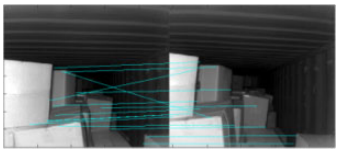
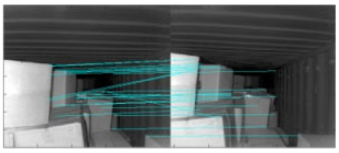
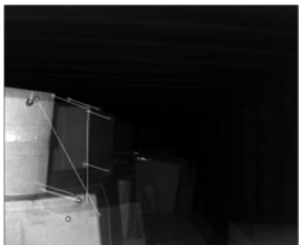
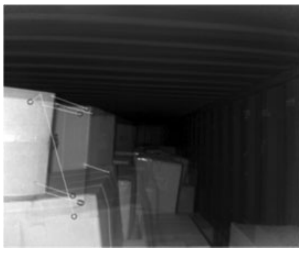
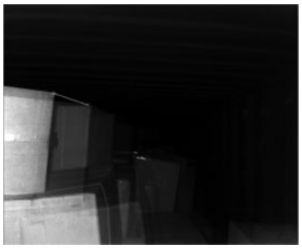
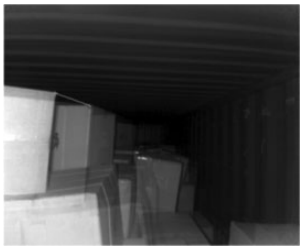
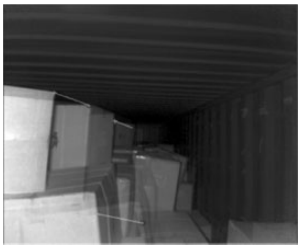
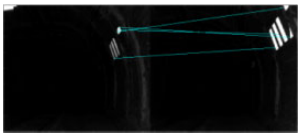
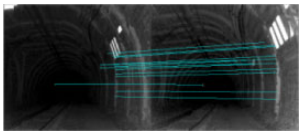
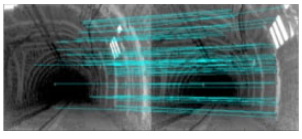
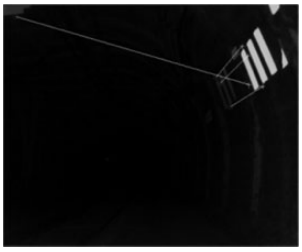
	Default	Gamma correction	Depth correction
SIFT			
SURF			
Harris			

Table 3. Matching of neighborhood figures of coal mine.

	Default	Gamma correction	Depth correction
SIFT			
SURF			
Harris			

detecting objects—deformed cardboard boxes on the captured image in the shortest possible time in order to create a trajectory of approaching the manipulator end point to the position of the currently unloaded box. Additional functionalities of the SW application will depend on the overall concept of the automatic unloading system and the strategy for boxes unloading from the container.

In order to create 3-D maps and their visualizations in the mining environment, it will be necessary to verify the computational demands of the process and its eventual application in real time.

An important decision-making factor for the Kinect v2 sensor applicability will be to verify its reliability in operating conditions in conjunction with the whole system and any measurement errors depending on the length of the operating time.

Conclusion

The contribution of our research is the creation of the correction filter that improves the infrared image taken by the Kinect v2 sensor based on data from the depth map. Application of the filter allows getting a black-and-white image of the scene without the use of conventional black and white or RGB camera and artificial light in lowlight environments. The gamma correction value of the infrared pixel has been determined at its distance from the sensor for the whole operating range of the sensor.

When accepting limiting properties of the sensor such as a relatively limited default operating range (800–8000 mm), a 30 min time interval to achieve the right operating temperature and the lengthy process time of joining data from the RGB camera to other data, Kinect v2 is a solution that allows to get an appropriate black-and-white image for further image processing by SW tools such as OpenCV, or custom algorithms for objects and key points detection. The Kinect v2's biggest benefits include its low cost, easy accessibility, and relatively low power consumption.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This article has been elaborated under support of the project Research Centre of Advanced Mechatronic Systems, reg. no. CZ.02.1.01/0.0/0.0/16_019/0000867 in the frame of the Operational Program Research, Development and Education. This article was also supported by the specific research project SP2017/143 and financed by the state budget of the Czech Republic.

ORCID iD

Zdenko Bobovský https://orcid.org/0000-0003-4134-5251
Václav Krys http://orcid.org/0000-0002-4840-6961
Vladimír Mostýn http://orcid.org/0000-0002-2588-0804

References

1. Doliotis P, McMurrough CD, Criswell A, et al. A 3D perception-based robotic manipulation system for automated truck unloading. In: *2016 IEEE international conference on automation science and engineering (CASE)*, Fort Worth, TX, 21–25 August 2016, pp. 262–267. DOI: 10.1109/COASE.2016.7743416.
2. Raaj Y, Nair S, and Knoll A. Precise measurement of cargo boxes for gantry robot palletization in large scale workspaces using low-cost RGB-D sensors. In: *Computer vision—ACCV 2016: 13th Asian conference on computer vision*, Taipei, Taiwan, 20–24 November 2016, pp. 472–486. DOI: 10.1007/978-3-319-54190-7_29.
3. Schwarz M, Milan A, Periyasamy AP, et al. RGB-D object detection and semantic segmentation for autonomous manipulation in clutter. *Int J Robot Res* 2017; 17: DOI: 10.1177/0278364917713117.
4. Cadena C and Košecká J. Semantic parsing for priming object detection in indoors RGB-D scenes. *Int J Robot Res* 2014; 34: 582–597. DOI: 10.1177/0278364914549488.
5. He L, Dong Q, and Wang G. Fast depth extraction from a single image. *Int J Adv Robot Syst* 2016; 13: 11. DOI: 10.1177/1729881416663370.
6. Scholz-Reiter B, Echelmeyer W, and Wellbrock E. Development of a robot-based system for automated unloading of variable packages out of transport units and containers. In: *2008 IEEE international conference on automation and logistics*, Qingdao, 1–3 September 2008, pp. 2766–2770. DOI: 10.1109/ICAL.2008.4636644.
7. TEUN.com—TEUN takes a load off your hands. http://www.teun.com/en/ (accessed 31 July 2017).
8. Wong U, Morris A, Lea C, et al. Comparative evaluation of range sensing technologies for underground void modeling. In: *2011 IEEE/RSJ international conference on intelligent robots and systems*, San Francisco, CA, 25–30 September 2011, pp. 3816–3823. DOI: 10.1109/IROS.2011.6094938.
9. Leung K, Lühr D, Houshiar H, et al. Chilean underground mine dataset. *Int J Robot Res* 2017; 36(1): 16–23. DOI: 10.1177/0278364916679497.
10. Wasenmüller O and Stricker D. Comparison of Kinect V1 and V2 depth images in terms of accuracy and precision. In: *Computer Vision—ACCV 2016 workshops: ACCV 2016 international workshops*, Taipei, Taiwan, 20–24 November 2016, pp. 34–45. DOI: 10.1007/978-3-319-54427-4_3.
11. Canny J. A computational approach to edge detection. *IEEE Trans Pattern Anal Mach Int* 1986; 8(6): 679–698. DOI: 10.1109/TPAMI.1986.4767851.
12. David G. Lowe. Object recognition from local scale-invariant features. In: *International conference on computer vision*, Corfu, Greece, 20–27 September 1999, pp. 1150–1157.