Comparison of Kinect v1 and v2 Depth Images in Terms of Accuracy and Precision

Oliver Wasenmüller and Didier Stricker

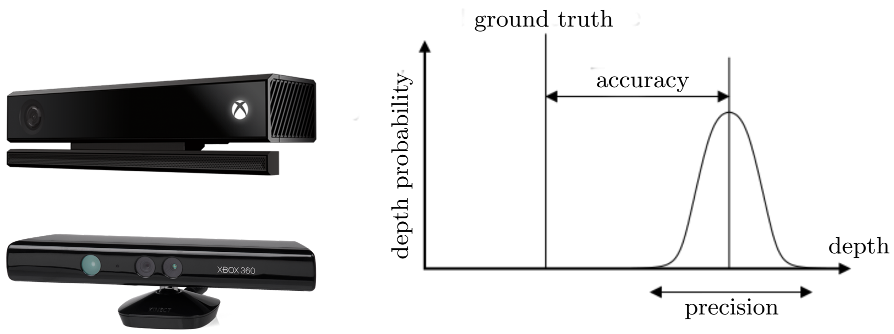
German Research Center for Artificial Intelligence (DFKI)
oliver.wasenmueller@dfki.de, didier.stricker@dfki.de

Abstract. RGB-D cameras like the Microsoft Kinect had a huge impact on recent research in Computer Vision as well as Robotics. With the release of the Kinect v2 a new promising device is available, which will – most probably – be used in many future research. In this paper, we present a systematic comparison of the Kinect v1 and Kinect v2. We investigate the accuracy and precision of the devices for their usage in the context of 3D reconstruction, SLAM or visual odometry. For each device we rigorously figure out and quantify influencing factors on the depth images like temperature, the distance of the camera or the scene color. Furthermore, we demonstrate errors like *flying pixels* and *multipath interference*. Our insights build the basis for incorporating or modeling the errors of the devices in follow-up algorithms for diverse applications.

1 Introduction

Since a couple of years RGB-D cameras have a huge impact on the research in the Computer Vision community as well as on related fields like Robotics and Image Processing. These cameras provide dense depth estimations together with color images at a high frame rate. This considerably pushed forward several research fields such as: 3D reconstruction [1, 2], camera localization and mapping (SLAM) [3, 4], gesture and object recognition [5, 6], bilateral filtering [7, 8], and many more. Recently, several algorithms have been developed using the Microsoft Kinect v1, since it is one of the most common RGB-D devices. With the release of the Microsoft Kinect v2 a new promising device is available, which uses a new Time-of-Flight (ToF) camera and will – most probably – be the basis for the development and evaluation in many future research.

Our contribution in this paper is a rigorous evaluation and comparison of the depth images of Kinect v1 and Kinect v2. We concentrate on the depth images of the two devices, since they are the core input for many algorithms. The gained results on accuracy and precision can be incorporated or modeled in numerous follow-up algorithms [9]. This includes especially RGB-D 3D reconstruction, SLAM or visual odometry, since their accuracy is directly related to the inaccuracies and noise in the used depth images. We analyze in this paper the influence of temperature, camera distance and scene color on the depth values of both devices. Furthermore, we analyze errors like *flying pixels* and *multipath*



(a) Microsoft Kinect v1 vs. Kinect v2

(b) Accuracy vs. Precision

Fig. 1. We present our systematic comparison and evaluation of the Microsoft Kinect v1 and Kinect v2. More precisely, we investigate the accuracy and precision of the captured depth images.

interference. We hope to provide a fruitful basis for future research and development with the devices. We also summarize and illustrate all our results in the supplementary video.

Because of its recent release, only little work has been published on the Kinect v2 [10]. The precision of the depth images of the single sensors (Kinect v1 or Kinect v2) was already assessed in some publications [10–13] by analyzing the noise properties addressing special applications. Other publications comparing the two devices target towards special application fields of the Kinect like motion tracking [14], face tracking [15] or multimedia [16]. We compare the two sensor in identical environments and in identical experiments in order to draw repeatable conclusions on precision and accuracy of the captured depth images. To the best of our knowledge the accuracy in terms of a metrically correct depth estimation was not assessed so far. State-of-the-art papers measure the distance from the camera case to a seen object with a tape [10] or a laser [11]. But, depth is defined from the camera center to an object, which is hard to measure with their approaches. In our approach we determine ground truth depth estimation for planar surfaces with a checkerboard. This delivers accurate results and enables easy repetition for other researcher using their own Kinect sensors or even other cameras. Our experiments enable us to directly compare the results for the two devices.

2 Preliminaries

We evaluate and characterize in this paper the Microsoft Kinect v1 and Kinect v2, which are RGB-D cameras consisting of one depth and one color camera. The depth image records in each pixel the distance from the camera to a seen object.

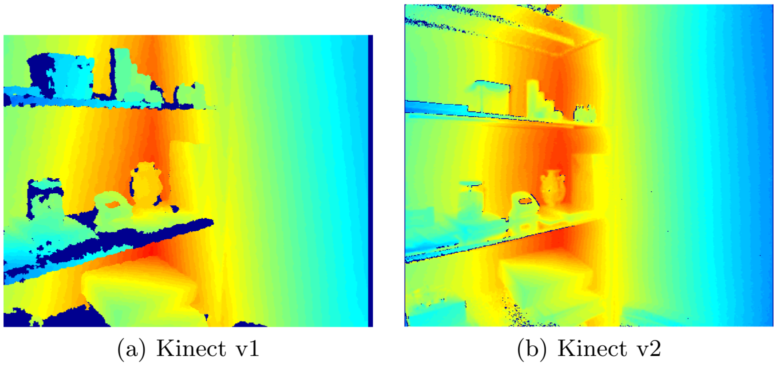


Fig. 2. Captured depth images of the same scene for the Kinect v1 and Kinect v2.

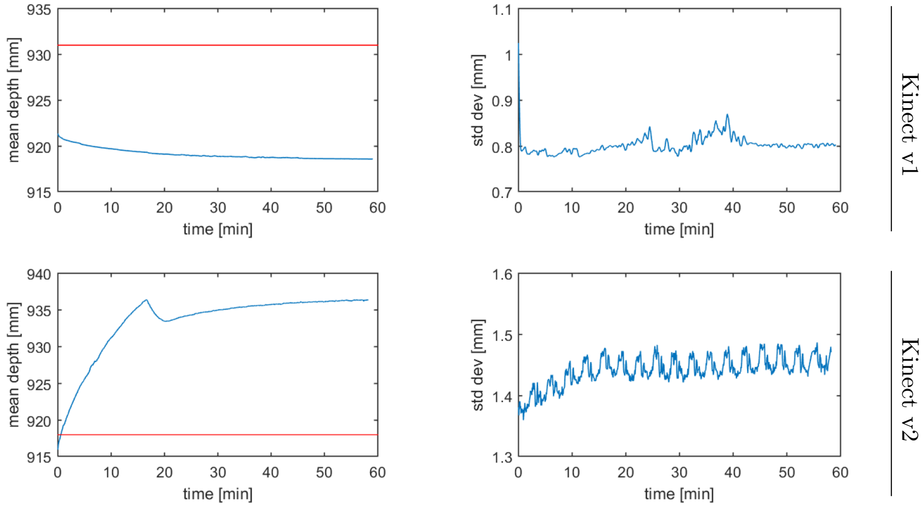
	Kinect v1		Kinect v2	
	Resolution [Pixel × Pixel]	Frame Rate [Hz]	Resolution [Pixel × Pixel]	Frame Rate [Hz]
color	640 × 480	30	1920 × 1080	30
depth	640 × 480	30	512 × 424	30
infrared	640 × 480	30	512 × 424	30

Table 1. Resolution and frame rate of the images captured by a Microsoft Kinect v1 and Kinect v2.

The Kinect v1 measures the depth with the Pattern Projection principle, where a known infrared pattern is projected into the scene and out of its distortion the depth is computed. The Kinect v2 contains a Time-of-Flight (ToF) camera and determines the depth by measuring the time emitted light takes from the camera to the object and back. Therefore, it constantly emits infrared light with modulated waves and detects the shifted phase of the returning light [17, 18]. In the following, we refer to both cameras (Pattern Projection and ToF) as depth camera. We recorded all images in the raw output conditions using the *OpenNI* driver [19] for Kinect v1 and the unofficial *libfreenect2* driver [20] for Kinect v2. This means we recorded the images with the resolutions and frame rates of Table 1. We performed all recordings in an air-conditioned room with constant temperature and without direct sunlight illumination to assure reliable results.

3 Evaluation

In this section, we analyze and describe the properties of Kinect v1 and Kinect v2 depth images. The goal is to investigate the accuracy and precision of the two devices, because this information is required for algorithms like 3D reconstruction, SLAM or visual odometry. Accuracy is defined as the difference/offset of a measured depth value compared to the ground truth distance. Precision is defined as the repeatability of subsequent depth measurements under unchanged



(a) mean depth (b) standard deviation

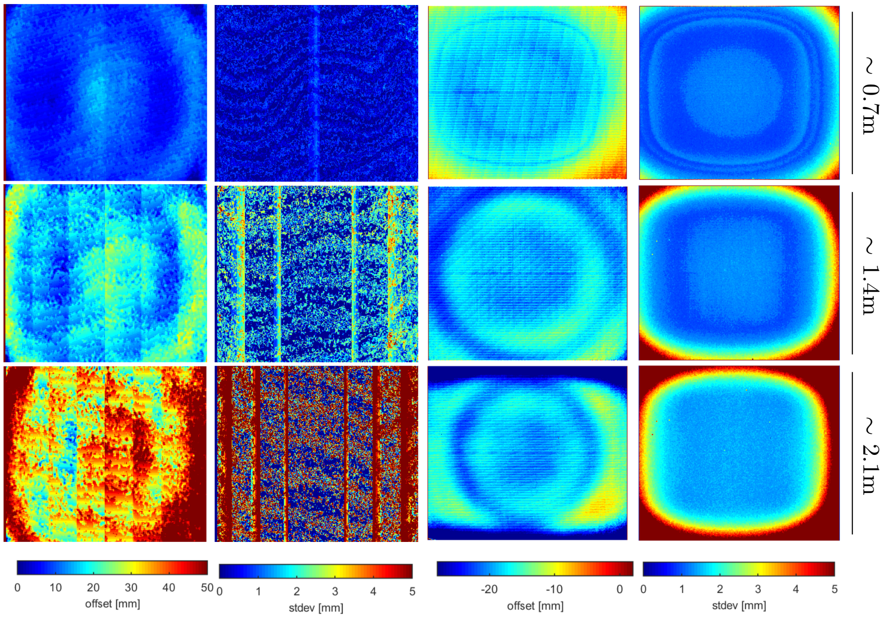
Fig. 3. Evaluation of depth values over time, while the camera heats up and captures a flat wall. For the Kinect v1 (top) the depth values are slightly deceasing but almost constant over time. For the Kinect v2 (bottom) the depth values strongly correlate to the device temperature. The red line depicts the ground truth distance.

conditions. The two definitions are also illustrated in Figure 1b. For determining the accuracy we need to know the ground truth depth. Therefore, we estimate the pose of planar surfaces relative to the camera center and compute the ground truth depth from that relation. Other than state-of-the-art papers [10, 11] we generate the ground truth precisely with the help of a 12 × 10 checkerboard as visible in Figure 1b. The corners of the board can be easily detected in the captured infrared images within subpixel precision [21]. Since the dimensions of the checkerboard are known, we can apply the PnP algorithm [22] to estimate the relative pose of the board. With this information we can describe the wall as a plane and compute a ground truth depth value for each pixel individually. While capturing depth images for the evaluation, the checkerboard was not visible in the scene to avoid its influence (cp. Section 3.3).

Since the images of both cameras exhibit a relative high level of noise, we want to be robust against it and on the other hand describe it. Therefore, we always capture a set of 300 depth images – unless otherwise mentioned – while the camera stands on a stable tripod. For the evaluation of absolute depth values we use the mean depth of the image set in each pixel. The standard deviation is computed based on the deviation in an image set.

3.1 Influence of Temperature

First, we investigate the influence of temperature on the captured depth images. Especially the Time-of-Flight (ToF) camera – or more precisely the infrared

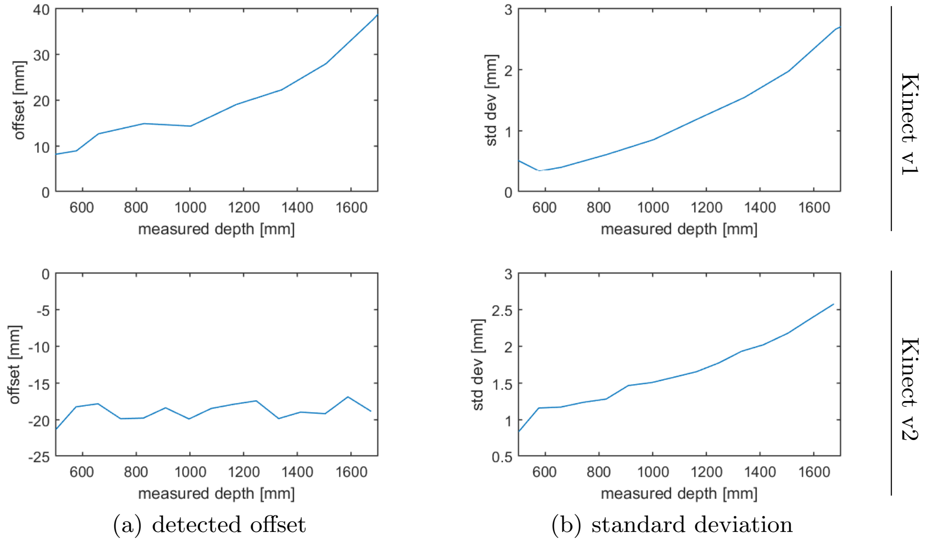


(a) Kinect v1 (b) Kinect v2

Fig. 4. Evaluation of the per-pixel error in depth images in 0.7m, 1.4m and 2.1m distance for (a) Kinect v1 and (b) Kinect v2.

emitter – is getting warm while capturing. Therefore, the Kinect v2 has an integrated fan with a non-influenceable control. Nevertheless, the temperature of the device varies. We mounted cold and recently unused Kinect v1 and Kinect v2 on a stable tripod facing a flat white wall. Then we captured all depth images for a period of one hour and analyzed them. The results of processing these 108,000 images are depicted in Figure 3 showing the mean measured distance and the mean standard deviation over time for both cameras.

The Kinect v1 shows a weak correlation to the temperature. While capturing the measured depth values are decreasing for less than 2 mm. The standard deviation is on an almost constant level of 0.8 mm. In contrast, for the Kinect v2 the distance measurements exhibit a strong correlation to the temperature. In the first 16 min the distance increases constantly for around 20 mm. Then, the fan starts to rotate leading to a distance decrease for 4 min of around 3 mm. Afterwards, the distance increases again in a converging manner of around 3 mm. The standard deviation correlates only weakly to the temperature. It slightly increases until the fan starts to rotate and stays on an almost constant level afterwards. Concluding, we recommend to run the Kinect v2 for at least 25 min before capturing in order to avoid temperature influences. Kinect v1 can already be used after a short warm-up and constant depth values are delivered. The measured distances will be compared to the ground truth distance in Section



(a) detected offset

(b) standard deviation

Fig. 5. Comparison of the captured depth with the ground truth distance for Kinect v1 (top) and Kinect v2 (bottom). (a) While the Kinect v2 has a (almost constant) offset of -18 mm, the Kinect v1 has an exponentially increasing offset of up to 40mm in the analyzed distances. (b) The standard deviation is exponentially increasing for both cameras. Please not the different scales on the y-axis.

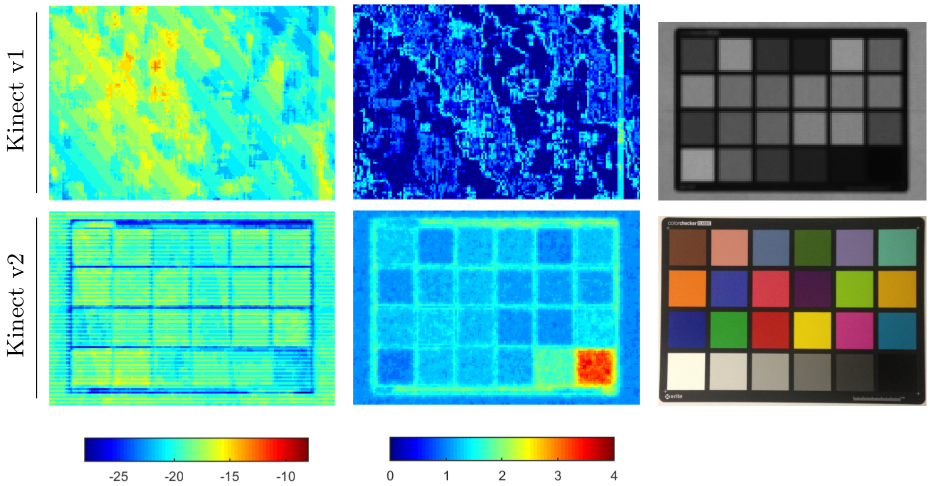
3.2 in order to draw conclusions on the absolute accuracy and precision. For the remaining experiments of this paper we let both cameras warm up for at least one hour.

3.2 Influence of Camera Distance

In this section, we investigate the influence of the camera distance to the scene. Therefore, we again capture a flat wall in several distances with a warm Kinect v1 and Kinect v2 standing on a stable tripod.

The left column of images in Figure 4a and 4b show the offset of depth pixels to the ground truth for three different distances. For the Kinect v1 we detected a stripe pattern in the depth images. The number of stripes increases with the distance to the wall. The stripes lead to an irregular and difficult to model offset in the depth images. In addition, pixels in the image corners have a huge offset. For the Kinect v2 we detected a variable per-pixel offset, which mainly depends on the distance of the pixel to the image center. The corner pixels have a much higher offset than the inner. The reason for it is the infrared light cone, which does not illuminate the scene homogeneously. The infrared light cone and the offset pattern coincide.

Furthermore, we detected for both cameras a mean offset as more detailed in Figure 5a. In this figure only the central pixels are considered, since outer



(a) offset [mm]

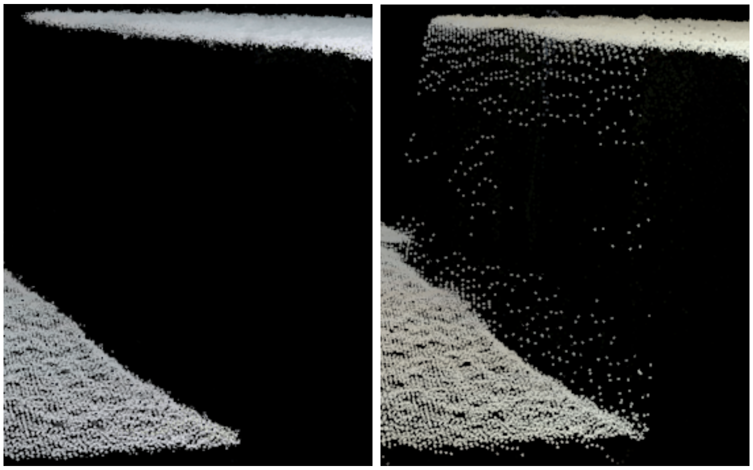
(b) standard deviation [mm]

(c) infrared and color reference images

Fig. 6. Evaluation of the influence of the scene color on the depth values. Whereas Kinect v1 (top) is not influenced by the color, the Kinect v2 (bottom) is affected in terms of (a) offset and (b) standard deviation. As a reference (c) shows the infrared and color image captured by Kinect v2.

pixels are too unreliable. We define the central pixels as a circle with a radius of 300 pixels around the camera center. For the Kinect v1 we detected an exponentially increasing offset for increasing distances. While the offset for 0.5m is below 10mm, the offset increases more than 40mm for 1.8m distance. In contrast, Kinect v2 we detected a offset of on average -18 mm. This means the measured depth values of the Kinect v2 are too deep respectively long. The slight variation is negligible compared to other influence factors.

Next we have a look on the standard deviation, which is shown in the right column of Figure 4a and 4b. For Kinect v1 the standard deviation contains again the stripe pattern and increases with the distance. In contrast, for Kinect v2 the standard deviation in the central pixels is almost constant and increases considerably for the outer pixels. As shown in Figure 5b the standard deviation correlates to the distance for both cameras. However, the standard deviation is lower for Kinect v1 than for Kinect v2 in given distances. Summarized, the precision and accuracy of Kinect v1 decreases with increasing distances. In contrast, the accuracy of Kinect v2 is almost constant over different distances, whereas the precision is also decreasing. Another property, which is visible best in the supplementary video, is the noise behavior. The Kinect v2 incorporates a per-pixel noise, meaning that in case of imprecise measurements the depth values of neighboring pixels strongly differ. The Kinect v1 shows in contrast a per-patch noise, meaning that neighboring pixels have similar values and errors.



(a) Kinect v1 (b) Kinect v2

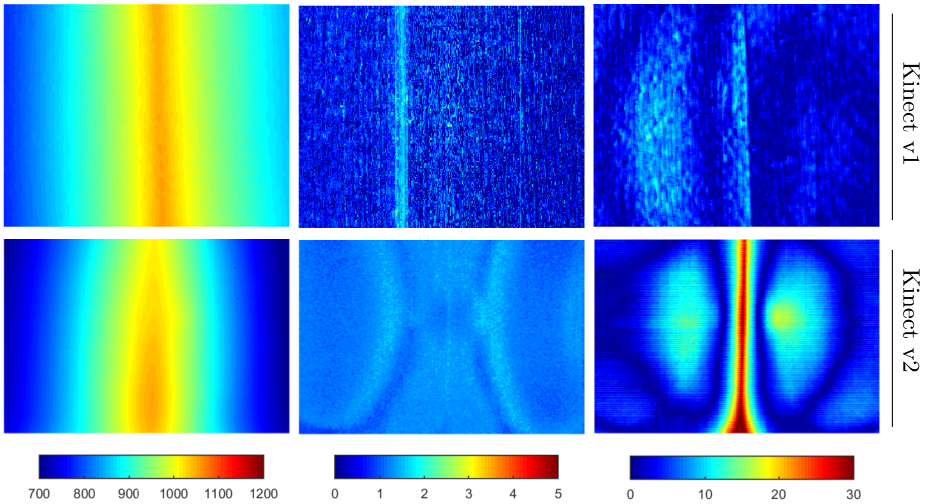
Fig. 7. To show the *flying pixel* effect we recorded two boards lying upon each other. The erroneous pixels in between for the Kinect v2 are *flying pixel.* The effect is even more noticeable in our supplementary video.

3.3 Influence of Object Color

In this section, we evaluate how the color of a scene influences the depth estimation of the two devices. Therefore, we capture a planar x-rite ColorChecker [23] with 24 different colors in around one meter distance. Figure 6 shows the offset from the ground truth and the standard deviation of the depth images for both cameras. It can be clearly seen that the depth estimation of Kinect v2 depends on the scene color, whereas Kinect v1 does not. Dark colors have an up to 10 mm higher depth value than lighter colors for Kinect v2. The scene color has an even more obvious influence on the standard deviation of the depth values (see Figure 6b). The black surface has a standard deviation of up to 4 mm, whereas light colors have a deviation of around 1 mm. By trend, for scene parts with less reflective colors it is less reliable to estimate the depth with the Kinect v2. Thus, for the evaluation of depth offsets and variations it is required to use only colors with similar reflectivity.

3.4 Flying Pixel

In this section, we analyze the so-called *flying pixel*; a well-known artifact for Time-of-Flight (ToF) cameras [24], since all ToF cameras suffer from this problem. *Flying pixels* are erroneous depth estimates which occur close to depth discontinuities as visible in Figure 7b and also on the image boundaries. In this experiment we placed two boards upon each other with a distance of around 200 mm. We captured the scene with a Kinect v1 and a Kinect v2 perpendicular to the boards. Although there should be no geometry in between, there



(a) captured depth [mm] (b) standard deviation [mm] (c) deviation from reference [mm]

Fig. 8. We captured two perpendicular walls in order to analyze the *multipath interference* effect. For Kinect v1 (top) this effect is not visible; only the usual noise is visible (cp. Fig. 4). In contrast, for Kinect v2 (bottom) on each wall a clear bulge with up to 19 mm deviation is visible, caused by reflected light of the other wall.

are several 3D points captured with Kinect v2. The effect is also noticeable in Figure 1 and even more in our supplementary video. In contrast, Kinect v1 does not contain any flying pixels, which makes it much more precise close to depth discontinuities.

3.5 Multipath Interference

In this section we analyze the so-called *multipath interference* effect. Since rays of light are being sent out from the cameras, light can reflect off surfaces in numerous ways and a particular pixel may receive light originally sent out for other pixels as well [25]. This appears often in concave geometries, even without highly reflective surfaces. Both cameras might be affected by this interference, since both cameras send out light in order to measure the depth. To analyze this effect we captured images of two perpendicular walls leading to the results of Figure 8, where we considered only central pixels in order to eliminate other effects. In Figure 8c we compare the depth values with the reference values. For Kinect v1 this effect is not visible at all. Only the usual noise is visible that we already investigated in Section 3.2. In contrast, for Kinect v2 on both walls a clear bulge is visible with up to 19 mm deviation. This effect is even more noticeable in our supplementary video. The standard deviation of both cameras is negligibly influenced by this artifact as shown in Fig. 8b. For Kinect v2 we detected a huge offset in the intersection of the two walls, which is caused by smoothing algorithms inside the camera hardware that can not be influenced. Kinect v1 does not show this smoothing effect and captures a sharp edge.

4 Conclusion

In this paper, we systemically evaluated and analyzed the two versions of Microsoft Kinect. We concentrated on the depth images, since they are the core input for algorithms like 3D reconstruction, SLAM or visual odometry. The goal was to investigate the accuracy and precision of the depth images of both devices in order to give suggestions for research in follow-up algorithms.

First of all we figured out a strong correlation of depth accuracy and temperature for the Kinect v2, resulting in the recommendation to pre-heat the device for at least 25 min in order to achieve reliable results. The Kinect v1 captures reliable images already after a few initial images. With our precise ground truth generation using a checkerboard we are able to proof the accuracy and precision of the captured depth images in different distances. While the accuracy decreases exponentially with increasing distance for Kinect v1, Kinect v2 has a constant accuracy in form of an offset of -18 mm. This is a very important fact for the above mentioned applications, since a constant offset can be easily modeled. In addition, Kinect v1 incorporates the stripe pattern in the depth images, which is difficult to compensate. For Kinect v2 all central pixels show a similar accuracy; only the image corners deviate. On the other hand, the precision of the depth images is higher for Kinect v1. This holds for flat surfaces, but especially for depth discontinuities, where *flying pixels* occur for Kinect v2. In follow-up algorithms these imprecisions must be modeled or compensated. The respective literature presents several approaches for that such as bilateral filtering [7, 8] or fusion of subsequent depth images [2, 26]. Furthermore, the depth estimation of Kinect v2 is influenced by the scene color, whereas Kinect v1 is unaffected. This has to be considered in bilateral filtering approaches, since they rely on coinciding color and depth changes [7, 8]. In contrast to Kinect v1, Kinect v2 depth images are influenced by the *multipath interference* effect, meaning that concave geometry is captured with bulges. The respective literature proposes approaches to compensate for this effect [25].

Summarized, we recommend to use Kinect v2 in the context of 3D reconstruction, SLAM or visual odometry. Kinect v2 has the higher accuracy, which is difficult to enhance in an algorithmic way. However, due to the lower precision we recommend to apply many pre-processings on the depth images before using them. This includes compensations of random noise, *flying pixels* and *multipath interference*. These pre-processings are not necessary (in that extend) for Kinect v1, which makes it suitable for fast prototypes. The results of our evaluation can be used to incorporate or model the errors of the respective device in follow-up algorithms. We hope to provide a fruitful basis that considerably pushes forward further research with the devices.

References

1. Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohi, P., Shotton, J., Hodges, S., Fitzgibbon, A.: Kinectfusion: Real-time dense surface mapping and tracking. In: IEEE International Symposium on Mixed and Augmented Reality (ISMAR). (2011)
2. Wasenmüller, O., Meyer, M., Stricker, D.: Augmented reality 3d discrepancy check in industrial applications. In: IEEE International Symposium on Mixed and Augmented Reality (ISMAR), IEEE (2016) 125–134
3. Kerl, C., Sturm, J., Cremers, D.: Robust odometry estimation for rgb-d cameras. In: IEEE International Conference on Robotics and Automation (ICRA), IEEE (2013) 3748–3754
4. Wasenmüller, O., Meyer, M., Stricker, D.: CoRBS: Comprehensive RGB-D benchmark for SLAM using kinect v2. In: IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE (2016)
5. Chen, C., Jafari, R., Kehtarnavaz, N.: Utd-mhad: a multimodal dataset for human action recognition utilizing a depth camera and a wearable inertial sensor. In: IEEE International Conference on Image Processing (ICIP), IEEE (2015) 168–172
6. Chandra, S., Chrysos, G.G., Kokkinos, I.: Surface based object detection in rgbd images. In: Proceedings of the British Machine Vision Conference (BMVC), BMVA Press (2015) 187.1–187.13
7. Vianello, A., Michielin, F., Calvagno, G., Sartor, P., Erdler, O.: Depth images super-resolution: An iterative approach. In: IEEE International Conference on Image Processing (ICIP). (2014) 3778–3782
8. Wasenmüller, O., Bleser, G., Stricker, D.: Combined bilateral filter for enhanced real-time upsampling of depth images. International Conference on Computer Vision Theory and Applications (2015)
9. Wasenmüller, O., Ansari, M.D., Stricker, D.: Dna-slam: Dense noise aware slam for tof rgb-d cameras. In: Asian Conference on Computer Vision Workshop (ACCV workshop), Springer (2016)
10. Fankhauser, P., Bloesch, M., Rodriguez, D., Kaestner, R., Hutter, M., Siegwart, R.: Kinect v2 for mobile robot navigation: Evaluation and modeling. In: International Conference on Advanced Robotics (ICAR). (2015)
11. Lachat, E., Macher, H., Mittet, M., Landes, T., Grussenmeyer, P.: First experiences with kinect v2 sensor for close range 3d modelling. In: International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (ISPRS). (2015)
12. Butkiewicz, T.: Low-cost coastal mapping using kinect v2 time-of-flight cameras. In: Oceans-St. John's, 2014, IEEE (2014) 1–9
13. Fürsattel, P., Placht, S., Balda, M., Schaller, C., Hofmann, H., Maier, A., Riess, C.: A comparative error analysis of current time-of-flight sensors. IEEE Transactions on Computational Imaging **2** (2016) 27–41
14. Samir, M., Golkar, E., Rahni, A.A.A.: Comparison between the kinect v1 and kinect v2 for respiratory motion tracking. In: IEEE International Conference on Signal and Image Processing Applications (ICSIPA), IEEE (2015) 150–155
15. Amon, C., Fuhrmann, F., Graf, F.: Evaluation of the spatial resolution accuracy of the face tracking system for kinect for windows v1 and v2. In: Proceedings of the 6th Congress of the Alps Adria Acoustics Association. (2014)
16. Zennaro, S., Munaro, M., Milani, S., Zanuttigh, P., Bernardi, A., Ghidoni, S., Menegatti, E.: Performance evaluation of the 1st and 2nd generation kinect for

multimedia applications. In: 2015 IEEE International Conference on Multimedia and Expo (ICME), IEEE (2015) 1–6

17. Sell, J., O'Connor, P.: The xbox one system on a chip and kinect sensor. IEEE Micro (2014) 44–53
18. Lefloch, D., Nair, R., Lenzen, F., Schäfer, H., Streeter, L., Cree, M.J., Koch, R., Kolb, A.: Technical foundation and calibration methods for time-of-flight cameras. In: Sensors, Algorithms, and Applications: Time-of-Flight and Depth Imaging. Springer (2013) 3–24
19. (OpenNI) http://www.openni.org.
20. Blake, J., Echtler, F., Kerl, C.: (libfreenect2) https://github.com/OpenKinect/libfreenect2.
21. Rufli, M., Scaramuzza, D., Siegwart, R.: Automatic detection of checkerboards on blurred and distorted images. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE (2008) 3121–3126
22. Lepetit, V., Moreno-Noguer, F., Fua, P.: Epnp: An accurate o (n) solution to the pnp problem. International Journal of Computer Vision (IJCV) **81** (2009) 155–166
23. x rite: ColorChecker Classic. (http://xritephoto.com/ph_product_overview.aspx ?ID=1192) Accessed 2016.
24. Gottfried, J., Nair, R., Meister, S., Garbe, C., Kondermann, D.: Time of flight motion compensation revisited. In: IEEE International Conference on Image Processing (ICIP), IEEE (2014) 5861–5865
25. Freedman, D., Smolin, Y., Krupka, E., Leichter, I., Schmidt, M.: Sra: Fast removal of general multipath for ToF sensors. In: European Conference on Computer Vision (ECCV). (2014) 234–249
26. Cui, Y., Schuon, S., Thrun, S., Stricker, D., Theobalt, C.: Algorithms for 3d shape scanning with a depth camera. IEEE transactions on pattern analysis and machine intelligence **35** (2013) 1039–1050