2019 16th International Conference on Ubiquitous Robots (UR)
Jeju, Korea, June 24-27, 2019

Analysis and Noise Modeling of the Intel RealSense D435 for Mobile Robots

Min Sung Ahn, Hosik Chae, Donghun Noh, Hyunwoo Nam, and Dennis Hong[1]

Abstract— **Cameras that provide distance measurement along with RGB data have increasingly been appearing in the market as alternatives to the more expensive setup of LIDARs and webcams. While products such as the Kinect have existed in the past, its weight and form factor have been demanding constraints for mobile robots, specifically legged robots that are sensitive to payload. Recently Intel released a new lineup of Intel RealSense RGB-D cameras that have favorable characteristics for legged robots, specifically in terms of resolution, frames per second, form factor, weight, and price range. However, because these active stereo sensors are noisy for reasons such as non-overlapping image regions or lack of texture, it is beneficial to empirically model the noise. Systematic errors, specifically the distance inhomogeneity and depth bias, are observed to recognize and verify the limitations of the camera. We also analyze the non-systematic error by modeling both the axial and lateral noise as a function of distance and angle of incidence using a Gaussian distribution for its versatile applicability for mobile robots in mapping.**

I. INTRODUCTION

As mobile robots transition from a laboratory setting to the outside world, perceiving its surrounding environment becomes necessary. Autonomous vehicles mostly rely on LIDARs to sense their surroundings. However, noticeable downsides are their price range, accuracy in the short range, and sometimes their form factor. Consequently, although some mobile robots do make use of LIDARs, the release of affordable depth cameras such as the Microsoft Kinect and the Asus Xtion Pro have seen many modern mobile robots adopt these depth sensors [1]. The price range of some depth sensors such as RGB-D cameras are typically magnitudes less than that of the LIDARs, while also providing superior accuracy in the short range. Additionally for mobile robots, characteristics such as a camera's form factor, weight, and field of view (FOV) are important as for example, although numerous successful research has been done using the Microsoft Kinect, its bulky size and weight may be infeasible for a mobile robot that cannot carry much payload.

In January 2018, Intel released a new line-up of RGB-D cameras under its RealSense family. Effectively replacing all previous generations, the Intel RealSense D400 was announced with two active stereo cameras in its family to begin with. The D415 has a higher pixel density due to its higher resolution but lower FOV with a rolling shutter, making it a sensor suitable for a static environment. On the other hand, the D435 has a wider FOV with a global shutter, making it relatively better for capturing a moving environment. Recently, extensive analysis on the D415 has been done. One work has tested the D415 in near range situations (100∼1000 mm) and also compared it with other sensors designed for close range [2]. Similarly, another conducted experiments on not only the D415, but also for the Microsoft Kinect v2, the Orbbec Astra S, and partially for the D435 [3]. Pixel-wise characterization was done by pointing the sensors at a white planar target to capture the pixel-level uncertainty. Sensor-wise characterization was done by assessing the sensors ability to measure a known geometry. While these two works do report depth accuracies, for mobile robots, a more informative model of the noise can be useful for potential post-processing purposes.

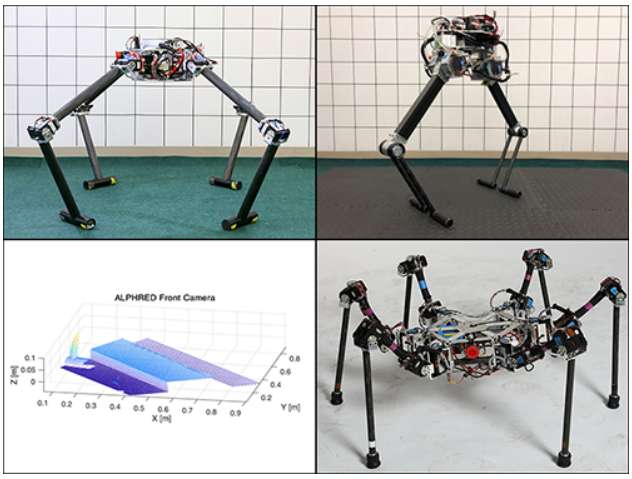

Fig. 1. Legged robots ALPHRED (top left), NABi (top right), and SiLVIA (bottom right) that use the Intel RealSense D435 for mapping its local surroundings. An example of ALPHRED's view is shown in the bottom left.

Work has been done in modeling the Microsoft Kinect v2 for mobile robot navigation purposes [4]. Systematic and non-systematic errors were assessed, with specific focus on depth and amplitude errors. The modeling approach follows the methods taken in [5], where the axial and lateral noise of the Kinect v2 was measured and a model that is a function of distance and angle was fitted for both. Both methods propose using these models to enhance the performance of existing 3D reconstruction strategies.

In a similar fashion, with the Intel RealSense D435, we attempt to provide metrological characterization of the

*This research was supported by a grant (code R2016001) from Gyeonggi Technology Development Program funded by Gyeonggi Province.

[1]All authors are with the Robotics and Mechanisms Laboratory (RoMeLa) at the University of California, Los Angeles (UCLA), Los Angeles, CA 90095, USA. The corresponding author can be reached at aminsung@ucla.edu.

TABLE I

DETAILED COMPARISON BETWEEN DEPTH SENSORS.

Model	D435	D415	Kinect v2
Resolution	848x480	1280x720	512x424
Frame Rate	90	90	30
Shutter Type	Global	Rolling	Global
Dimensions (mm) LxWxH	90x25x25	99x23x20	249x66x67
Weight (g)	72	72	970

sensor, as well as its axial and lateral noise models at varying angles of incidence. The key difference between the Intel RealSense D435 and the Microsoft Kinect and the Kinect v2 is that aside from the extreme difference in form factor and weight, the D435 is equipped with a stereo camera while the Kinects do not. Thus, in scenes where the structured light technology from the Kinect and the time-of-flight (ToF) principle used in Kinect v2 falls short from its fragile nature and susceptibility to interference from sunlight, an active stereo camera can still manage to compute reasonable depth values.

Thus, this paper's primary contribution is in that to the authors' best knowledge, no such analysis has been done for what is a promising sensor for mobile robots such as those shown in Figure 1. This paper first briefly introduces the Intel RealSense D435 technology in Section II. Section III presents the systematic error analysis, while Section IV explains the noise modeling approach used. Section V analyzes this captured data and presents an axial and lateral noise model for the D435. Section VI concludes the paper.

II. REALSENSE D435

The Intel RealSense D435 is an active stereo depth camera that uses Intel's custom ASIC, the Intel RealSense Vision Processor D4, to conduct a custom variant of the Semi Global Matching algorithm to compute the depth. It also has an optional infrared (IR) projector that assists in improving the depth accuracy by projecting a non-visible static IR pattern when the scene's texture is low. Additionally, unlike the Kinect v2, which can get depth at a resolution of 512×424 at up to 30 frames per second (FPS), the D435 can get up to 848×480 at up to 90 FPS. Its form factor and weight is also significantly smaller and less. Compared to the D415, the D435 has a global image shutter which makes it a better candidate for mobile robots that perceives the environment as moving. A more detailed comparison between the two D400 cameras and the Kinect v2 is shown in Table I.

Our data collection setup uses Intel's official software development kit[1], which supports wrappers to various different languages. In our case, Python was used to record the data.

III. ERROR ANALYSIS

Although the Intel RealSense improves on previous technology by using a stereo matching algorithm assisted with an IR projector to make the sensor less susceptible to interference from different lighting conditions, it is still vulnerable due to a multitude of reasons. Intel claims that its built-in algorithm uses a complex combination of all the parameters in the "Advanced Mode" to calculate the depth stream and suggests using one of the pre-defined depth settings available in the SDK. Nonetheless, we attempted to modify each parameter extensively, but usually saw minimal gain in performance. Therefore, this work uses the default settings after calibrating the cameras using Intel's Dynamic Calibration tool and verifying the depth quality using the depth quality tool provided by the SDK.

[1]https://github.com/IntelRealSense/librealsense

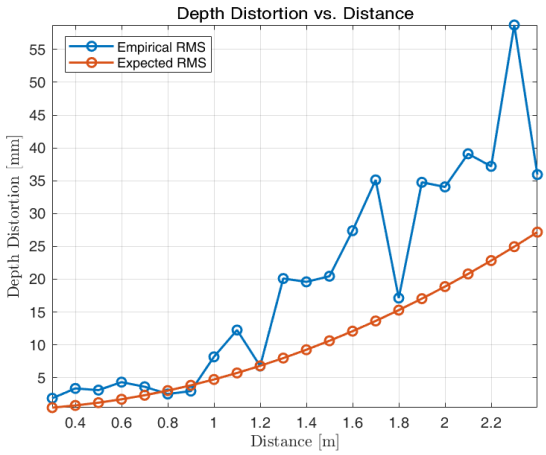

Fig. 2. Comparison between empirical and expected RMS depth error as a function of distance.

In a calibrated state, we analyze and model the systematic errors and non-systematic errors respectively. The most prominent type of error of the former is what is known as "distance inhomogeneity." It refers to the potential depth error at different distances.

To observe this error, the sensor's front glass is positioned in parallel with a flat white plane, and distance measurements to the plane are collected at a known distance. The average of the distance measurements are then compared with the ground truth value to compute the error. In fact, with stereo sensors such as the D435, an expected RMS error e exists, and can be calculated by:

$$e = \frac{d^2 \times s}{f \times b}$$
$$f = \frac{0.5 \times p_x}{tan(\frac{V_h}{2})} \quad (1)$$

where f is the focal length, p_x is the X resolution in pixels, V_h is the horizontal field of view, d is the distance in millimeters, s is the sub-pixel error, and b is the baseline.

After calibrating the camera and checking the depth quality by confirming its subpixel RMS error, the camera was positioned from 0.3 m to 2.4 m at 0.1 m intervals away from a white wall, which is verified using the Bosch GLM35 laser distance measuring device which has an accuracy of ± 1.5 mm. Its disparity shift was set to 0 and the laser power was set at 150 mW. For fair comparison with the expected RMS error, the RMS error of the difference between the average of

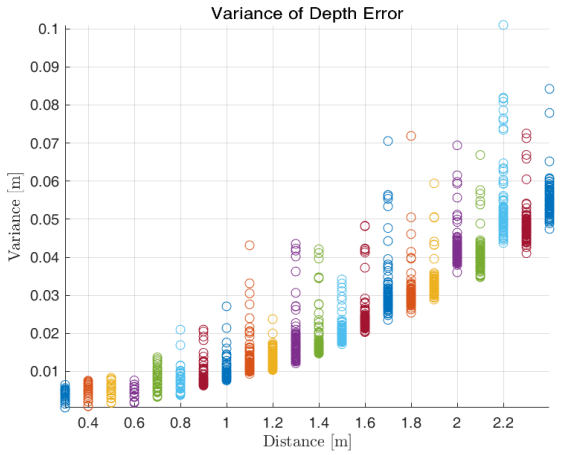

Fig. 3. Variance of 100 samples at different distances.

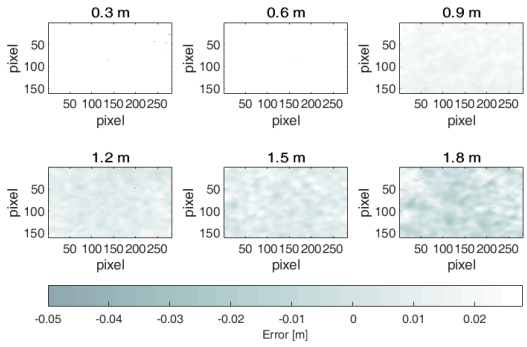

Fig. 4. Distance between each pixel and a fitted plane for a subset of distances to observe the depth bias.

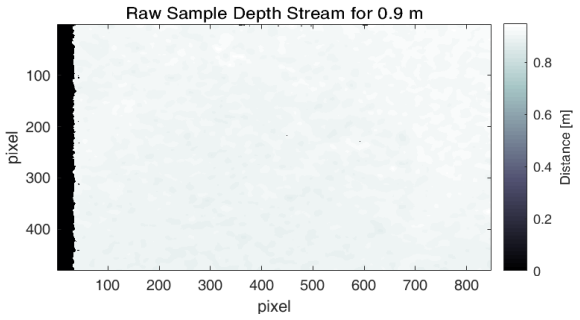

Fig. 5. No data or "shadowing" region on the left of a raw depth stream due to a non-overlapping region.

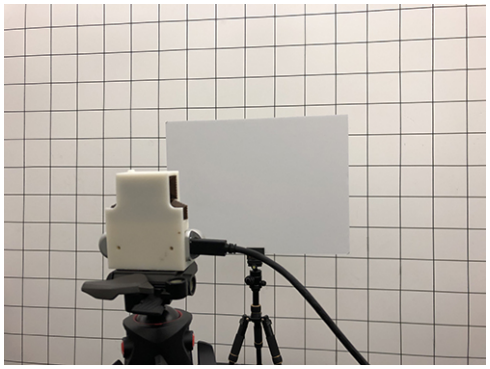

Fig. 6. Actual setup used to gather data at different distances and angles. A distance measurement device with 1.5 mm accuracy was used to measure distances and a video tripod precisely rotated the planar target to the desired angles.

100 samples and the known distance is computed. Figure 2 shows a comparison between the empirical and the expected error and the empirical data shows that in reality, more error can be expected, but not by much from the expected values. Qualitatively, it can be observed that the error is bound under 5 mm up to $\sim$1000 mm, but grows quadratically afterwards. Interestingly, the variance of the error also grows quadratically as shown in Figure 3.

Additionally, systematic bias observation is attempted by fitting a plane to the captured depth values of the white wall and computing each depth value's distance to it. Unlike a systematic circular shaped bias observed towards the center of the D415 [3], as seen in Figure 4, a clear pattern does not exist. We suspected averaging the 100 samples as potentially squashing the behavior and conducted the same calculations for a single frame, but no clear pattern was observable.

IV. MODELING APPROACH

Beyond the systematic errors, we are interested in capturing the D435's 3D noise distribution (i.e. non-systematic error) by measuring its axial and lateral noise, for use in various applications such as mapping for legged robots. Because the D435 has two IR imagers for stereo vision, with the image stream from the left used as reference, the resulting depth data has a section with no data due to a non-overlapping region, as seen on the left of a raw depth stream in Figure 5. When the scene is closer, this region is greater, as opposed to when it is farther away.

We follow an approach traditionally done on the Kinect in the axial direction [6] [7] which has also been extended for the lateral direction [5]. Following the common convention shown in Figure 7, the direction the camera is facing is designated as the Z axis and the rotating axis is the Y axis. The sensor was mounted on a tripod with a leveling base, facing a flat white board mounted on a video tripod that could rotate about the vertical axis at precise angles, as seen in Figure 6.

The camera took multiple rectangular samples sufficiently inside the planar board at different evenly spaced θ_y were also taken. z distances away from it. Additionally, as the angle the robot is looking at its surrounding can also increase the depth noise, a noise model as a function of the angle of incidence was also sought for. To do this, multiple samples at different

In our work, the axial noise is considered to be the pixel-wise standard deviation of the distance between the depth value and a plane fitted on the region. Similarly, the lateral noise is computed as the variation of the pixels about a line fitted along the left and right vertical edges of the planar target. Data collection was completely automated bar

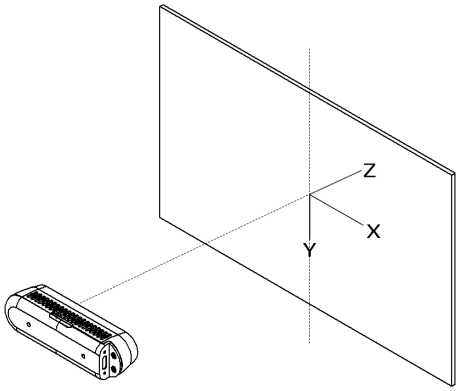

Fig. 7. The flat white target is positioned a known z distance rotated θ_y degrees. The coordinates follow the common camera convention.

the manual moving of the camera to different z distances and rotating the planar target to different θ_y angles in a clutter-free room. A series of bilateral filtering, Canny edge detection, and shape detection methodologically found the planar target. Following this, vertical edges were found with the corners of the planar board removed, while a rectangle in the center of the target and sufficiently away from the edges to avoid potential corruption from the edge noises was also found. This data was recorded 100 times per combination of z distance and θ_y rotation.

V. NOISE MODELING

Considering the dimensions of mobile robots such as ALPHRED [8], NABi [9], and SiLVIA [10], a distance range of 0.4 m to 2.1 m was selected with angles from $0°$ to $75°$ at $15°$ intervals. Note that because of the aforementioned "shadowing" effect due to the depth image being about the left IR imager, during rotation from $0°$ to $75°$, shadowing at the left edge of the planar target noticeably worsens relative to the right edge. Consequently, two independent lateral noise analyses are done for the left and right edge.

We are interested in modeling the noise under a Gaussian distribution for its versatile applicability in a variety of different algorithms. Observing the axial noise distributions in Figure 8, it is clear that a Gaussian distribution can be fit to the data. For verification purposes, an identical analysis was done on the data obtained from the white wall, which also showed a Gaussian distribution.

Based on the curves fit about the obtained standard deviations, an approximation of the noise σ_z, as a function of distance z and angle θ_y, is obtained and shown in Figure 9 and 10. Axial noise results in a quadratic behavior as a function of z. To incorporate the noise model exploding to infinity towards 90, a hyperbolic term with heuristically obtained values of 0.022 and $z^{3/2}$ are included as in previous works to embed this phenomena.

$$\sigma_z(z,\theta_y) = \; 0.001063 + 0.0007278z + 0.003949z^2 + 0.022z^{3/2}\frac{\theta_y}{(\pi/2-\theta_y)^2} \quad (2)$$

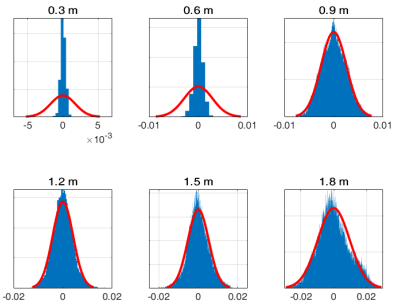

Fig. 8. Axial depth measurements for a subset of the distances are shown. It is clear that the axial noise can be modeled as a Gaussian.

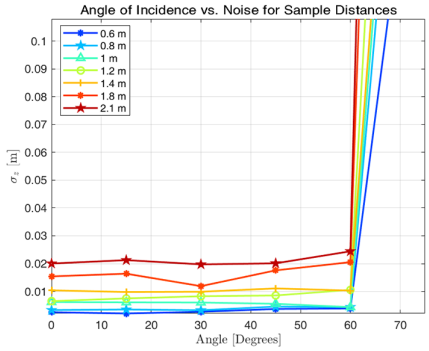

Fig. 9. For selected distances, axial noise as a function of angle is shown. As the angle approaches $90°$, noise approaches infinity.

Contrary, the lateral noise, possibly due to the limitations imposed by the shadowing effect, did not exactly have a normal distribution. Nonetheless, we force a fit to see if a qualitative trend exists among the different distances over the angles. However, as seen in Figure 11, no qualitative behavior is recognizable between the noise values obtained at different distances. Thus, a conservative 90th percentile of the sampled lateral noise values are selected for the left edges (σ_L) and the right edges (σ_R).

$$\sigma_L = 0.0432\,\text{m}$$
$$\sigma_R = 0.0407\,\text{m}$$

VI. CONCLUSION

The Intel RealSense D435 camera is a promising depth sensor for mobile robots, unlike LIDARs which are often magnitudes more expensive for only superior performance in long range sensing, while the Kinect v2 is considerably bulky and heavy. The D435's form factor, weight, and price range is especially more attractive for legged robots that are sensitive to additional payload it needs to carry on top of the bare minimal hardware required for locomotion.

With this in mind, in this paper, we empirically collected depth data of a flat white wall as well as depth data of a planar target at every combination of a chosen array of distance and angles. We were able to verify the quadratic nature of

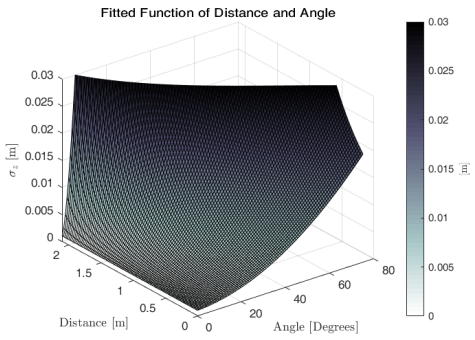

Fig. 10. The axial noise can be seen as a surface that peaks towards infinity regardless of the distance value as the angle approaches 90°. The drawn surface is cut off at 0.03 to show the quadratic nature at angles less than 90.

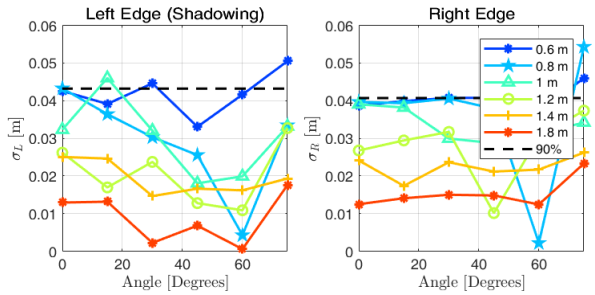

Fig. 11. Lateral noise for selected distances are shown for both the left edge of the planar target, where shadowing exists, and the right edge. No identifiable behavior exists.

depth error as a function of distance, and also observed that its variance also grows quadratically. Furthermore, models of axial noise, as a function of distance z and angle θ_y, as well as a conservative constant value for the left and right lateral noise were found. We believe that this noise model can be used for mobile robots using the D435 for applications that assume Gaussian noise. Our team plans to apply this model in refining our mapping algorithms for our legged platforms.

REFERENCES

[1] M. S. Ahn, H. Chae, and D. W. Hong, "Stable, autonomous, unknown terrain locomotion for quadrupeds based on visual feedback and mixed-integer convex optimization," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 3791–3798.

[2] M. Carfagni, R. Furferi, L. Governi, C. Santarelli, M. Servi, F. Uccheddu, and Y. Volpe, "Metrological and critical characterization of the intel d415 stereo depth camera," *Sensors*, vol. 19, no. 3, p. 489, 2019.

[3] S. Giancola, M. Valenti, and R. Sala, "Metrological qualification of the intel d400 active stereoscopy cameras," in *A Survey on 3D Cameras: Metrological Comparison of Time-of-Flight, Structured-Light and Active Stereoscopy Technologies*. Springer, 2018, pp. 71–85.

[4] P. Fankhauser, M. Bloesch, D. Rodriguez, R. Kaestner, M. Hutter, and R. Siegwart, "Kinect v2 for mobile robot navigation: Evaluation and modeling," in *2015 International Conference on Advanced Robotics (ICAR)*. IEEE, 2015, pp. 388–394.

[5] C. V. Nguyen, S. Izadi, and D. Lovell, "Modeling kinect sensor noise for improved 3d reconstruction and tracking," in *2012 second international conference on 3D imaging, modeling, processing, visualization & transmission*. IEEE, 2012, pp. 524–530.

[6] J. Smisek, M. Jancosek, and T. Pajdla, "3d with kinect," in *Consumer depth cameras for computer vision*. Springer, 2013, pp. 3–25.

[7] K. Khoshelham and S. O. Elberink, "Accuracy and resolution of kinect depth data for indoor mapping applications," *Sensors*, vol. 12, no. 2, pp. 1437–1454, 2012.

[8] J. Hooks and D. Hong, "Implementation of a versatile 3d zmp trajectory optimization algorithm on a multi-modal legged robotic platform," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 3777–3782.

[9] J. Yu, J. Hooks, X. Zhang, M. S. Ahn, and D. Hong, "A proprioceptive, force-controlled, non-anthropomorphic biped for dynamic locomotion," in *2018 IEEE-RAS 18th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2018, pp. 1–9.

[10] X. Lin, H. Krishnan, Y. Su, and D. W. Hong, "Multi-limbed robot vertical two wall climbing based on static indeterminacy modeling and feasibility region analysis," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 4355–4362.