Velocity Estimation Using an RGB-D Sensor

Everett Bryan & Bryce Pincock

Abstract—We characterized the noise in Kinect range data in order to improve the accuracy of simulated Kinect range measurements. We apply a Minimum Mean Squared Error (MMSE) linear estimator to the simulated range measurements to estimate an Autonomous Ground Vehicle’s (AGVs) velocity. We successfully characterized the noise and can simulate Kinect range measurements. We also estimate an AGV’s velocity with less than 5% error.

I. INTRODUCTION

An AGV must be able to estimate its states in order to safely navigate an environment. AGVs estimate their states by interpreting data provided by a GPS sensor. When a GPS signal is unavailable, other proximity sensors must be used for state estimation. The Microsoft Kinect sensor is a proximity sensor well suited for state estimation and navigation. We will use the Kinect to estimate one of the vehicle state variables, velocity.

A. The Kinect Sensor

The Microsoft Kinect combines 3D depth map information with traditional RGB color camera data. Although the technology is still in its early stages, this low-cost sensor has taken the place of the more commonly used stereo cameras or scanning laser range finders. The advantages of this sensor over stereo cameras is the low computational expense in calculating depth and the accuracy of the range measurements. It is unknown to the authors if any research has been done to characterize the noise of depth measurement of this sensor. A more thorough understanding of the noise within the Kinect enables the development of accurate and sophisticated algorithms for use with this sensor. The majority of our work is the characterization of the Kinect’s noise as well as one application of using this information to develop an accurate algorithm.

II. NOISE CHARACTERIZATION

The core of our work is to characterize the noise in the Kinect. Once this noise is known, suitable algorithms can be designed to optimize the use of the depth maps returned by the Kinect. In this section, we describe our process in characterizing the noise. We describe our data collection procedure, error analysis and noise analysis.

A. Data Collection

In our experiment, we pointed the Kinect at a flat wall such that the Kinect was parallel to the wall, as shown in Figure 1. We moved the Kinect away from the wall at 1cm increments. We began collecting data at 50cm from the wall and ended at 377cm. The total data collected includes more than 100 million individual data points. With this number of data we aim to correctly model Kinect range measurement noise.

All of this data is included in our analysis even though the specified practical range is 0.8 – 3.5m [1]. The Field-Of-View (FOV) of the Kinect is 59° in the horizontal and 45° in the vertical. This FOV is so wide that at the largest range the depth map returns range measurements from the ceiling and floor. We therefore truncate the depth map to ignore the regions where the floor and wall exist in the maps. This truncation step simplifies our analysis and ensures that our results are accurate and consistent.

B. Error Analysis

Our experimental setup was motivated by our desire to measure truth data. Because the Kinect was pointed at a flat wall, we were able to measure range truth. The error calculation was easily computed as,

\[ E = M_{measured} - M_{truth}, \hspace{1cm} (mm) \]  

where \( M_{measured} \) is the depth map returned by the Kinect and \( M_{truth} \) is the depth map with no noise present. One of these error maps can be seen in Figure 2. From the error maps we concluded that the noise is a deterministic function of distance added to the random noise present in all systems. We modeled each measurement from the Kinect as,

\[ M_{measured} = M_{truth} + f(z) + \nu, \hspace{1cm} (mm) \]  

where \( f(z) \) is the deterministic noise in the system and \( \nu \) is the random noise in the system. The deterministic noise of the system, \( f(z) \), is a function of the measured range.
C. Noise Analysis

In order to characterize the random noise in the system, we use Equation 1 and Equation 2 to get

\[ \nu = M_{\text{measured}} - M_{\text{truth}} - f(z) = E - f(z). \]  

(3)

In Equation 3 we see that \( \nu \) is a function of the deterministic noise. Therefore, we must first find the deterministic noise, \( f(z) \), to find the random noise \( \nu \). Using the least squares method, we fit a polynomial to \( f(z) \). We chose a 6th order polynomial to represent \( f(z) \) because the lower order polynomials were not flexible enough to follow the behavior of \( f(z) \). The deterministic behavior of the error for each pixel is expressed by,

\[ f(z) = c_0 + c_1 \cdot r + c_2 \cdot r^2 + \cdots + c_6 \cdot r^6. \]  

(4)

The values of the coefficients exhibit a linear behavior with range indicating that the deterministic function for error depends on the range reading. For simplicity, the coefficients were estimated using a least squares error algorithm.

The behavior of \( f(z) \) can be seen in Figure 3. Given the \( f(z) \) for this same pixel, its noise \( \nu \) was computed, as shown in Figure 4. From the histogram of the noise, we see that the noise is gaussian, as shown in Figure 5. The gaussian distributions fit to each pixel are each pixel’s marginal pdf.

D. Noise Characterization Results

If the noise characterization has been done correctly then the simulated Kinect data should look something like actual Kinect range data. As you can see from Figure 6, the simulated error does in fact look much like the actual Kinect error data in Figure 2.

In order to completely characterize the noise of the Kinect, we must now find the joint pdfs.

III. VELOCITY ESTIMATION

By characterizing the noise in the Kinect range measurements, we used this information to obtain accurate velocity
estimates. The velocity of an AGV can be estimated by

\[ \hat{V} = Hz \cdot (\hat{r}_t - \hat{r}_{t-1}) \quad m/s, \]  

(5)

where \( \hat{V} \) is the estimated velocity, \( Hz \) is the scan frequency of the Kinect measured in hertz and \( \hat{r}_t \) and \( \hat{r}_{t-1} \) are successive range measurements obtained at times \( t \) and \( t-1 \) respectively. The range measurements \( \hat{r}_t \) and \( \hat{r}_{t-1} \) correspond to ranges of some feature. Equation 5 can easily be extended to handle multiple features. Since the purpose of this paper is to illustrate the development of a MMSE linear estimator to correct for the noise characteristics of the Kinect, we do not focus on the difficult task of feature extraction and data correspondence. Therefore, in order to simplify feature extraction from the Kinect data, we assumed the \( 10 \times 10 \) pixels in the center of each depth map to be a feature. We then took the average range value from this feature to estimate the velocity of the AGV. However, since Kinect range data is noisy, as seen in Figure 2, we created an optimal linear estimator in the Minimum Mean Squared Error (MMSE) sense to improve our range measurements. The MMSE linear estimator is

\[ \hat{Y}(X) = AX, \]  

(6)

where

\[ A = E \left[ YX^t \right] R_X^{-1}. \]  

(7)

From our characterization of the noise in the Kinect range data, we estimated what should be the optimal depth map.

A. Velocity Estimation Results

The estimated velocity results were very close to actual velocity truth. The results were typically better using a linear MMSE estimator over no estimator. Figure 7 shows the results of the linear estimator versus truth and no estimator.

IV. CONCLUSION

We successfully characterized the noise of the Kinect depth sensor in an effort to estimate AGV velocity. We characterized the noise by collecting data, finding the deterministic and random noise of the system and verified that this characterization was correctly performed through simulation. When compared with actual Kinect data, the simulated error resembles actual Kinect range measurement error. The simulated depth map may serve as a simulation test bed for filter testing and estimation and is useful for research in the area of RGB-D filtering.

REFERENCES