The Measurement of Carried Cartons using Multiple Kinect Sensors

Yuka Kohno Masaya Maeda Tomoyuki Hamamura Bunpei Irie
Toshiba Corporation
Power and Industrial Systems R&D Center
{ yuka.kohno, masaya.maeda, tomoyuki.hamamura, bunpei.irie }@toshiba.co.jp

Abstract

We propose a system to measure the dimensions of carton boxes carried on a conveyer belt with an accuracy of 5mm, required in logistic applications. Our system estimates the dimensions through a series of depth images with relatively low resolution taken by unsynchronized multiple sensors. To achieve the accuracy, we first obtain the dimensions via the distance between parallel planes instead of edge lengths, then correct errors caused by time lags among sensors using timestamps, and finally correct other measurement errors by regression analysis. We evaluated our method on three types of cartons carried on a conveyer belt in 0.9m/s, and confirmed the required accuracy.

1. Introduction

In logistics and distribution industry, an efficient measurement system of carton box dimensions is in demand for automatic sorting, calculation of transportation charge, and making layout plans of cargos. To meet the needs, scanning systems which obtain 3D shape of objects using line laser scanners to output their length, mass, and other measurements are provided by measuring instrument manufacturers. They are capable of handling various shapes with measurement precision of 5mm, which is required in logistics applications. However, such devices are not popular in relatively small-scale facilities because of their cost and space restriction. Though the devices have function to handle various shapes, in logistic use, most of the objects being carried are in box-shape. In such case, a low-cost system which handles limited types of objects is preferred. The system should be designed to measure objects being carried on a conveyer belt, as well as the ones placed still manually, and the measurement process should be carried out online in order to sort objects immediately in the next process. To meet such needs, we have developed a system to measure the dimensions of box-shaped cartons, using Kinect, a low-cost range sensor originally developed as a computer game input device.

Kinect sensors have been used in many applications such as gesture recognition and 3D modeling since their launch in 2010. There are many researches to find its features or the ability as a 3D scanner. Khoshelham et al. have investigated the precision of the depth measurement based on a mathematical model and verified it by experiments [1]. According to their experiments, the size of the quantization steps in depth value could be approximated by the quadratic polynomial of its value, which is about 7mm around the depth of 1.5m, the distance in which we plan to place the sensors relative to the objects. Andersen et al. had also performed similar experiment and evaluated the accuracy by the number of levels the depth values vary among pixels capturing a plane of the same distance [2]. They found that in most of the cases they vary within 2 to 4 levels, which roughly correspond to 20mm in distance of 1.5m. They have also evaluated the spatial precision by observing the noise on the contours of objects in the depth images. Placing a rectangular plane orthogonal to the sensor, the precision of its edge turned out to be in magnitude of 4 pixels, which is about 10mm at a distance of 1.5m. Empirically, there would be larger noise and more missing pixels when the plane is not orthogonal to the sensor or is moving. According to their studies, naive method is not sufficient to measure an object dimension in precision of 5mm.

There are researches to use Kinect sensors or other low-resolution depth sensors to acquire fine 3D shape data [3][4]. Their method is very useful in making fine 3D data from single Kinect sensor, but not suitable for processing a number of cargos carried one after another on a conveyer belt. Since their algorithms rely on large number of steps such as SfM and super-resolution of 3D data, the process time reduction would be difficult.

Our system uses fixed multiple sensors to capture both front and back side of the object, instead of moving a single sensor and combining frames afterwards. Box-shaped cartons are carried on a conveyer belt in constant velocity, to be captured as a sequence of depth image, using sensors set on both sides of the belt. The dimensions of the cartons are then calculated according to the depth images. Without any synchronization among the sensors, the capturing time of each sensor has some time lags to each other, which results in some spatial gaps among the data from each sensor. Our algorithm achieves the accuracy of 5mm using depth image with relatively low resolution, and taken by unsynchronized multiple sensors. Our principles are:

i. The dimension of each side of a carton is calculated as the distance between two parallel planes, which reduce the error induced by local noise and missing pixels.
ii. Multiple frame images with their time stamps are used to detect pairs of parallel planes as much as possible and to calculate the edge length without the effect of time lags in capture among sensors.
iii. The measurement result in each frame is corrected by regression analysis, to reduce the errors arising from the quality of sensors.

The next section describes our algorithm in detail. Then in Section 3, we discuss the verification of the algorithm by the experiment using actual cartons carried on a conveyer belt, and finally conclude with some remarks and future works in Section 4.

2. Measurement of carton dimensions

In this Section we explain our algorithm to measure the dimensions of carton box. Figure 1 shows the flowchart of the whole process, and the following sections explain each step in the flowchart. The input of the process is a sequence of depth images and their timestamps taken by multiple sensors, placed as shown in Figure 2 for example. The intrinsic and extrinsic parameters of each sensor are obtained beforehand in the calibration phase, and used to generate and align point cloud in measurement process. Planes are then detected from the point cloud data, out of which the floor plane, the top plane, and the four side planes of the carton box are selected, and three dimensions of the box are calculated as the distance between a pair of planes. After some correction steps, the results from multiple frames and multiple pairs of sensors are averaged without outliers to output one set of measurement values for the carton box.

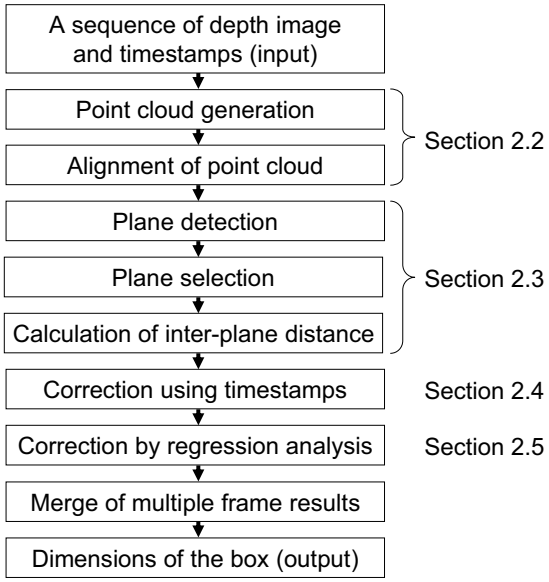

Figure 1. Flowchart of measurement process

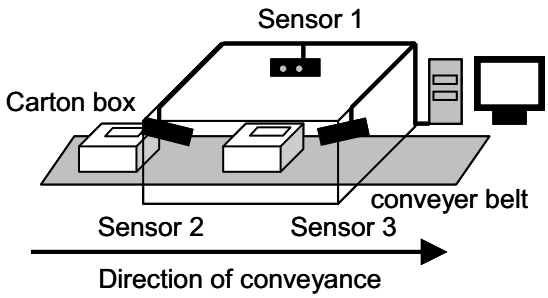

Figure 2. An example of sensor settings

2.1. Calibration

Before the explanation of the generation and alignment of the point cloud, we first explain the calibration step to obtain the camera parameters.

Intrinsic camera parameters of each sensor are obtained using checker board images following the method introduced by Smisek [3]. The capture of the checker board patterns by IR camera is carried out during the daytime while there are IR light from the sun, with the IR projector covered, to get clear checker board image by IR camera.

Extrinsic parameters, or relative positions and angles among sensors, are obtained using a colored box with its dimensions known. A box is placed in the capture area of both sensors and captured as both depth image and color image. Point cloud data is generated from the depth image by the method described in Section 2.2, and colored with corresponding pixels in color image. Plane detection is then implemented using the method in Section 2.3, followed by conveyer belt plane and side plane selection according to their size and hue. Finally, the rotation and translation is calculated so that corresponding planes, or a plane and its corresponding plane on the opposite side would overlap to each other.

2.2. Point cloud generation and alignment

For all depth images in input, a set of point cloud, representing the 3D shape of the captured area, is generated. In our algorithm, we first reduce the resolution of the original VGA depth image by $1/2^3$, before the point cloud generation. This is to smooth the surface of the 3D shape despite the low resolution of depth values, and to accelerate the following processes. In the generation process, each point (x, y, z) in the point cloud is calculated by the following formula using the intrinsic parameters obtained in the calibration phase:

$$\begin{cases} x = \dfrac{d}{fx}(u - cx), \\ y = \dfrac{d}{fy}(v - cy), \\ z = d, \end{cases}$$

where (x, y, z) represents the position in 3D space, (u, v) the position of each pixel in the depth image, d the depth value of the image, (cx, cy) the principle point, and (fx, fy) the focal length. Then the depth images obtained by sensor 2 and 3 are affine transformed by extrinsic camera parameters to be described in the uniform coordinate system based on sensor 1.

2.3. Plane detection, selection, and measurement

Plane detection is implemented according to the normal vectors of each points and their continuity. The normal vectors are calculated according to the distribution of neighboring points. Then the points with similar normal vectors are extracted by voting method, and the continuous areas in the extracted point clouds are detected as points composing planes, from which the plane parameters are calculated by LSM. After the detection process, the planes representing the floor, the top, and the sides of the cartons, as shown in Figure 3, are selected according to their positions and normal directions. To be in detail, a plane parallel to and in close distance to the conveyer belt plane, calculated in the phase of calibration, is selected as the floor plane. A plane parallel to and with some distance from the floor plane is se-

lected as the top plane. The two planes are paired to calculate the height of the carton as the distance between the planes. Then the planes orthogonal to the floor is extracted as the candidate side planes of the carton, and confirmed to be one if they are parallel or orthogonal to each other among the candidate planes. Finally, the width and the depth are calculated from two pairs of parallel side planes.

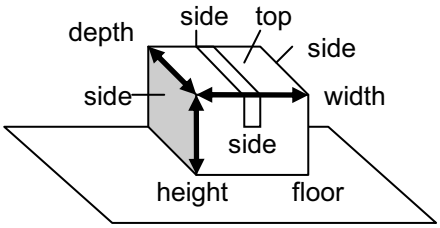

Figure 3. Three measurements of cartons and the pairs of planes corresponding to them

2.4. Correction using time stamp

Capturing depth images of moving cartons using unsynchronized multiple sensors, there may be some gaps among point clouds taken by different sensors, due to the time lags. That is, point cloud taken by one sensor may shift in the direction of the conveyance compared to another, as shown in Figure 4. To remove the errors caused by the gaps, we use timestamps, obtained from the sensors and accompanied to each image, to correct the measurements. We have found that Kinect sensors keep capturing depth images in a constant frame rate, while the PC fetches the image from the sensor whenever the process is ready for it. Therefore, we could get the initial time lags among the sensors by some calibration method before the measurement process starts. In the measurement process, the actual time difference among sensors could be calculated as the sum of initial time lag and the difference in timestamps of the frames to be processed. Using the time difference in each frame, the correction value is calculated by the function:

$$CorrectionValue = dt \cdot dZ \cdot \cos\theta,$$

where dt is the time difference of the capture between two depth images from which the plane is detected, dZ the velocity of conveyance, and θ an angle between the orientation of the measurement and the direction of the conveyance.

Using time stamps, pairs of parallel planes detected in different frames are also used for the calculation of the plane distance. Using more pairs of planes contributes in detecting and measuring carton box more robustly.

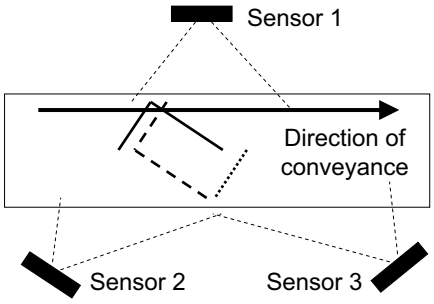

Figure 4. The planes captured by unsynchronized sensors

2.5. Correction by regression analysis

After correcting the measurements using timestamps, we found that there are some non-random errors which seem to depend on the orientation of measurement, and on the positions of the cartons on the conveyer belt, as shown in Figure 5. To reduce the error, we estimate error curves by regression analysis and correct the measurements by the estimated error. The error curve approximates the difference of the measurements to the actual dimensions, obtained by processing some model carton boxes with ground truth measurements. The form of the error curves are a trigonometric function, assumed to be approximating the sensor position errors,

$$CorrectionValue = a \cdot \sin\theta + b \cdot \cos\theta + c,$$

with θ an orientation of measurement, and polynomials, including linear functions approximating sensor rotation, and higher orders for other distortion errors,

$$CorrectionValue = a_0 + a_1 Z + a_2 Z^2 + a_3 Z^3,$$

with Z the position of the carton. The parameters of trigonometric functions are estimated first, as it seems more dominant. Then the polynomial parameters are estimated for each set of sensors by which the two planes are captured, using the measurements corrected by the trigonometric function. The error curve with the coefficient of determination, or R^2, larger then 0.5 is adopted.

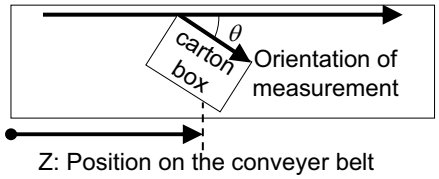

Figure 5. θ: The orientation of measurement, Z: the position of carton box on the conveyer belt.

3. Experiments

For evaluation of our method, we made an experiment using 54 sequences of depth image data, three sequences each for three types of carton boxes placed on a conveyer belt in six different orientations. The carton boxes are shown in Figure 6 and their dimensions in Table 1. The six orientations are 0, 30, 60, 90, 120, and 150 degrees to the directions of conveyance. The velocity of the conveyer belt was 0.9m/s. Three sensors were set roughly forming a equilateral triangle, as shown in Figure 2, to capture the side planes of the cartons. One additional sensor was set to capture depth images from above to obtain the floor plane and the top plane. Since we had not taken the timestamps in collecting the experimental data, we used the number of frame missed by the PC in fetching images from the sensors, input manually according to the human observation. This should be more accurate with original timestamps. We evaluated the result in terms of success rate of extraction and average error, maximum error, and standard deviation of errors in measurements. The results of width and depth are evaluated together as they could be replaced with one another. The two correction steps, the correction using timestamps and the correction by regression, were processed only in the measurement of width and depth, and

not for the heights whose errors were below 5mm without correction. The evaluation was performed with and without the correction processes to verify their effects. The results are shown in Table 2. The number of frames used per carton box varies from 5 to 17 with average of 13 for width and depth, and from 5 to 14 with average of 10 for heights.

Figure 6. The cartons used in the experiment

Table 1. dimensions of cartons used in experiment

Carton	Width	Height	Depth
1	550	400	365
2	455	295	325
3	780	100	190

Table 2. The result of the experiment

Result without correction

	W/D	Height	Average
Detection rate [%]	100	100	100
Average error [mm]	15.1	1.70	10.6
Max error [mm]	82.8	4.87	82.8
SD of error [mm]	13.7	1.21	12.9

Result after correction using time stamps

	W/D	Height	Average
Detection rate [%]	100	100	100
Average error [mm]	9.04	1.70	6.60
Max error [mm]	27.8	4.87	27.8
SD of error [mm]	6.32	1.21	6.26

Result after correction by regression

	W/D	Height	Average
Detection rate [%]	100	100	100
Average error [mm]	3.08	1.70	2.61
Max error [mm]	11.6	4.87	11.6
SD of error [mm]	2.49	1.21	2.24

From the result, we can observe that two correction processes are both effective in decreasing the measurement errors. Without the correction using time stamps, the gap among sensor data turned out to be up to 144mm, with the maximum number of missing frame be 3 and the gap per missing frame 48mm. Though the effects of the displacements are weakened by averaging among multiple frames, they still make the measurements unreliable.

In correction by regression, the coefficients of the trigonometric function were estimated to be a=-29.3, b=0.69, and c=19.9, with a confidence of R^2=0.71, for all pairs of sensors. The coefficients of polynomials were estimated separately for all three pairs of sensors, out of which a linear function with a_0=7.56, a_1=0.035, and R^2=0.79 was adopted for sensor 2 and 3. Note that, since the regression analysis is assumed to correct errors caused by relative positions and angles among sensors, these values would differ in different sensor settings. We have also examined the regression by separating the estimation step in more detail, such as trigonometric function approximation for each pairs of sensors, and polynomial approximation for each orientation. Though these approaches seem effective in theory, the result turned out to be slightly worse than that without correction, because of the over-fittings to small data.

As a final result, the average error is 2.61mm, about half of the required precision, and though the maximum error is 11.6mm, 85.2% of the total measurement result was in error less than 5mm. Considering the precision of Kinect sensors as a depth sensor, described in Section 1, and the fact that the object is moving, it could be said that our algorithm is effective in making a good measurements using low-resolution depth image.

The reason why the errors of height measurement were smaller than the other two could be explained by the absence of sensor placement errors and distortion from object movement. Since the floor and top planes are both extracted from same depth image, it does not contain error caused by the relative angles and positions of the sensors. The normal vectors of the two planes, both nearly orthogonal to the direction of movement, also contribute to the precision as the shape of such plane is less likely to be distorted by the movement.

As to the process time, the average process time per carton is 1530msec altogether, using Intel® Core™ 2 CPU (2.66GHz) with 4GB RAM, with no parallel process nor acceleration. By performing the point cloud generation and plane detection in parallel process, using one core for each sensor data, the total process time would be 620msec per carton on average.

4. Conclusion

We have introduced a method to measure the dimensions of box-shaped cartons carried on a conveyer belt in constant velocity, using unsynchronized multiple depth image sensors with low-resolution. The accuracy of measurement was achieved by calculating distance between pairs of planes as the dimensions, correction processes using timestamps, and by regression analysis. The precision of our system proved to be below 5mm, the required precision in logistics applications, in 85% of an experimental data, including three types of cartons in six orientations, carried in a velocity of 0.9m/s.

As a future work, automation of calibrations for two correction steps is necessary. After the development for practical use, we will extend our algorithm to handle luggage other than box-shape.

References

[1] K. Khoshelham, "Accuracy analysis of kinect depth data," ISPRS workshop laser scanning 2011, 2011.

[2] M.R. Andersen, T. Jensen, P. Lisouski, A.K. Mortensen, M.K. Hansen, T. Gregersen and P. Ahrendt, "Kinect Depth Sensor Evaluation for Computer Vision Applications," Electrical and Computer Engineering Technical Report ECE-TR-6, 2012.

[3] J. Smisek, M. Jancosek and T. Pajdla, "3D with Kinect," ICCV Workshops 2011, 2011.

[4] Y. Cui, D. Stricker, "3D Shape Scanning with a Kinect," Siggraph 2011, 2011.