# Calibration of a Network of Kinect Sensors
# for Robotic Inspection over a Large Workspace

Rizwan Macknojia, Alberto Chávez-Aragón, Pierre Payeur, Robert Laganière
School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, ON, Canada
[rmack102, achavez, ppayeur, laganier]@uottawa.com

## Abstract

*This paper presents an approach for calibrating a network of Kinect devices used to guide robotic arms with rapidly acquired 3D models. The method takes advantage of the rapid 3D measurement technology embedded in the Kinect sensor and provides registration accuracy within the range of the depth measurements accuracy provided by this technology. The internal calibration of the sensor in between the color and depth measurement is also presented. The resulting system is developed to inspect large objects, such as vehicles, positioned within an enlarged field of view created by the network of RGB-D sensors.*

## 1. Introduction

Efficient methods for representing and interpreting the surrounding environment of a robot require fast and accurate 3D imaging devices. Most existing solutions make use of high-cost 3D profiling cameras, scanners, sonars or combinations of them, which often result in lengthy acquisition and slow processing of massive amounts of information.

The ever growing popularity and adoption of the Kinect RGB-D sensor motivated its introduction in the development of a robotic inspection station operating under multi-sensory visual guidance. The extreme acquisition speed of this technology supported the selection of Kinect sensors in the implementation to handle the requirement for rapidly acquiring models over large volumes, such as that of automotive vehicles. The method presented in this work uses a set of Kinect sensors to collect 3D points as well as texture information over a vehicle bodywork. A dedicated calibration methodology is presented to achieve accurate alignment between the respective point clouds and textured images acquired by Kinect sensors that are distributed in a collaborative network of imagers to provide coverage over large volumes.

The paper is organized as follows. Section 2 describes related work. Section 3 explains the internal and external calibration of Kinect devices. Section 4 presents an experimental setup, an analysis of calibration parameters, as well as some results and their evaluation. Finally, Section 5 presents some conclusions and future work.

## 2. Related work

In 2010 Microsoft introduced the *Kinect for Xbox 360* sensor as an affordable and real-time source for medium quality textured 3D data dedicated to gesture detection and recognition in a game controller. Since then, numerous researches have recognized the potential of this RGB-D imaging technology, especially because of its speed of acquisition, and attempted to integrate it in a broad range of applications.

Among the numerous examples of applications for the Kinect technology that rapidly appeared in the literature, Zhou *et al*. [1] propose a system capable of scanning human bodies using multiple Kinect sensors arranged in a circle. Maimone and Fuchs [2] present a real-time telepresence system with head tracking capabilities based on a set of Kinect sensors. They also contribute an algorithm for merging data and automatic color adjustment between multiple depth data sources. An application of Kinect in the medical field for position tracking in CT scans is proposed by Noonan *et al*. [3]. They track the head of a phantom by registering Kinect depth data to high resolution CT template of a head phantom. Rakprayoon *et al*. [4] use a Kinect sensor for obstacle detection of a robotic manipulator.

On the other hand, the depth data of the Kinect sensor is also known to suffer from quantization noise [5] [6], that increases as the distance to the object increases. The resolution also decreases with the distance [6]. The depth map may also contain occluded and missing depth areas mainly due to the physical separation between the IR projector and the IR camera, and to the inability to collect sufficient IR signal reflection over some types of surface. These missing values can however be approximated by filtering or interpolation [2] [7].

In order to merge data collected from different Kinect sensors, various approaches have been proposed for simultaneous calibration of Kinect's sensors. Burrus [8] proposes to use traditional techniques for calibrating the Kinect color camera involving manual selection of the

four corners of a checkerboard for calibrating the depth sensor. Zhang *et al.* [9] automatically sample the planar target to collect the points for calibration of depth sensor and used manual selection of corresponding points between color and depth images for establishing the extrinsic relationship within a single Kinect sensor. Gaffney [10] describes a technique to calibrate the depth sensor by using 3D printouts of cuboids to generate different levels in depth images. The latter however requires an elaborate process to construct the target. Berger *et al.* [11] use a checkerboard where black boxes are replaced with mirroring aluminum foil, therefore avoiding the need for blocking the projector when calibrating the depth camera.

The work presented here introduces a different calibration technique for Kinect devices, which specifically addresses both its internal and external calibration parameters. Internal calibration corresponds to estimating the intrinsic parameters for the color and IR cameras and also the extrinsic calibration between them. A method to relate color and depth pixels is proposed. The external calibration between multiple devices is also presented, without the need to cover their IR projectors, and that achieves accuracy compatible with that of the depth data available. The method is experimentally validated with a network of calibrated sensors that work together to accurately acquire a 3D profile over objects of a large dimension, here automotive vehicles.

## 3. Kinect sensors calibration

The Kinect technology consists of a multi-view system that provides three outputs: an RGB image, an infrared image and a depth image for each sensor. Therefore, when these devices are grouped and operated as a collaborative network of imagers in order to enlarge the overall field of view and allow for modeling of large objects, such as automotive vehicles, precise mapping between color and infrared of all RGB-D sensors must be achieved. For this purpose an internal calibration procedure that estimates the intrinsic parameters of each camera within every device as well as the extrinsic parameters between the RGB and the IR cameras inside a given Kinect is developed, along with an external calibration process that provides accurate estimates of the extrinsic parameters in between the respective pairs of Kinect devices.

### 3.1. Internal calibration

**3.1.1. Intrinsic parameters estimation for built-in Kinect cameras.** The internal calibration procedure includes the estimation of the respective intrinsic parameters for the color and the IR sensors, which are: the focal length ($f_x$, $f_y$), the principal point ($O_x$, $O_y$), and the lens distortion coefficients ($k_1$, $k_2$, $p_1$, $p_2$, $k_3$) [12]. Because the RGB and IR cameras exhibit different color responses,

the proposed calibration technique uses a regular checkerboard target of size 9x7 that is visible in both sensors' spectra. During internal calibration the Kinect's IR projector is blocked by overlapping a mask on the projector window since it cannot be turned off by the driver software. The IR projector otherwise introduces noise over the IR image as shown in Fig. 1(a), and without projection, the image is too dark as shown in Fig. 1(b). Therefore standard external incandescent lamps are added to illuminate the checkerboard target, Fig. 1(c). The color image is not affected by the IR projection and creates a clear pattern, Fig. 1(d).
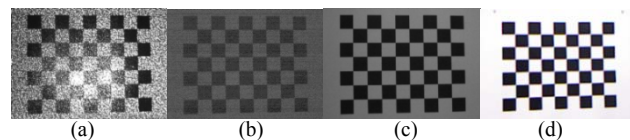


| (a) | (b) | (c) | (d) |

Fig. 1. Views of the checkerboard in different configurations. a) IR image with IR projector, b) IR image without IR projector, c) IR image with incandescent lighting and without projector, and d) color image.

The checkerboard is printed on a regular A3 size paper, which does not reflect back the bright blobs due to the external incandescent lamps in the IR image plane. To ensure the best calibration results, 100 images are collected from both the color and the IR cameras. Both images are synchronized in each frame, so that they can be used for extrinsic calibration between the cameras (next section). To estimate the intrinsic parameters, each Kinect is calibrated individually using Zhang's camera calibration method [12]. The method is applied 10 times on 30 images randomly selected among the 100 captured images. The reprojection error is also calculated for each iteration, which is a measure of the deviation of the camera response to the ideal pinhole camera model. The reprojection error is calculated as the RMS error of all the target calibration points. The results of the calibration for the least reprojection error are shown in Table 1 for five Kinect sensors involved in the network.

TABLE 1 : Internal intrinsic calibration of embedded sensors

| Intrinsic Parameters of IR camera in pixels | | | | |
|---|---|---|---|---|
| sensor | $f_{x\ IR}$ | $f_{y\ IR}$ | $O_{x\ IR}$ | $O_{y\ IR}$ | Error |
| K0 | 584.2 | 582.6 | 326.7 | 233.5 | 0.136 |
| K1 | 585.9 | 583.8 | 325.2 | 242.3 | 0.148 |
| K2 | 597.7 | 595.7 | 322.2 | 232.1 | 0.131 |
| K3 | 599.0 | 597.1 | 331.5 | 240.3 | 0.157 |
| K4 | 581.7 | 579.5 | 319.6 | 246.3 | 0.145 |

| Intrinsic Parameters of RGB camera in pixels | | | | |
|---|---|---|---|---|
| sensor | $f_{x\ RGB}$ | $f_{y\ RGB}$ | $O_{x\ RGB}$ | $O_{y\ RGB}$ | Error |
| K0 | 517.9 | 516.7 | 321.0 | 245.6 | 0.127 |
| K1 | 518.8 | 517.0 | 331.1 | 261.4 | 0.124 |
| K2 | 535.7 | 537.3 | 336.2 | 252.8 | 0.129 |
| K3 | 525.1 | 523.0 | 322.1 | 255.1 | 0.153 |
| K4 | 517.2 | 515.2 | 319.7 | 254.8 | 0.146 |

| Distortion Parameters of IR camera | | | | | |
|---|---|---|---|---|---|
| sensor | $k_{1\ IR}$ | $k_{2\ IR}$ | $p_{1\ IR}$ | $p_{2\ IR}$ | $k_{3\ IR}$ | Error |
| K0 | -0.1193 | 0.5768 | 0.0011 | 0.0037 | -0.8692 | 0.136 |
| K1 | -0.1323 | 0.6297 | -0.0004 | 0.0028 | -0.9595 | 0.148 |
| K2 | -0.1279 | 0.7134 | 0.0003 | 0.0014 | -1.2258 | 0.131 |
| K3 | -0.1505 | 0.6235 | 0.0004 | 0.0033 | -0.9402 | 0.157 |
| K4 | -0.1394 | 0.7395 | 0.0019 | 0.0018 | -1.2704 | 0.145 |

| Distortion Parameters of IR camera | | | | | |
|---|---|---|---|---|---|
| sensor | $k_{1\ RGB}$ | $k_{2\ RGB}$ | $p_{1\ RGB}$ | $p_{2\ RGB}$ | $k_{3\ RGB}$ | Error |
| K0 | 0.2663 | -0.8656 | 0.0015 | -0.0053 | 1.0156 | 0.127 |
| K1 | 0.2918 | -1.0374 | -0.0012 | -0.0056 | 1.4310 | 0.124 |
| K2 | 0.2914 | -1.1027 | -0.0002 | -0.0009 | 1.5614 | 0.129 |
| K3 | 0.2516 | -0.9045 | -0.0015 | 0.0017 | 1.1420 | 0.153 |
| K4 | 0.2380 | -0.8270 | -0.0010 | 0.0020 | 1.0251 | 0.146 |

After calibration, both the RGB and IR cameras achieve reprojection error between 0.12 and 0.16 pixel, which is better than the original performance given by the Kinect sensor. The reprojection error without calibration of the IR camera is greater than 0.3 pixel and that of the color camera is greater than 0.5 pixel. The focal length of the IR camera is larger than that of the color camera, i.e. the color camera has a larger field of view. It is also apparent that every Kinect sensor has slightly different intrinsic parameters. This confirms the need for a formal intrinsic calibration to be performed on every device individually to support accurate data registration.

**3.1.2. Extrinsic parameters estimation for built-in Kinect cameras.** The respective location of the color and IR cameras within each Kinect unit is determined by stereo calibration. The camera calibration method proposed by Zhang [12] also provides the location of the checkerboard target with respect to the camera coordinate system. If the target remains fixed for both cameras then the position between both cameras is defined by Eq. (1).

$$H = H_{RGB} H_{IR}^{-1} \qquad (1)$$

where $H$ is the homogenous transformation matrix (consists of 3x3 rotation matrix $R$ and 3x1 translation vector $T$) from the RGB camera to the IR camera, $H_{IR}$ is the homogenous transformation matrix from the IR camera to the checkerboard target, and $H_{RGB}$ is the homogenous transformation from the RGB camera to the checkerboard target. The translation and rotation parameters between the RGB and IR sensors are shown in Table 2 for five Kinect sensors. The internal extrinsic calibration parameters allow to accurately relate the color and depth data collected by a given Kinect device.

TABLE 2 : Internal extrinsic calibration of embedded sensors

| Translation (cm) and Rotation (degree) between RGB and IR | | | | | |
|---|---|---|---|---|---|
| sensor | $T_x$ | $T_y$ | $T_z$ | $R_x$ | $R_y$ | $R_z$ |
| K0 | 2.50 | 0.0231 | 0.3423 | 0.0017 | 0.0018 | -0.0082 |
| K1 | 2.46 | -0.0168 | -0.1426 | 0.0049 | 0.0032 | 0.0112 |
| K2 | 2.41 | -0.0426 | -0.3729 | 0.0027 | 0.0065 | -0.0075 |
| K3 | 2.49 | 0.0153 | 0.2572 | -0.0046 | 0.0074 | 0.0035 |
| K4 | 2.47 | 0.0374 | 0.3120 | 0.0052 | 0.0035 | 0.0045 |

**3.1.3. Registration of color and depth within a given Kinect device.** The Kinect sensor does not provide the registered color and depth images. Once the internal intrinsic and extrinsic parameters are determined for a given Kinect device, the procedure to merge the color and depth based on the estimated registration parameters is performed as follows. The first step is to properly relate the IR image and the depth image. The depth image is generated from the IR image but there is a small offset between the two, which is introduced as a result of the correlation performed internally during depth calculation. The offset is 5 pixels in the horizontal direction and 4 pixels in the vertical direction [5] [13]. After removing this offset using Eq. (2), each pixel of the depth image exactly maps the depth of the corresponding pixel in the IR image. Therefore, the calibration parameters of the IR camera can be applied on the depth image.

$$depth(x, y) = depth\_o(x - 5, y - 4) \qquad (2)$$

where $x$ and $y$ is the pixel location, $depth\_o(x,y)$ is the offseted depth value affecting the Kinect depth sensor and $depth(x,y)$ is the corrected depth value.

The second step is to transform the color and the corrected depth images to compensate for radial and tangential lens distortion using OpenCV [14]. The geometric transformation on the images is estimated using the distortion parameters and provides undistorted color image and depth image ($depth\_ud(x, y)$). The next step is to determine the 3D coordinates corresponding to each point in the undistorted depth image, using Eq. (3) to (5).

$$X_{IR} = \frac{(x - O_{x\_IR})\, depth\_ud(x,y)}{f_{x\_IR}} \qquad (3)$$

$$Y_{IR} = \frac{(y - O_{y\_IR})\, depth\_ud(x,y)}{f_{y\_IR}} \qquad (4)$$

$$Z_{IR} = depth\_ud(x,y) \qquad (5)$$

where ($X_{IR}$, $Y_{IR}$, $Z_{IR}$) are the 3D point coordinates of a depth image with respect to the IR camera reference frame, $(x,y)$ are the pixel location, $(f_{x\_IR}, f_{y\_IR})$ are the focal length of the IR camera, $(O_{x\_IR}, O_{y\_IR})$ are the optical center of the IR camera and $depth\_ud(x, y)$ is the depth of a pixel in the undistorted depth image.

Next, the color is assigned from the RGB image to each 3D point $P_{IR}(X_{IR}, Y_{IR}, Z_{IR})$. The color is mapped by transforming the 3D point $P_{IR}$ into the color camera reference frame using the internal extrinsic camera parameters and then reprojecting that point on the image plane of the RGB camera using the intrinsic parameters to find the pixel location of the color in the undistorted color image using Eq. (6) to (8).

$$P_{RGB}(X_{RGB}, Y_{RGB}, Z_{RGB}) = R \cdot P_{IR} + T \qquad (6)$$

$$x = \left(\frac{X_{RGB} f_{x\_RGB}}{Z_{RGB}}\right) + O_{x\_RGB} \qquad (7)$$

$$y = \left(\frac{Y_{RGB} f_{y\_RGB}}{Z_{RGB}}\right) + O_{y\_RGB} \qquad (8)$$

where $P_{RGB}$ is the 3D point with respect to the color camera reference frame, $R$ and $T$ are the rotation and translation parameters from the color camera to the IR camera, and $(x,y)$ is the location of color in the undistorted color image.

Fig. 2(a) shows the portion of a car as imaged from the color camera, Fig. 2(b) shows the colored depth information in the interval 0-2.5 m from the slightly different point of view of the IR camera contained in the same Kinect device, while keeping the Kinect sensor static with respect to the car. The difference in position and orientation between the two cameras contained in the Kinect unit is accurately compensated by the estimated extrinsic parameters obtained from internal calibration.



(a)                              (b)

Fig. 2. Accurate registration of color and depth images: a) color image, b) colored depth image.

## 3.2. External Calibration

The last set of parameters estimated in the calibration process are the extrinsic ones that are the relative position and orientation between every pair of Kinect sensors. The external calibration is performed between pairs of IR cameras over the network of sensors because depth information is generated with respect to these cameras. The concept behind the proposed method consists in determining, for every pair of sensors, the position and orientation of a fixed planar checkerboard in real world coordinates. Knowing the orientation of the plane from two different points of view (i.e. two Kinect sensors), it is possible to estimate the relative orientation and position change between the sensors.

The procedure developed for external calibration consists in positioning a standard planar checkerboard target within the visible overlapping regions of any two Kinect sensors. Unlike most calibration techniques in the literature, there is no need to move the checkerboard to image it from multiple views. On the contrary, a fixed target increases the performance of the method. The result is a rigid body transformation that best aligns the data collected by a pair of RGB-D sensors. A best-fit plane calibration method is applied. It takes advantage of the rapid 3D measurement technology embedded in the sensor and provides registration accuracy within the range of the depth measurements accuracy. An important advantage of this method is the fact that it is unnecessary to cover the Kinect infrared projector to perform this phase of the

calibration, which facilitates manipulations when remotely dealing with the network of Kinect devices.

The method consists in finding a normal vector and the center of the checkerboard plane, which define the relative orientation and translation of the checkerboard plane. The first step is to compute the 3D coordinates of the corners on the checkerboard with respect to the IR camera frame, using Eq. (3) to (5). When the checkerboard target is positioned in front of a Kinect sensor, the IR projector pattern appears on the checkerboard target as shown in Fig. 3(a). This pattern creates noise and makes it difficult to extract the exact corners using OpenCV [14]. Since the noise is similar to salt and pepper noise, a median filter of size 3x3 provides a good reduction in the noise level without blurring the image, as shown in Fig. 3(b).
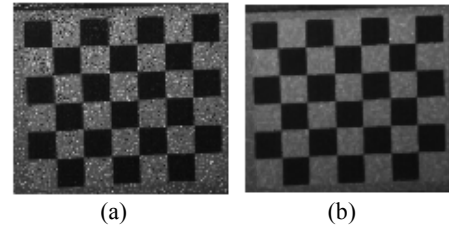


(a)                              (b)

Fig. 3. IR image of the checkerboard target for external calibration: a) effect of the projected IR pattern, b) filtered image using a Median filter of size 3x3.

Moreover, the extracted points are not entirely mapped over a single plane because of quantization effects in the Kinect depth sensor. Therefore, the corner points are used to estimate the three dimensional plane, Eq. (9), that minimizes the orthogonal distance between that plane and the set of 3D points. The equation of the plane then permits to estimate the orientation in 3D space of the target with respect to the IR camera.

$$z = Ax + By + C \qquad (9)$$

Let the 3D coordinates of the $n$ corners of the checkerboard target be $P_1(x_1,y_1,z_1)$, $P_2(x_2,y_2,z_2)$,…, $P_n(x_n,y_n,z_n)$, then the system of equations for solving the plane equation are $Ax_1+By_1+C=z_1$, $Ax_2+By_2+C=z_2$,..., $Ax_n+By_n+C=z_n$. These equations can be formulated into a matrix problem.

$$\begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ \vdots & \vdots & \vdots \\ x_n & y_n & 1 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \qquad (10)$$

This over determined system is solved for the values of $A$, $B$, and $C$ with an orthogonal distance regression approach [15], which provides the best fit plane on those points. All the 3D points, $P_n$, are projected on the best fit plane as $P'_n$. These points serve to define the center and the normal vector of the plane. However, projected points, $P'_n$, do not represent the exact corners of the checkerboard. Therefore the center of the plane cannot be defined only by the intersection of two lines passing close to the center. Fig. 4(a) shows the set of possible lines

passing close to the center. The closest point to all intersections between these lines is selected as a center point *O*. Two points *X* and *Y* are selected on the plane to define vectors $\overrightarrow{OX}$ and $\overrightarrow{OY}$. The normal to the plane is defined by the cross product $\vec{n} = \overrightarrow{OX} \times \overrightarrow{OY}/|\overrightarrow{OX} \times \overrightarrow{OY}|$

The orientation and the translation between two Kinect's IR cameras are calculated from the normal vectors and the centers of the checkerboard target defined with respect to both IR cameras. Let $\overrightarrow{n_1}$ and $\overrightarrow{n_2}$ be the two normal vectors, and $O_1$ and $O_2$ be the estimated centers of the target with respect to Kinect's IR cameras 1 and 2 respectively. If $\overrightarrow{n_2}$ is mapped on camera 1's frame then the rotation between the two vectors can be defined by the axis angle representation. The angle between two vectors is defined by Eq. (11) and the rotation axis is normal to both vectors and defined by Eq. (12).
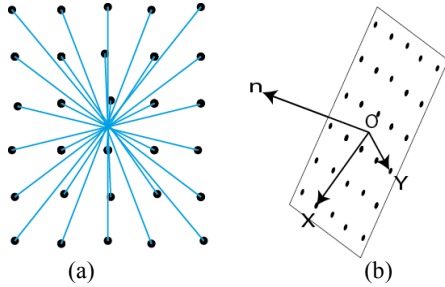


(a)             (b)

Fig. 4. a) Possible combination of lines passing through the center of the checkerboard, b) the normal vector and the center of a checkerboard target.

$$\theta = \cos^{-1}(\overrightarrow{n_1}.\overrightarrow{n_2}) \qquad (11)$$
$$\overrightarrow{V_n} = \overrightarrow{n_1} \times \overrightarrow{n_2} \qquad (12)$$

The axis-angle representation can be defined in quaternion form as:

$$q = [q_0 q_1 q_2 q_3] \qquad (13)$$

where $q_0 = cos(\theta/2)$, $q_1 = \overrightarrow{V_{nx}}sin(\theta/2)$, $q_2 = \overrightarrow{V_{ny}}sin(\theta/2)$, and $q_3 = \overrightarrow{V_{nz}}sin(\theta/2)$. The quaternion representation can be converted into a rotation matrix *R*, defined as:

$$R = \begin{bmatrix} 1-2q_2^2-2q_3^2 & 2q_1q_2+2q_0q_3 & 2q_1q_3-2q_0q_2 \\ 2q_1q_2-2q_0q_3 & 1-2q_1^2-2q_3^2 & 2q_2q_3+2q_0q_1 \\ 2q_1q_3+2q_0q_2 & 2q_2q_3-2q_0q_1 & 1-2q_1^2-2q_2^2 \end{bmatrix} \quad (14)$$

where *R* is the rotation matrix between two camera axes and $q_0$, $q_1$, $q_2$, $q_3$ are the quaternion coefficients. The translation between two camera frames is calculated using the centers of the checkerboard target as:

$$T = O_1 - RO_2 \qquad (15)$$

# 4. Experimental results

## 4.1. Setup

The imaging system designed to assist the navigation of a robotic arm for the inspection of a vehicle from color

and depth information is shown in Fig. 5. Five Kinect for Xbox 360 sensors are positioned to cover the complete side and partial front and back of a vehicle. The setup covers a 180 degrees field of view around a vehicle and can be replicated on the other side for a complete 360 degrees view. The sensors are positioned *1.0 m* above the ground and parallel to the floor. Kinects *K3* and *K4* are rotated towards the vehicle about 65 degrees with respect to Kinects *K0*, *K1* and *K2*. This configuration permits to meet the following requirements: 1) a minimum coverage area of the setup, 2) the collection of depth readings in the range of 0.8 to 3 m, which is the range where Kinect performs properly, with a standard deviation of 2 cm and a quantization error around 2.5 cm, and 3) an overlapping area of 0.5 m to 1 m between contiguous sensors to ensure accurate point cloud alignment and to support the external calibration process.
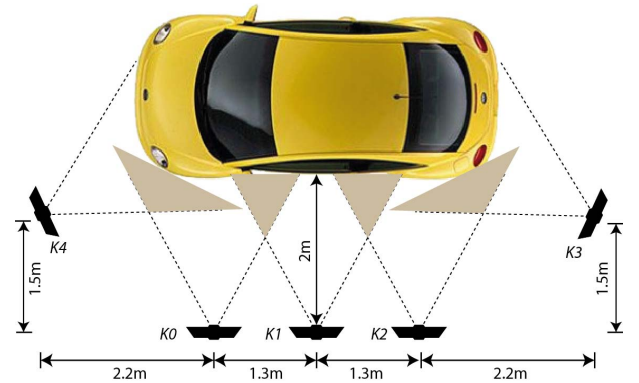


Fig. 5. Experimental configuration for scanning a vehicle.

## 4.2. Evaluation of Intrinsic Parameters

The quality of the intrinsic calibration method is measured by the reprojection error. We performed some experiments to observe the effect of inaccurate intrinsic parameter estimates during reconstruction. Kinect provides the depth of each pixel captured by the IR image, but the exact location of the pixel in the X and Y directions depends on the intrinsic parameters. In the experiments, the Kinect sensor is placed in front of a rectangular and planar object of size 60x25cm that is kept parallel to the Kinect IR camera image plane. Depth data is more accurate in close range, therefore the object is placed at a distance of 60cm where the quantization step size is less than 1mm [16]. The object is projected into the world coordinates, using Eq. (3) to (5), with the acquired depth and the intrinsic parameters obtained by calibration. The reconstructed object is shown in Fig. 6(a), where the red silhouette defines the actual size of the object, which is approximately the same size as that of the reconstructed object. The same experiment is also performed using the default intrinsic parameters encoded in OpenNI [17] and the result is shown in Fig. 6(b). In this case the

reconstructed object is significantly enlarged as compared to the red silhouette of the original object. The blue silhouette shows the scaled size, which is increased by 8.6mm in height and 6.2mm in width with the default intrinsic parameters used by OpenNI [17], as shown in Fig. 6(c). Therefore, a formal estimation of the intrinsic parameters within any Kinect sensor helps improve the accuracy on the scale of the reconstruction.
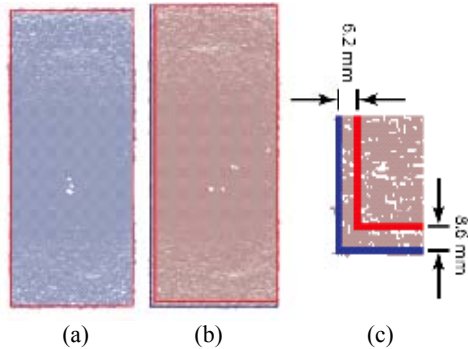


(a)　　　　(b)　　　　(c)

Fig. 6. Reconstruction of a planar target. Red silhouette shows the original size: a) using experimental calibration parameters, b) using OpenNI default parameters, blue silhouette shows the extended size, c) differences in size.

## 4.3. Network Calibration

The camera arrangement shown in Fig. 5 includes overlapping regions between contiguous sensors marked in gray. During the calibration phase, the checkerboard target is successively placed within those areas for external calibration between every pair of neighbor Kinect IR sensors. Fig. 7 shows the calibration target placed in the overlapping region between Kinect $K0$ and $K4$ during an experimental calibration procedure. External calibration is performed in pairs using the proposed method discussed in section 3.2. The center Kinect, $K1,$ is set as a base of reference for the setup. The relative calibration is then calculated between ($K1, K0$), ($K1, K2$), ($K2, K3$) and ($K0, K4$).
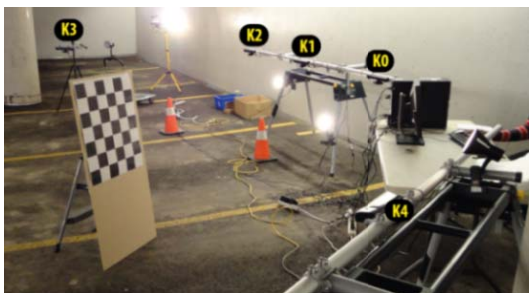


Fig. 7. Placement of calibration target during calibration of Kinects $K0$ and $K4$.

## 4.4. Data Collection and Results

After calibration, the data collection with the system is performed in a sequence. The overlapping regions

between two contiguous Kinect sensors might contain interference, since all Kinect devices project a pattern of infrared points at the same wavelength to create their respective depth map. This produces small holes over the depth maps of overlapping sensors. To prevent this problem, the data is collected sequentially over different time slots. During the first time slot, sensors $K0$ and $K2$ simultaneously collect their respective information. Then, sensors $K1$, $K3$ and $K4$ scan the corresponding regions over the vehicle. The delay between the shots is the time needed to shut down the devices and initialize the next devices. This process is performed by the Kinect drivers from the OpenNI [17] framework and takes between 1 and 2 seconds to initialize each device. Fig. 8 shows a vehicle standing in front of the setup for rapid 3D modeling that will drive the robotic inspection. The reconstruction for two different types of vehicles is shown in Fig. 9. These models are obtained by only merging the raw depth and color images via the experimentally estimated calibration parameters. No filtering or smoothing is applied on data.



Fig. 8. Capturing 3D data over a vehicle with the network of Kinect sensors.



Fig. 9. Six different views of two reconstructed vehicles.

The windshield, lateral windows, and part of headlamps and rear lamps are missing in the depth map because the IR energy generated by the Kinect devices passes through the transparent surfaces or is deflected in other directions.

However, the rear window of the larger vehicle, which is made of tinted glass, is partially captured. All of the main areas of the vehicle body and wheels, including dark rubber tires, are accurately reconstructed and sections of the model acquired from the five viewpoints are correctly aligned, even over narrow roof supporting beams and highly curved bumpers areas. Table 3 presents a comparison between the characteristics of the reconstructed vehicles and their actual dimensions. The Kinect depth quantization introduces scaling errors of about 1cm in height and width and a depth error of about 2.5 cm at 3m distance. Each sensor covers the full height of the vehicle and the average error on height is under 1%. The estimation of the length of the vehicle and the wheel base (i.e. the distance between the centers of the front and back wheels) involve all the calibration parameters. The error on the length is under 2.5%, which is relatively minor given the medium quality of data provided by Kinect at a depth of 3m and in proportion to the large working volume. For further assessment of the algorithm, an ICP algorithm [18] was further applied on the final results, but it did not significantly improve the registration.

TABLE 3 : Reconstruction compared with ground truth

|     |             | Height | Length | Wheel base |
|-----|-------------|--------|--------|------------|
| Car | Actual (mm) | 1460   | 4300   | 2550       |
|     | Model (mm)  | 1471   | 4391   | 2603       |
|     | Error (%)   | 0.75   | 2.11   | 2.07       |
| Van | Actual (mm) | 1748   | 5093   | 3030       |
|     | Model (mm)  | 1764   | 5206   | 3101       |
|     | Error (%)   | 0.91   | 2.21   | 2.34       |

## 5. Conclusion and future work

In this work a calibration methodology for the Kinect sensor and for networking such sensors is presented on an application for collecting 3D data over large workspaces. The best-fit plane calibration method takes advantage of the 3D measurement technology embedded in the sensors and provides registration accuracy within the range of the depth measurements accuracy provided by the Kinect technology. The proposed calibration technique opens the door to a great number of real-time 3D reconstruction applications over a large workspace using low-cost RGB-D sensors.

## References

[1] J. Zhou, L. Liu, Z. Pan and H. Yan, "Scanning 3D Full Human Bodies Using Kinects," *IEEE Transactions on Visualization and Computer Graphics,* vol. 18, no. 4, pp. 643-650, 2012.

[2] A. Maimone and H. Fuchs, "Encumbrance-free Telepresence System with Real-time 3D Capture and Display using Commodity Depth Cameras," in *IEEE International Symposium on Mixed and Augmented Reality*, pp.137-146, 2011.

[3] P. J. Noonan, T. F. Cootes, W. A. Hallett and R. Hinz, "The Design and Initial Calibration of an Optical Tracking System using the Microsoft Kinect," in *IEEE Nuclear Science Symposium and Medical Imaging Conference*, pp.3614-3617, 2011.

[4] P. Rakprayoon, M. Ruchanurucks and A. Coundoul, "Kinect-based Obstacle Detection for Manipulator," in *IEEE/SICE International Symposium on System Integration*, pp.68-73, 2011.

[5] J. Smisek, M. Jancosek and T. Pajdla, "3D with Kinect," in *IEEE International Conference on Computer Vision Workshops*, pp.1154-1160, 2011.

[6] K. Khoshelham, "Accuracy Analysis of Kinect Depth Data," in *ISPRS Workshop on Laser Scanning*, pp.1437-1454, 2011.

[7] S. Matyunin, D. Vatolin, Y. Berdnikov and M. Smirnov, "Temporal Filtering for Depth Maps Generated by Kinect Depth Camera," in *3DTV Conference on The True Vision - Capture, Transmission and Display of 3D Video*, pp.1-4, 2011.

[8] N. Burrus, "RGBDemo," [Online]. Available: http://labs.manctl.com/rgbdemo/.

[9] C. Zhang and Z. Zhang, "Calibration Betweeen Depth and Color Sensors for Commodity Depth Cameras," *in IEEE International Conference on Multimedia and Expo,* pp. 1-6, 2011.

[10] M. Gaffney, "Kinect/3D Scanner Calibration Pattern," [Online]. Available: http://www.thingiverse.com/thing:7793.

[11] K. Berger, K. Ruhl, Y. Schroeder, C. Bruemmer, A. Scholzk and M. Magnor, "Markerless Motion Capture using Multiple Color-Depth Sensors," in *Proc. Vision, Modeling and Visualization*, pp. 317–324, 2011.

[12] Z. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, no. 11, pp. 1330-1334, 2000.

[13] K. Konolige and P. Mihelich, "Technical description of Kinect calibration," [Online]. Available: http://www.ros.org/wiki/kinect_calibration/technical.

[14] "OpenCV," [Online]. Available: http://opencv.willowgarage.com/wiki/.

[15] "Real-time computer graphics and physics, mathematics, geometry, numerical analysis, and image analysis. Geometric tools," [Online]. Available: http://www.geometrictools.com/LibMathematics/Approximation/Approximation.html.

[16] R. Macknojia, A. Chávez-Aragón, P. Payeur and R. Laganière, "Experimental Characterization of Two Generations of Kinect's Depth Sensors," *in IEEE International Symposium on Robotic and Sensors Environments,* pp. 150-155, 2012.

[17] "OpenNI," [Online]. Available: http://openni.org/.

[18] P.J.Besl and N.D.McKay, "A Method for Registration of 3-D Shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 14, pp. 239-256, Feb 1992.