Proceedings of the 2011 IEEE
International Conference on Mechatronics
April 13-15, 2011, Istanbul, Turkey

Altitude Control of a Quadrotor Helicopter Using Depth Map from Microsoft Kinect Sensor

John Stowers[#1], Michael Hayes[#2] and Andrew Bainbridge-Smith[#3]
[#]Electrical & Computer Engineering, University of Canterbury, New Zealand.
[1]john.stowers@ieee.org
[2]michael.hayes@canterbury.ac.nz
[3]andrew.bainbridge-smith@canterbury.ac.nz

***Abstract*—Reliable depth estimation is a cornerstone of many autonomous robotic control systems. The Microsoft Kinect is a new, low cost, commodity game controller peripheral that calculates a depth map of the environment with good accuracy and high rate. In this paper we calibrate the Kinect depth and image sensors and then use the depth map to control the altitude of a quadrotor helicopter. This paper presents the first results of using this sensor in a real-time robotics control application.**

***Index Terms*—quadrotor, visual flight control, Microsoft Kinect, depth map**

I. Introduction

The Microsoft Kinect (Figure 1) is a low cost peripheral, released November 2010, for use as a game controller with the Xbox 360 game system. The device can be modified to obtain, simultaneously at 30 Hz, a 640 × 480 pixel monochrome intensity coded depth map and a 640 × 480 RGB video stream.

Fig. 1. The Microsoft Kinect sensor. From left to right, the sensors shown are; the IR projector, the RGB camera, the monochrome camera used for the depth computation.

Computation of depth maps are common in visual robotic control systems. They are used in autonomous navigation [1], map building [2], [3] and obstacle avoidance [4].

Due to the importance of depth maps in robotics, this paper attempts to quantify the accuracy, performance and operation of the Kinect sensor, as no public information is available. Furthermore, we test the Kinect sensor and its suitability for use in dynamic robotic environments by using the computed depth map to control the altitude of a flying quadrotor helicopter.

This paper will proceed as follows. Section I introduces the Kinect sensor hardware and the use of depth maps in research. Section II describes the calibration procedure and calibration results. Section III introduces quadrotor helicopters and the experimental platform and control system against which the Kinect was tested. The paper concludes with Section IV, a discussion of experimental flight results using the sensor.

A. Computation of Depth Maps

Attempts to compute depth maps can be grouped into passive or active methods [5]. Passive depth sensing tries to infer depth from 2D images from multiple cameras, for example through stereo correspondence algorithms [6] or optical flow [7]. Active methods usually employ additional physical sensors such as lasers, lighting or infra–red illumination cast on the scene. Structured light [9] based sensors use triangulation to detect the ranges of points within their field of view [10]. This solves the correspondence problem of stereo vision via the constraints induced by the structure of the light source. Once determined that a camera pixel contains primary laser return (e.g. not laser light returned from secondary reflections), the range of the reflecting surface viewed in the direction of the pixel is immediately determined to within the resolution capabilities of the system. Thus the correspondence problem which consumes a great deal of the CPU time in stereo vision algorithms is replaced with the computationally much simpler problem of determining which pixels of the sensor detect the primary laser return. Time–of–flight (TOF) cameras too avoid the correspondence problem, instead utilising the time–of–flight principle. They illuminate the scene for a short period of time, for example by using a brief pulse of light, and measure the duration before the illumination pulse is reflected back and detected on the image sensor. TOF cameras typically have high power consumption due to the high illumination switch currents required, while only achieving moderate resolution [8]. The Primesense[1] chipset in the Kinect uses a form of structured light; a proprietary Light Coding™technique to compute depth. The Kinect sensor consists of an infrared laser projector combined with a monochrome CMOS camera, and a second RGB video camera. Both cameras provide 640 × 480 pixel images at 30 Hz.

Little information is available on the Light Coding™technology, or the accuracy of the depth map

[1] http://www.primesense.com/?p=535

978-1-61284-985-0/11/$26.00 ©2011 IEEE

from the Kinect. This paper quantifies the absolute accuracy of the Kinect depth map and verifies its performance in the dynamic environment of quadrotor helicopter flight.

B. Kinect Hardware Details

The Kinect sensor connects to a PC/Xbox using a modified USB cable[2]. The physical USB interface remains unchanged, however subsequent to the Kinect release the protocol[3] was decoded and software to access the Kinect was enabled.

The Kinect features two cameras, a Micron MT9M112 640×480 pixel RGB camera, and a 1.3 megapixel monochrome Micron MT9M001 camera fitted with an IR pass filter. Accompanying the monochrome IR camera is a laser diode for illuminating the scene. Through reverse engineering it was determined the depth map has 11-bit resolution, and the video 8-bit. Despite the monochrome IR camera having higher resolution, both cameras only deliver 640×480 pixel images. Both image sensors have an angular field of view of 57° horizontally and a 43° vertically. The Kinect also features a microphone array and a motorized pivot, although neither of these features were required for visual flight control nor subsequently tested as part of this evaluation.

II. Calibration of the Kinect Sensors

A. Depth Camera Calibration

The depth camera returns an 11–bit number (raw values in the range $0 \dots 2047$) which needs further processing in order to extract the true depth from the sensor. A calibration procedure was performed whereby a number of reference images were captured at known distances (Figure 2).

This process was repeated a multiple times over varied ambient light conditions in order to check the insensitivity of the depth measurement to environmental conditions. The results of this calibration procedure is shown in Figure 3.

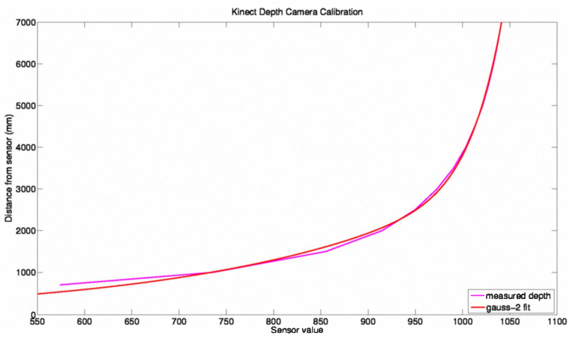


Fig. 3. Kinect depth camera calibration results and line of best fit.

A second order Gaussian model was found to be an appropriate fit ($r^2 = 0.9989$) for the data over the calibrated range.

[2]necessary to provide additional current.

[3]gratefully started by the OpenKinect project; https://github.com/OpenKinect.

Let $f(x)$ be the true depth from the image sensor, and x the raw range value, then

$$\begin{aligned} f(x) &= a_1 * exp(-((x-b_1)/c_1)^2) \\ &+ a_2 * exp(-((x-b_2)/c_2)^2), \end{aligned} \tag{1}$$

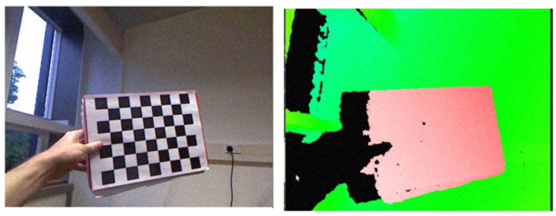
where

$$\begin{aligned} a_1 &= 3.169 \times 10^4 \\ b_1 &= 1338.0 \\ c_1 &= 140.4 \\ a_2 &= 6.334 \times 10^{18} \\ b_2 &= 2.035 \times 10^4 \\ c_2 &= 3154.0. \end{aligned}$$

It can be seen from the calibration results (Figure 3) that the depth map is accurate and repeatable over the $0.4 \dots 7.0$ m range. Additionally, if the Kinect is unable to estimate the depth to certain regions in the image, those pixels are filled with the value 2047, making it easy to ignore these pixels from further image analysis.

B. Camera Calibration and Alignment

Future research may involve combining the images from both cameras. In order to do so accurately, the intrinsic and extrinsic parameters of both cameras must be known so their images may be represented in a single co-ordinate system. Standard stereo computer vision techniques illustrated in Figure 4 were used to perform this calibration [11].

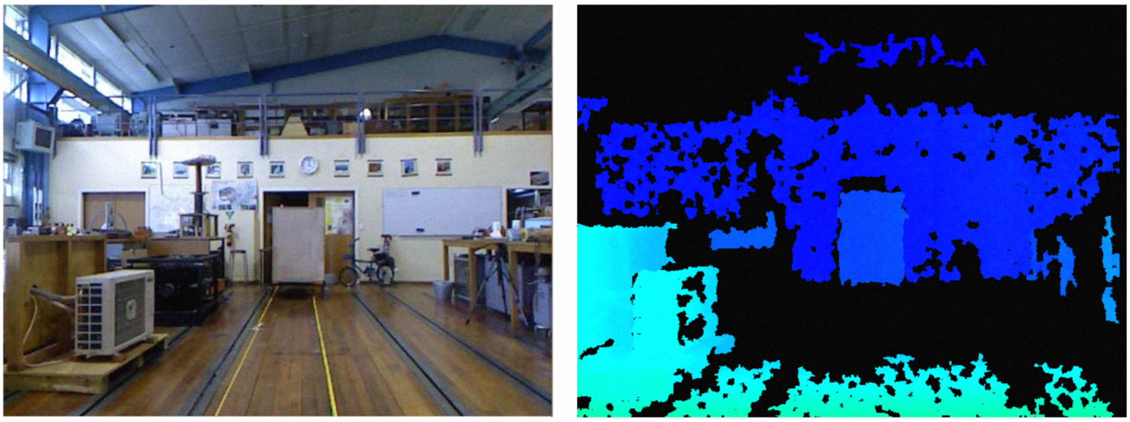


(a) (b)

Fig. 4. The standard chessboard approach for calculating the camera intrinsic parameters for the two cameras. (a) and (b) Manual matching of 4 points in both images (the corner of the chessboard) in order to calculate R and T matrices.

The RGB camera intrinsic parameters was calculated using the 'chessboard' approach and the implementation from OpenCV (*cvFindChessboardCorners* was used). Calibration of the intrinsic parameters for the depth camera was performed by manually picking out the chessboard corners from the depth image. The calibration results for the two cameras are shown in Table I.

The extrinsic parameters, the physical relationship between the cameras, was computed by manually matching the outline of the chessboard between the frames. The rotation (R) and translation (T) matrices were thus computed to be;



(a) (b)

Fig. 2. The calibration environment for testing. The board in the centre of the frame is placed 650 mm from the image plane. (a) Image captured from RGB camera. (b) False coloured depth map from depth camera.

	RGB Camera	Depth Camera
c_x	2.62×10^{2}	3.51×10^{2}
c_y	3.29×10^{2}	3.02×10^{2}
f_x	5.22×10^{2}	5.80×10^{2}
f_y	5.25×10^{2}	5.38×10^{2}
k_1	2.45×10^{-1}	-2.01×10^{-1}
k_2	-8.39×10^{-1}	9.82×10^{-1}
p_1	-2.05×10^{-3}	-7.72×10^{-4}
p_2	1.49×10^{-3}	4.89×10^{-3}
k_3	8.99×10^{-1}	-1.38×10^{0}

TABLE I
INTRINSIC CALIBRATION VALUES FOR THE KINECT RGB AND DEPTH CAMERAS.

$$R = \begin{bmatrix} 9.99\times10^{-1} & 1.39\times10^{-3} & -1.83\times10^{-2} \\ -1.88\times10^{-3} & 9.99\times10^{-1} & -1.32\times10^{-2} \\ 1.74\times10^{-2} & 1.20\times10^{-2} & 9.99\times10^{-1} \end{bmatrix}$$

$$T = \begin{bmatrix} 2.09\times10^{-2} \\ -7.12\times10^{-4} \\ -1.34\times10^{-2} \end{bmatrix}.$$

III. Quadrotor Helicopter Experimental Platform

The first quadrotors for UAV research were developed by Pounds *et al.* [12], Bouabdallah *et al.* [13], and are now a popular rotorcraft concept for unmanned aerial vehicle (UAV) research platforms. The vehicle consists of four rotors in total, with two pairs of counter-rotating, fixed-pitch blades located at the four corners of the aircraft (Figure 5).

Due to its specific capabilities, quadrotors also provide a good basis for visual flight control research. First, quadrotors do not require complex mechanical control linkages for rotor actuation, relying instead on fixed pitch rotors and using variation in motor speed for vehicle control. This simplifies both the design and maintenance of the vehicle. Second, the use of four rotors ensures that individual rotors are smaller in diameter than the equivalent main rotor on a helicopter, relative to the airframe size. The individual rotors, therefore, store less kinetic energy during flight, mitigating the risk posed by the rotors should they collide with people or objects. These combined, greatly accelerate the design and test flight process by allowing testing to take place indoors, by inexperienced pilots, with a short turnaround time for recovery from incidents. Finally, the improvement of Lithium-polymer battery technology has enabled longer flight times with heavier payloads, increasing the computational power that can be carried onboard, thus increasing the complexity of visual algorithms that can be experimented in real time.

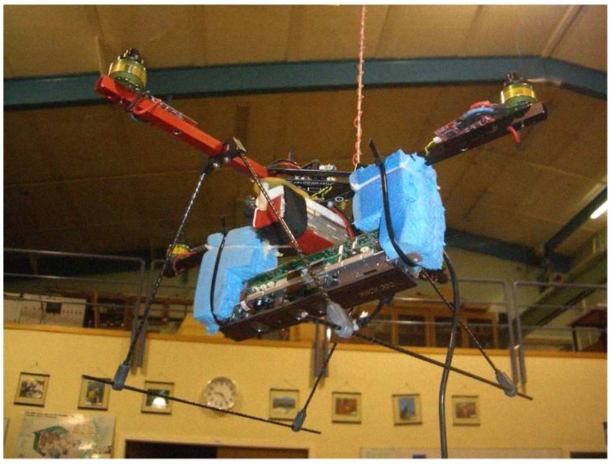


Fig. 5. Quadrotor helicopter in flight. The Kinect sensor is seen mounted below the centre of the craft, pointing towards the ground.

A. Experimental Hardware

The quadrotor is of custom design and construction [14]. It features a real time embedded attitude controller running on a 32-Bit ARM 7 Microprocessor at 60 MHz. An inertial measurement unit (IMU) is also present, and contains a 3-axis accelerometer, 3x single-axis gyroscopes, and a 3-axis magnetometer.

The quadrotor hardware consists of a cross-frame made of square aluminum tubes joined by plates in the center. This design has proved to be robust, usually only requiring a propeller replacement after a crash. On this frame are mounted four brushless motor / propeller combinations, four brushless motor controllers, the avionics and the battery. The two pairs of opposite rotating propellers are mounted at a distance of 400 mm.

The Kinect sensor is mounted under the craft, pointing towards the ground (Figure 5). Using the calibrated output from the depth camera, a control system was developed to maintain a constant altitude during flight of the quadrotor helicopter and thus evaluate the suitability of the Kinect sensor in dynamic environments. The visual altitude controller runs on a standard laptop PC. The Kinect is connected to the PC via USB which means the quadrotor is limited to operation within cable range of the laptop.

The embedded attitude controller[4] running on the quadrotor hardware will maintain attitude stability, while all control of the altitude will be handled by the visual controller using the Kinect depth map. The attitude state (roll, pitch and yaw) is continuously sent from the attitude controller over a wireless serial link, over which commands are also returned to the quadrotor.

Consider Figure 6, Let θ be the pitch angle of the quadrotor,

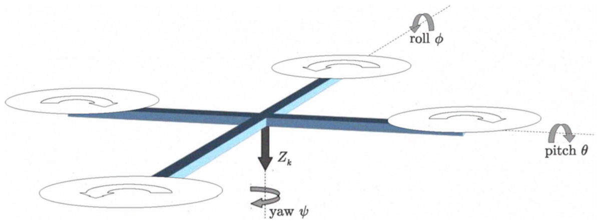


Fig. 6. Co-ordinate system for the quadrotor helicopter and Kinect camera (represented here by the arrow adjacent to Z_k).

and ϕ be the roll angle. Let Z_k be the depth observed in the Kinect body frame, that is the depth perpendicular to the image sensor. The true depth, Z_b, corrected for the craft attitude is given by;

$$Z_b = Z_k \cos\theta \cos\phi. \tag{2}$$

A proportional-integral (PI) controller was implemented to control Z_b, the quadrotor altitude. The commanded output, c, from the controller is given by;

$$c = K_P\Delta + K_I \int \Delta\, dt, \tag{3}$$

where $K_p = 5$ and $K_I = 1$ are the proportional and integral control gains determined experimentally. The control system is discrete, a 4th–order Runge-Kutta (RK4) integrator is used and dt, the frequency of update of the control system, is $\frac{1}{20\,\text{Hz}} = 5$ ms.

IV. Control Results

The visual flight controller was given complete authority to command the quadrotor thrust, and hence its altitude. Figure 7 shows the performance of the control system. A video[5] is also available.

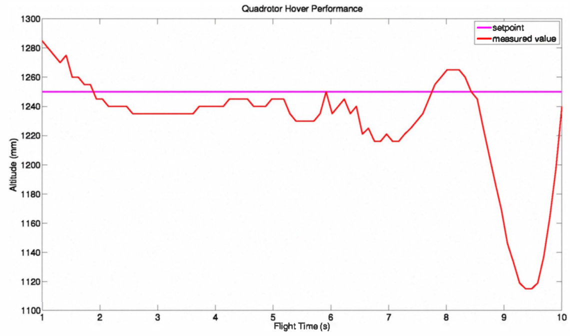


Fig. 7. Flight performance of the quadrotor attempting to maintain an absolute altitude of 1300 mm using the Kinect sensor provided depth map.

The quadrotor was commanded to hover at 1300 mm above the laboratory floor. While the PID controller is not optimal for this task (note the constant offset from the set-point due to insufficient integral action), it was suitable and allowed the quadrotor to maintain altitude until reflection from the wooden floor gave incorrect depth measurements and caused the craft to become unstable.

V. Conclusion

The successful control of quadrotor altitude using the Kinect depth map demonstrates that the sensor is capable of operation in dynamic environments. Its low cost, high frame rate and absolute depth accuracy over a useful range make it suitable for use on robotic platforms.

Further work will involve integrating the Kinect into the navigation layer of the quadrotor system. It will likely be moved into a traditional forward pointing orientation and the depth and RGB images combined in the manner described in Section II-B. The forward orientation of the depth camera will require more robust methods to detect the ground plane; the Hough transform and RANSAC will be explored for this purpose.

References

[1] D. Murray and J. J. Little, "Using real-time stereo vision for mobile robot navigation," *Autonomous Robots*, vol. 8, no. 2, p. 161, 2000.

[2] D. Wooden, "A guide to vision-based map building," *Robotics Automation Magazine, IEEE*, vol. 13, no. 2, pp. 94–98, june 2006.

[3] R. Sim and J. Little, "Autonomous vision-based robotic exploration and mapping using hybrid maps and particle filters," *Image and Vision Computing*, vol. 27, no. 1-2, p. 167, 2009.

[4] M. Kumano, A. Ohya, and S. Yuta, "Obstacle Avoidance of Autonomous Mobile Robot using Stereo Vision Sensor," in *Proceedings of the 2nd International Symposium on Robotics and Automation*, 2000, pp. 497–502.

[5] S.-Y. Kim, E.-K. Lee, and Y.-S. Ho, "Generation of ROI Enhanced Depth Maps Using Stereoscopic Cameras and a Depth Camera," *Broadcasting, IEEE Transactions on*, vol. 54, no. 4, pp. 732–740, dec. 2008.

[6] L. Nalpantidis, D. Chrysostomou, and A. Gasteratos, "Obtaining Reliable Depth Maps for Robotic Applications from a Quad-Camera System," p. 906, 2009.

[4] http://www.waspuav.org/

[5] http://www.waspuav.org/resources/icm2011/kinect-hover-video

[7] S. C. Diamantas, A. Oikonomidis, and R. M. Crowder, "Depth estimation for autonomous robot navigation: A comparative approach," in *Imaging Systems and Techniques (IST), 2010 IEEE International Conference on*, july 2010, pp. 426–430.
[8] A. Medina, F. Gayá, and F. d. Pozo, "Compact laser radar and three-dimensional camera," *J. Opt. Soc. Am. A*, vol. 23, no. 4, pp. 800–805, Apr 2006.
[9] S. Yi, J. Suh, Y. Hong, and D. Hwang, "Active ranging system based on structured laser light image," in *SICE Annual Conference 2010, Proceedings of*, aug. 2010, pp. 747–752.
[10] E. G. H. Ilstrup, David, "Single Frame Processing for Structured Light Based Obstacle Detection," in *Proceedings of the 2008 National Technical Meeting of The Institute of Navigation*, San Diego, CA, 2008, pp. 514–520.
[11] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Second ed. Cambridge University Press, ISBN: 0521540518, 2004.
[12] P. Pounds, R. Mahony, P. Hynes, and J. Roberts, "Design of a Four-Rotor Aerial Robot," Auckland, New Zealand, November 2002.
[13] S. Bouabdallah, P. Murrieri, and R. Siegwart, "Design and control of an indoor micro quadrotor," vol. 5, apr. 2004, pp. 4393–4398.
[14] J. Stowers, M. Hayes, and A. Bainbridge-Smith, "Quadrotor Helicopters for Visual Flight Control," in *Proceedings of Electronics New Zealand Conference 2010*, Hamilton, New Zealand, 2010, pp. 21–26.