Online available since 2013/Dec/06 at www.scientific.net
© (2014) Trans Tech Publications, Switzerland
doi:10.4028/www.scientific.net/AMM.479-480.677

Calibration of RGB-D sensors for Robot SLAM

Chen-Tung Chi[1,2,a], Shih-Chien Yang[2,b] and Yin-Tien Wang[2,c]

[1]Dept. of Mechanical Eng., Taipei Chengshih U. of Science and Technology, Taipei 112, Taiwan

[2]Dept. of Mechanical and Electro-Mechanical Eng., Tamkang U., New Taipei City 251, Taiwan

[a]cdchi@tpcu.edu.tw, [b]600370422@s00.tku.edu.tw, [c]ytwang@mail.tku.edu.tw

Keywords: RGB-D sensor, Sensor calibration, Simultaneous localization and mapping (SLAM)

Abstract. This paper presents a calibration procedure for a Kinect RGB-D sensor and its application to robot simultaneous localization and mapping (SLAM). The calibration procedure consists of two stages: in the first stage, the RGB image is aligned with the depth image by using the bilinear interpolation. The distorted RGB image is further corrected in the second stage. The calibrated RGB-D sensor is used as the sensing device for robot navigation in unknown environment. In SLAM tasks, the speeded-up robust features (SURF) are detected from the RGB image and used as landmarks for the environment map. The depth image could provide the stereo information of each landmark. Meanwhile, the robot estimates its own state and landmark locations by mean of the extended Kalman filter (EKF). The EKF SLAM has been carried out in the paper and the experimental results showed that the Kinect sensors could provide reliable measurement information for mobile robots when navigating in unknown environment.

Introduction

When a robot is navigating in unknown environment, it relies on various sensors to recognize outer world and to estimate state of robot itself in order to achieve the task of autonomous navigation. Commonly used sensors include laser range finder (LRF) and vision sensor. LRF can offer high precision measuring data, but it is too expensive to be extensively used. The vision sensor has relatively wide range of cost, from cheap low end to expensive high-order products, being generally applied in robot's sensing devices. Vision sensors use two-dimensional image measurement without depth information of environmental objects. While recovering depth information for visual measurement, one must spend a lot of calculation cost. Microsoft officially released in early 2012 commercial Windows version of Kinect (Kinect for Windows) sensor [1], with price about US $ 150~250. Kinect sensor, equipped with RGB lens, can capture color images of the environment in RGB format. It also has an infrared transmitter and a CMOS camcorder, which can detect the corresponding depth of images. Such kind of sensor is suitable for robot patrol in the environment for its functions and low price. The color images and 3D depth images are utilized in this study as measurement data for the robot navigation.

Related work

Microsoft Kinect sensor has several advantages, including low price, multiple sensing capability, and free software development toolkit. This study uses Kinect-provided depth and color images as environmental information for mobile robots. The Kinect 3D depth imaging sensor mainly uses light coding, which has key technology of using Laser Speckle. When laser light reaches rough objects or penetrates through frosted glass, the measured random scattering or reflection spots are called Laser Speckle. The 3D depth information can be generated by recording Laser Speckle in different position and distance by an IR camcorder, followed by comparative and statistical analysis [2].

The color images and depth images captured by Kinect need to be calibrated before they can be applied as measuring devices because of the distortion away from the original. Since distortion of color images depends on the lens structure and different parameter setting, it is necessary to find camera intrinsic parameters. The right image can be correctly restored according to these parameters. Correction can be classified into around two general types [3]: photogrammetric calibration and self-calibration. The photogrammetric calibration, taking photography against the object with known high precision locations or dimensions, is fast, but usually requires expensive calibration equipment. Self-calibration, moving camera or objects freely, and utilizing conversion relationship of different images to get intrinsic parameters of camera and external transformation parameters, has become increasingly common, and its accuracy has been greatly improved. Zhang [3] made complete derivation of conversion relationship between the images and objects, and then used Levenberg-Marquardt optimization method to get accurate parameters. Heikkila and Silven [4] established the image correction model. Bouguet [5] released the calibration toolbox based on their algorithms. Our study used these methods to deal with the calibration of color images.

The calibration of 3D depth image depended on the color camera lens and the IR depth lens. Since these lenses do not locate at the same position, we must transfer the depth information into color image at the same image location so that the depth information can be applied. In this paper, the transformation is simulated by using a bilinear interpolation model.

The image features detected by Kinect sensor can be used to represent landmarks in the environment and provide necessary environment message for robot navigation. Because the camera is moving, it is not easy to detect and recognize imaging characteristics. A detection method based on scale-invariant feature was developed by Lindeberg [6]. He established image Hessian matrices, whose elements are the convolution of the Gaussian second order derivative (Laplacian of Gaussian, LoG) with the image. An image feature is selected by examining the determinant of Hessian matrix based on the non-maximum suppression rule. The advantages of this method include that the scale-invariant features can be captured and has high stability and repeatability. On the other hand, the disadvantage is that the computation is intensive. For the scale-invariant feature issues, Bay *et al*. [7] replace the matrix of Gaussian second order derivative with the box filter, calculate the approximate of determinant of the Hessian matrix, and then couple with integral imaging method. This method that significantly reduces calculation time is called SURF (Speeded-Up Robust Features). SURF is applied in this study to detect features from the Kinect color images. The feature map of environmental image is therefore established by combining Kinect's depth information. Environmental map are used in recursively updating and estimating the positions and speeds of the robot in the environment. The estimation is achieved by using Extended Kalman Filter (EKF) [8].

Alignment of RGB and Depth Images

The Kinect captures both the color and depth images, as shown in left and middle panels of Fig. 1, to construct the stereo coordinates of environmental objects. However, the pixel coordinates of a corresponding object in these two images could not be matched correctly because the RGB and depth cameras are not located at the same position. There exists unknown transformation between the RGB and depth images. We need to determine the relation between these two image planes in order to correct the image slant.

The depth image is a grayscale image converted from the depth datum of the original IR image, as shown in right panel of Fig. 1. There is a deviation of 8 pixels between the depth and IR images. We implement the alignment procedure for the RGB and IR images using the method of geometry transformation [9]. Taking the geometry transformation of a quadrangle area as an example, Fig. 2 shows the corresponding sketch map of two planes, where the vertices of the quadrangle are the corresponding connection points. We use a pair of bilinear equations to represent the pixel coordinate transformation between RGB and IR images

$$x' = Ax+By+Cxy+D. \quad (1)$$

$$y' = Ex+Fy+Gxy+H. \quad (2)$$

where x and y are the coordinates on the RGB image; while x' and y' are the coordinates on the IR image. The 8 parameters can be solved from the equations if there are 4 corresponding points.

Constructed by the 8 parameters, the transform model can be used to transform all the pixels within the quadrangle area defined by the 4 vertices. By substituting the coordinates of the object features in the RGB image into Eqn. (1) and (2), the coordinates of object features in the IR image can be determined. After the alignment of the RGB and IR images, the 8-pixels deviation is added to the IR image in order to obtain the pixel location of depth image.

Two real images of the chess board are taken separately by utilizing RGB lens and IR lens for calibration. Twenty feature points are fetched from two planes and the corresponding pixel coordinates are recorded respectively. In order to determine the depth values of the object features in the RGB image, the 20 pixel coordinates, which come from the RGB image, are substituted into the right side of Eqn. (1) and (2). The 20 pixel coordinates from IR image are substituted into left side of the same equations. The 8 parameters are solved and listed as follows:

A= 1.0878	B= -0.0100	C= 0.0001	D= -27.4165
E= -0.0029	F= 1.0970	G= -0.0000	H= -21.2797

These 20 pixel coordinates on the IR image, RGB image and coordinates transformed from RGB image via the equations are charted in Fig. 3, where green triangles are the original coordinates on the RGB image, red squares are the coordinates on the IR image, blue rhombi are RGB coordinates after calibration, coordinates of blue rhombi are near to the coordinates of red squares.

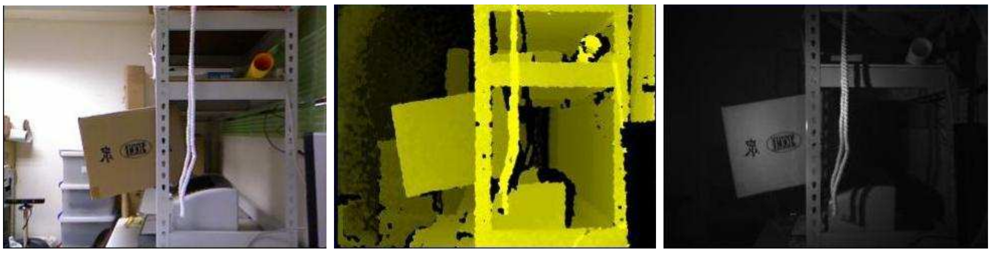

Fig.1 RGB image (left), depth image (middle) and IR image (right)

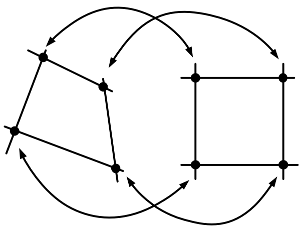

Fig.2 Sketch map of two responding planes

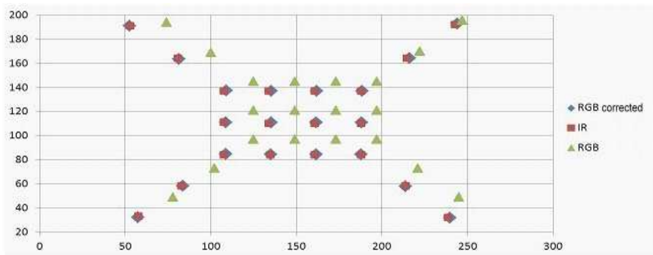

Fig.3 Chart of calibrated pixel locations

Calibration of the RBG image

The procedure of the RGB camera calibration is as follows. 20 photos of black-and-white chess board are taken, as shown in Fig. 4. Each photo is shot in different directions, whose angle is always less than 90 degree. All photos are read by using the MATLAB program. Pick four corners of each photo sequentially and the intrinsic parameters of the camera can be calculated. Another eight inverse parameters can be determined by operating the intrinsic parameters of the camera. The images can be revised by using the inverse parameters [4].

Calibration Results

Comparison of 20 point coordinates before and after transformation is listed in Table 1. All the pixel coordinates on RGB image in Fig. 1 are put into the equations for transformation. The grayscale values of the transformed coordinates on IR image are superimposed on the original RGB. Two images coincide each other roughly except the edge area with no depth information, indicating good calibration, as shown in Fig. 5.

Table 1 Comparison of 20 point coordinates before and after transformation

type	Coordinate					RMSE
IR image	58,33	239,32	53,191	243,192	83,58	0
	214,58	81,164	215,164	108,84	188,84	
	108,137	188,137	135,84	161,84	108,111	
	134,110	161,110	188,110	134,137	161,137	
RGB image	78,49	245,49	74,194	247,196	102,73	17.85
	221,73	100,169	222,170	125,97	197,97	
	125,145	197,145	149,97	173,97	125,121	
	149,121	173,121	197,121	149,145	173,145	
RGB and image corrected	57.3,32.2	239.8,31.7	52.5,191.3	244.1,193.0	83.5,58.5	0.97
	213.8,58.1	81.3,163.8	216.1,164.5	108.8,84.7	187.8,84.5	
	108.9,137.4	188.2,137.2	135.1,84.6	161.4,84.6	108.8,111.0	
	135.2,111.0	161.6,110.9	188.0,110.8	135.3,137.3	161.8,137.2	

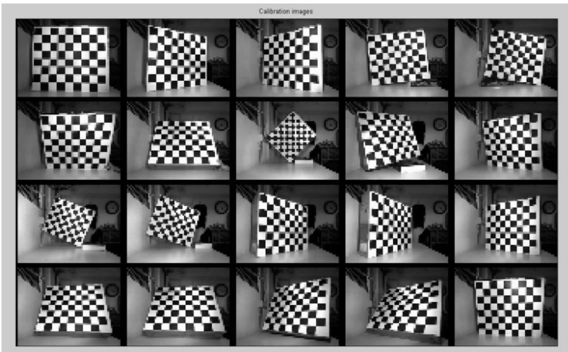

Fig. 4 20 photos of chess board

Fig. 5 Overlap image of Kinect RGB image and depth image

Robot SLAM

While the robot is cruising, we can get measuring information by Kinect sensor, and utilize the above-mentioned methods to rectify the information. After getting correct RGB image and depth data, the SURF algorithm is applied to detect the features and landmarks on the image. Coping with the depth data, 3D coordinates of robot or landmark can be calculated and applied in the update step of the Extended Kalman filter. The purpose of simultaneously tracking robot position and estimating environment state can be achieved. In this paper, a self-location experiment for a moving robot in the indoor environment was completed.

The experimental results are so presented that the status of features is represented by means of different colors. If new feature is detected, it is shown by a yellow square. The unstable feature is represented by a blue square. If the feature is confirmed as stable one after calculating, red square is used. The green color is depicted if there is misjudge, which always takes place in camera moving or vibration. In this paper, the origin of the world coordinate in the map is defined as the initial position of SLAM system. The images of detecting routs and localization are captured in Fig. 6 and Fig. 7, which show the RGB image, depth image and trajectory of camera drew by MATLAB in initial and latter stages of experiment respectively.

Conclusion

We developed a procedure to calibrate the RGB and depth images of Kinect sensors. The calibrated RGB-D sensor was used as a measuring device for robot navigation in unknown environment. The SURF features were detected from the corrected RGB images. For each feature, the pixel coordinate obtained from the RGB image was combined with the depth information from the IR image to calculate the 3D coordinate. Two experiments were carried out to demonstrate the performance of the RGB-D sensors for robot SLAM systems. The results showed that the RGB-D sensor can provide reliable measurement information when robot navigating in unknown environment.

Fig. 6 Image of the initial stage of experiment (new features were detected)

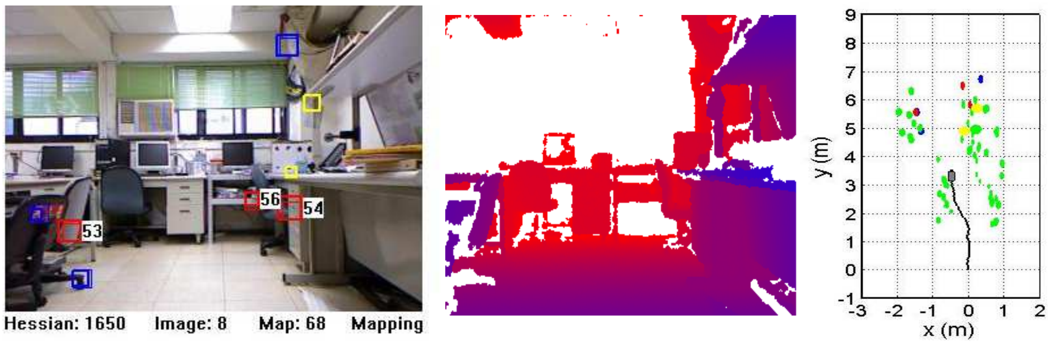

Fig. 7 Image of the latter stage of experiment (stable features were ascertained)

Acknowledgments: This paper was partially supported by the National Science Council in Taiwan under grant no. NSC101-2221-E-032-017 and NSC101-2632-E-032-001-MY3.

References

[1] Information on http://www.microsoft.com/en-us/kinectforwindows/

[2] Information on http://www.techbang.com/posts/2936-get-to-know-how-it-works-kinect

[3] Z. Zhang, in: A flexible new technique for camera calibration, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.22 no.11 (2000), p. 1330

[4] J. Heikkila and O. Silven, in: A Four-step Camera Calibration Procedure with Implicit Image Correction, Proc. IEEE CVPR (1997), p. 1106

[5] Information on http://www.vision.caltech.edu/bouguetj/calib_doc/

[6] T. LIndeberg, in: Feature detection with automatic scale selection, IJCV Vol. 30 (2) (1998), p. 79

[7] H. Bay, A. Ess, T. Tuytelaars and L.V. Gool, in: Speeded-Up Robust Features (SURF), CVIU Vol.100 (2008), p.346

[8] G. Welch and G. Bishop, An Introduction to Kalman Filter, UNC-Chapel Hill, TR 95-041 July 24 (2006), p. 14

[9] R. C. Gonzalez and R. E. Woods, Digital Image Processing, Prentice Hall (2007)

Applied Science and Precision Engineering Innovation
10.4028/www.scientific.net/AMM.479-480

Calibration of RGB-D Sensors for Robot SLAM
10.4028/www.scientific.net/AMM.479-480.677