A Novel Method To Collect Kinect-Comparable RGB-D Outdoor Image

Xinliang Zhu{}, Shuming Tang{}

High-tech Innovation Center, Institute of Automation, Chinese Academy of Sciences, 95 Zhongguancun East Road, Beijing, China.

Abstract

RGB-D based visual computing has received increasing attention among researchers, such as pose estimation, segmentation, people detection, object labelling and so on. However, almost all the existed RGB-D based researches are using indoor RGB-D datasets. The main reason is that there's no available sensors for collecting outdoor RGB-D images. In this paper, we combine a ToF camera with a RGB camera and solve the image registration problem based on pin-hole camera model. The initial experiments show that our method can perform as well as Kinect indoors but surpass Kinect outdoors.

Keywords

RGB-D; Sensor Fusion; Camera Model; Kinect

Introduction

Computer vision plays more and more important role in intelligent systems such as autonomous vehicle, surveillance, and so on. Most of those computer vision technologies are based on RGB camera, which often suffers from the light condition, occlusion, high computing complexity, etc.

To solve those problem and improve the performance of computer vision algorithms in object detection and recognition, many researchers tried to employ RGB-D images [Di, Spinello, Sung], which combines the depth information with RGB images. The performance of RGB-D based algorithms were improved significantly in some applications such as people detection, gesture recognition [Di, Spinello, Luber] and so on. There are many existed RGB-D based dataset available to the researchers [Lai, Luber, Singh, Spinello].

However, almost every current existed RGB-D dataset are collected indoors using Kinect. The most significant limit of Kinect now is that it can't be used outdoors, which keeps it away from the usage in some important scenes like pedestrian detection.

In this paper, we propose a new method, which combines Time-of-Flight (ToF) camera (depth camera) with common RGB camera (Fig 1), to solve the problem above and generate outdoor available RGB-D images.

FIG. 1 TOF CAMERA WITH RGB CAMERA

In the left parts of this paper, we first solve the problem of image registration based on pin-hole camera model in image registration section. Then the experimental results are compared with Kinect. Finally, we conclude that our method is comparable with Kinect though with some limits.
Image Registration

To fuse the two different cameras is the first step of generating RGB-D image. Image registration is to register depth information into every pixel in the RGB image.

In this section, we will introduce a novel method to solve the problem based on the pin-hole camera model. In pin-hole camera model, a 3D point (X,Y,Z) in the world coordinate space can be projected into the image plane using a perspective transformation:

\[
sm' = A[R|t|M']
\]

where, R, t are the extrinsic parameters of the camera, A is a camera matrix, \((c_x, c_y)\) is a principal point that is usually at the image center and \(f_x, f_y\) are the focal lengths expressed in pixel units.

Considering every pixel’s value in a depth image is the depth value, which represents the Z value, in a real-world, we proposed an image registration method which employing this feature. The process of image registration is shown in Fig. 2. Take one point \((u, v)\) with value of \(z\) from depth image for example. Firstly, project it inversely to the depth camera’s camera coordinate system following the formula:

\[
p_{uv} = \frac{z(u - c_x)/f_x}{z(v - c_y)/f_y}
\]

where, \(p_{uv}\) is the coordinate in depth camera’s camera coordinate system, \(c_x, c_y, f_x\) and \(f_y\) are the intrinsic parameters of ToF camera. The goal is to assign every depth value in depth image to a point in RGB image. We calculate the relative extrinsic parameter between the ToF camera with the RGB camera instead of projecting the points in depth camera’s camera coordinate system into the world coordinate system to finish the image registration. In the following, the relative extrinsic parameter is calculated. Assume one point \(p\) in real world coordinate system, \(X_d, X_e\) and \(X_w\) are the coordinates in depth camera’s camera coordinate, RGB camera’s camera coordinate and world coordinate system. Then, \(X_d, X_e\) can be expressed in the following formula:

\[
X_d = R_dX_w + T_d
\]

\[
X_e = R_eX_w + T_e
\]

where, \(R_d, T_d, R_e, T_e\) are the extrinsic parameters of depth camera and RGB camera. Eliminate the \(X_w\),

\[
X_c = R_cR_d^{-1}X_d + T_c - R_cR_d^{-1}T_d
\]

Then, the relative extrinsic parameter can be expressed as follows:

\[
R_r = R_eR_d^{-1}
\]

\[
T_r = T_c - R_cR_d^{-1}T_d
\]

Now, project the point \(p_{uv}\) into RGB camera’s image plane and finish the process of image registration using formula:

\[
q_{xy} = R_r(p_{uv} - T_r)
\]

where, \(q_{xy}\) is the depth value of pixel \((x,y)\) in RGB image plane.

![Fig. 2 The Process of Image Registration](image)

Experimental Results

We compare our experimental results with Kinect in the same scene. The scenes we choose includes indoor scene and outdoor scene. From the comparable results, we can conclude that the RGB-D images acquired by our method are comparable with Kinect indoors and better than Kinect outdoors.

We also do linear interpolation for our depth image, because the initial size of depth image is 144 x 176, which is near 12
times smaller than the RGB image. Fig 3 shows the results of our method applied to image registration indoors and outdoors. The first row in (a), (b) shows the result of our method and the second row shows the result of Kinect. (a) and (b) compare the performance indoors and outdoors separately. From it, we find our method is totally comparable to the performance of Kinect within 2.5 meters, while with high resolution in the distance over 2.5 meters. The Kinect is totally unavailable outdoors in sunny weather, while our method solves the problem very well.

However, there are some limits within our method comparing to Kinect. Firstly, since the ToF camera is with rather less pixels compared to RGB camera, it often needs to do interpolation to make it look well, which will usually enlarge the noise of the depth image and the computation time of image registration. Secondly, the denoising method needs to be employed to make a better RGB-D image.

**Conclusions**

We have proposed a novel method to collect Kinect-comparable RGB-D outdoor image using a ToF camera and a RGB camera. We solve the problem of image registration problem which is important to fuse two cameras based on the pin-hole camera model. A concise theory and elegant model is proposed. We also discuss the limits of our methods. In the future, we will focus our research on denoising the RGB-D image and analyze the performance of image registration quantitatively.

**ACKNOWLEDGEMENT**

This research work is supported in part by National Natural Science Foundation of China under Grants 91120008 and 61220002. We want also to show our great appreciation to Professor Jinqiao Wang for his help and discussions in image registration.

**REFERENCES**


http://docs.opencv.org/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html


Kubacki J. “Learning and detecting objects in combined range and color images based on local feature frames.” Bielefeld: Bielefeld University; 2012.


**Xinliang Zhu** is born in Siyang, Jiangsu Province, on February 18, 1991. Xinliang Zhu is pursuing his Master of Science degree in control theory and control engineering in Institute of Automation, Chinese Academy of Sciences, Beijing, China now. Xinliang Zhu focuses his research on computer vision, machine learning and intelligent robots.

He has published 1 paper on the cognition of autonomous vehicle and 1 paper on the subset selection, which was accepted by AAAI.

**Shuming Tang** received her Ph.D degree in control theory and control engineering from the Graduate University of Chinese Academy of Sciences, Beijing, China. Currently, she is an associate research professor in the Institute of Automation, Chinese Academy of Sciences. Her research interests are focused on intelligent transportation systems (ITS) and intelligent vehicles. She has published extensively in those areas.

She is a committee member of Robot Sports in China, an associate editor of the IEEE Transactions on ITS, a Co-Chair of the Technical Committee on ATS of Intelligent Transportation Systems Society (ITSS). She was a registration Co-Chair of the 2003 IEEE International Conference on ITS, an AdCom member of IEEE Intelligent Transportation Systems Council (ITSC) in 2004, a Publicity Chair of the 2005 IEEE International Conference on Vehicular Electronics and Safety, and a member of the Board of Governors of the IEEE ITSS from 2005 to 2006.