An Algorithm for Obstacle Detection based on YOLO and Light Filed Camera

1st Rumin Zhang
Medialab
University of Electronic Science and Technology of China
Chengdu, China
rm_zhang@uestc.edu.cn

2nd Yifeng Yang
Medialab
University of Electronic Science and Technology of China
Chengdu, China
jervint@hotmail.com

3rd Wenyi Wang
Medialab
University of Electronic Science and Technology of China
Chengdu, China
wangwenyi@uestc.edu.cn

4th Liaoyuan Zeng
Medialab
University of Electronic Science and Technology of China
Chengdu, China
lyzeng@uestc.edu.cn

5th Jianwen Chen
Medialab
University of Electronic Science and Technology of China
Chengdu, China
chenjianwen@uestc.edu.cn

6th Sean McGrath
Marconi Center
University of Limerick
Limerick, Ireland
sean.macgrath@ul.ie

***Abstract*—This paper presents a novel obstacle detection algorithm in the indoor environment. The algorithm combines the YOLO object detection algorithm and the light field camera which is more simple than normal RGB-D sensor and acquires depth image and high-resolution images at the same in one exposure. The RGB Image rendered by the light filed camera is taken as an input of the YOLO model which was trained base on nearly 100 categories of common objects. According to the object information and the depth map, the obstacle was accurately calculated including its size and position. Experimental results demonstrate that the proposed method can provide higher detection accuracy under indoor environment.**

***Index Terms*—Obstacle detection, light field camera, YOLO, depth map**

I. Introduction

Obstacle detection takes an important role in autonomous and semi-autonomous vehicles. Detecting obstacles especially minor objects and reacting to the changes of environment efficiently in time is still an open problem. This paper focus on detecting obstacles in path of the vehicle, which operating like humanoid system.

Several researchers have investigated various sensing technologies to detect obstacles in the vicinity of vehicles. Kang et al. [1] described a method by minimizing an energy function to remove the unconcerned background from the image base on IR line. But this method has a limitation of measuring distance and IR is susceptible by intense of illumination. Aminogram et al [2] designed an obstacles detection algorithm fusing the laser data with vision. But this scheme is able to offer only probable orientation of the object. Besides, there are researches on operating deep learning on object recognition and obstacle detection. Nguyen et al. [3] presented a system combining deep learning with detection, recognition and tracking on-road vehicles and pedestrians.

However, this method need multiple sources of deep information and local patterns which increase the cost of whole system. This solution is not suitable for indoor environment. Wang et al. [4] proposed an approach to recognize brakelight to operate vehicle following for an autonomous driving system based on deep learning. Asif et al. [5] researched on performing objects recognition and grasp and minimizing the uncertainty of the class labels based on training an novel objective function which extracted from RGB-D point cloud. The problem of the method is that the analysis of the point cloud data is cost.

In this paper, we propose an obstacle detection algorithm by combining deep learning with the light field camera. The algorithm will classify objects into categories and mark them in the image. The exact size and position of the obstacle can be calculated based on the recognition result and the depth map.

The remainder of the paper is organized as follows. Section 2 describes the proposed algorithm explaining the obstacle detection process. In Section 3, experimental results are shown. Section 4 concludes the paper and identifies some directions for future work.

978-1-5386-5147-6/18 $31.00

II. Proposed Method

Our system is implemented with a powerful processor as well as a light field camera. In order to extract metric information from 2D image, the calibration of the light field camera is a necessary step in our algorithm. The RGB Image and depth map can be rendered base on the 4D information captured by the light field camera. Then the RGB image is taken as input of YOLO and the YOLO detection algorithm outputs object recognition results the scene. Then the filter process is applied on the outputs. Finally, the obstacles are detected based on depth map and the object information.

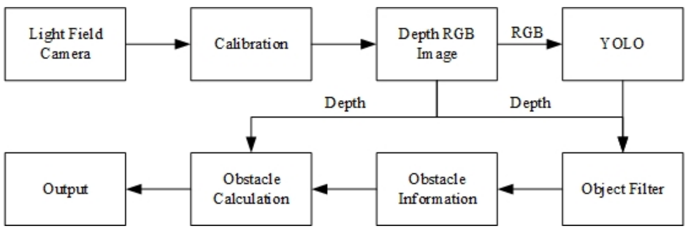


Fig. 1. Framework of the proposed method

A. Light Field Camera

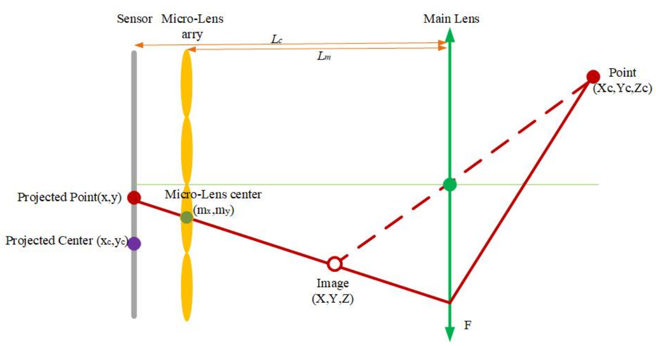


Fig. 2. Projection model of the micro-lens-based light field camera. The thin-lens model and the pinhole model are used for the main lens and the micro-lenses, respectively.

Compared with the traditional camera, the light field camera can capture the 4D light field including 2D spatial information and 2D directional information in a single snapshot. Capturing images with the light field camera solves many of the problems faced by the conventional photography. Rendering refocused images, all-in-focus images, depth map, 3D scene reconstruction, HDR and super-resolution are just a few of the emerging technique [6]–[8].

B. Calibration

Let F be the focal length of the main lens. Assume an arbitrary point (X_c, Y_c, Z_c) in the camera coordinate system, all the rays from the point pass through the main lens and head to a point which can be called image.Let (X_w, Y_w, Z_w) be one of object in the word coordinate system. It should be transformed into the camera coordinate system by an transform matrix with a $3*3$ rotation matrix R and a $3*1$ translation vector t:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = R \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} + t = \begin{bmatrix} r_{11}X_w + r_{12}Y_w + t_1 \\ r_{21}X_w + r_{22}Y_w + t_2 \\ r_{31}X_w + r_{32}Y_w + t_3 \end{bmatrix} \quad (1)$$

Where r_{ij} and t_i are the elements of R and t at the i-th row and j-th column.

C. Depth Estimation

Based on the SAD (sum of Absolute Difference) algorithm, the disparity between the sub-images can be obtained [9].

$$\sigma(\Delta_n) = \frac{1}{M} \sum_{i_n=1}^{i_n=m'-\Delta_n} |r_n(i_n + \Delta_n) - r_{n+1}(i_n)| \quad (2)$$

Where Δ_n is the disparity of pixels between adjacent sub-images. M is the number of the matching pixels. m' represents pixels number of one sub-image. $r_n(i_n)$ is the radiance function of the sub-image, where $0 < n < N$, $0 < i_n < m$ and m is the number of the pixels corresponding to each micro-lens and $m = d/\delta$ (δ-the size of each pixel).

Disparity of the sub-image can be found by minimizing the above equation through changing the Δ_n. As the Fig.3 illustrates, the depth a could be calculated with the optimal Δ_n.

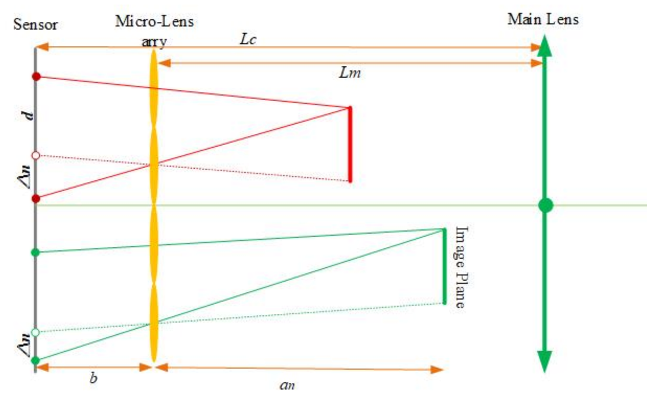


Fig. 3. The disparity of the object between the neighbor views formed by micro-lens.

According to the principle of similar triangle, we can derive the following equation.

$$\frac{d}{\Delta_n \sigma} = \frac{a_n}{b} \quad (3)$$

$$a_n = \frac{bd}{\Delta_n \sigma} = \frac{m}{\Delta_n} b \quad (4)$$

where δ is the pixel diameter and f is the focal length of the micro-lens. As the Fig.4 depicts, the micro-lens width is d. An describes the object length corresponding to the n-th micro-lens. And the position of the sub-image depends on a_n. Let b be the distance between sensor and micro-lens plane.

TABLE I
COMMON INDOOR OBJECT CLASSES

Bedroom		Utility room		Living room	
glasses	socket	skateboard	ball	microphone	coin
table lamp	piano	wrench	screwdriver	air conditioner	fan
guitar	hat	broom	mop	frame	sofa
stationery box	pen	hammer	racket	coffe table	remote
ruler	laptop	dustpan	iron	wallet	book
eraser	wire	suitcase	kite	TV	vase
Bathroom		Kitchen		Common	
shampoo	soap	trash can	gas tank	cash	sliper
razor	toilet	cup	spoon	person	table
towel	washing machine	microwave oven	plate	chair	dog
toothbrush	toothpaste	bowl	fork	cat	potted plant
hair drier	rinse cup	knife	winglasses	umbrella	bottle
washboard	washbasin	chopsticks	gas stoves	key	shelf

The distance Z_w of the object point from the lens is computed using equation (4) and the lens equation, as follows:

$$\frac{1}{z_w} + \frac{1}{L_m - a_n} = \frac{1}{F} \tag{5}$$

D. RGB Image Rendering [10]

Let $r(q,p)$ be the radiance at a given plane perpendicular to the optical axis, where q and p represent the position and direction in ray space, respectively. The intensity of an image at a given spatial point, denoted $I(q)$ is the integral of the radiance over all of the rays incident at that point, as following:

$$I(q) = \int_p r(q,p)\, dp \tag{6}$$

E. YOLO

YOLO is a new approach to object detection based on the deep learning [11] as showed in Fig.4, it analyses $S * S$ rectangles to get all the probability of every possible object and finally reach the point for segmentation and grasp of objects applied in many ways such as face detection or object recognition, and the precise of some experiment is above 75%, and even reach higher in some experimental environment such as now we know as SSD [12] and DSSD [13]. So we propose a heuristic method to use deep learning for obstacles recognition and segmentation in RGB image in complex indoor environment.

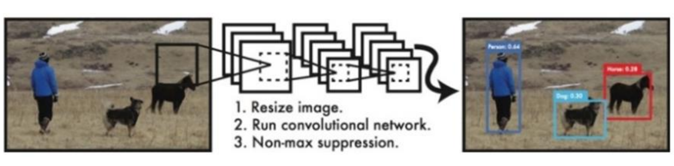


Fig. 4. The YOLO Detection System. Processing images with YOLO is simple and straightforward.

F. Object Filter

When the RGB image was taken as input of the YOLO system, there would be multiple objects in the output recognition result. But the objects far away the vehicle or drew by two or more bounding boxes should be removed from the recognition result. So, the object filter is necessary in our algorithm.

We design our filter based on the depth map rendered in subsection B. Assuming k is the threshold and the central coordinate of one object is (x_c, y_c) in depth map, we can obtain the following equation.

$$\begin{cases} if\ d(x_c, y_c) > d, object\ removed \\ if\ d(x_c, y_c) < d, object\ reserved \end{cases} \tag{7}$$

G. Obstacle Calculation

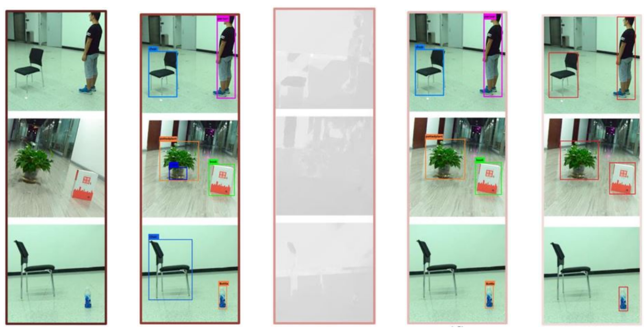
Using the object detection result based on $YOLO$ detection algorithm, the obstacle can be derived and its information such as its location, size and distance from the camera would be calculated. We define the format of the obstacle information as {Obstacle: x, y, $width$, $height$}, where (x, y) and $(width, height)$ represent the coordinates of the pixel which locates at the upper left corner of the contour and width and height of the contours, respectively. According to the camera parameters and the original depth map, the physical dimensions and distance from the camera of the obstacles in the scene can be obtained.

III. EXPERIMENTAL RESULTS

The algorithm used for obstacle detection was written using the C++ programming language with CMake. The implementation extensively makes use of OpenCV, OpenGL and CUDA. Besides the OpenCV, we also work with some external libraries including CUDNN, CUDA, OpenGL, GLFW3 and GLEW. The obstacle detection algorithm was tested on a computer system with dual-core Intel core i7-7700 processors running at 3.20 GHz and a 16GB main memory.

In order to detect the common obstacles in the indoor environment, we download the images of the common obstacles from network and labeled them for training YOLO. The common objects are listed as in table 1.

As the Fig.5 denotes, the left column is the RGB images rendered by the light field camera. Then the recognition results of the YOLO model are described in the second column on the left. Depth map calculated based on the 4D light field information is used to remove the objects which is recognized by the YOLO algorithm but unconcerned for the obstacle detection, as shown in (c) and (d). Finally, the obstacles are detected and its location and size information are also calculated us the parameters of the light field camera and the depth map.



(a) (b) (c) (d) (e)

Fig. 5. The performance of our algorithm. (a) RGB Image rendered by the light field camera. (b) Recognition result using the YOLO model. (c) The depth map calculated based on the 4D information of the scene. (d) The recognition result filtered by our filter. (e) Output.

IV. Conclusion

In this paper, we proposed a new method to detect obstacles in the indoor environment that introduces deep learning technology and the light field camera to recognize the obstacle and perceive its information. According to the recognition result and depth map, the object filter is applied to remove the unconcern obstacle. To demonstrate the performance of our method, different types of scene, including pedestrian, chair, book and so on, are demonstrated. The experimental results prove the effectiveness of our obstacle detection algorithm. In the future, we will further investigate how to detection the obstacles in real time base on the light filed camera and the deep learning technology.

Acknowledgment

This work is partially supported by the research pro-gram of The application of VR based motion capture techniques in Sichuan Provincial Race-walking Teams training (Grant Number WQ21SS1115Y17001).

References

[1] Kang M C, Kim K S, Noh D K, et al. A robust obstacle detection method for robotic vacuum cleaners[J]. IEEE Transactions on Consumer Electronics, 2015, 60(4):587-595.

[2] F. Utaminingrum et al.A laser-vision based obstacle detection and distance estimation for smart wheelchair navigation," 2016 IEEE International Conference on Signal and Image Processing (ICSIP), Beijing, 2016, pp. 123-127

[3] Nguyen V D, Nguyen H V, Tran D T, et al. Learning Framework for Robust Obstacle Detection, Recognition, and Tracking[J]. IEEE Transactions on Intelligent Transportation Systems, PP(99):1-14.

[4] J. G. Wang et al., Appearance-based Brake-Lights recognition using deep learning and vehicle detection, 2016 IEEE Intelligent Vehicles Symposium (IV), Gothenburg, 2016, pp. 815-820.

[5] Asif U, Bennamoun M, Sohel F A. RGB-D Object Recog-nition and Grasp Detection Using Hierarchical Cascaded Forests[J]. IEEE Transactions on Robotics, PP(99):1-18.

[6] Zhang R, Liu P, Liu D, et al. Reconstruction of refocusing and all-in-focus images based on forward simulation model of plenoptic camera[J]. Optics Communications, 2015, 357:1-6.

[7] Bishop T E, Favaro P. The Light Field Camera: Extended Depth of Field, Aliasing, and Superresolution[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2011, 34(5):972-986.

[8] Ng R. Digital light field photography[M]. Stanford University, 2006.

[9] Zhang R, Yu R, Liu D, et al. All-Focused Light Field Image Rendering[M]// Pattern Recognition. Springer Berlin Heidelberg, 2014:32-43.

[10] Georgiev T G. Focused plenoptic camera and rendering[J]. Journal of Electronic Imaging, 2010, 19(2):021106.

[11] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, Real-Time Object Detection[J]. 2016:779-788.

[12] Liu W, Anguelov D, Erhan D, et al. SSD: Single Shot MultiBox Detector[J]. 2016:21-37.

[13] Fu C Y, Liu W, Ranga A, et al. DSSD : Deconvolutional Single Shot Detector[J]. 2017.