International Conference on Computer and Communication Engineering (ICCCE 2012), 3-5 July 2012, Kuala Lumpur, Malaysia

Stereo Vision Images Processing for Real-time Object Distance and Size Measurements

Yasir M Mustafah, Rahizall Noor, Hasbullah Hasbi, Amelia Wong Azma
International Islamic University Malaysia,
P.O. Box 10, 50728, Kuala Lumpur

Abstract— **Human has the ability to roughly estimate the distance and size of an object because of the stereo vision of human's eyes. In this project, we proposed to utilize stereo vision system to accurately measure the distance and size (height and width) of object in view. Object size identification is very useful in building systems or applications especially in autonomous system navigation. Many recent works have started to use multiple vision sensors or cameras for different type of application such as 3D image constructions, occlusion detection and etc. Multiple cameras system has becoming more popular since cameras are now very cheap and easy to deploy and utilize. The proposed measurement system consists of object detection on the stereo images and blob extraction and distance and size calculation and object identification. The system also employs a fast algorithm so that the measurement can be done in real-time. The object measurement using stereo camera is better than object detection using a single camera that was proposed in many previous research works. It is much easier to calibrate and can produce a more accurate results.**

Keywords-component; stereo vision; distance measurement; size measurement

I. INTRODUCTION

Object distance and size measurement is becoming more essential in many applications especially in mobile autonomous system. Information on the distance and size of the surround objects is useful for the navigation and localization of the mobile system. Since most of autonomous systems nowadays are equipped with vision sensors or cameras, it is very beneficial that the vision information is utilized to obtained distance and size information that can be used to assist the system.

Many research works have been done to obtain a particular object distance from an image. Initially most of the proposed method utilizes only a single vision sensor. For example, Rahman et al, proposed a person to camera distance measuring system using a single camera based on the eye-distance [1]. The distance is obtained by calculating the variation in eye-distance in pixels with the changes in camera to person distance. Meanwhile, Wahab et al. proposed a monocular vision system which utilizes Hough transforms [2]. Their system also depends on the relative size of the targeted object to determine the distance. A different approach utilizing a single camera was proposed by Kim et al. They proposed a distance measurement method using a camera and a rotating mirror [3]. A camera in front of a rotating mirror captures a sequence of reflected images. The images are then analyzed to obtain the distance information. The distance measurement is based on the idea that the corresponding pixel of an object point at a longer distance moves at a higher speed in a sequence of images.

Recently, several research works have started to utilize multiple vision sensors for the purpose of object distance and size measurement. Most of the works propose to utilize stereo configuration of cameras. One example is the work of A-Lin et al. where they utilize stereo vision to measure the safe driving distance of a vehicle [4]. The distance measurement is based on the disparity of the front car in the two frame capture by the stereo camera. Than in the work publish by Baek et al. an improvement in the calculation of the disparity is proposed so that a more accurate object distance can be measured [5]. In their work they had shown that by using their method the disparity of the object in a larger view area can be obtained.

Most of the work proposed in the literature focused only on the distance measurement. As we have stated before, the object size information is also very useful especially in a navigation and localization of an autonomous system. Moreover, the size information also can be used for a short term object identification which is useful for autonomous system. In this paper we proposed to utilize a method of measuring the distance of an object as well as the size of the object using stereo vision sensor. The method also employs a much faster algorithm so that the measurement can be done in real-time.

II. OBJECT SIZE AND DISTANCE MEASUREMENT

The flow of the object size measurement system that we proposed started with stereo vision image capture. Then, on both images, a preprocessing will be applied followed by object detection and segmentation. Finally the object distance and size will be calculated. From the object size, object identification can be done. The flow of the object size measurement system is as illustrated in the Figure 1 below.

A. Stereo Image Capture

Stereo image capture is done by using two video cameras which are aligned in parallel in fixed position. Both cameras are calibrated so that they have matching image properties such as the size, color space and lighting. The object of interest can be measured for its distance and size when it enters the overlapping view of the two cameras. Figure 2 below illustrates the stereo vision setup.

978-1-4673-0479-5/12/$31.00 ©2012 IEEE

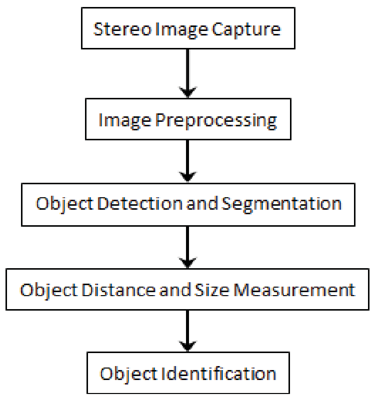

Figure 1. The flow of the proposed object size measurement system utilizing stereo camera.

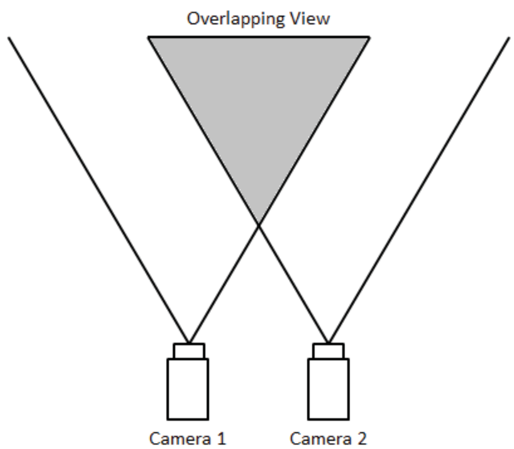

Figure 2. Stereo image capture

B. Pre Processing

Improving Speed. Preprocessing is the set algorithm applied on the images to enhance their quality and improve the computational efficiency. It is an important and common in any computer vision system. In the stereo vision system, a good selection of preprocessing stages could greatly improve the speed of the system. For our system, after the images are captured, the size of the image is down-scaled to a quarter of its original size to improve the computation speed. For the proposed system, the original size which is 1280x720 pixels is downscaled to 640x360 pixels. This reduction of size does not affect the accuracy of the system since the size of the object of interest in the original image is so much bigger than 2x2 pixels.

To improve the speed further we then converted the images from the RGB (Red, Green and Blue) color space to the grayscale color space. RGB color space consist of 3 channels which means that it requires 3 times more computation and memory space compared to grayscale which only consist of one channel. By applying these two preprocessings, we theoretically improve the speed of the system by 24 folds (2 images x 4 down-scaling factor x 3 color channels reduction).

Removing Noise. The downscaling applied earlier help to reduce the noise in the image since it average a group of pixels together which inherently produce a smoothing effect. We further remove the noise by applying a median filtering on the images. Median filter is selected since it is fast and able to produce sufficient smoothing. Figure 3 shows grayscaling and median filtering on a captured image.

C. Object Detection and Segmentation

Object Detection. Different most of the other stereo vision works [4][5], the stereo matching process that we utilizes is only applied to the object of interest which can be detected using an object detection algorithm. This approach is much faster compared to the stereo matching using features similarity in the two images. In our proposed system, the object of interest is basically a foreign moving object that enters the view of the stereo vision system. We assume that the cameras of the system are always fixed to one position. Hence, the detection of the object of interest is done using two operations, the pixel to pixel background subtraction and thresholding.

The background subtraction is done by taking the difference of every pixel, I_T in the image to its respective reference pixels in the background model, I_{BG}. The difference value is then thresholded with a value, T_R to determine if the pixel belongs to the object of interest or not. If the pixel does not belong to the object of interest, it will be used to update the background model.

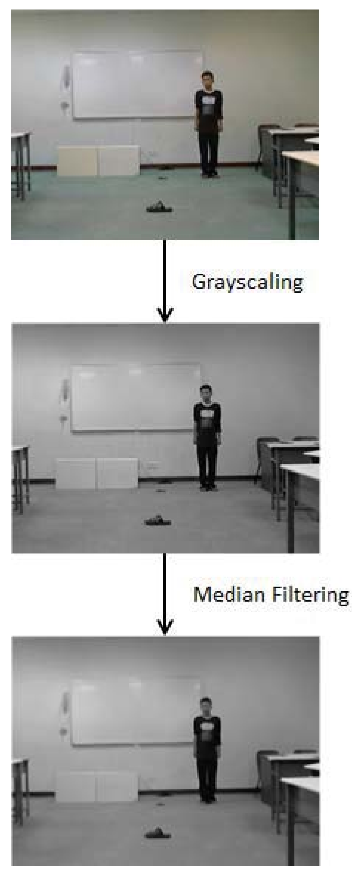

Figure 3. Grayscaling and median filtering

$$I_T = object\ if,$$
$$|I_T - I_{BG}| > T_R$$

Background model is initialized by assuming that the first frame in the video is the background. The background model is updated by averaging the new intensity with the background model intensity value. An on-line cumulative average is computed to be a new background as:

$$\mu_T = \alpha I + (1 - \alpha)\ \mu_{T-1}$$

where μT is the model intensity, I is the intensity and α is update weight. The value for α is set to be 0.5. Figure 4 shows the background subtraction and thresholding on a captured image.

Object Segmentation. The binary image resulting from the background subtraction and thresholding stage is then processed for object of interest segmentation. Firstly, a quick morphology is applied on the binary image to improve the subtraction result. A sequence of erode and dilate operation are involve in the morphology where the effect is to remove smaller detected regions usually due to noise and to enlarge the areas of object of interests and to close any holes within them.

We then applied a connected component analysis on the image to segment the object of interests on the image. Connected component analysis locates separated regions in the binary image and labels them as different objects. A one pass connected component analysis is applied to improve the speed of the system. From the connected component analysis results, blob extraction is done by drawing the bounding box around every object of interests. Figure 5 shows the morphology and blob extraction on a captured image.

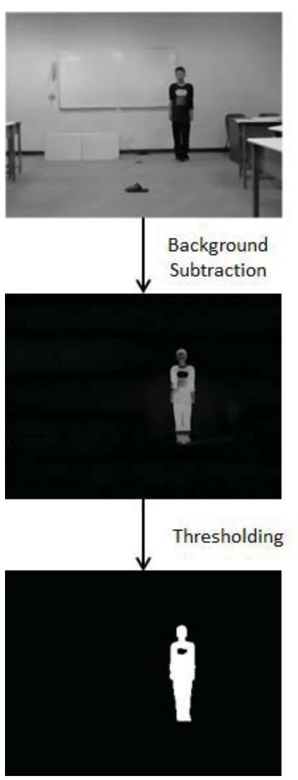

Figure 4. Background subtraction and thresholding

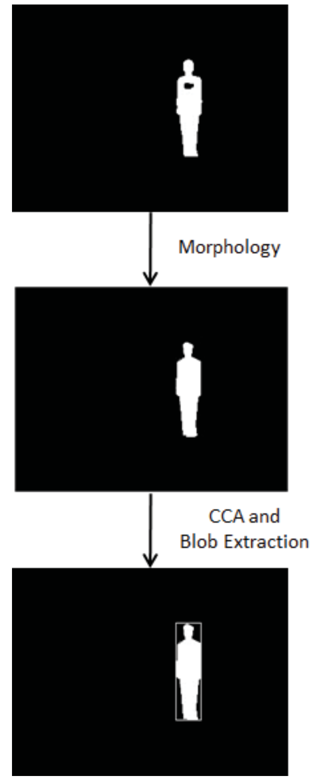

Figure 5. Morphology and connected component analysis (CCA) and blob tracking

D. *Object Distance and Size Measurement*

Up to this point, all the processing was done in parallel on two images captured from the two cameras. Distance measurement and size calculation use the information extracted from the object of interest information on both images.

Distance measurement. Distance measurement is done by using the using the disparity value of the object of interest in the two image, d in pixel. Since the camera is aligned in parallel, we can simply take the pixel difference between the two width center lines of the object of interest as the disparity as shown in Figure 6.

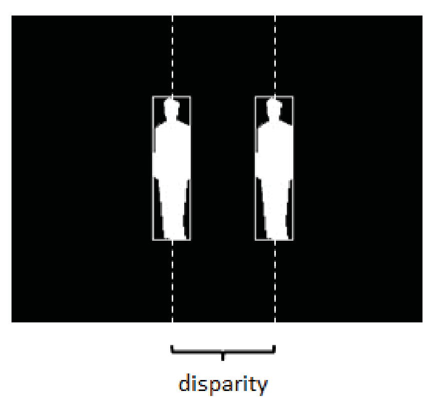

Figure 6. Object disparity

The distance, D of the object can be calculated using the following equation:

$$D = \beta\, d^{-1}$$

where β is a fixed parameter given by:

$$\beta = b\, f$$

where b is the separation distance between two cameras and f is focal length of the cameras.

Size Measurement. For size measurement, which consists of width and height, the calculations are done by using the disparity value in pixel, d and the pixel value of the width, w and the height, h of the blob. The calculations of actual object width, W and height, H are done using the relationship of between width per pixel, λ_w and height per pixel, λ_h to the disparity. From our experiment we found that width per pixel and the height per pixel value of an object are linear and inversely proportional to its disparity. The equation for the width, W and height, H calculation is as the following:

$$W = \lambda_w\, w$$

$$H = \lambda_h\, h$$

λ_w and λ_h are obtained from a linear equation of a graph of λ_w and λ_h against d as the following:

$$\lambda_w = m_w\, d + c_w$$

$$\lambda_h = m_h\, d + c_h$$

where m is the gradient of the plotted graph and c is the value of λ at $d = 0$.

The graph of λ_w and λ_h against d are plotted by experimenting on several samples with known actual width and height. For different disparity value (varied by changing the distance of the object), the value of the λ_w and λ_h is calculated and plotted. Figure 6 shows the graph of λ_w against d and its linear equation.

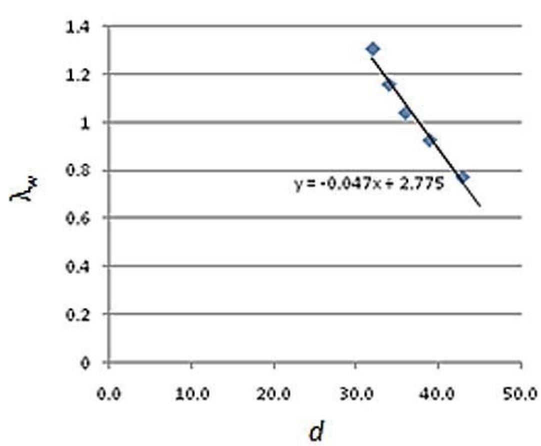

Figure 7. Graph of λ_w against d

E. *Object Identification*

Different objects can be distinguished from each other depending on their sizes. For example a car will have a different width compared to a pedestrian. For our system object is identified according to their height, h to width, w ratio and the height-width, $h\,w$ product (blob area).

A database is utilized to store the information and the associated label of known objects. If an object with matching information is found, the object will be identified. When a new object with no matching height to width ratio and height-width product in found, a new label will be instantiated for that object in the database.

III. RESULT AND DISCUSSION

The algorithm was developed using C++ programming language with OpenCV [6] library support and run on a single core Pentium 4 3.0 GHz processor. Tests were conducted to test the accuracy and the speed of the proposed object distance and measurement system. The results of some distance measurement tests are presented in Table 1. Results on object width and height measurement tests are presented in Table 2 and 3 respectively. Figure 7 shows some detection results on stereo images.

The results indicate that the measurement of distance is considerably accurate with a precision of ±25cm. Size measurement is also accurate with an error not more than ±3cm. Given that there is a single object in view, the average processing time for of distance and size measurement is about 67ms which means that the frame rate of the system can goes maximum up to 15 frames per second. 15 frames per second can be considered as acceptable for most autonomous system so that they can work in real time. The frame rate will drop if more objects are in the view of the cameras.

TABLE I. OBJECT DISTANCE MEASUREMENT RESULTS

Measured Distance (m)	Actual Distance (m)	Error (m)
3.101	3.000	-0.101
3.088		-0.088
2.901		0.099
3.112		-0.112
5.203	5.000	-0.203
4.880		0.120
5.204		-0.204
5.199		-0.199
7.016	7.000	-0.016
6.880		0.120
7.250		-0.250
6.810		0.190
9.917	10.000	0.083
9.780		0.220
10.211		-0.211
10.180		-0.180

TABLE II. OBJECT WIDTH MEASUREMENT RESULTS

Object	Measured Width (cm)	Actual Width (cm)	Error (cm)
Chair	52.065	51.000	-1.065
	52.970		-2.970
	49.830		1.170
	49.571		1.429
Beverage Box	9.612	9.000	-0.612
	10.510		-1.510
	7.786		1.214
	8.901		0.099
Table	67.860	68.000	0.140
	69.015		-1.015
	69.652		-1.652
	66.083		1.917
Pencil Box	31.250	30.000	-1.250
	32.208		-2.208
	28.786		1.214
	29.550		0.450

TABLE III. OBJECT HEIGHT MEASUREMENT RESULTS

Object	Measured Height (cm)	Actual Height (cm)	Error (cm)
Chair	80.784	80.600	-0.184
	78.645		1.955
	79.901		0.699
	82.020		-1.420
Beverage Box	20.196	19.000	-1.196
	19.780		-0.780
	17.661		1.339
	19.518		-0.518
Table	77.440	78.000	0.560
	77.480		0.520
	76.298		1.702
	79.519		-1.519
Pencil Box	5.852	5.000	-0.852
	5.198		-0.198
	3.986		1.014
	4.101		0.899

One disadvantage of the proposed method is that it depends strictly on constant environment lighting due to the use of background subtraction for object detection. One way to improve this is by using a more adaptive background subtraction method and background models. Another disadvantage is that the precision of the measurement is dependent on the resolution of the camera. Higher camera resolution would produce a more precise measurement. We chose to reduce the image resolution so that the process can be performed in real-time. Without the downscaling stage, the frame rate of the system would become too low which is about 0.15 frames per second. Our immediate future work, would involve studying the tradeoff between the resolution and the speed of the system.

IV. CONCLUSION

An object distance and size measurement using a stereo vision system is proposed. The method utilizes simpler algorithms so that a much faster processing speed can be achieved. The background subtraction is utilized to locate object of interest and the result is then used to calculate the disparity of the object. The distance and size measurements are considerably accurate and the average time it taken per cycle is 65ms.

REFERENCES

[1] Rahman K.A., Hossain M.S., Bhuiyan M.A.-A., Tao Z., Hasanuzzaman M., Ueno H., "Person to Camera Distance Measurement Based on Eye-Distance", 3rd International Conference on Multimedia and Ubiquitous Engineering, 2009 (MUE'09), pp.137-141, Qingdao, 4-6 June 2009

[2] Wahab M.N.A., Sivadev, N., Sundaraj, K., "Target distance estimation using monocular vision system for mobile robot", IEEE Conference on Open Systems (ICOS), 2011, pp.11-15, Langkawi, Malaysia, 25-28 Sept. 2011

[3] Kim H., Lin C. S., Song J., Chae H., "Distance Measurement Using a Single Camera with a Rotating Mirror", International Journal of Control, Automation, and Systems, vol.3, no.4, pp.542-551, 2005

[4] A-Lin H., Xue C., Ying G., Wen-Ju Y., Jie H, "Measurement of Safe Driving Distance Based on Stereo Vision", Proceedings of the 6th International Conference on Image and Graphics (ICIG), 2011, pp.902-907, Hefei, Anhui, 12-15 Aug. 2011

[5] Baek H. S., Choi J. M., Lee B. S., "Improvement of Distance Measurement Algorithm on Stereo Vision System(SVS)", Proceedings of the 5th International Conference on Ubiquitous Information Technologies and Applications (CUTE), 2010, pp.1-3, Sanya, 16-18 Dec. 2010

[6] Open Computer Vision Library (OpenCV), http://sourceforge.net/projects/opencvlibrary/, Last accessed 20th February 2012.

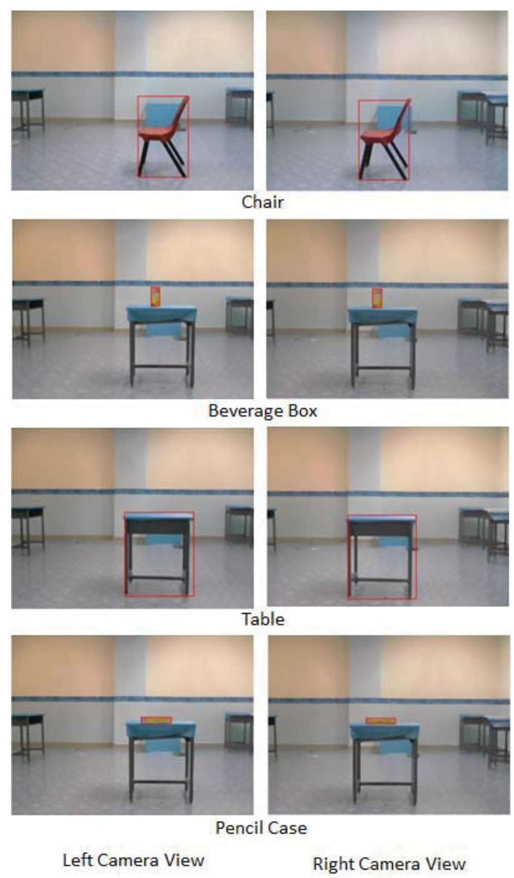

Figure 8. Object detection results on stereo images