Proceedings of the IASTED International Conference
Modelling and Simulation (MS 2013)
July 17 - 19, 2013 Banff, Canada

A REAL-TIME OBJECT DISTANCE MEASUREMENT USING A MONOCULAR CAMERA

Peyman Alizadeh [1], Meysar Zeinali [2]
[1]Student Member, IEEE,
School of Engineering, Laurentian University, Sudbury, ON, Canada
E-mail:Px_alizadeh@laurentian.ca
[2]Member, IEEE
School of Engineering, Laurentian University, Sudbury, ON, Canada
E-mail:Mzeinali@laurentian.ca

ABSTRACT
Distance measurement for moving object using image processing is one of the important research areas in robotics and computer vision. In the current paper, an improved method is proposed to compute the distance of the moving object using a monocular camera and based on the feature extraction method. First, the specific object is tracked by a camera and then the distance from the object to the camera is measured using a computationally efficient algorithm that is implemented in MATLAB/ SIMULINK environment. Experimental results show that this method is fast, efficient, and accurate and can be used for real-time robotic applications.

KEY WORDS
Desired Object Tracking; Distance Measurement; Image Processing; Single Camera.

1. Introduction

For decades, researchers have tried to develop different techniques of relating computer vision capabilities to applications and products, such as automotive safety, manufacturing, video surveillance, and visual servoing (one of the important related robotics applications). Visual servoing is a technique which controls the motion of the robot by using feedback information that is sent from a vision sensor. In visual servoing, it is difficult to track the moving object if the object distance is not accessible [1]. There are two different types of visual servo control: position-based visual servo control (PBVS) and image-based visual servo control (IBVS). The reference inputs for PBVS are the 3D relative position and orientation between the object and the robot end-effector in Cartesian space. In the image-based visual servo control, the reference input is the 2D object position which can be obtained from the image plane. The IBVS approach fails without having an accurate estimation of object distance and motion, especially when it is carried out in dynamic environments [2]. Because of error propagation, when a camera is employed in PBVS servo method, the small measurement error would highly affect the servoing accuracy [2].

To estimate the distance from the object, the tracking of an object must be addressed first. Object tracking can be classified into four main categories: model-based, appearance-based (region-based), feature-based, and contour-based methods. Model-based tracking methods need previous knowledge of the object shapes for the matching process and in order to find the object in the scene; Model-based tracking methods apply the exact geometrical models of the object. Model-based tracking methods have two shortcomings: 1) an object that is not in the database cannot be recognized by using these models. And 2) establishing this model is complicated and sometimes impossible [3]. Appearance-based methods track an object using the 2-D shape of the connected region. This tracking approach relies upon the information that is provided by the entire region pixels. Motion, color, and texture are examples of such information. These methods are not robust with complex deformation [4]. Feature-based methods track the specific features of each object. Many feature-based matching techniques have been developed although these techniques are not specialized for video object tracking. Advantages of these methods are their simplicity and stability for tracking algorithms. However, clustering of these features to the same group of an object is costly, specifically in dealing with a large number of features. Contour-based methods track the contour of the object rather than tracking the whole pixels that make up the object. In contour-based methods, the contour for the next frame is determined using the motion information about the object. Thus, the shape and the position of that contour are enhanced to fit into the object. Furthermore, the motion information is updated by the change in the contour location [3]. There are some advantages for using the contour-based representation. Merit of such methods is that they can cut the computational complexity. Another advantage of such techniques is their ability to track rigid and non-rigid objects. However, the contour-based methods fail to track the objects that are partly occluded [4].

To measure the distance of a moving object several methods presented in the literature. Zhang et al. [5] presented the 3D positions of a target in the camera coordinate frame by measuring the distance between feature point and principal point based on the calculated area in the image. The Zhang et al. [5] process is divided into three stages. The first stage is the camera calibration in order to calibrate the intrinsic parameters. The second stage is to set up a model for distance measurement along the optical axis direction according to the mapping relationship between the

DOI: 10.2316/P.2013.802-048

targets in the camera coordinates frame, and its projection in the pixel coordinate frame. The last stage is the absolute distance measurement. Yang and Cao [6] proposed a 6-D object localization method based on a monocular vision system which is obtained through decomposing the homography matrix and finally refining the result using the Levenberg–Marquardt algorithm. This approach showed an absolute accuracy of 0.3 mm and the repeatability accuracy of 0.1 mm for their robot's arm pose estimation. Coman and Balan [7] mentioned that their application of object distance measurement can be a starting point of complex applications as long as the object geometry remains a square. Coman and Balan [7] also indicated that the surface area and a line in the surface area for different objects are correlated to each other. Yamaguti et al. [8] found the distance to the texture surface of the object from the ratio of two images that are taken by a monocular camera at different locations, and then they used complex log mapping of these two images. In their method, the processing time is long and that the object is restricted to a plane which is vertical on the optical axis. Zhang et al. [9] proposed a method of object distance measurement based on a particle filter. This method is reported to be highly reliable and have strong real-time performance. However, the accuracy of the technique proposed by Zhang et al. [9] is not mentioned. Dixon et al. [10] presented a new continuous observer for determining the range information of an object moving with affine motion dynamics and known motion parameters using one camera. The impact of this experiment was that a continuous observer can be employed to make a monocular vision system in order to determine the range parameter for a moving object which moved with an affine or Riccati motion dynamics even in presence of sensor noise.

In the present research, the distance to the object is calculated using a feature extraction method. The proposed algorithm is robust and reliable with respect to the orientation and position of the object and could be used for any moving objects of different size, and in any direction on the ground.

The remainder of this paper is organized as follows: Section 2 presents the problem formulation, Section 3 describes the Image-processing algorithm, experimental setup is explained in Section 4, the experimental results and detailed discussion are also presented in Section 5, and Section 6 concludes the article.

2. Problem Definition

The main image formation technique is mostly based on the pinhole lens model. In this model, the mapping relationship between the 3-D world coordinates and the 2-D image coordinates plays an important role. Let's consider *M* in Figure 1 to be a point in the world coordinate relative to the camera frame. In this case, *m* would be the projection of *M* onto the image plane. Under the pinhole assumption, these two points and the origin of the camera frame (the center of the projection) that is in the back of the image plane at a distance μ will be collinear [11]. This model of image formation is applied in the present research. Now, the problem of mapping the object's location is described using measured distance and frame transformation.

For visual servoing purposes, there are five "standard" frame names associated with a robot and its workspace: the base frame, the station frame, the wrist frame, the tool frame, and the goal frame. The base frame is shown as **{B},** and it is the fixed part of the robot which is sometimes called the Link 0. The station frame **{S}** is called the universe

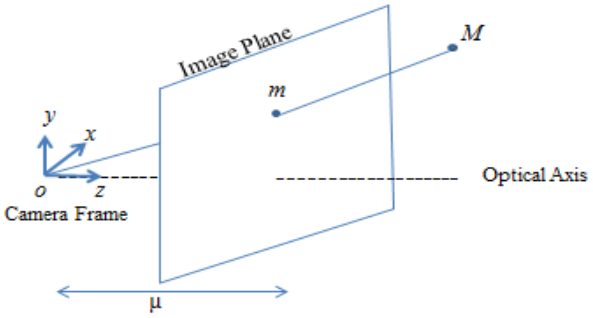

Figure 1. Camera coordinates frame

frames since all actions of the robot are made corresponding to this frame. The wrist frame **{W}** is the last link of the manipulator, and it is normally defined relative to the base frame. The tool frame **{T}** is assigned at the end of any tool that a robot is holding. Finally, the goal frame **{G}** is the location to which the tools need to move [12]. Figure 2 shows entire frame assignments to find the distance of the moving object with respect to a robot hand. In Figure 2, the object distance $\overline{AG}$ can be measured through the proposed algorithm in the present paper. The camera's location w.r.t. base frame is shown with vector $\overrightarrow{BA}$ (known by measurement), and $\overrightarrow{BT}$ is obtained through forward kinematics and using DH parameter of the robot. Using this information, it will be easy to find vector $\overrightarrow{TG}$, which is the distance of the moving object with respect to the tool frame of the robot, and can be calculated by equation (1).

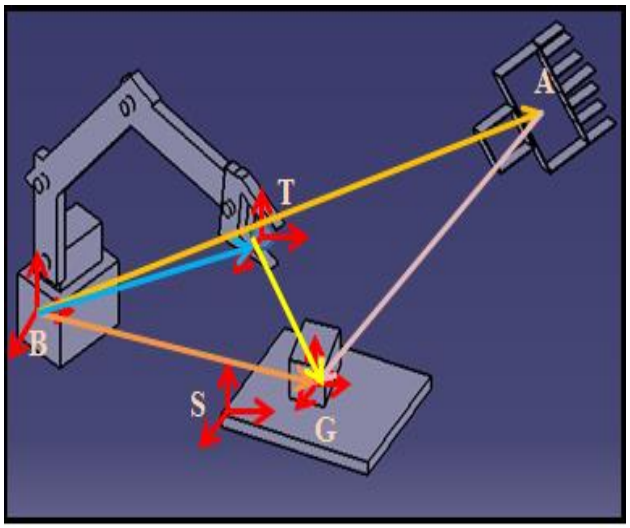

Figure 2. Standard frame assignment

$$\overrightarrow{BA} + \overrightarrow{AG} = \overrightarrow{BG}$$
$$\overrightarrow{BT} + \overrightarrow{TG} = \overrightarrow{BG} \quad (1)$$
$$\overrightarrow{TG} = \overrightarrow{BG} - \overrightarrow{BT}$$

There are three ways of using image based distance computation techniques: 1) Stereo vision based technique; 2) Mono vision based technique, and 3) Time-of-Flight camera technique. The stereo vision based method use two cameras to find the depth and the disparity map using a complex method. This technique is highly accurate, but it is time consuming because of many images of the same object needed to be processed at the same time. In addition, implementing such a technique is expensive as it requires two cameras. Moreover, the stereo vision accuracy fails with the increase of the distance to the object compared to the baseline distance between the two different views. On the other hand, mono vision based approach are comparatively less expensive as it requires only one camera[13].The Time-of-Flight depth estimation technique is used to find the depth information by measuring the total time required in the light to transmit and reflect from the object. It is hard to separate the incoming signal since the signal depends on many parameters such as intensity of the reflected light, intensity of the background light, and the dynamic range of the sensor which makes detecting the incoming light a hard task [3]. Besides, the accurate range of distance measurements for the time-of-flight camera is usually between 1– 4 m [14].

Therefore, researchers are looking for inexpensive, uncomplicated, and accurate techniques. Applying such superior techniques requires researchers to tackle with several other challenges such as object detection, obstacle avoidance, and location finding. There are two approaches to estimate the location of any object: contact and non-contact methods. Nowadays, the non-contact distance measurement system becomes useful in a wide variety of applications, though, most of the time it is impossible to have physical contact with the object to make the measurement [15]. High accuracy and time saving are advantages of using a non-contact measurement technique. Clarke and Williams [16] discuss the benefits of using a non-contact measurement system. These benefits include lower inspection costs, better quality control, faster production, smaller tolerances, and fewer defects. The object distance is defined as the distance of the desired object from the center of the lens along the optical axis. In addition, the image distance is defined as the distance from the focused image to the center of the lens.

The proposed object distance measurement is based on finding the closest point from the object to the bottom-center of the camera's field of view. Equation (2) is used to calculate the distance from the object to the camera where, *a*, is known value and resulted by measurement, *d* is the object distance and *h* is the height of the camera from the ground. Figure 3 shows, the coordinate system and the camera's field of view on the ground.

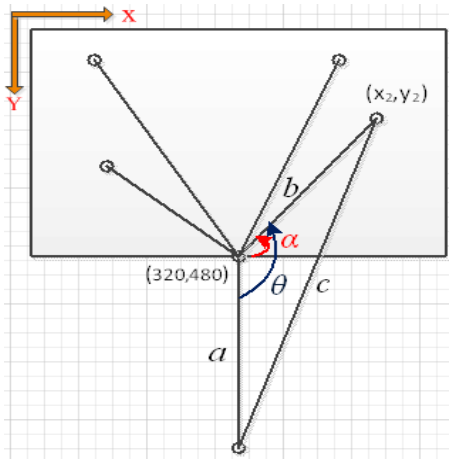

Figure 3. Object distance calculation

$$\alpha = \arctan\left(\frac{480 - y_2}{x_2 - 320}\right)$$
$$\theta = \frac{\pi}{2} + \alpha$$
$$b = \sqrt{(320 - x_2)^2 + (480 - y_2)^2} \quad (2)$$
$$c = \sqrt{a^2 + b^2 - 2 \times a \times b \times \cos(\theta)}$$
$$d = \sqrt{c^2 + h^2}$$

3. The Image-Processing Algorithm

In this section, some parts of the image-processing algorithm are explained. The color space conversion is used to change the color information to different color spaces. Figure 4(a) demonstrates the R. G. B. - three primary color of red, green and blue- image. Auto threshold technique is applied in the system in order to identify different objects in the camera's field of view. The Otsu method [17] is one of the widely referenced threshold techniques. As a result, this method would be satisfactory especially when dealing with a large number of pixels within each class with close pixel's value [17]. The image of a sample threshold by the auto-threshold technique is given in Figure 4(b).

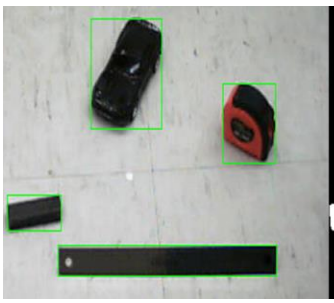

Figure 4(a) R. G. B. image.

Figure 4(b). Image after the threshold

3.1 2-D FIR Filter Filtering is another important technique that has to be accurately chosen to remove the noise from the imag e without reducing its sharpness.

3.2 Image Complement

Image complement is used to compute the complement of a binary or intensity image. The image complements switches the values of the zero and one valued pixels.

3.3 Averaging Subsystem

The averaging subsystem method is used to stabilize the system by applying four consecutive measurement values for the system. This technique would mediate the fluctuation arising from the noise into the system [18]. The completed algorithm to the system is shown in Figure 5.

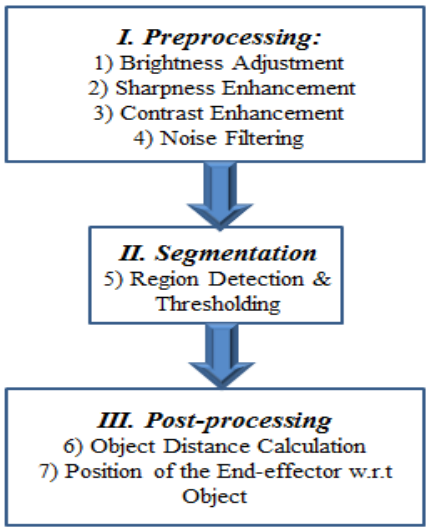

Figure 5. The Algorithm for object distance measurement

4. Experimental Setup

In this research, Logitech *Quick Cam® Communicate STX™* is used to get the R. G. B. image with resolution of 640 × 480 pixels. The captured image is then converted to intensity by color space conversion since most of the applications require an intensity of the image. To reduce the noise from the image of the desired object, a 2-D FIR Filter is applied. MATLAB Functions are used to carry out the required code in the subsystem for the desired object tracking algorithm. In the next step which is implemented in the subsystem, desired object distance is measured using the resulting image data. From implementing the object distance measurements that are shown as d_1 and d_2 for moving object at time t_0 and t in Figure 6, the horizontal distance traveled by the object on the ground could be calculated using the law of cosines suggested in Equation (3). The initial values are d_1=1160 mm and $\theta_0 = 55°$.

$$d_2^2 - d_1^2 - d^2 + 2\,d_1 d\cos(\theta) = 0$$

$$d = \frac{2\,d_1\cos(\theta) \pm \sqrt{4\,d_1^2\cos^2(\theta) - 4(1)(d_1^2 - d_2^2)}}{2} \qquad (3)$$

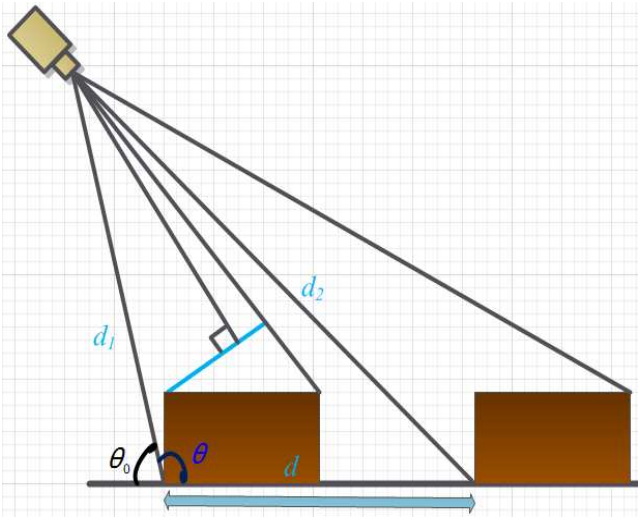

Figure 6. Distance moved on the ground by the object

5. Experimental Result and Discussion

In this experiment, the distance of the desired object is extracted and used for the second subsystem to find the position and orientation of the end-effector with respect to the object. The tracked object is shown with green bounding box in Figure 4(a). Tables 1 and 2 compare the results from the experiment. These tables respectively show the object distance average errors for two different samples of one rectangular block of size 25.43 mm×79.22 mm (W×L) and one irregular toy car of size 70.21 mm ×145.54 mm (W×L) which are measured at any location in the camera's field of view. In addition, Figures 7 and 8 show the tabulated results in Tables 1 and 2 in the graphical form. Some related works for object distance measurement are shown in Table 3. It should be mentioned that this experiment is implemented with the low resolution camera to check the proposed algorithm. The accuracy of the measurement could improve by using a high-resolution camera.

Image noise is the main unavoidable reason for producing errors during the image acquisition stage. Such error can occur in finding the exact point of contact with the object on the ground. Another potential cause of error would be the variation in image illumination across the camera's field of view. Using the method presented by [19], it is possible to measure an object distance when the optical axis of the camera is parallel to the ground. However, in case of dealing with finding object distance anywhere in the field of view of camera, Joglekar et al [20] method is found to be the possible solution.

Joglekar et al [20] define the in-path object distance and the oblique object distance as the distance to the object that is on the optical axis and as the distance from the object that is not on the optical axis, respectively.

Table 1
Average distance error measurements for the rectangular block

Trial	*Real distance (mm)*	*Proposed measured distance (mm)*	*Joglekar et al [20] measured distance (mm)*	*Proposed distance error (mm)*	*Joglekar et al[20] distance error (mm)*
1	1181	1176	1186	5	5
2	1213	1231	1131	18	82
3	1215	1190	1150	25	65
4	1272	1302	1622	30	350
5	1320	1340	1134	20	186
6	1323	1320	1399	3	76
7	1367	1414	2275	47	908
8	1432	1438	1424	6	8
9	1436	1468	1209	32	227
10	1481	1529	1433	48	48
11	1513	1521	1306	8	207
12	1548	1580	1183	32	365
13	1585	1617	1227	32	358
14	1621	1633	1156	12	465
15	1630	1615	1200	15	430
Average absolute error				23.43	269.64

Table 2
Average distance error measurements for the toy car

Trial	*Real distance (mm)*	*Proposed measured distance (mm)*	*Joglekar et al [20] measured distance (mm)*	*Proposed distance error (mm)*	*Joglekar et al[20] distance error (mm)*
1	1179	1173	1184	6	5
2	1206	1183	1160	23	46
3	1250	1262	1155	12	95
4	1253	1259	1204	6	49
5	1263	1238	1196	25	67
6	1295	1328	1953	33	658
7	1363	1352	1341	11	22
8	1373	1404	1196	31	177
9	1420	1458	2057	38	637
10	1480	1489	1344	9	136
11	1495	1542	1370	47	125
12	1503	1529	1202	26	301
13	1527	1570	1310	43	217
14	1580	1583	1236	3	344
15	1583	1613	1164	30	419
Average absolute error				16.50	381.50

We have investigated whether this method can be used to estimate the object distance anywhere on the camera's field of view with the provided setup. The results from some experiments using the Joglekar et al [20] method shows that as the object gets closer to the vanishing point, the object distance would surprisingly rise, and the error would significantly increase. Therefore, when the camera has an angle with optical axis, [19-20]'s methods are no longer valid as the object distance measurement would be a non-linear function of distance. Moreover, due to the fact that the oblique object distance is calculated from the result of the in-path object distance divided by the cosine angle between the optical axis and the line joining the camera point to the object's point of contact. Hence, the error is relatively similar for the in-path object distance as obtained. The proposed vision-based object distance measurement technique is principally different from above-mentioned optical techniques.

Table 3
Comparison of different methods for object distance measurement

Method	*Characteristics (in terms of distance)*	*Nature of the method*
Gat et al[19] (Optical axis is parallel to the ground)	Along the optical axis	Linear
Joglekar et al[20] (Optical axis is parallel to the ground)	In-path and oblique	Non-Linear
Proposed Method (Optical axis is not parallel to the ground)	Any-where on the field of view of camera	Non-Linear

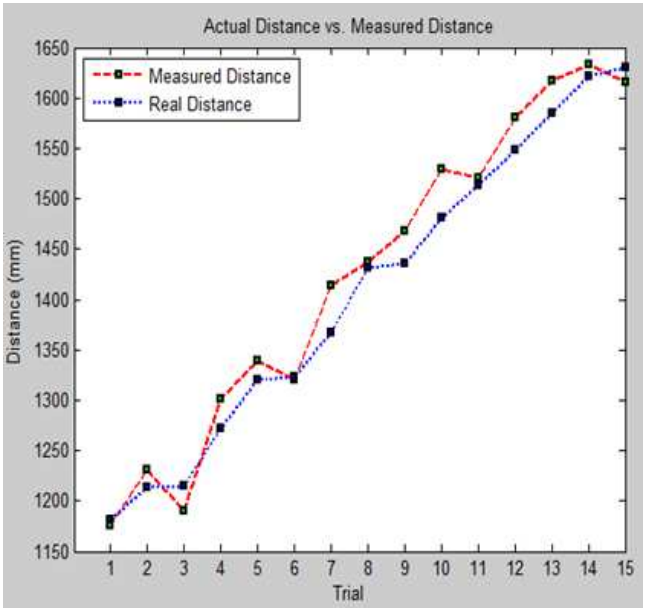

Figure 7. Object distance measurement for the rectangular block

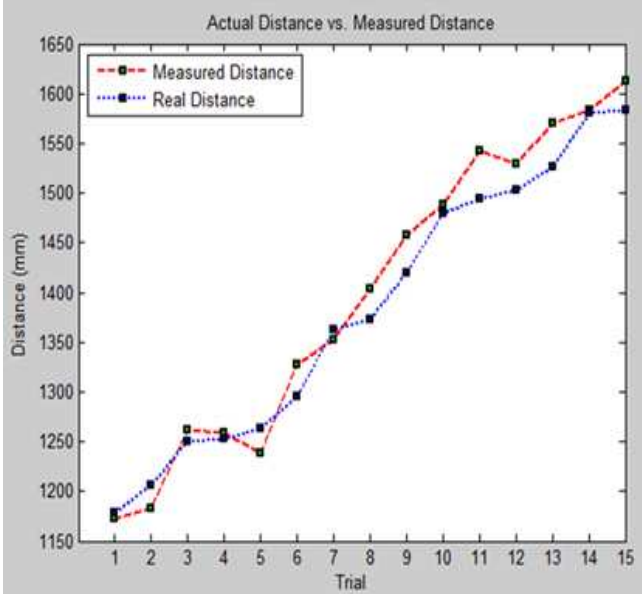

Figure 8. Object distance measurement for the toy car

6. Conclusion

In this paper, an improved method was proposed to calculate the object distance using a single camera even if the surface of an object is not parallel to the camera and is not restricted to be vertically intersecting the optical axis. The proposed method is able to identify the desired object and extract the object feature for a moving and a static object. The experimental results shows that the object distance average error for the rectangular block, and the irregular toy car are 23.43 mm and 16.50 mm, respectively. Moreover, the discrete sample time to run the proposed algorithm is estimated to be 0.033. This inexpensive distance measurement method could be applied to many applications such as pick and place task in robotics.

References

[1] A. Goto, & H. Fujimoto, Proposal of 6 DOF Visual Servoing for Moving Object Based on Real-Time Distance identification. *SICE.* Annual Conference, Japan, 2008, 3208-3213.

[2] H. Firouzi, & H. Najjaran, Real-time monocular vision-based object tracking with object distance and motion estimation. *Advanced Intelligent Mechatronics (AIM), IEEE/ASME International Conference on*, 2010, 987-992.

[3] K. M. Shaaban, & N. M. Omar, 3D information extraction using Region-based Deformable Net for monocular robot navigation. *J. Visual Communication and Image Representation, vol. 23*, 2012, 397-408.

[4] A. Cavallaro, O. Steiger, & T. Ebrahimi, Tracking video objects in cluttered background. *Circuits and Systems for Video Technology, IEEE Transactions on, 15(4)*, 2005, 575– 584.

[5] Z. Zhang, Y. Han, Y. Zhou, & M. Dai, A novel absolute localization estimation of a target with monocular vision, *Optik - Int. J. Light Electron Opt*, 2012, 1218-1223.

[6] Y. Yang, & Q. Z. Cao, Monocular vision based 6D object localization for service robots intelligent grasping. *Computers Mathematics with Applications, Volume 64, Issue 5,* 2012, 1235-1241.

[7] M. Coman, & R. Balan, Video Camera Measuring Application Using matlab. *Solid State Phenomena Vols.166-167*, 2010, 139-144.

[8] N. Yamaguchi, S. Oe & K. Terada, A Method of Distance Measurement by Using Monocular Camera. *Proceedings of the 36th SICE Annual Conference International Session*, Tokushima, 1997, 1255-1260.

[9] L. Zhang, Y. Bo, J. Zhang, & W. Zou, Study on Distance Measure Based on Monocular Vision. *Energy Procedia 13,* 2011, 9533-9542.

[10] W. E. Dixon, Y. Fang, D. M. Dawson, & T. J. Flynn, Range identification for perspective vision systems. *IEEE Trans. Automat. Contr., 48(12)*, 2003, 2232–2238.

[11] M. W. Spong, S. Hutchinson, & M. Vidyasagar, *Robot Modeling and Control*, (John Wiley and Sons, Inc., 2006).

[12] J. J. Craig, *Introduction to Robotics Mechanics & Control*, 3rd edition, NJ: Prentice-Hall, 2004.

[13] A. Rahman, A. Salam, M. Islam, and P. Sarker, An Image Based Approach to Compute Object Distance. *International Journal of Computational Intelligence Systems, Vol. 1, No. 4,* 2008, 304- 312.

[14] S. Y. Kim, E. K. Lee, & Y. S. Ho, Generation of ROI enhanced depth maps using stereoscopic cameras and a depth camera. IEEE Transactions on Broadcasting 54 (4), 2008, 732–740.

[15] M. C. Lu, and C. C. Hsu, & Y. Y. Lu, Image-Based System for Measuring Objects on an Oblique Plane and Its Applications in 2-D Localization *IEEE Sensors Journal, vol. 12, NO. 6*, 2012, 2249-2261.

[16] T. A. Clarke, & M. R Williams, Buyers guide to six non-contact distances measuring techniques. *Quality Today, Buyers Guide*, 1998, 145-149.

[17] M. Sezgin & B. Sankur, Survey over Image thresholding techniques and Quantitative performance evaluation. *Journal of Electronic Imaging, vol.13*, 2004, 146 –165.

[18] M. Coman, S.D. Stan, M. Manic, & R. Balan, Application of Distance Measuring with Matlab/Simulink. IEEE-HSI 2010, *the 3rd International Conference on Human System interaction*, Rzeszow, Poland, 2010, 113 – 118.

[19] I. Gat, M. Benady, & A. Shashua, A Monocular Vision Advance Warning System for the Automotive Aftermarket, *SAE Technical Paper*, 2005, 8 pages.

[20] A. Joglekar, D. Joshi, R. Khemani, S. Nair, & S. Sahare, Depth Estimation Using Monocular Camera. *International Journal of Computer Science and Information Technologies, Vol. 2 (4),* 2011, 1758-1763.