Determining Subject Distance Based on Face Size

Mohamed Tahir Ahmed Shoani
Faculty of Electrical Engineering
Universiti Teknologi Malaysia
Johor, Malaysia
tamohamed2@live.utm.my

Shamsudin H.M. Amin
Centre for Artificial Intelligence and Robotics
Universiti Teknologi Malaysia
Johor, Malaysia
sham@fke.utm.my

Ibrahim M. H. Sanhoury
Centre of Artificial Intelligence and Robotics
Universiti Teknologi Malaysia
Johor, Malaysia
Sanhoury124@yahoo.com

***Abstract*—This paper presents a novel method that can estimate the distance of a person up to 6 meters away. The proposed approach is based on face images extracted from a video of a monocular camera through the use of the Viola-Jones approach of OpenCV to detect the faces. The results of the experiments show the practicality of this approach in calculating the distances of the subjects in the camera's view.**

Keywords— Subject Distance, Face Detection, Face Size, OpenCV

I. INTRODUCTION

In many robot applications, it is important to discover the distance of the person interacting with the robot [1][2][3]. Many approaches were presented using different kinds of sensors.

There are generally two approaches for measuring the distance between an object and a robot; reflective and non-reflective [4]. In reflective (active) measuring methods, a sensor is used to send a signal and a receiver is used to detect the reflection of the signal on the surface, while in non-reflective (passive) measuring methods a camera or other sensor is used to measure a signal naturally occurring such as sun-light [5]. In ultrasonic based system an ultrasonic signal is used [6][7]. A laser is used in laser based systems [5][8], and an infrared transmitter and receiver are used in systems utilizing the infrared signal [9]. In reflective methods for distance measurement, surface uniformity and material may affect the accuracy of the measurement [4], hence the non-reflective (passive) approach has the advantage of using less energy to conduct its operation as well as being invariant to the surface uniformity and its material.

In non-reflective approaches to measuring distance, a camera is used to capture an image of the object and the distance is calculated based on a certain criteria. In [10], a stereo vision camera is used to calculate the distance of all types of obstacles at different distances, and in [11] stereo fisheye cameras are utilized to calculate the distance with high accuracy for car driving assistance. Although these approaches provide for good accuracy, they requires the use of two cameras rather than one as well as correct calibration for successful operation. In [12], a monocular camera is employed to calculate the relative distance between robots using the perspective projection image ratio of the robots' labels. While this approach requires low computational cost and can be applied on low performance embedded systems, it requires a certain label to be used on the devices (robots) in order to calculate the distance appropriately. In [13] the authors tilt the sensor of a monocular camera to achieve distance measurement based on focus information, however using this approach will severely limit the cameras capability of taking pictures with usable clarity of all sections of the image. In [14] a monocular camera is also used to calculate the distance of a face based on the distance between the two eyes of the person, however the distance calculation is limited to less than one meter (31 inches).

The methods used in non-reflective (passive) distance measurement vary in their complexity, accuracy and the distance that can be measured. In this work, a monocular camera is used to measure the distance of a person by calculating the size of the person's face. The results confirm a good accuracy for distances of up to 6 meters from the camera without the use of any zooming feature or extra optics or markings to conduct the measurements. The presented approach can be used with robots to interact with people at practical distances without using any extra sensors such as infrared, laser or ultrasonic.

The rest of this paper is organized as follows: Section 2 outlines the methodology of the approach. Section 3 provides an overview of the system used for conducting the experiment and performing the calculations. In section 4 the experiments are outlined and section 5 presents a discussion of the results. Finally, the paper concludes with a general overview of the work carried out.

II. THE METHODOLOGY

The method for calculating the subject distance from the camera is based on the subject's face size. Previous methods have used the Euclidian distance between the eyes of the subject [11], however they were able to measure distance of up to 31 inches (0.787 m), with an accuracy of 88.96% at that distance. In this approach, which is based on face size to calculate the subject distance, the authors managed to successfully measure distances of up to 6 meters with an overall average accuracy of 96.23 %.

The approach consisted of two steps, one to detect and record the face sizes and the second to actually use the data to calculate a face size in real time. For the first step, several experiments were performed, in which different subjects from

978-1-4799-7862-5/15/$31.00 © 2015 IEEE

diverse races and different genders would stand at pre-determined distances from the camera. A program written with C++ and utilizing the face detection method in OpenCV which is based on the work of Viola-Jones [15] would read the images from the camera and then the detected face and its size would be saved to the hard disk. Afterwards an equation would be formulated based on the gathered data. Using Microsoft Excel's Scatter graph and Trend-line functions; selecting the power function calculations yields equations (1 – 3) which were used for calculating the subject distance based on face image sizes for resolutions of (1) 320x240, (2) 640x480 and (3) 1280x720.

$$y = 53.648\, x^{-0.988} \tag{1}$$

$$y = 127.14\, x^{-1.001} \tag{2}$$

$$y = 253.62\, x^{-1.007} \tag{3}$$

In the second step, a program, also written in C++ and uses OpenCV was used to read the camera video stream and search for face images in the video frames. When a face image is detected, one of the equations, which were calculated in the first step, would be used to find the distance of the subject (face) in the video. By replacing 'y' with the face size in pixels, x would yield the distance.

The applied equation is chosen with respect to the video resolution used. The video resolution would be automatically detected using the OpenCV picture property ".cols". After calculating the distance of each detected face in the frame image, a box is drawn around the face of the subject and the distance is written under the box to indicate the distance of the subject (person).

III. The System

A. 3.1 System Description

The distance calculating system consists of two parts, software and hardware. Following is an outline of each:

B. 3.1.1 The Software

In implementing the approach, the freely available OpenCV [13, 14] library was used, which contains functions for detecting faces as well as other functions that deal with digital images and video. The OpenCV functions can be used with C, C++, Java and Python, and provide for a convenient way of implementing computer vision applications.

The program captures a frame from the video camera or video file and then, through the use of the frontal face Haar cascade classifier and the OpenCV function detectMultiScale, the faces of subjects are detected, and then analyzed for distance calculation. C++ of Microsoft Visual Studio 2010 [15] was used as the programming language to implement the software part of the system.

C. 3.1.2 The Hardware

Three cameras with different resolutions were used to capture the same video in resolutions ranging from low to high. The cameras used were two mobile phone cameras and the laptop's built in camera. One of the two mobile phone cameras captured video at a low resolution of 320x240 at 15 frames per second (fps) while the other one captured video at a high density resolution of 1280x720 at 30 fps. The laptop camera was capturing frames from the camera video stream at a resolution of 640x480. The laptop used was based on an Intel i5 CPU, with 4 GB of memory and running a 64 bit version of Windows 7 home premium.

IV. The Experiments & Formula

The experiments were carried out in a lab environment with overhead florescent lighting. In order to cover a broader spectrum of faces, the chosen subjects were of different nationalities and of both genders.

A. The Measurement Experiments

On the floor of the lab a distance of six meters was measured and twelve marks were laid out 0.5 meters apart. The laptop and the two mobile phones (cameras) were placed on a table at one end, and a subject performing the experiment would stand at the floor marks, holding cue cards to indicate the standing distance. The three cameras would be used simultaneously to recording the different distances of the subject. Fig.1 (a) shows the floor markings and (b) shows one of the subjects standing at a distance of 6.0 meters from the camera and holding a cue card to indicate the distance.

(a)

(b)

Fig.1, (a) Markings on the lab floor. (b) A subject standing at a distance of 6 meters and holding a cue card

During the experiment, the person performing the trial would have to stand at the marking for an adequate period of time to allow the three cameras to take a sufficient number of images to be later used for analysis and formulation of the equation. Depending on the facial features, the software would capture 3 - 20 images of the subject's face at each position, and as can be seen from Fig.1, the lighting conditions were sufficient for video recording and image capturing.

The Results

Tables (I – III) show the results gathered from the experiments, and Fig. 2 shows them graphically.

TABLE I. FACE IMAGE SIZES FOR DISTANCES OF 320X240

Dst	S1	S2	S3	S4	S5	S6	S7	S8	Avg
0.5	128.5	129.0	139.4	114.3	106.8	144.0	130.6	152.1	130.6
1.0	63.8	65.6	65.1	56.1	53.0	68.0	70.5	72.5	64.3
1.5	44.8	45.3	41.0	44.9	38.0	44.0	46.0	44.8	43.6
2.0	27.6	27.1	29.5	29.4	25.9	31.8	31.3	32.5	29.4
2.5	22.8	23.4	24.3	22.5	22.5	24.4	24.0	25.3	23.6
3.0	20.3	21.0	22.0	-	-	22.1	21.8	21.3	21.4

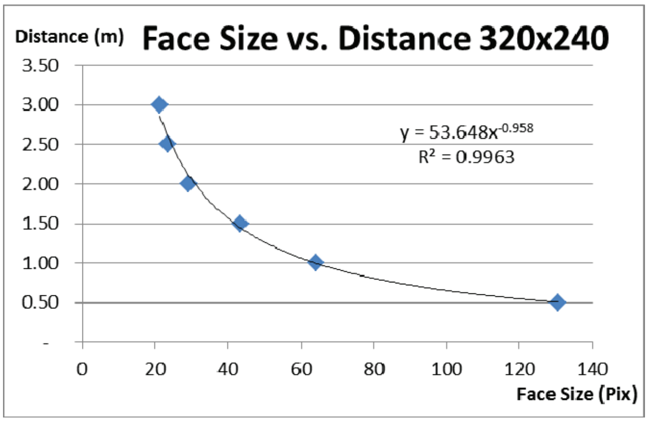
TABLE II. FACE IMAGE SIZES FOR DISTANCES OF 640X480

Dst	S1	S2	S3	S4	S5	S6	S7	S8	Avg
0.5	234.8	255.1	240.1	189.1	205.3	256.1	222.5	266.3	233.7
1.0	125.9	123.0	120.6	117.6	105.8	128.4	124.1	125.0	121.3
1.5	79.5	77.3	77.4	77.8	72.5	86.0	86.8	84.9	80.2
2.0	59.5	58.8	57.4	56.0	53.6	63.0	63.9	63.0	59.4
2.5	48.1	49.2	47.4	46.6	45.8	50.8	51.6	50.4	48.7
3.0	43.4	42.8	41.5	40.2	40.0	44.4	44.5	44.6	42.7
3.5	37.0	38.0	35.4	33.8	34.0	38.6	40.3	39.2	37.0
4.0	31.5	31.0	29.8	26.0	28.0	35.3	35.0	31.4	31.0
4.5	26.4	27.7	26.2	24.5	24.5	29.8	28.9	27.8	27.0
5.0	25.1	25.0	22.6	22.0	22.5	25.3	25.4	25.8	24.2
5.5	22.7	21.7	-	-	-	24.3	23.6	23.0	23.0
6.0	22.0	-	-	-	-	21.6	22.9	22.0	22.1

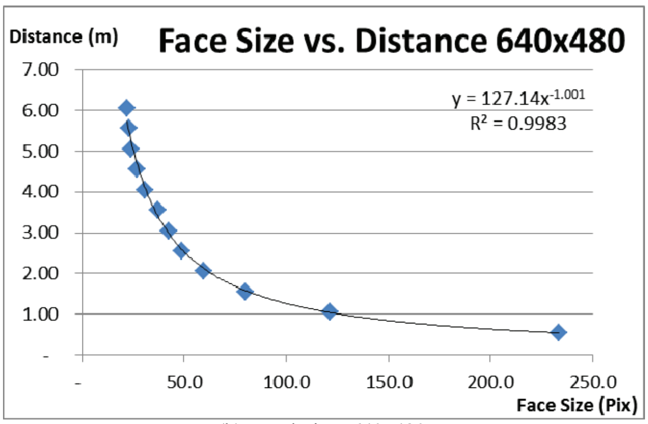
As can be seen from the results, the software was unable to detect the face image for subjects standing further than three meters for the low resolution of 320x240; the main reason being the vagueness of the facial features at this distance. To clarify this issue, Fig.3 shows the facial images of three subjects at a distance of 3 meters in the three different resolutions of the three cameras. As can be seen from Fig.3, the facial features at a distance of 3 meters in a resolution of 320x240 are unclear, standing further away from the camera results in the inability of detecting the subject's face. Although not all the subjects were detected, the resolution of 640x480 results in detecting most of the faces at 6 meters, as can be seen from Table 2. Using the higher resolution of 1280x720 resulted in detecting almost all the faces at the distance of 6 meters. Using a higher resolution permits face detection at further distances, but reduces the overall speed of the system as the frame size increases by four times whenever the resolution is doubled. Table III shows the results from the HD camera which was recording at a resolution of 1280x720.

TABLE III. FACE IMAGE SIZES FOR DISTANCES OF 1280X720

Dst	S1	S2	S3	S4	S5	S6	S7	S8	Avg
0.5	465.6	503.8	465.6	429.4	445.3	521.1	522.1	566.5	489.9
1.0	276.9	249.6	227.6	218.4	224.3	255.6	275.0	264.1	248.9
1.5	155.1	158.1	159.3	161.6	160.8	175.1	170.3	163.4	163.0
2.0	132.8	122.1	114.9	104.5	111.9	133.5	131.9	130.4	122.7
2.5	91.4	92.9	93.4	92.0	86.1	101.6	107.4	105.0	96.2
3.0	76.8	78.3	77.9	76.6	68.0	85.3	86.6	80.0	78.7
3.5	70.0	64.8	66.8	62.1	59.6	71.1	67.4	68.4	66.3
4.0	58.1	58.5	58.3	59.0	53.3	64.1	62.1	60.5	59.2
4.5	53.4	54.0	53.0	54.3	49.3	57.0	54.9	54.8	53.8
5.0	49.3	49.5	46.9	49.5	47.3	51.6	53.8	52.5	50.0
5.5	47.3	46.9	44.5	46.5	44.5	46.8	49.0	47.9	46.7
6.0	43.1	45.6	42.0	-	41.9	44.9	47.5	45.6	44.4

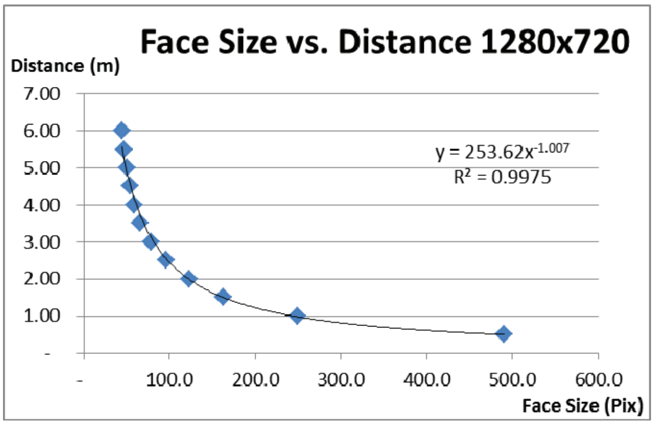


(a) Resolution: 320x240



(b) Resolution: 640x480

Fig.2, a graph showing the face size versus distance



(c) Resolution: 1280x720

Fig.2, a graph showing the face size versus distance

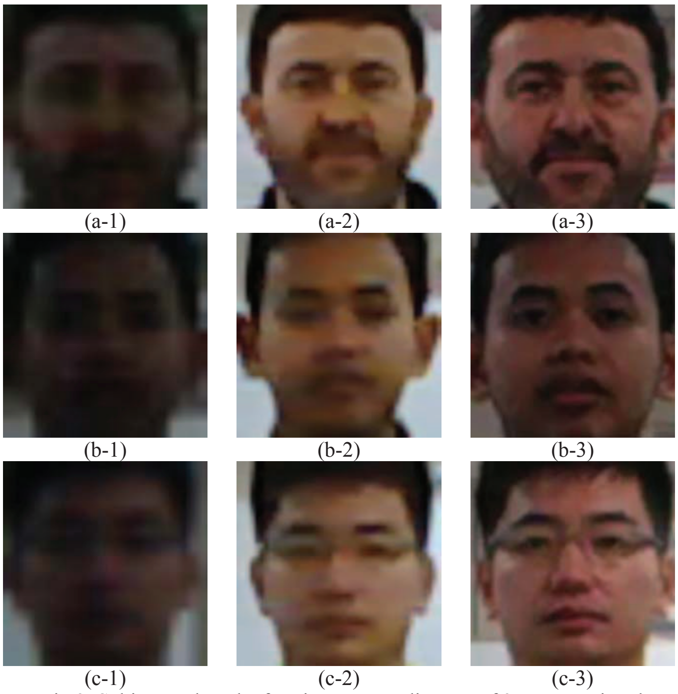


Fig.3, Subjects a, b and c face images at a distance of 3 meters using three cameras with resolutions of:

1) 320x240, 2) 640x480, 3) 1280x720

B. *Formulation of the Equation*

After gathering the readings of the different experiments, three different equations were formulated to be used to calculate the distance of any subject from the camera provided the resolution of the camera allows for the detection of the subject's face at that distance as shown in equations (1-3). These equations are based on the mean values of the different face sizes at the specific distances of 0.5 to 6.0 meters in intervals of 0.5 meters. It was noted that different faces diverge in sizes, and that the captured face size for the same subject may give different readings from one video frame to another; therefore by using the mean value of the subjects' face sizes it was possible to conclude a suitable criteria.

C. *The Testing Experiments*

In the testing phase, three different cameras with different resolutions were used. A mobile phone camera with a resolution of 320x240, a built-in laptop camera with a resolution of 640x480 and the third camera was also a mobile phone camera with a resolution of 1280x720. The video from the two mobile cameras was recorded to be later used for analysis and calculation, while the video stream from the laptop camera was used in live streaming and real-time distance calculations were performed. The testing experiments were done in two modes. In the first mode, the same images and videos captured during the measurement phase were used. This was essential in order to verify the validity of the equations against the actual distance as was shown on the cue cards. In the second phase, several different subjects would stand at random distances from the cameras and within the cameras' views. A measuring tape was used to measure the actual distance between each subject and the cameras. All the readings were recorded to be later compared with the readings from the cameras. Finally, it is important to point out that the cameras used for the testing experiment, were the same camera used in the measuring experiment, and that they were placed at the same elevation in both experiments.

The Results

The results gathered from the different experiments were tabulated and an accumulated accuracy value was calculated. Tables IV and V show the results from the first and second mode of experiments. They present a comparison between the actual distances and the output from the computer using the formulated equation. Fig. 4 shows a screenshot of the readings in both modes of the experiment. The overall accuracy for the first mode experiments was 96.23%, while the overall accuracy for the second mode experiments was 97.51%.

TABLE IV. THE RESULTS FROM THE FIRST MODE OF THE EXPERIMENT

Actual Dst	Avg. Calc. Dst	Avg. Accuracy %
0.5	0.53	93.80%
1.0	1.05	94.81%
1.5	1.60	93.13%
2.0	2.13	93.44%
2.5	2.54	98.28%
3.0	2.92	97.44%
3.5	3.46	98.71%
4.0	4.07	98.17%
4.5	4.69	95.82%
5.0	5.07	98.55%
5.5	5.42	98.55%
6.0	5.64	94.06%

TABLE V. THE RESULTS FROM THE SECOND MODE OF THE EXPERIMENT

Subject	Actual Dst	Avg. Calc. Dst	Avg Accuracy %
1	1.2	1.26	93.64%
2	1.9	1.92	99.21%
3	1.8	1.82	97.46%
4	3.2	3.28	97.01%
5	2.6	2.56	98.27%
6	1.8	1.81	99.44%

(a)

(b)

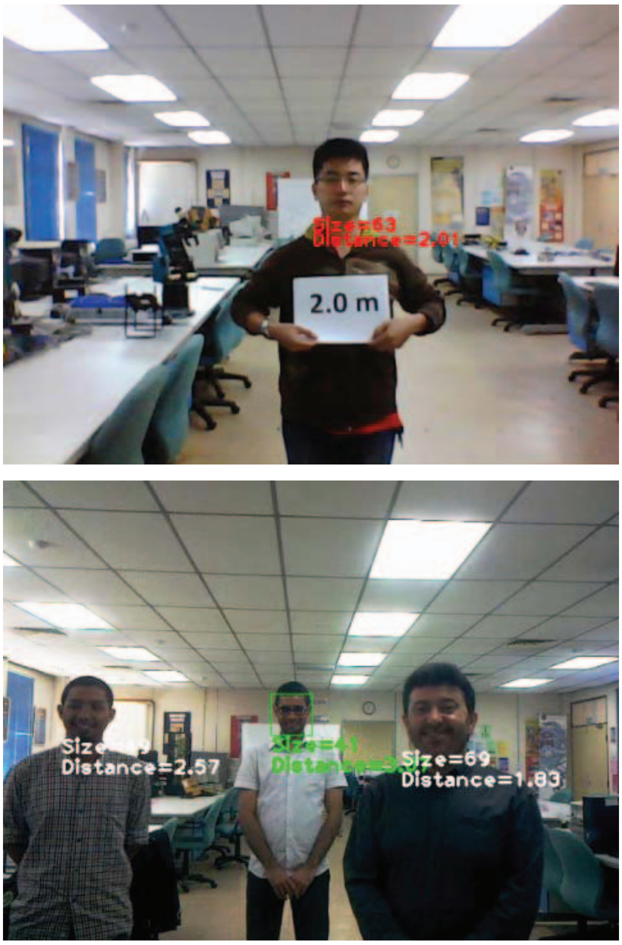


Fig.4, the testing experiments: (a) 1st mode, (b) 2nd mode

V. DISCUSSION

By examining the results of the experiments, it can be seen that the distance of the subjects can be calculated with a good accuracy. In addition to other factors, such as illumination and skin color, the resolution of the camera or the video plays a vital role in determining the furthest possible distance at which the subject's face can be detected and the distance could be calculated by the approach.

From Fig. 2 it can be seen that, as the subject moves further from the camera, the differences in the face size become less significant. This strongly affects the ability to precisely determine the actual distance when the subject is far away from the camera. However, at close proximity to the camera the face size changes rapidly and this results in a more volatile output; resulting in a rapidly fluctuating distance reading.

Moreover, since the face size differs from one person to another, the approach used the average values of all the participating subjects' faces to calculate the result which affected the distance calculation accuracy. During the second mode of the testing experiments the system managed to determine the distances of several different subjects simultaneously without a drop in accuracy or speed. Finally, by examining Table 4 that contains the results for a single subject distance calculation and Table 5, which contains the results for multiple subjects, it can be seen that calculating the distance of multiple subjects simultaneously did not affect the accuracy of the calculation.

VI. CONCLUSION

A system consisting of software and hardware was used to detect human faces and calculate their relative distances from a monocular camera without using extra distance or range measuring hardware. The software program utilizes an equation that links the face size with its distance from the camera. The software design allowed for simultaneous detection of more than one face and the calculation and output of the respective distances. The hardware design was simplistic, incorporating only a camera and a computer to perform the calculations, which can be replaced with a micro-controller for embedded system design. The results obtained from this study confirm the robustness of the approach and the possibility of utilizing it in many applications such as in a human-robot interface. Future work may focus on finding a relation between the calculated distance and the camera's field of view.

REFERENCES

[1] S. Satake, T. Kanda, D. F. Glas, M. Imai, H. Ishiguro, and N. Hagita, "How to Approach Humans□? – Strategies for Social Robots to Initiate Interaction," pp. 109–116.

[2] H. Hüttenrauch, K. S. Eklundh, A. Green, and E. A. Topp, "Investigating spatial relationships in human-robot interactions."

[3] T. Tasaki, S. Matsumoto, H. Ohba, M. Toda, K. Komatani, T. Ogata, and H. G. Okuno, "Dynamic Communication of Humanoid Robot with multiple People Based on Interaction Distance."

[4] C. Chen, C. Hsu, T. Wang, C. Huang, T. K. Rd, and T. County, "Three-Dimensional Measurement of a Remote Object with a Single CCD camera," pp. 141–146, 2007.

[5] R. A. Jarvis, "A Perspective on Range Finding for Computer Vision," no. 2, 1983.

[6] Hawk Measurement Systems, "Ultrasonic Distance Measuring.pdf," WO 90/08966, 1990.

[7] M. M. Saad, C. J. Bleakley, S. Dobson, and S. Member, "Robust High-Accuracy Ultrasonic Range Measurement System," vol. 60, no. 10, pp. 3334–3341, 2011.

[8] H. Kikuta, K. Iwata, and R. Nagata, "Distance measurement by the wavelength shift of laser diode light," Appl. Opt., vol. 25, no. 17, 1986.

[9] G. Benet, F. Blanes, J. E. Simó, and P. Pérez, "Using infrared sensors for distance measurement in mobile robots," Rob. Auton. Syst., vol. 40, no. 4, pp. 255–266, Sep. 2002.

[10] S. Nedevschi, R. Danescu, and D. Frentiu, "High accuracy stereo vision system for far distance obstacle detection," IEEE Intell. …, pp. 292–297, 2004.

[11] X. Bai, B. Wang, H. Wang, Z. Tan, and A. Higashi, "High Accuracy Distance Measurement System with Stereo Fisheye Cameras," no. 1155.

[12] Y. Yu, S. Kodagoda, and Q. P. Ha, "Relative Distance Estimation for Indoor Multi-robot Control Using Monocular Digital Camera," 2007.

[13] L. Xiaoming, Q. Tian, and C. Wanchun, "Real-Time Distance Measurement Using a Modified Camera," pp. 1–5, 2009.

[14] K. A. Rahman, M. S. Hossain, M. A.-A. Bhuiyan, T. Zhang, M. Hasanuzzaman, and H. Ueno, "Person to Camera Distance Measurement Based on Eye-Distance," 2009 Third Int. Conf. Multimed. Ubiquitous Eng., pp. 137–141, Jun. 2009.

[15] P. Viola and M. Jones, "Robust real-time face detection," Int. J. Comput. Vis., 2004.

[16] Itseez. OpenCV. 2013 [cited 2014 08/12/2014]; Available from: http://opencv.org/.

[17] Liu, Y. Face Detection System Design Based on OpenCV. in Proceedings of the 9th International Symposium on Linear Drives for Industry Applications, Volume 2. 2014. Springer.

[18] Microsoft. Visual Studio. 2013 [cited 2014 08/12/2014]; Available from: http://www.visualstudio.com/en-us.