Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Design of Real Time Facial Tracking and Expression Recognition for Human-Robot Interaction

IOP Conf. Series: Journal of Physics: Conf. Series **1090** (2018) 012044 doi:10.1088/1742-6596/1090/1/012044

Design of Real Time Facial Tracking and Expression Recognition for Human-Robot Interaction

W.S. Mada Sanjaya[1,2], Dyah Anggraeni[1,2], Atip Juwardi[1,2], and Madinatul Munawwaroh[1,2]

[1]Department of Physics, Faculty of Science and Technology, Universitas Islam Negeri Sunan Gunung Djati Bandung, Indonesia
[2]Bolabot Techno Robotic Institute, CV. Sanjaya Star Group, Bandung, Indonesia

E-mail: madasws@gmail.com

Abstract. In this study, we investigate the facial tracking and expression recognition to developed Social Robot called SyPEHUL (System of Physics, Electronics, HUmanoid robot and machine Learning). This social robot contains by 12 Degree of Freedom (DoF) for actuating robotic head based on Arduino microcontroller to display four type different facial expressions and tracking of the human face. In this research, expression recognition processed by an algorithm based on Python 2.7 (with OpenCV library) using Cascade Classification and LBPH Face Recognizer. The result shows that the implementation of facial tracking and expression recognition to Social Robot show a good accuracy of recognition rate and works well for Human-Robot Interaction.

1. Introduction

Social Robot is an interactive robot which can communicate with the human. Researchers mostly developed an automation expressive Social Robot to entertaining or help human's works, for example; KISMET [1], Flobi [2], Eddie [3], Probo [4], Muecas [5], and other [6]. To build the constructions of Social Robot have many component and complex methods. One of the most important parts of Social Robot is the camera as its eyes (vision sensor) and some face detection and recognition program needed to make the robot can looking and knowing something.

Image processing need to be able to recognize an objects, whether in the form of colors, shapes, etc. Especially to detect human face, need some face detection program. Many method to get face detection, for example: Viola-Jones method [7] [8] [9], Roberts Cross method [10], and other [11] [10]. And many classifier to recognize a face detection, such as; Local Binary Pattern (LBP) [12] [13], Support Vector Machine (SVM) [14], AdaBoost [15], Hidden Markov Model (HMM) [16], Bayesian [17], and other [18] [19]. Implementation of face detection they are; social robot [1] - [6], automation attendance [20], home security [21] [22] [23], game [9], and other.

In this paper will be described a real-time face detection and recognize human expression by using Cascade Classification method (Viola-Jones method) and Local Binary Pattern Histogram (LBPH) Face Recognizer Method based on Python 2.7 and OpenCV Library. Finally, this research will be implemented to 12 Degree of Freedom (DoF) Social Robot named SyPEHUL (System of Physics, Electronics, HUmanoid robot and machine Learning) to recognize, imitating,

IOP Conf. Series: Journal of Physics: Conf. Series **1090** (2018) 012044 doi:10.1088/1742-6596/1090/1/012044

and tracking human four type of different human expression (such as: smile, sad, angry, and surprise) based on Arduino microcontroller for Human-Robot Interaction.

The paper is organized as follows. In section 2, described the theoretical background of Face Recognition and Expression Recognition details. In section 3, describe the method and system design. In section 4, described a hardware design of Social Robot. In section 5, described the application of Face Tracking and Expression Recognition in detail. Finally, in Section 6 the concluding remarks are given.

2. Theoretical Background

2.1. Face Detection using Viola-Jones Method

The face detection in this study used the method by Paul Viola and Michael Jones published in 2001 [15] usually referred as the Viola-Jones method or only Viola-Jones. an approach done to detect objects in an Image is to combine the following four key concepts:

2.1.1. Simple rectangular feature often referred to as Haar features Images of human faces grouped on the bright side and the dark side. The existence of Haar features is determined by subtracting the average of the dark-region pixels by the average of the pixel light-region, show at Figure 1. If the difference is above the threshold, then the Haar features are said to be "exist".

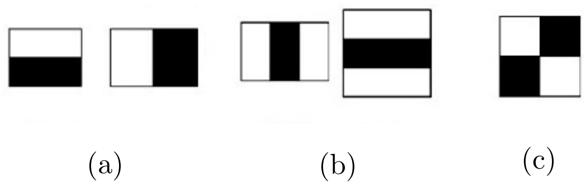


Figure 1. Haar Rectangular Feature [15]: (a)Edge, (b)Line, (c)Four Rectangle

2.1.2. An integrated Image for quick feature detection The integrated image is the addition of pixel values into the original image. In general, "integration" means to add all the smallest units. In this case, the smallest unit is the value. The already integrated value represents the sum of all pixels above the threshold and is to the left of the image. By starting the above procedure from the top left of the image to the bottom right of the image, the entire image can be integrated with only a few operations per pixel, show at Figure 2.

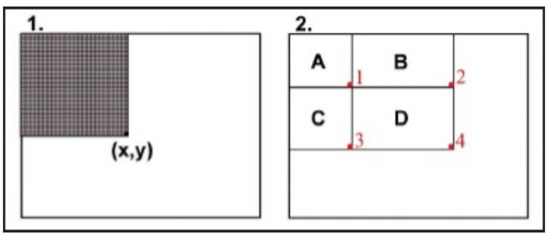
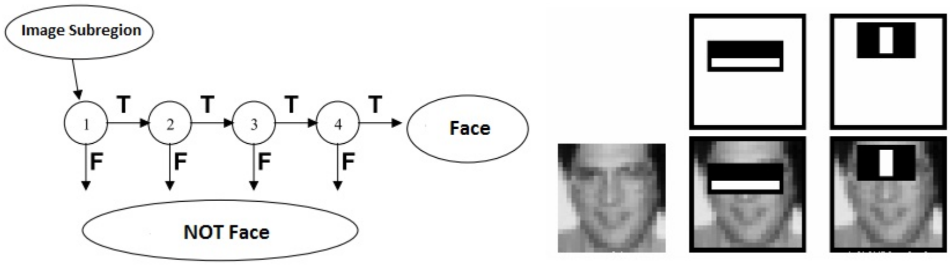


Figure 2. (1) Integration Process (2) Rectangle Divided into Multiple Segments [15]

2.1.3. Methods AdaBoost machine-learning Viola and Jones use a machine-learning method called AdaBoost method. AdaBoost classifier incorporates many "weak" to make it into a classifier that "strong". AdaBoost selects a set of "weak" classifier to combine and assign weights for each.

IOP Conf. Series: Journal of Physics: Conf. Series **1090** (2018) 012044 doi:10.1088/1742-6596/1090/1/012044

2.1.4. A Cascade Classifier that combines many features efficiently Viola and Jones combine a series of AdaBoost classifier as a filter chain shown in Figure 3(a). The order of filters in this cascade is based on the weight of results from the AdaBoost. The larger the filter weights in the first chain, then to eliminate the non-face image area will also be done quickly. Figure 3(b) shows the first two features of the original Viola-Jones cascade placed on a face. The first key is on the cheek area which is brighter than the eye area and the second is that the connection between the forehead and the nose is brighter than the eye.



(a) (b)

Figure 3. Cascade Classifier [15]; (a)Method, (b)Result

2.2. Face Recognizer using Local Binary Pattern (LBP) Method
Local Binary Pattern (LBP) is a texture descriptor that can be used for faces. LBP is defined as the comparison of the pixel binary value at the center of the image with 9-pixel values around it [19]. For example, in a 3x3 image, the binary value in the image center is compared with the surrounding value. By subtracting the pixel value at the center of the image by its surrounding pixel value, if the result is more or equal to 0 then it is given a value of 1 and if the result is less than 0 then it is given a value. After that, arrange 8 binary values clockwise or otherwise and convert 8 binary bits into decimal values to replace the pixel value in the center of the image, as illustrated in Figure 4.

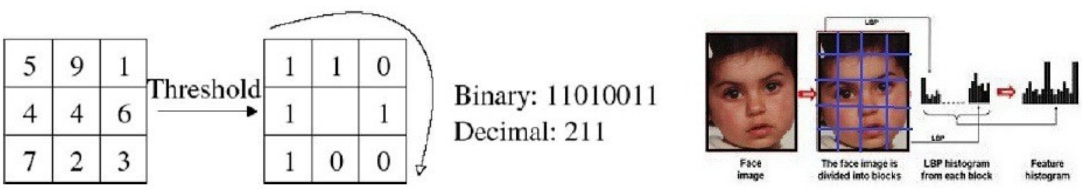


Figure 4. Illustration of Local Binary Pattern (LBP) [19]

The decimal value of 8 bits of (LBP code) can be expressed in the following equation 1 [24]:

$$LBP_{P,R}(x_0, y_0) = \Sigma_{p=0}^{p-1} s(g_p - g_0)2^p \quad (1)$$

From the function $s(x)$, define:

$$s(x) = \{^{1, x \geq 0}_{0, x < 0} \quad (2)$$

Equation 2 aims to eliminate the variability caused by the contrast illumination so that the facial image with various variations of the illumination will get almost similar output. By

IOP Conf. Series: Journal of Physics: Conf. Series **1090** (2018) 012044 doi:10.1088/1742-6596/1090/1/012044

extracting pixels using LBP, we get a new matrix value which will be changed to a histogram for facial vector features.

The LBP method can be implemented on face images to extract features which can be used to get a measure of the similarity between images. LBP code is calculated for every pixel of an image. The histogram of these patterns from the image called labels. Lable forms a feature vector and is represented the image texture. These histograms calculating the distance between the histograms can then be used to measure the similarity between the images. [24]

3. Method and System Design
In general, the main tools and component used in this research are: Personal Computer, Arduino microcontroller, Social Robot (SyPEHUL), camera, connections, and others. Program written in Arduino IDLE and Python 2.7 (with OpenCV Library). And Figure 5 shown generally the process of Facial Tracking and Expression Recognition of Social Robot SyPEHUL, below:

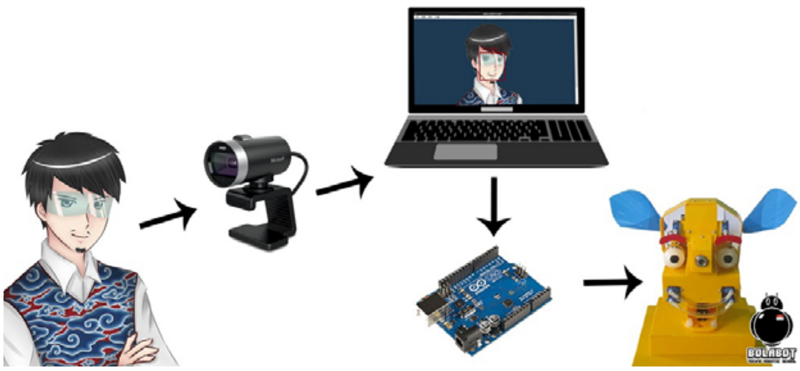


Figure 5. General Research Scheme

From the Figure 6 can be explained that the camera as a Social Robot vision sensor. The human face detection process by algorithm on the computer using Cascade Classifier Method. Then the process divided into 2 processes: the first process is making training data, consist of features extraction process using LBPH Feature and using LBPH Face Recognizer Method for classifying face expression. Second, is testing process, such as LBPH features extraction, then matching with Trained Data. The matching data be processed to obtain expression classification and face coordinates. For coordinate data is used to move the robot's neck so that the robot seems to follow our face. While the classification process Social Robot will make interaction expression (happy, sad, angry, and surprise) like the facial human expression. All processes work in real-time based on Python 2.7 (with OpenCV) and Arduino. microcontroller.

4. Hardware Design
To design a system of social robots it takes a major component especially the motor servo that serves as a robot facial driver, such as; Eyebrows, eyes, mouth, ears and neck. And in general, the materials needed the camera as a visual sensor and the Arduino microcontroller to serves as the motor servo to control the facial expression of robot. Figure 7 are the design and realization of our Social Robot named SyPEHUL which used in this research:

SyPEHUL consists of twelve motor servo (12 Dof motor servo) component which connects with Arduino microcontroller. SyPEHUL mechatronic consists of; eyebrow using 2 servos, eye using 2 servos, mouth using 4 servos, ear using 2 servos, and neck using 2 servos. From the Figure 8, each servo supplied used by the 5-volt battery which has 100 mA of current, and each ground of servo connected to the ground of Arduino Board in order get a better working. And

IOP Conf. Series: Journal of Physics: Conf. Series **1090** (2018) 012044 doi:10.1088/1742-6596/1090/1/012044

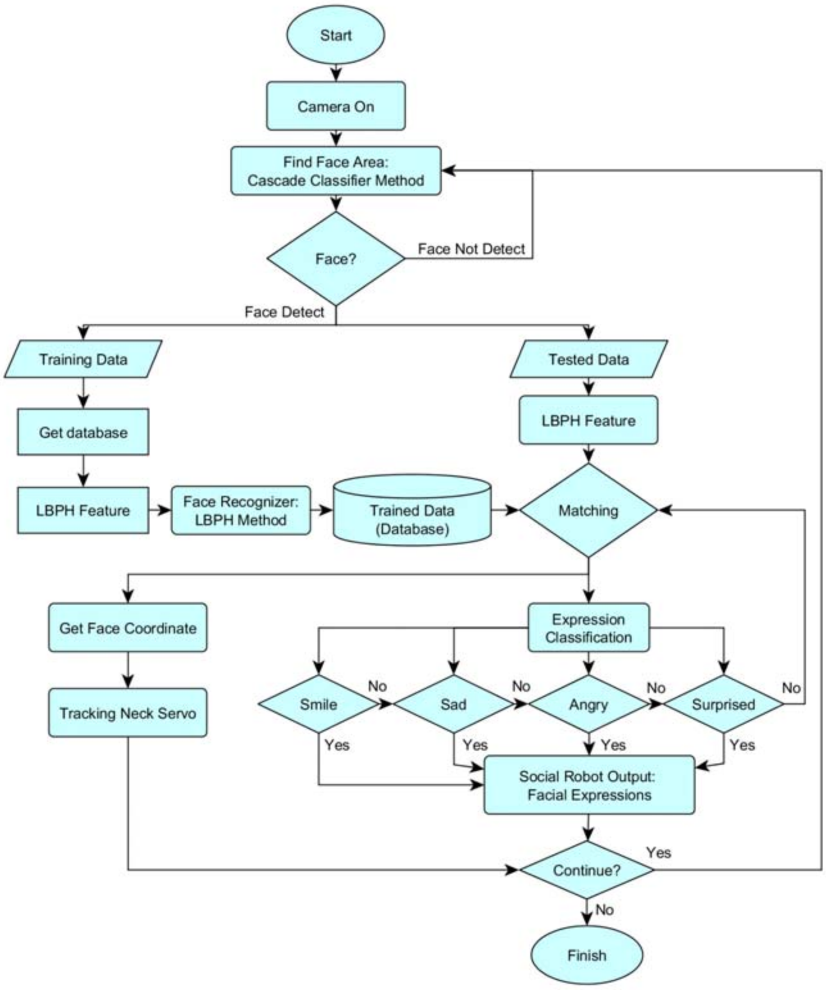
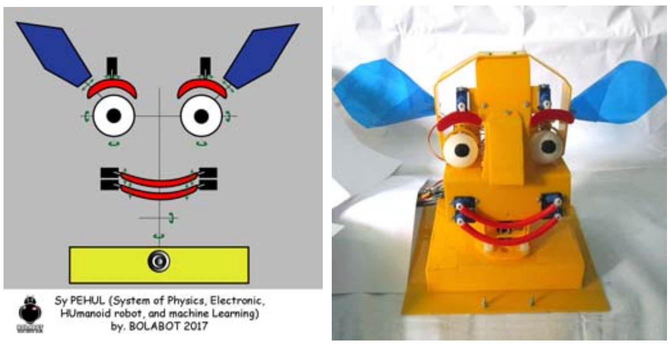


Figure 6. General System Scheme of Social Robot Algorithm



(a) (b)

Figure 7. SyPEHUL Design; (a)Design, (b)Realization

from the mechatronics, SyPEHUL Expressions divided by 4 interaction, such as; Smile, Sad, Angry, and Surprise, shown as the Figure 9.

IOP Conf. Series: Journal of Physics: Conf. Series **1090** (2018) 012044 doi:10.1088/1742-6596/1090/1/012044

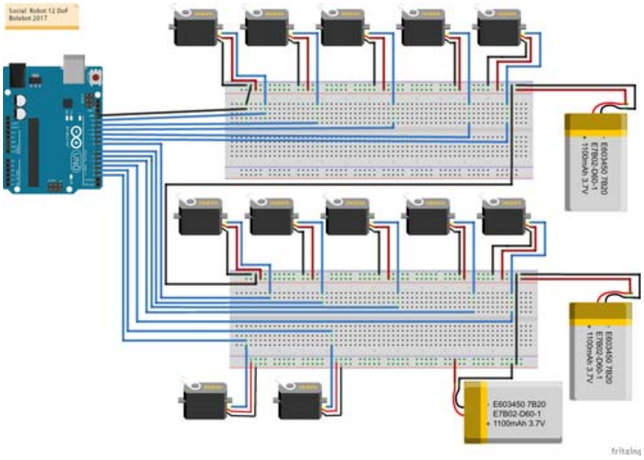
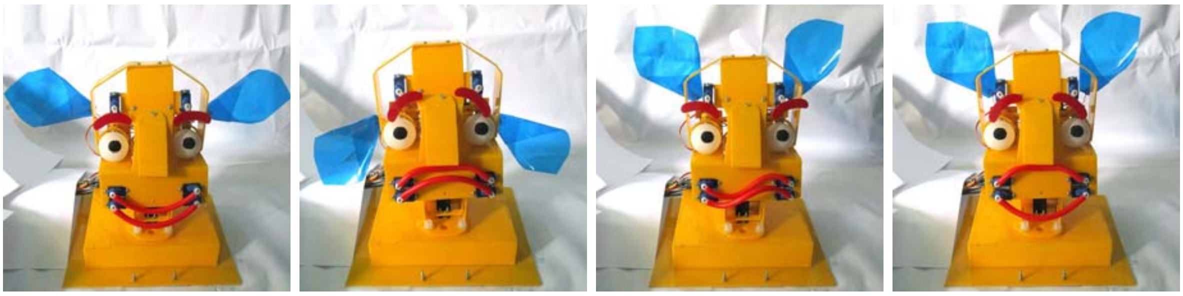


Figure 8. SyPEHUL Schematic



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

Figure 9. SyPEHUL Expression; (a)Smile, (b)Sad, (c)Angry, (d)Surprise

5. Result and Discussion

5.1. Expression Recognition Database and System Test

The beginning of the research process is formulated ideas, concepts, objectives, problem formulation, until the study of making Social Robot. To get a Social Robot can recognize human faces, in this research used Cascade Classification method for face detection and Local Binary Pattern Histogram (LBPH) Face Recognizer to recognize the expression of human faces based on Python 2.7 and OpenCV library. Figure 10 is an example of collecting databases of some people for making training data.

Database made from 5 expression image, they are; Smile, Sad, Angry, Surprise, and Neutral. From the obtained group of data, needed for classification process using LBPH Feature Extraction and LBPH Face Recognizer, then will be called the trained data. First, the trained data is tested by trained respondent for data clarification. The test results are shown on Table 1 is the accuracy average rate of expression recognition by trained respondents (in the database). And Figure 11 show the examination of trained data by trained respondent.

Table 1 explain that the database clarification by trained respondent for human facial expression recognition shows a good accuracy. Second, the trained data is tested to the not trained respondents (outside database). The expression recognition accuracy average rate of not trained respondents shown on Table 2 and the examination by not trained respondent shown as Figure 12. And from Table 2 show that the accuracy of expression recognition by not trained data is well for some expression.

Figure 10. Database Example

Table 1. The expression recognition accuracy rate of trained respondents in Percent

Test	Smile	Sad	Surprise	Angry	Neutral
Smile	**100**	0	0	0	0
Sad	0	**90**	0	0	10
Surprise	0	0	**100**	0	0
Angry	0	0	0	**100**	0
Neutral	0	12.5	5	0	**82.5**

(a)Smile (b)Sad (c)Angry (d)Surprise (e)Neutral

Figure 11. Tested the expression recognition of trained respondent

5.2. Facial Tracking

From the programming system of face detection can get the face coordinate. In this research, the face coordinate needed to control neck motor servo of Social Robot SyPEHUL in order to track human face. The first step is set the resolution of the video image, we define as 480 x 640 pixel. Then, Equation 3 and Equation 4 shows an algorithm to find a center coordinate of the human face from camera capture.

$$xx = x + (w/2) \tag{3}$$

$$yy = y + (h/2) \tag{4}$$

IOP Conf. Series: Journal of Physics: Conf. Series **1090** (2018) 012044 doi:10.1088/1742-6596/1090/1/012044

Table 2. The expression recognition accuracy rate of respondents not trained data in Percent

Test	Smile	Sad	Surprise	Angry	Neutral
Smile	**80**	0	0	0	20
Sad	0	**50**	22.5	0	27.5
Surprise	0	0	**100**	0	0
Angry	0	0	0	**77.5**	12.5
Neutral	5	0	35	0	**60**

(a)Smile (b)Sad (c)Angry (d)Surprise (e)Neutral

Figure 12. Tested of not-trained respondents

x is initial horizontal face coordinate, w is width coordinate of face, xx is a center value of horizontal face coordinate, y is initial vertical face coordinate, h is height coordinate of face, yy is a center value of vertical face coordinate, and units in pixel. Figure 13 show the section of camera coordinate divided into 4 section, are; Up, Down, Right and Left.

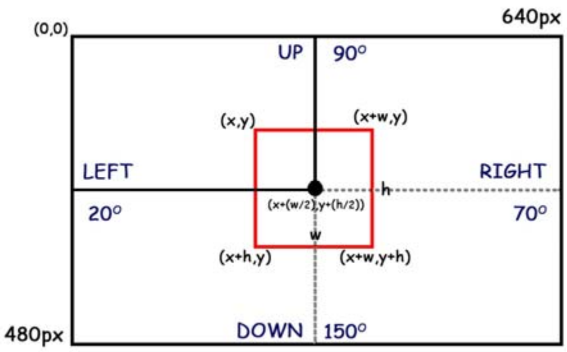


Figure 13. Camera Coordinate and angle of motor servo

Before converting the coordinate value become an angle of the motor servo, we must know the minimum and maximum angle of Social Robot neck motor servo to look up/down (vertically) or look right/left (horizontally). In this research, motor servo arrange are: (1) arrange of horizontal angle is 90^o - 150^o, (2) arrange of vertical angle is 20^o - 70^o. Figure 13 illustrated the distribution of camera pixel, direction, and angle of Social Robot neck motor servo. Thus, after the center coordinate value sent from Python to Arduino IDLE, then process to control the Social Robot neck motor servos.

5.3. Implement The System to Social Robot

The identifying and categorizing the trained data of expression recognition are successfully, then applied the trained data to be a control expression system of Social Robot SyPEHUL. Figure 14 shown a training of the Expression Recognizer:

(a)Smile (b)Sad (c)Angry (d)Surprise

Figure 14. Examination of human expression recognition to Social Robot expression interaction

In addition to the classification of human facial expressions, faces detected by the camera obtained coordinates of the face position. The coordinates can define as x coordinate (as the horizontal neck) and y coordinate (as the vertical neck). Then, the coordinates can be converted as the angle to control motor servo of Social Robot by using Python and Arduino serial communication was described before. Figure 15 show the result of tracking and expression recognition for Human-Robot Interaction.

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

Figure 15. Tracking and following human face; (a)Smile, (b)Sad, (c)Angry, (d)Surprise

6. Conclusion

In this research have been presented and developed a Social Robot SyPEHUL which can recognize, imitating, and tracking human face. The result of expression recognition processed by the algorithm based on Python 2.7 (with OpenCV library) using Cascade Classification and LBPH Face Recognizer method has a good accuracy rate. Also, the implementation of facial tracking and expression recognition to control 12 DoF of Social Robot SyPEHUL based on Arduino microcontroller works effectively to make four type imitating facial expression. Future works will focus on the combination of speech recognition and facial expression to enhance emotional expression of SyPEHUL for Human-Robot Interaction.

Acknowledgement

The authors would like gratefully acknowledge the financial support from DIPA UIN Sunan Gunung Djati Bandung.

References

[1] Breazeal C 2003 *Robotics and Autonomous System* **42** 167–175
[2] Hegel F, Eyssel F and Wrede B 2010 *IEEE International Symposium on Robot and Human Interactive Communication* **19** 107–112
[3] Buss M, Sosnowski S, Bittermann A and Kolja K 2006 *International Conference on Intelligent Robots and Systems* 3113–3118
[4] Doroftei I, Adascalitei F, Lefeber D, Vanderborght B and Doroftei I A 2016 *7th International Conference on Advanced Concepts in Mechanical Engineering* **147**
[5] Cid F, Moreno J, Bustos P and Nunez P 2014 *Sensors* **14** 7711–7737
[6] Shayganfar M, Rich C and Sidner C L 2012 *Int. Conf. of Intelligent Robots and Systems* **12**
[7] Tikoo S and Malik N 2016 *International Journal of Computer Science and Mobile Computing* **5** 288–295
[8] Boda R and Priyadarsini M J P 2016 *ARPN Journal of Engineering and Applied Sciences* **11** 13472–13476
[9] Zhan C, Li W, Ogunbona P and Safaei F 2008 *International Journal of Computer Games Technology* **2008** 1–7 ISSN 1687-7047
[10] Das S 2016 *International Journal of Signal Processing, Image Processing and Pattern Recognition* **9** 143–158
[11] Lakshmi H C V and PatilKulakarni S 2010 *International Journal of Computer Theory and Engineering,* **2** 552–558
[12] Shan C, Gong S and Mcowan P W 2009 *Image and Vision Computing* **27** 803–816 ISSN 0262-8856 URL `http://dx.doi.org/10.1016/j.imavis.2008.08.005`
[13] Sanjaya W S M, Anggraeni D, Zakaria K, Juwardi A and Munawwaroh M 2017 The Design of Face Recognition and Tracking for Human-Robot Interaction *ICITISEE* (Yogyakarta, Indonesia: IEEE) pp 1–6
[14] Sadek N O, Hikal N A and Zaki F W 2017 *International Journal of Advanced Computer Science and Applications* **8** 303–311
[15] Viola P and Jones M J 2001 Robust Real-time Object Detection Tech. Rep. February Cambridge Research Laboratory The Cambridge, Massachusetts
[16] Punitha A and Geetha M K 2013 *International Journal of Emerging Technology and Advanced Engineering* **3** 180–185
[17] Hedaoo S V, Katkar M D and Khandait S P 2014 *International Journal of Engineering Trends and Technology* **8** 517–521
[18] Khan M I and Bhuiyan M A A 2009 *International Journal of Computer Science and Network Security* **9** 300–306
[19] G M R and Rajesham J 2013 *International Journal of Scientific & Engineering Research* **4** 1540–1546
[20] Thakare N, Shrivastava M and Kumari N 2016 *International Journal of Computer Science and Mobile Computing* **5** 74–78
[21] Sharma N and Thanaya I 2016 *International Journal of Innovative Research in Science, Engineering and Technology* **5** 10357–10362
[22] Hajari P V and Andurkar A G 2015 *International Journal of Advanced Research in Computer and Communication Engineering* **4** 232–234
[23] Manjunatha R and Nagaraja R 2017 *International Research Journal of Engineering and Technology* **4** 437–442
[24] Rahim M A, Hossain M N, Wahid T and Azam M S A 2013 *Global Journal of Computer Science and Technology Graphics & Vision* **13**