**Proceedings of the**
**International Conference on Mechanical Engineering and Renewable Energy 2015**
**(ICMERE2015) 26 – 29 November, 2015, Chittagong, Bangladesh**

**ICMERE2015-PI-148**

# DESIGN A ROBOTICS HEAD TO INTERACT WITH HUMANS USING NON-VERBAL GAZE FOLLOWING

**Mohd. Habib Ullah[1,*], Rahat Hossain Samit[2] and Mohammed Moshiul Hoque[3]**

[1-3]Department of Computer Science & Engineering (CSE), CUET, Chittagong, Bangladesh
[1,*] habibcse06@gmail.com,[2]rahat106@yahoo.com,[3]moshiul_240@cuet.ac.bd

***Abstract***- *Designing a Social Humanoid Robot is an intelligent based an autonomous technology. It is an advance research field in present stage. Basically, we would like to design a robot head to show gaze following expression with natural interaction. We know that 60% human communication is non-verbal that is an important research field topic that is how interfaces of this communication can be designed and implemented. To achieve this goal, several robotic head have been developed to interact human as like as social being. But it is unclear what the knowledge to integrate in robot head. Mainly, our aim is to design and implement a social humanoid robot head based on gaze following knowledge and faceAPI algorithm. Our system is cheap, reliable, well performance and robust because of it is built as like as retro projection based concept. Most important and enrich information of non-verbal cues are gaze and eye contact which are known as mutual gaze following. Face Tracking finds face that is face detection and tracks the face in real time video sequence. To implement face tracking concept, we use faceAPI algorithm from Seeing Machines Technology that is facial feature based face detection and tracking algorithm which provides 6 degree-of-freedom (ROLL,PITCH,YAW angle and X,Y,Z position). Therefore, we perform a face tracking task using a web camera mounted on a pan-tilt-zoom unit using this algorithm. Finally, we are able to design a 3D face and also track face that is built as gaze following robot.*

**Keywords:** Social Robot, Face Detection, Face Tracking, faceAPI, Non-verbal Communication and Gaze.

## 1. INTRODUCTION

Social robot is an intelligent based robot which follows natural human machine interface and attractive social behaviors i.e. eye contact, gaze following, join attention, share object etc. Basically, we designed a social robot head to interact human and follow human attention with non-verbal gaze following. Service or communication robots are required to be able to perform not only verbal communication but also non-verbal communication and for human beings, one of the most important aspects of non-verbal communication is gaze [1]. Application of robot extended from home to industries to automate every sphere of task. Day by day, we see that many robotic head had been developed to implement human to robot interaction. For example: Avatar, iCub, LightHead, Mask-Bot, Mask-Bot 2i, Curved Screen Face, KISMET and so on. iCub robotic head which is included in the European Project RobotCub. “Mask-Bot 2i” is an interchangeable robot face where different faces can be projected onto the active head system. [2] How to replace mechatronic and android faces head into retro faces to cost effective, flexibility and robustness potential developments, it is learned from F. Delaunary et al.[3]. A live generated video is retro projected into a semi-transparent mask which is known as retro-projected animated face (RAF) technology. Use of servo motor to design articulated arm and mouth to express emotion which was studied from Albert Version of HUBO humanoid robot [4]. Using this knowledge in our project, we designed our pan tilt module (PTM). KISMIT was also a social emotional robot and had a simple mouth to display emotional expression which is developed into MIT Lab. Distance, direction, and head visibility and pose estimation are taken account for appropriate interactive behaviors to communicate face to face social interaction [5]. Face detection is the psychological process by which we locate and attend to faces in a visual scene [6]. Research shows that our ability to detect faces is affected by a range of visual properties such as color and orientation. And also face tracking enables us to achieve real time surveillance. With the help of face detection and face tracking concept, we were designed face detection and tracking module (FDTM). From this module, we got roll, pitch, and yaw angle measurement and also three position vectors. By using, robot behavior module, we controlled our robot and thus we implemented non-verbal gaze following behavior on our designed social robot head.

In this paper, we introduce our non-verbal social robotics head. The system follows the retro based "Mask-bot" communication system for gaze following non-verbal communication. In the next section, we discuss about previous related works which are influence us. In section 3, we discuss in details about system overview with the system specification for this social robotic head. In section 4, we design our experiment. Finally, the results evaluation and the conclusion are shown in section 5, 6.

## 2. RELATED WORK

In the present stage, we examined huge robot head to interact human naturally. However, we see that it is not enough to cost effective, flexible and has also limitation on behavior control to design a social robotic head. The mechanical design of the iCub head is divided in three major parts: Neck Mechanism, Eyes Mechanism and Cover (Face). To allow the robot to interact with other people and to have all desired behavior several sensors were applied. For vision, the main sensory modality, two DragonFly cameras with VGA regulations and 30 fps speed was considered [7]. But this project is also too costly and complicated to be developed for everyday purpose. In Robovie R-2, the unit of its eye is composed of a CMOS camera, infra-red pass film, and servo motors for eye movement [8].The infrared pass film is attached with the center of a white, hollow eyeball. The robot drives the motors of its neck and eyes based on the sensory data from its CMOS cameras in both the left and right eyes. Host computer (PentiumIV, 2.8 GHz) on the robot receives the image of the person's frontal view and the gaze data including his/her focusing point from the gaze direction device. This robot was developed in a more effective way to actively participate in a gaze tracking with the human participants. Other than that the goal of this project could not be achieved. Eye Contact Robot, detects the eyes (pupils) and the nostrils in the zoomed-in image [9].They use the feature extraction module in the face recognition software library by Toshiba [10]. In this project, they used an algorithm which is not always capable of finding human face properly in all condition. So it needed an extra feature to efficiently detect faces. Roboceiptionsist is a conversational robot not having a proper structure to make an interaction successful. It operates with a monitor on the top of them as a face which is pretty disturbing when communicating with human and also works on complex algorithm which is not suitable in real life environment. The Mask-Bot system is a life-size, retro projected face shape display system with the ability to show realistic talking which uses pre-calibrated 3D face animation, auditory and speech motion output [11,12]. Distortion resulting comes out from mismatch between calibrated 3D face model and 3D mask feature because of 3D mask shape is fixed and also an error is occurred for different face models and morphed faces to use specific text-to-speech output. To overcome these errors, still images were used without head motion for stimuli in the experiment. Mask-Bot 2i has been developed as a stand along platform for performing research into human-robot interaction (HRI) whose main feature is the avatar animated onto a mask using projection system and other feature is fits into the footprint of an average adult human. It is based on three hardware components:1) projector, 2) optics, and 3) mask. This system reduces the lens distortion using Guhring's line-shift method which is created in 1st generation Mask-Bot system. Fig.1 shows a snapshot of Mask-Bot system architecture.

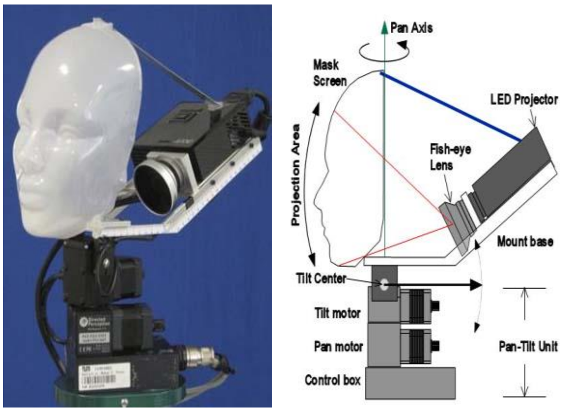


Fig.1: Mask-Bot

We also studied a few algorithms to implement face detection and tracking system i.e. faceAPI algorithm [13], AdaBoost Haar Classifier [14], POSIT algorithm [15].

## 3. PROPOSED SYSTEM OVERVIEW

The system consists with three main components. They are Head Detection and Tracking Module (HDTM), Pan-Tilt Control Module (PTM) and Behavior Control Module (BCM). In this phase, we discuss about the system architecture and also its implementation procedure. The system basically designs to display human like face for gaze following. It provides effective non-verbal communication clues. Fig.2 shows a conceptual design of the proposed system.

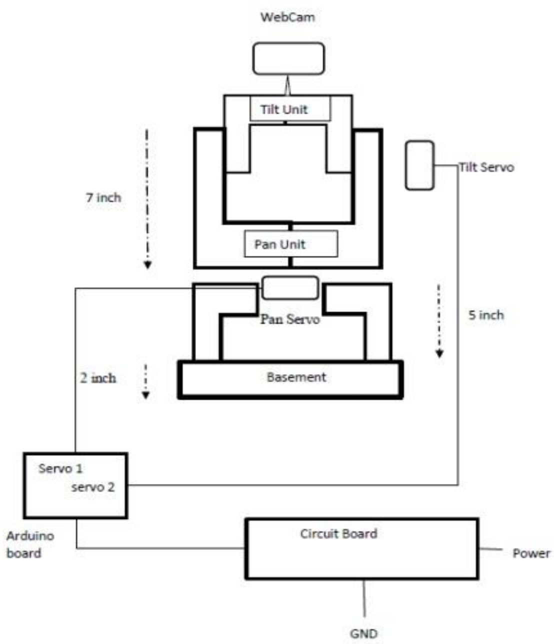


Fig.2: Conceptual System Architecture

The Head consists with a 3D Face MASK, a LED Projector, 2 (two) servo motor, a face detector camera (USB Web Camera). 3D Mask, LED Projector, USB Camera is mounted on pan-tilt servo motor. Pan-Tilt servo motor is controlled by PTM. A RS-232 communication port with an Arduino Mega board is used to connect between Pan-Tilt Module and PC (Windows7, 32bit). LED Projector (3M Mpro 120 Projector) is used to project face animation (eye image & face database) computer graphics on the 3D face mask. A USB camera is used to detect target human head.

### 3.1 Face Detection and Tracking Module

Face Detection and Tracking Module (FDTM) separate into two sub module those are (a) Face Detection Module (FDM), and (b) Face Tracking Module (FTM).

- **Face Detection**: To detect face, we use Viola & Jones AdaBoost learning algorithm using Haar Features. Its output is used in FDTM to track the face position. We use a USB web camera (Logitech Web CAM) to detect face.
- **Tracking System**: To implement of this phase, we can choice one of two methods: POSIT algorithm or faceAPI algorithm. Here, we implement faceAPI algorithm to get 6DOF (3 Rotation and 3 Translation) [16] because the result of faceAPI algorithm is more efficient that POSIT algorithm. From this algorithm, we easily get Roll, Pitch, Yaw angle. But we use only pitch and yaw angle to control pan-tilt movement because roll angle is less essential to non-verbal gaze following communication.

### 3.2 Pan-Tilt Control Module

To implement of this phase, we needed a powerful motor to move 3D mask along with LED projector and subordinate structure. Basically, we could use one of three motors i.e. DC motor, Servo Motor, Stepper Motor to control the head movement. In our project, we used two servo motor to control pan-tilt movement that is head horizontal and vertical movement respectively. The payload capacity of MG955 Servo Motor is about 10KG/CM and rotation speed is 0.20Second/$60^0$. It is enough to carry load of our head structure. It is basically design to operational control module. It is precise motion control, so it may not suitable as like as human real time head movement.

### 3.3 Behavior Control Module

To establish gaze following concept, we designed a few rule. It is essential a social robot [17] to communicate each-other that is human to machine interaction. Following are the rule by which we established the non-verbal gaze following communication between human and robot. If it is called Behavior Control Module (BCM), but it also acts as like as Robot Control Module (RCM).

The procedures or rules or steps of Behavior Control Module [18, 19] are

- Step (1): Robot detects and track the target human in its field of view (FOV) that is $0^0$ to $180^0$ (left-right and up-down).
- Step (2): Robot turns its head to the target human and calibrated its head position to establish gaze crossing.
- Step (3): Robot detects human face to establish gaze flowing with calibration pan-tilt angle. If pan-tilt angle >=$15^0$, and then change gaze. Thus our system establishes gaze flowing system.

## 4. EXPERIMENTAL SETUP

We designed our system with a 3D robot head who attracts a target human in its whole field of view that is in between $0^0$ (zero degree) to $180^0$ both pan and tilt movement. In Fig.3, we show the experimental block diagram of our system.

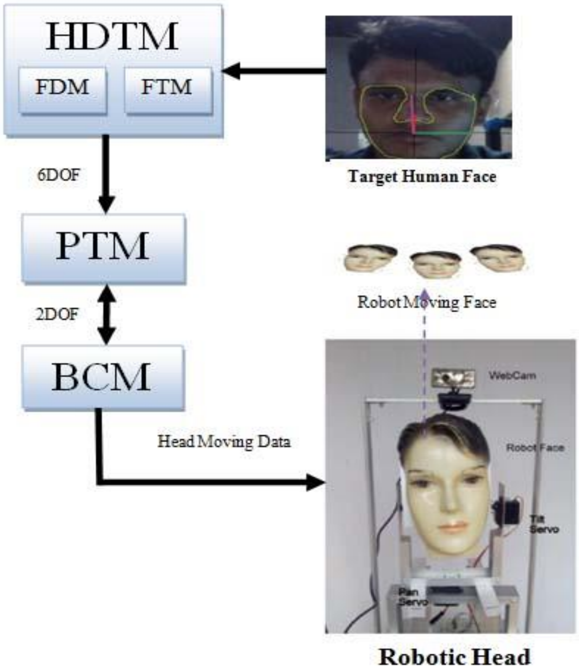


Fig.3: Experimental set up.

From HDTM, we get six degree of freedom (6DOF) those are three position point and three rotation angles. Mainly, we needed head orientation angle to establish human-robot gaze following system those are pan, tilt and zooming angle. But, in pan-tilt module, we only worked with pan and tilt angle. So, we no needed zooming angle in the behavior control system. Therefore, pan tilt movement controlled by PTM and gives two degree of freedom (2DOF) (pan, tilt angle) as an input of behavior control module. Behavior Control Module acts as a Robot Control Module. So, it is interchangeable to each-other. BCM is activated by pan-tilt angle where consider $15^0$ angle different within both pan-tilt angle changes. When the target human move any sideline i.e. left to right or up to down and if its anyone angle difference(pan or tilt ) is greater than or equal to $15^0$ and then robot follows the human to establish gaze following system.

## 5. EVALUATION RESULTS AND DISCUSSION

For the evaluation of our developed social robot head, we have included two analyses: 1) Quantitative Evaluation and 2) Subjective Evaluation.

### 5.1 Quantitative Evaluation

Speed: We have considered 12 steps to move whole field of view from $o^0$ to $180^0$ both left-to-right and up-to-down and for accurate calibration with real time object that is for smooth movement, we have added a delay of 100ms for each step. It keeps track of the inertia of the hardware structure and makes the system efficient. Therefore, the speed of robot head movement is Hardware Speed + Step Delay * 12 =1.71sec/$180^0$, where Hardware Speed=.51sec/$180^0$ and Step Delay=.1sec/$15^0$
Movement Range: pan is $40^0$ to $140^0$ and tilt is $50^0$ to $140^0$.
*Load Capacity*: The robot weight around 1.2 kg including the power source. It is weight up to 4kg that is enough for our robot head.
*Accuracy:* We have collected data from our subjective experiment where target human face position varies and the robot keeps tract it. The experiment result of face detection and tracking along PTM verses time is given in Figs. 4, 5, and 6.

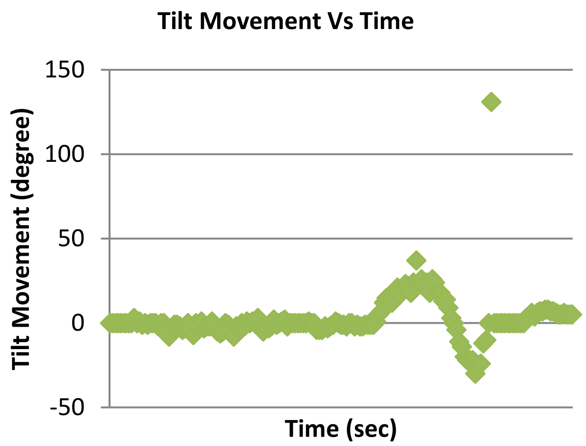


Fig. 4: Accuracy of face detection and tracking along tilt movement vs. time.

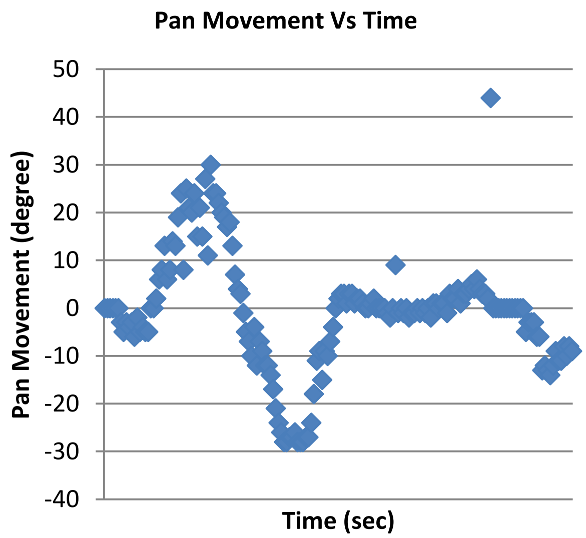


Fig. 5. Accuracy of face detection and tracking along pan movement vs. time.

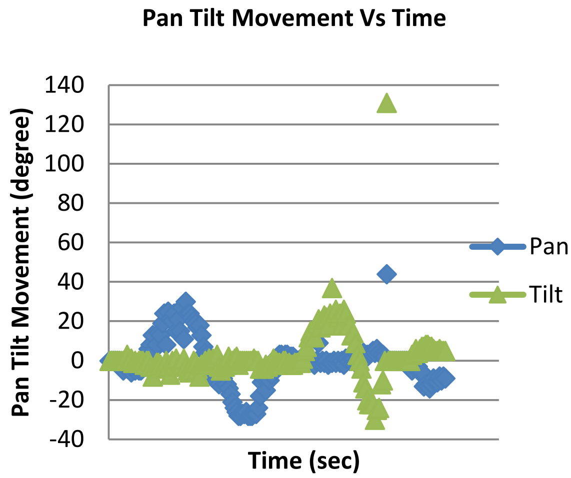


Fig. 6. Accuracy of face detection and tracking along pan tilt movement vs. time.

In Fig. 4, we show that the experience of face detection and tracking vertically. The starting point of central field of view with frontal face is directly denoted by $o^0$ (zero degree) at the web camera. The positive (+) values means the deviation to up and the negative (-) value means the deviation to down. Overall movement of up to down in increasing time, the range of tilt orientation values between $-32^0$ (lowest) to $30^0$ (above).

In Fig. 5, we show that the experience of face detection and tracking horizontally. The starting point of central field of view with frontal face is directly denoted by $o^0$ (zero degree) at the web camera. The positive (+) values means the deviation to left side and the negative (-) value means the deviation to right side. Overall movement of left to right in increasing time, the range of pan orientation values between $-30^0$ (lowest) to $31^0$ (above).

In Fig. 6, we show that the experience of face detection and tracking module accuracy between pan tilt movement versus time. In blue line, the positive (+) values means the deviation to left side and the negative (-) value means the deviation to right side. In acid green line, the positive (+) values means the deviation to up and the negative (-) value means the deviation to down. When we get constant continuous values of pan and random values of tilt and that period we defined it as a steady stage position of pan and vice versa.

### 5.2 Subjective Evaluation

*Experiment Design*: In our multimedia lab in Computer Science and Engineering department, we have arranged an evaluation session for our project. We placed our robot and laptop and participants stood around two meter away from the robot and the robot followed the participants face. For each participant we first told them our robot's functionality, usage and our motivation of the project. They interacted in various time periods. During this time period, they shared their experience with the system.

Fig 6. Scene of Human-Robot interaction experiment.

*Measurements*: The measurement was a simple rating on a Likert scale of 1 to 5. The values of the Likert scale are 1-Very Bad, 2-Bad, 3-Usual, 4-Good, 5-Excellent. The questionnaire contained four items. They are the following:

Table 1: Questionnaire for Subjective Evaluation Results

| Question | Items |
|---|---|
| Q1 | Are the movements generated by the robot accurate? |
| Q2 | Do you think the robot follows your face when moves horizontally left to right or right to left? |
| Q3 | Do you think the robot follows your face when moves vertically up to down? |
| Q4 | Do you think that the robot can follow your gaze accurately? |

We conducted the measures of analysis of variance (ANOVA) for all questions. We observed a total of 40 (10 *4) questionnaire for all participants. There are significant difference was found among questionnaire responses ($F(3, 39) = 3.07$, $p = 0.04$) of participants. Fig. 7 indicates the mean and standard error values for all questions. Thus, the result shows that participants feel that the robotic system can produce interesting behaviors while interacting with it and their questionnaire respond revealed that the robotic system can be able to follows the participant's gaze or face quite satisfactory. Specifically, it can follow while the human turns their face vertically or horizontally with reasonable accuracy.

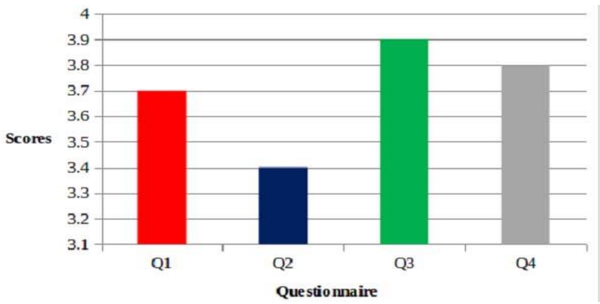


Fig 7. Subjective Evaluation Result

*Discussion*: Both quantitative and subjective evaluation result shows that the robot performs fine behaviors to participants except the speed of the robot. These are the features a gaze or face following robot should have. In this experiment, we observed the efficiency of these features in our robot. We found satisfactory results with our robot.

## 6. CONCLUSION

We have developed a social robotic head which is able to follow human gaze with non-verbal behavior. The system is based on a FaceAPI algorithm from Seeing Machines Technology for face detection which also performs a face tracking task using a web camera mounted on a pan-tilt module. We are capable to mount pan-tilt servo movement smoothly with the help of behavior control module. In future, If we extent our project with 3D face animation of computer graphic on our active robot head with the help of LED Projector (3M Mpro 120 Projector), we shall be able to replace our system into retro-projected robotic face fully. We evaluate our robot in both simulated and real environment. The overall experimental result including subjective and quantitative evaluation shows that the project is functioning quite well. We tried to provide the most appropriate response from the robot which is gassed by the user. The feedback and response from the user end is satisfactory. So, the system has encouraged us to further development.

## 7. ACKOWLEDGEMENT

This work is supported by Directorate of Research & Extension, Chittagong University of Engineering & Technology (CUET), Chittagong, Bangladesh.

## 8. REFERENCES

[1] M.M Hoque, D. Das, T. Onuki, Y. Kobayashi, and Yoshinori Kuno, "Robotic System Controlling Target Human's Attention" in ICIC, 2012, Huangshan, China.

[2] B. Pierce, T. Kuratate, C.Vogl, and G.Cheng ," Mask-Bot 2i: An active customisable Robotic Head with Interchangeable Face" in Business Innovation Center Osaka, Japan: 12th IEEE-RAS International Conference on Humanoid Robots, 2012.

[3] F. Delaunay, J. De Greeff, T. Belpaeme, "Towards Retro-projected Robot Faces: an Alternative to Mechatronic and Android Faces" in Robot and Human Interactive Communication, RO-MAN: The 18th IEEE International Symposium, IEEE, 2009, Toyama.

[4] J. Oh, D. Hanson, W. Kim, Y. Han, J.-y Kim, and I. Park, Design of Android type Humanoid Robot Albert HUBO in 2006 IEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, Oct. 2006, pp. 148-1433.

[5] P. Marek, S. Michalowski, S. Sabanovic, R. Simmons, "A Spatial Model of Engagement for a Social Robot" in Advanced Motion Control: 9th IEEE International Workshop, IEEE, 2006, Istanbul.

[6] H. Lewis, and Ellis, "How to detect a face: A survey of the psychological evidence", International Journal of Imaging Systems and Technology, no. 13, pp. 3-7, 2003.

[7] R. Beira, M. Lopes, M. Praca, J. Santos-Victor, A. Bernardino, G. Metta, F. Becchi, R. saltaren, "Design of the Robot-Cub (iCub) Head" in IEEE International Conference on Robotics and Automation, 2006.

[8] Y. Yoshikawa, K. Shinozawa, H. Ishiguro, N. Hagita,

and T. Miyamoto, "Responsive robot gaze to interaction partner" in Proceedings of robotics: Science and systems, 2006.

[9] D. Miyauchi, A. Nakamura, and K.Yoshinori, "Bidirectional eye contact for human robot communication", IEICE transactions on information and system, vol. 88. no. 11, pp. 2509-2516, 2005.

[10] K. Fukui and O. Yamaguch, "Facial feature point extraction method based on combination of shape extraction and pattern maching", System and Computers in Japan, vol. 29, no. 6, pp.49-58, 1998.

[11] T. Kuratate, M. Riley, B. Perce, and G. Cheng, "Gender identification bias induced with texture images on a life size retro-projected face screen" in RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication., IEEE, 2012, pp. 43-48.

[12] A. Maejima, T.Kuratate, B. Pierce, S. Morishima, and G.Cheng, "Automatic Face Replacement for a Humanaoid Robot with 3D Face Shape Display" in Business Innovation Center Osaka, Japan :12th IEEE-RAS International Conference on Humanoid Robots , IEEE, 2012.

[13] Seeing Machines, "faceAPI"[Online], Available:http://www.seeingmachines.com/product/faceapi/.

[14] T. Vatahska, Maren Bennewitz, and Sven Behnke, "Feature-based Head Pose Estimation from Images" in Humanoid Robots: 7th IEEE-RAS International Conference, IEEE, 2007, Pittsburgh, PA.

[15] D. DeMenthon and L. S. Davis, "Model-Based Object Pose in 25 Lines of Code", International Journal of Computer Vision, 15, pp. 123-141, June 1995.

[16] Q. Ji, R. Hu, "3D Face pose estimation and tracking from a monocular camera" in ELSEVIER (www.elsevier.com/locate/imavis), Image and Vision Computing, Vol. 20, no. 7, 2002, pp. 499–511.

[17] P. Velez, A. Ferreiro, "social robotic in therapies to improve children's attentional capacities" in Review of the Air Force Academy, Ecuador, 2014.

[18] S. Satake, T. Kanda, Dylan F. Glas, M. Imai, H. Ishiguro, N. Hagita, "How to Approach Humans? Strategies for Social Robots to Initiate Interaction" in Human-Robot Interaction (HRI): 4th ACM/IEEE International Conference, IEEE, 2009, La Jolla, CA.

[19] B. Mutlu, T. Shiwa, T. Kanda, H. Ishiguro, N. Hagita, "Footing in Human-Robot Conversation: How Robots Might Shape Participant Roles Using Gaze Cues", HRI, ACM/IEEE international conference, 2009.