HRI for Interactive Humanoid Head Amir-II for Visual Tracking and Servoing of Human Face

Aseef Iqbal *aseef.iqbal@gmail.com*
Department of Mechatronics Engineering
Faculty of Engineering
International Islamic University Malaysia
Off Jalan Gombak, Kuala Lumpur, 53100
Malaysia

Amir A Shafie *aashafie@iium.edu.my*
Department of Mechatronics Engineering
Faculty of Engineering
International Islamic University Malaysia
Off Jalan Gombak, Kuala Lumpur, 53100
Malaysia

Md Raisuddin Khan *raisuddin@iium.edu.my*
Department of Mechatronics Engineering
Faculty of Engineering
International Islamic University Malaysia
Off Jalan Gombak, Kuala Lumpur, 53100
Malaysia

M Farid Alias *mfarid.alias@gmail.com*
Department of Mechatronics Engineering
Faculty of Engineering
International Islamic University Malaysia
Off Jalan Gombak, Kuala Lumpur, 53100
Malaysia

Jamil Radhi *en_radhi@hotmail.com*
Department of Mechatronics Engineering
Faculty of Engineering
International Islamic University Malaysia
Off Jalan Gombak, Kuala Lumpur, 53100
Malaysia

Abstract

In this paper, we describe the HRI (Human-Robot Interaction) system developed to operate a humanoid robot head capable of visual tracking and servoing of human face through image processing. The robotic humanoid head named Amir-II, equipped with a camera and servoing mechanism is used as the platform. The Amir-II tracks the human face within the field-of-vision (FOV) while the servoing mechanism ensures the detected human face remains at the center of its FOV. The algorithm developed in this research utilizes the capability offered by scientific computing program MATLAB along with its Image Processing Toolbox. The algorithm basically compares the locations of the face in the image plane that is detected from the static face image captured from real-time video stream. The calculated difference is then used to produce appropriate motion command for the servo mechanism to keep track of the human face moving within the range of its FOV.

Keywords: Humanoid Head, Human-Robot Interaction (HRI), Emotional Expression, Face Detection, Visual Servoing, SMQT, Split-up SNoW Classifier, Matlab, Image Processing.

1. INTRODUCTION

As robots have been predicted to become part of our everyday life, there has been a significant number of active research in the area of Human-Robot Interaction (HRI) for socially interactive humanoid robots. Among the innovative methods proposed includes Michalowski [1] that shows a rhythmic movement technique to engage a robot with human for an effective interaction. Cynthia [2] and Rosalind [3] suggests that HRI for applications like socially interactive machines can be very effective if it can exchange emotional expressions with the human counterpart. To date, robots have been studied in a variety of therapeutic application domains, ranging from using robots as exercise partners, using robots in pediatrics, robots as pets for children and elderly people, and robots in autism therapy. Researchers have developed robots engaged in social interaction with human using various modes of communication. Robots such as Paro [4], Robota [5], Keepon [6], Infanoid [7], Kismet [8] have been used successfully to emotionally engage with human very effectively via speech, vision, touch etc. as the channel for interaction.

The most common way of expressing emotional state of a human is via facial expression augmented with verbal cues and physical gestures. Some of the significant works in analyzing facial expressions are presented in [9-11]. Robots like Buddy [12], Kobian [13] and Kismet are developed as research platforms capable of displaying emotions through facial expressions towards their human operators.

An important step towards developing an emotionally responsive and intelligent robot is the capability of using visual cue as an input and analyzing it when interacting with human operator. In this type of application, capability of detecting and tracking a human face from video sequences is necessary. But locating and tracking of human face from visual input is particularly challenging because human faces have a very high degree of variability in terms of pose, scale and significant facial features.

This paper discusses the techniques used in the humanoid head AMIR-II [14] for detecting and tracking human face as a part of its HRI design. The following section discusses on the evolution of the robotic head AMIR-II. In section 3, the development of the graphical user interface (GUI) to operate AMIR-II is discussed with brief details on its different functional modules. Section 4 elaborates on the techniques used for face detection that also includes the results of using the techniques adopted and improvements achieved. Servoing and tracking of the detected face is explained in section 5. In section 6, some experimental results are presented to display the capacity of AMIR-II with the present implementation. The concluding remarks in section 7 summarize the achievements and scopes of further improvements.

2. AMIR: THE HUMANOID HEAD

The first prototype of the robotic head, named AMIR-I [15, 16], had Basic Stamp 2 microcontroller at its heart. The controller was linked to 17 parallax servo motors connected with different parts of the mechanical structure. The aim of AMIR-I was to head-start into this emerging field of research and create a test bed for development and iterative improvement towards developing an interactive and facially expressive humanoid head. AMIR – I was capable of displaying only 3 basic emotions and valid head movements (pan-tilt) with its limited Degree-of-Freedom. AMIR-I had a PING))) ultrasonic sensor attached for identifying the presence of any operator in front and was only a platform to initiate the research on developing an effective human-robot interaction system.

AMIR-II in figure 1 is an improvement over previous prototype replacing all the electronics and some minor changes in mechanism inside AMIR-I. Amir-II is capable of producing 5 different facial expressions, i.e. neutral, happy, angry, sad and disgust (figure 2). The facial expressions conveyed through facial features as the mouth shape together with the positioning of the eyebrows and the eyelids. A major upgrade is inclusion of a vision system for visual feedback to the system. The mechanical structure of Amir-II consists of 9 degrees of freedom (DOFs), in which 2 DOFs are for its neck (pan-tilt), 3 DOFs for its mouth, 1 DOF for each eyelids and 1 DOF for each eyebrow. In Amir-II, the parallax servo motors are replaced with more capable Dynamixel AX-12+ smart servo motors from Robotis. These motors have very unique characteristics. The AX-12+ robot servo has the ability to track its speed, temperature, shaft position, voltage, and load. Features also include 300 degree of movements in 1024 increments, 1,000 kbps communication speed in half-duplex mode and a huge 16.5 kg-cm holding torque at 12V operating voltage etc. – all in roughly the same size of a standard servo. These servo motor are controlled by a PC with USB2DYNAMIXEL – an interface between the PC and AX-12+ via high-speed USB2.0 port. Matlab and Image Processing Toolbox 2.0 was used to develop the controller for the system. Use of Matlab made the system integration very easy and allowed us to concentrate more on implementing efficient algorithm and rapid development of Graphical User Interface (GUI) [9] for an effective HRI for AMIR-II.

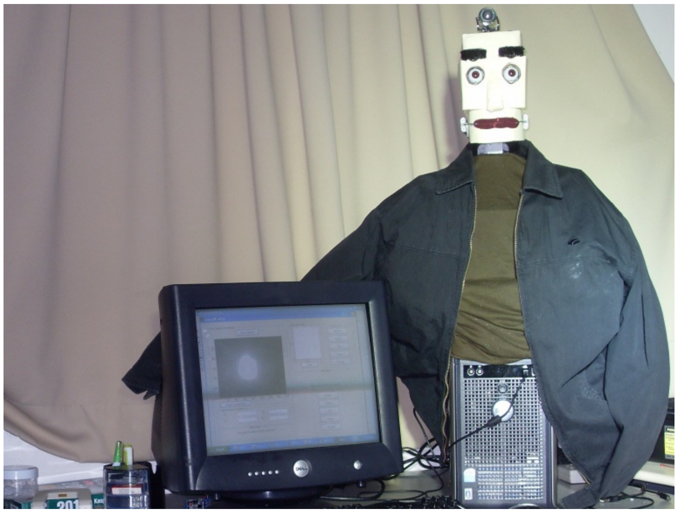


FIGURE 1: The robotic head AMIR-II

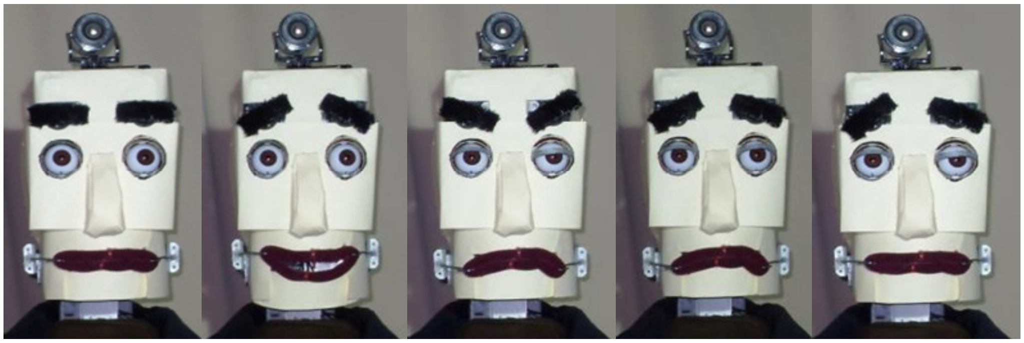


FIGURE 2: Five different facial expressions by Amir-II. From left to right: neutral, happy, angry, sad, disgusted.

3. INTERFACE DESIGN

Panel Design

The GUI program serves as the control and monitoring tool for the operation of Amir-II. In figure 3, the GUI contains 3 panels in which each panel carries its own functionality as follows:

1) *Live Camera Streaming Panel:* The Live Camera Streaming panel allows the user to monitor the live operation of the robot Amir-II operator can view the live image streaming from the camera by clicking the Start/Stop button within this GUI panel. The face detection module will automatically display a red bounding box as a human face is detected within the field of view (FOV) of the

camera. Meanwhile, the message box at the bottom of the panel will display the status of the USB2Dynamixel connection to the PC.

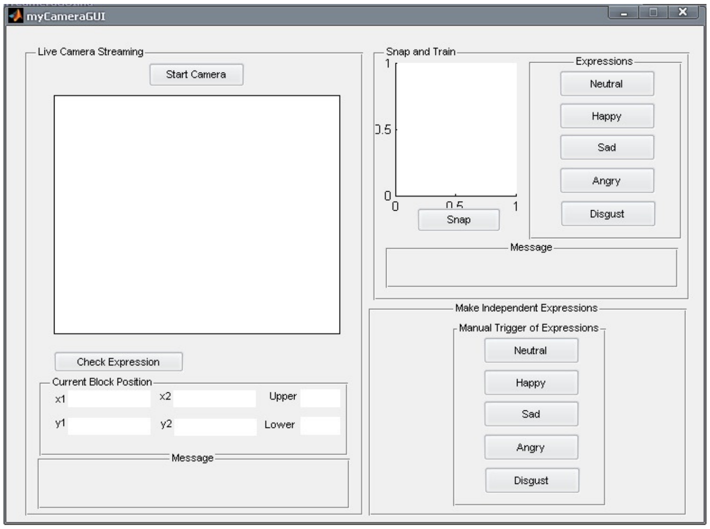


FIGURE 3: The graphical user interface for Amir-II

2) Snap Panel: The Snap panel provides the capturing, labeling and saving of the facial expression images into a database. The snap segment contains a snap button, a screenshot window of captured image and 5 buttons of the respective facial expressions. As the Snap button is clicked, the detected face area within the bounding box will be captured and displayed inside the screenshot window. After that the operator can click the related expression button to save the image and its expression label into the database.

3) Manual Trigger of Expressions: The Manual Trigger of Expression panel lets the operator to trigger the respective facial expressions for Amir-II. To demonstrate the capability of Amir-II to produce facial expression, its operator can select which expression is to be produced at a time by clicking the related expression button within this panel. Such manual trigger of expression is also useful in inspecting the condition of Amir-II mechanical structure.

For its back-end, the GUI program consists of several modules as shown in figure. 4. The functionality of each program module is described as follows:

1) Image Acquisition Module: In order to reduce the complexity of processing the image, the original RGB streaming image is converted to its grayscale format. The live streaming image is being displayed based on the default resolution of the camera used. With camera resolution set to its lowest acceptable value (at 120 pixels x 160 pixels), the image frame rate becomes near 30 frame per second.

2) Face Detection Module: The face detection module uses SNoW classifier algorithm [14] to detect a human face within the camera FOV. The live detected face area is marked with a bounding box on the image streaming window. High frame rate is crucial for the face to be kept tracked since it can consistently appear on the camera FOV in real time.

3) Face Tracking Module: Once the face is detected, the face tracking module requests the servo callback functions to trigger the servo movements. There is an acceptable boundary (AB) defined for a tracked face to move around within the visual input. As the tracked face moves beyond the AB, the servo callback function is called to send the related commands to the servos through USB2Dynamixel device [14]. This process involves the repositioning the neck servos through their pan-tilt movements to refocus the tracked face back within the AB. The location of the face within the image will determine the required direction and magnitude of rotation for the neck servos.

4) Face Image Capture Module: This module is the one functioning for the Snap Panel. The database directory and the method of labeling the expression image data are determined here.

5) Facial Expression Recognition Module: To generate the proper facial expression by Amir-II, the human facial expression within the camera FOV is first recognized. This information is needed in the decision-making process for human-robot interaction, to produce appropriate expression in response, based on the chosen behavior model.

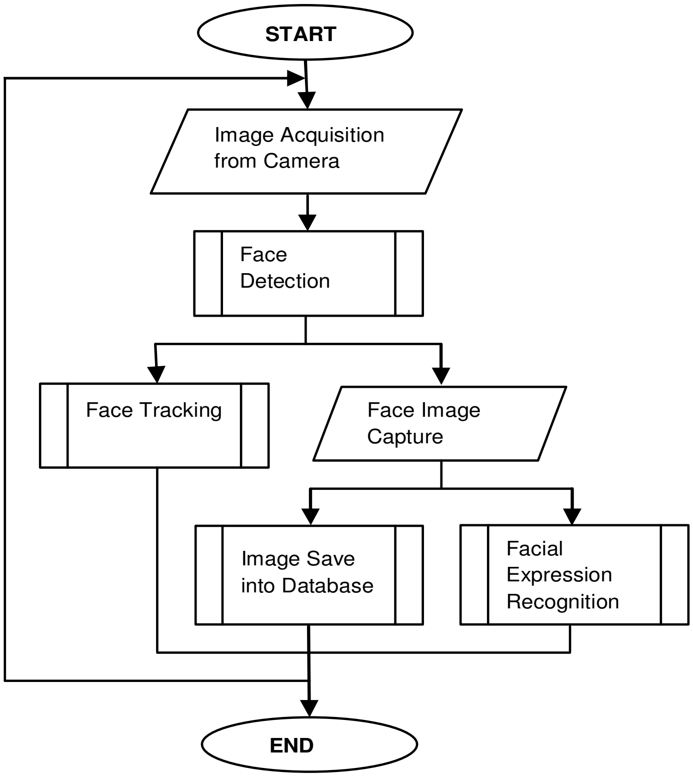


FIGURE 4: Flowchart of the interfacing program for AMIR-II

4. FACE DETECTION METHOD FOR AMIR-II

The foremost task of the system in terms of functionality is to identify a human face from the scene that is *face detection* or *face segmentation*. Functionality of face tracking and, as a whole, the success of entire system depends on the accuracy of the human face detection.

Initial face segmentation techniques were only capable of detecting single frontal-face from image with simple uncluttered background using neural network, template-matching or skin color properties [17]. Recent advancements in technology allowed researchers to attempt more computing-intensive techniques such as appearance-based or optical methods to increase the detection rate. Some of the established face-detection systems are eignefaces [18], which implements Principal Component Analysis (PCA), Fisherfaces using Linear Discriminant Analysis [19], Bayesian method using probabilistic distance metric [20], etc. These techniques proved to be efficient in segmenting multiple human faces even with partial occlusion and complex, cluttered background images.

Segmentation of moving faces from a video sequence requires a different approach. One method is detecting face in single frame of the video with any of the techniques applied on static image. Subsequent video frames are then compared using pixel-based change detection procedures based on difference images. More recent methods use optical-flow techniques for detecting human face from the video. They extract the color information from the video frames to identify possible location of human face in the image-plane. These methods can also be applied for segmenting multiple faces.

Face Detection in AMIR-II using skin color

In AMIR-II, we initially used skin color to find the face from an image [21]. Although there are people from different ethnicities, the color distribution of skin is relatively clustered together in a specific area of an image. Our algorithm captured an image frame from live video feed via camera in RGB color-space, converted it into HSV color-space and used a combination of Hue and saturation value to correctly identify the face skin from the captured frame. The assumption is that the facial skin is the only part of the body skin exposed within the field-of-view of the camera. The results are shown in figure 5 (a).

Even though this method is very easy to implement to cluster out skin in finding the face, the obvious problem comes when there is any other object in the scene which contains similar color tone. Variation of illumination also poses a great challenge for this method to work properly. Problems due to these limitations can be observed from figure 5(b).

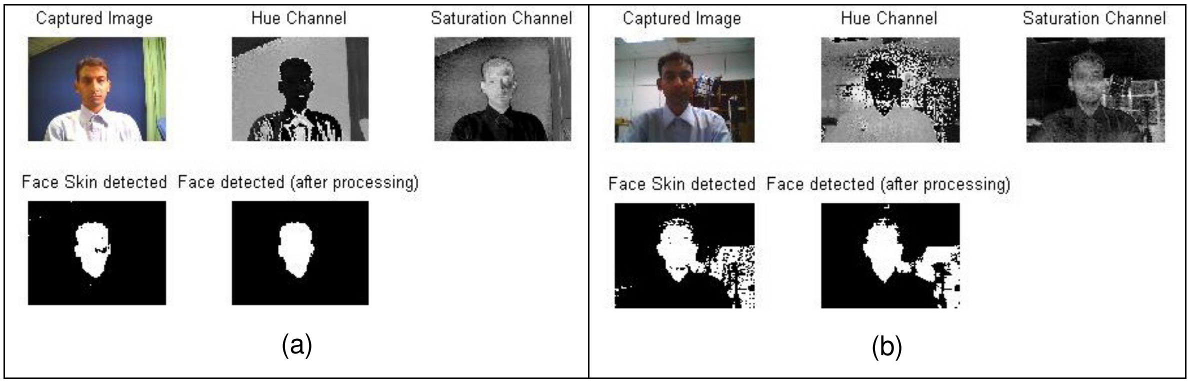


FIGURE 5: (a) Different stages of processing the input video frame to identify a human face using skin color information. (b) Error in face detection using skin color picked up some part of a bookshelf as part of detected face.

Improved Face Detection Using Appearance-based Method

To overcome the limitations of previous method, we adopted a new face detection technique using local Successive Mean Quantization Transform (SMQT) and the split-up Sparse Network of Winnows (SNoW) classifier [22]. Local SMQT features are used to extract the illumination-insensitive properties from image data. Split up SNoW classifier is an appearance-based pattern recognition method that can be utilized to find face object. Results of applying this method can be observed in figure 6.

5. VISUAL SERVOING AND TRACKING OF HUMAN FACE

Visual Servoing is the way of using vision data to control the motion of a robot while target tracking refers to constantly following a moving target body and adapt its own motion to maintain the target within its observation range. Visual servoing and tracking of an object, as the name implies, exploits computer vision for its input, manipulates the input data using different image processing techniques to convert it into an acceptable form to the system, and finally utilizes this information to control its own motion so that the target object remains within its field-of-vision. The target object in this case is the human face.

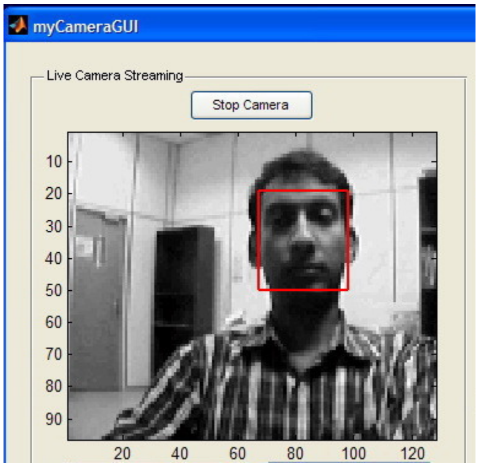


FIGURE 6: Face Detection of AMIR-II using local SMQT and split-up SNoW classifier.

As discussed in [23], visual servoing techniques can be categorized in two broad classes – Image-Based Visual Servo control (IBVS) and Position-Based Visual Servo (PBVS) control. In IBVS control, the parameters that are required for the system to control its motion are immediately available from the image data. In PBVS control, a set of intermediate 3D parameters are computed from the image measurement which would be used for control system.

Image-based visual tracking of human face can take two approaches – a) head tracking, tracking the motion of head as a rigid object [24], b) facial-features tracking, tracking the deformations of shape of the facial features, i.e. eyes, nose, lips, confined within the anatomical head region [25].

For AMIR-II, we implemented Image-Based Visual Servoing and Tracking of human head as a whole which also contains face. Mathematically, the process can be described as minimization of an error **e(t)**, where

$$e(t) = s[m(t),\ a] - s^{*} \qquad (1)$$

Here, ***s*** is the set of image-plane coordinates of the points within the FOV, ***m*** are the pixel coordinates constructing the box encircling the face detected (figure 4), and ***a*** is a set of camera dependant parameters which have to be input manually.

6. EXPERIMENTAL RESULTS AND ANALYSIS

Our system is executed on a Personal Computer with Intel Core2Duo processor running at 2.00 GHz speed, 1 GB DDR2 RAM under Microsoft Windows XP (SP3) as its operating system. The program is developed in Matlab with Image Processing Toolbox 2.0. To actuate the AX-12+ Dynamixel servo motors from the program, the supplied library file dynamixel.h from Dynamixel SDK was utilized.

Our experiments are constrained with following assumptions:

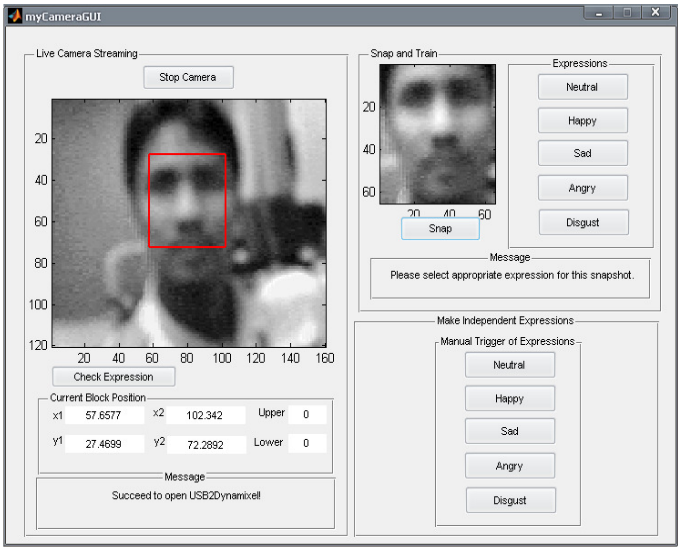
- The face should take up a significant area in the image.
- The system should be operated in indoor illumination condition.
- The face movement should be considerably slow for tracking to be effective.

The GUI program is executed with the USB camera and AX-12+ servos connected to the PC as seen in figure 7. Firstly the message box displays the status of the connection, i.e. "Succeeded to open the USBDynamixel", which means that the PC-servos connection is working properly. As the Start button is clicked the live image streaming starts and automatically the program detects the available human face within the camera FOV.

At the same time, the detected face is being tracked automatically by Amir-II through the movements of its neck joints. Whenever the tracked face moves beyond the AB, Amir-II will adjust its orientation accordingly so that the tracked face can be refocused back into the AB. However, if the tracked face moves too fast, Amir-II would lose its focus towards the face and revert back to the natural position of its neck. To deal with this condition, its face tracking module needs to be further developed in terms of its algorithm to become more adaptive.

To capture the snapshot of the face image, the Snap button is clicked by the robot operator. Satisfied by the human facial expression and the image quality, the relevant facial expression label for the snapshot is then clicked. The image file will be saved into the database with the related expression as part of its filename.

For face detection, we executed some experiments to identify the range of deviation of face view (frontal / partial), pose and the maximum distance from the robot vision system.



Current Block Position

x1 57.6577 x2 102.342 Upper 0

y1 27.4699 y2 72.2892 Lower 0

Message

Succeed to open USB2Dynamixel!

Message

Please select appropriate expression for this snapshot.

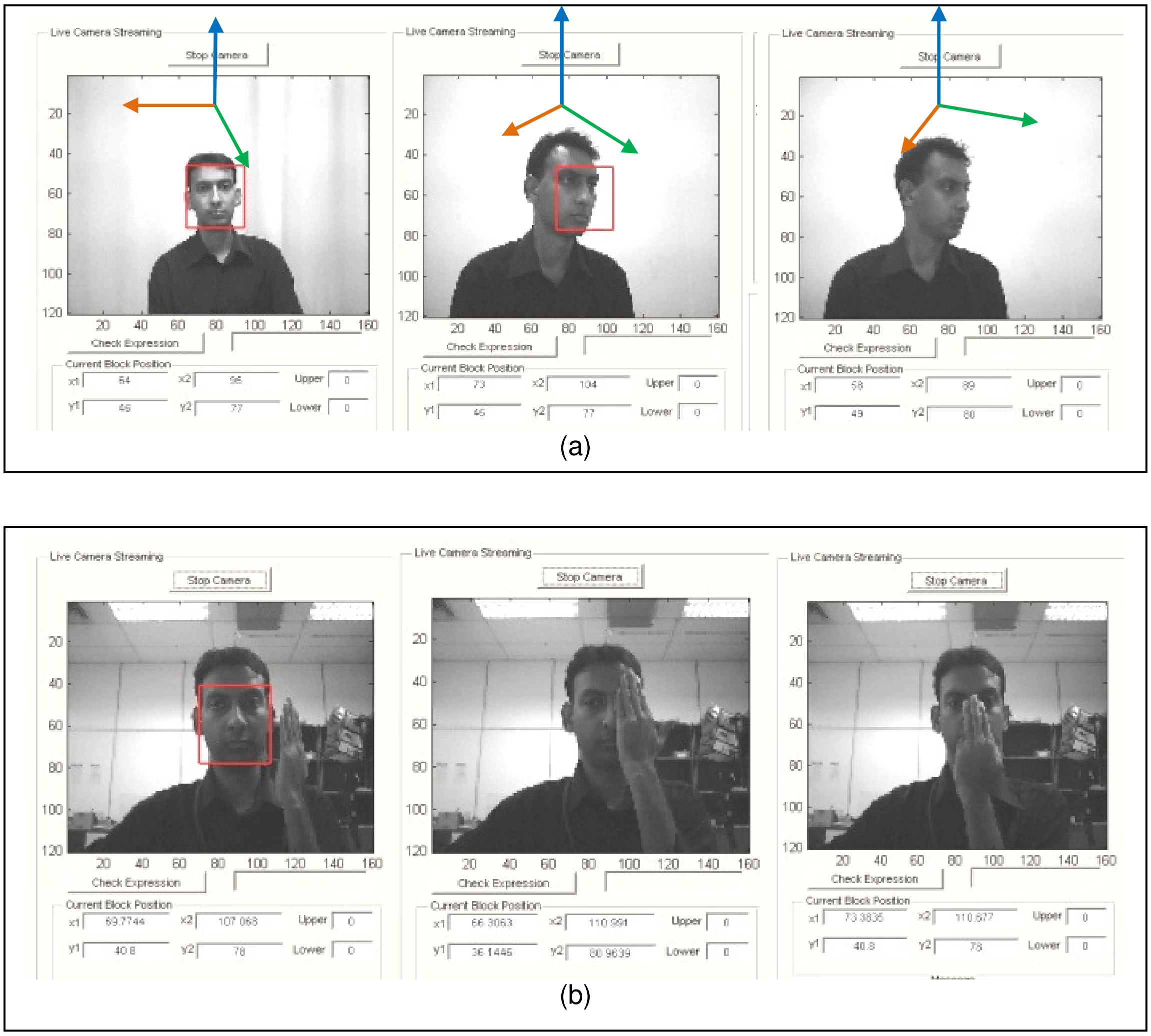
FIGURE 7: The operational GUI

Our face detector could properly identify a face in figure 8(a) as long as significant facial features (nose, two eyes etc) remain visible within the FOV of the camera. In our experiment, the system could detect face correctly with the face rotated by 60º- 65º about vertical axis, depending on how much the face occupies in the entire image. The system cannot detect faces rotated more than 65º about the vertical axis passing through the head.

Even in frontal view of the face, the detector could not identify the face when part of the face was obstructed in figure 8(b). This is because important facial features were missing from their expected symmetric position and hence in the principal component of the image, which was input to the classifier to find face in input image.

Similar limitation was also observed in the pose of frontal face view. The detector can find the face when the head is much deviated from it vertical orientation. In our experiment, the system could not identify a face when it was rotated more than 20º about the horizontal axis as seen in figure 9(a).

The face detector also needs the person whose face is to be detected and tracked be occupying at least 5% of the entire image frame. Beyond this, the size of face image compared to the entire image becomes small enough making the system fail to detect the face as some important facial features lose the details required. This is evident from figure 9(b).



x1 64 x2 96 Upper 0
y1 46 y2 77 Lower 0

x1 73 x2 104 Upper 0
y1 46 y2 77 Lower 0

x1 58 x2 89 Upper 0
y1 49 y2 80 Lower 0

(a)

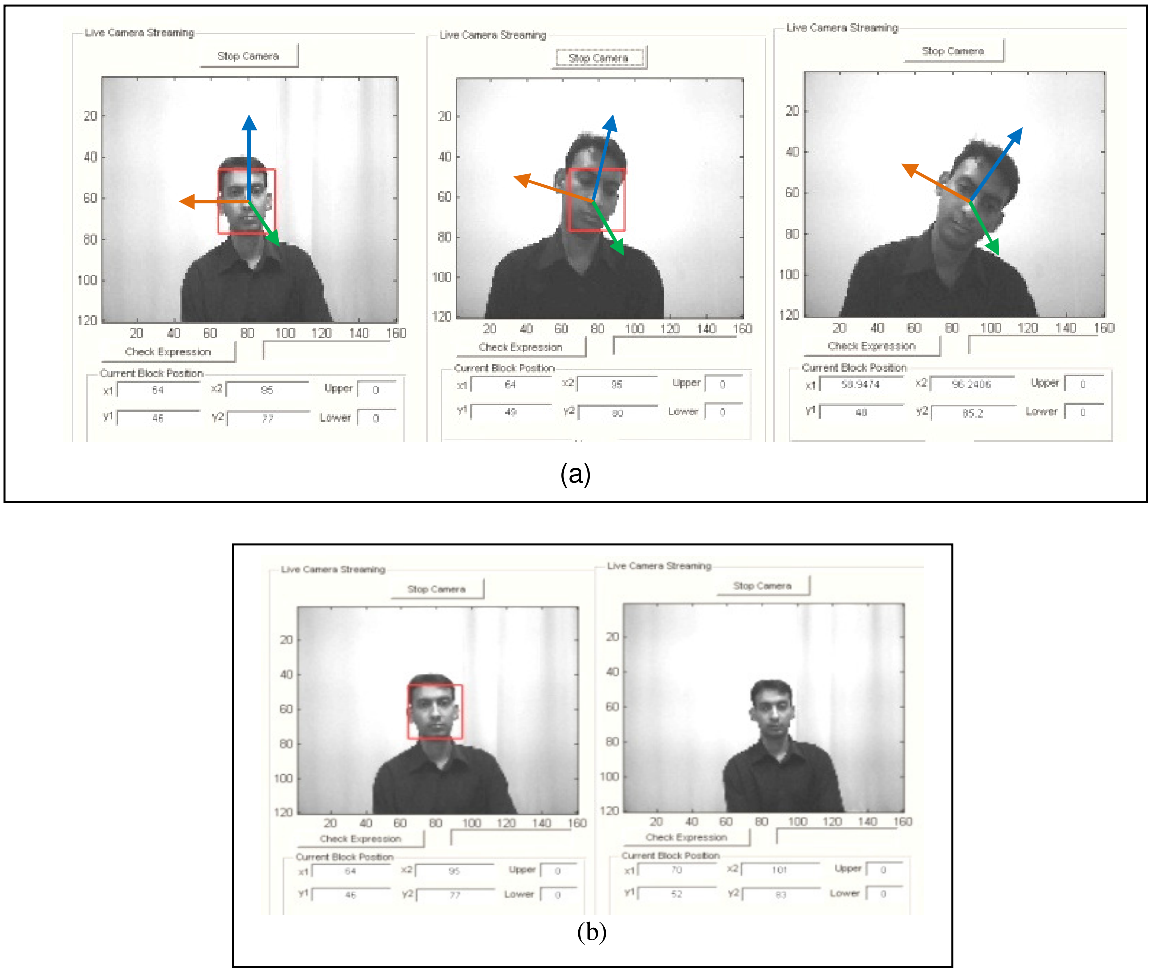
x1 69.7744 x2 107.068 Upper 0
y1 40.8 y2 78 Lower 0

x1 66.3063 x2 110.991 Upper 0
y1 36.1445 y2 80.9639 Lower 0

x1 73.3835 x2 110.677 Upper 0
y1 40.8 y2 78 Lower 0

(b)

FIGURE 8: (a) Detection of face partially occluded due to rotation in an image, (b) Face detection failed with partial obstruction.



x1 64 x2 95 Upper 0
y1 46 y2 77 Lower 0

x1 64 x2 95 Upper 0
y1 49 y2 80 Lower 0

x1 58.9474 x2 96.2406 Upper 0
y1 48 y2 85.2 Lower 0

(a)

x1 64 x2 95 Upper 0
y1 46 y2 77 Lower 0

x1 70 x2 101 Upper 0
y1 52 y2 83 Lower 0

(b)

FIGURE 9: (a) System detecting the face with limited change in pose, (b) the face was not detected even with its frontal view with no occlusion and vertical pose, because the ratio of the face to the entire image is high.

As for servoing and tracking, AMIR-II can track a face nicely as long as the face is detected. One of the significant observations is when more than one face is detected and they have different motions in different directions, AMIR-II vision system can track the faces successfully as long as they remain within the FOV of camera but cannot generate its motion for servoing as it does not have any information regarding which face to follow.

7. CONCLUSION AND FUTURE DIRECTION

The servoing and tracking of human face by AMIR-II enabled us to experiment new techniques for further development of the system. Current detection rate of the face detector of AMIR-II is acceptable according to recent findings where the system was able to successfully track human face as a single object. However, enabling the system to track the facial-features would make it more interactive as the system could be upgraded to recognize and track the facial expressions. The upgrade is already in progress for this project.

Range of servoing of tracked human face could also be improved. Also, the tracking actuation at the moment only involve servoing the motors from point to point so the tracked object remains within its field-of-view. The process could be further improved implementing PID within the loop to reduce the jitter in motion.

The next step forward to progress with this research is to use the face matrix for extracting facial features to identify the emotional states of the human.

8. ACKNOWLEDGEMENT

This work is funded by the Fundamental Research Grant Scheme (FRGS) of the Ministry of Higher Education Malaysia.

9. REFERENCES

[1] H. K. Marek P. Michalowski, "Rhythm in human-robot social interaction," *IEEE Intelligent Systems,* vol. 23, pp. 78-80, 2008.

[2] C. Breazeal, "Socially intelligent robots," *interactions,* vol. 12, pp. 19-22, 2005.

[3] W. P. Rosalind, "Affective computing: challenges," *Int. J. Hum.-Comput. Stud.,* vol. 59, pp. 55-64, 2003.

[4] K. W. Takanori Shibata, Tomoko Saito, Kazuo Tanie "Human Interactive Robot for Psychological Enrichment and Therapy," in *Social Intelligence and Interaction in Animals, Robots and Agents (AISB) 2005*, University of Hertfordshire, Hatfield, England, 2005, pp. 98-107.

[5] A. Billard, B. Robins, K. Dautenhahn, and J. Nadel, "Building Robota, a Mini-Humanoid Robot for the Rehabilitation of Children with Autism," *the RESNA Assistive Technology Journal,* vol. 19, 2006 2006.

[6] H. Kozima, M. Michalowski, and C. Nakagawa, "Keepon - A Playful Robot for Research, Therapy, and Entertainment," *International Journal of Social Robotics,* vol. 1, pp. 3-18, 2009.

[7] H. Kozima, "Infanoid: An experimental tool for developmental psycho-robotics," *International Workshop on Developmental Study,* 2000.

[8] C. L. Breazeal, *Designing sociable robots*. Cambridge, Mass.: MIT Press, 2002.

[9] P. Maja and B. Marian Stewart, "Machine Analysis of Facial Expressions," in *Face Recognition*, Kresimir Delac and M. Grgic, Eds., First Edition ed Viena, Austria: I-Tech Education and Publishing, 2007, pp. 377-416.

[10] Y. I. Tian, T. Kanade, and J. F. Cohn, "Facial Expression Analysis," in *Handbook of Face Recognition*, K. J. Anil and Z. L. Stan, Eds., ed: Springer, 2005, pp. 247-276.

[11] M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: machine learning and application to spontaneous behavior," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, pp. 568-573 vol. 2.

[12] M. S. J. K. G. Oh, S.-J. Kim and S. S. Park, "Function and driving mechanism for face robot, buddy," *The Journal of Korea Robotics Society,* vol. 3, pp. 270–277, 2008.

[13] M. Zecca, Y. Mizoguchi, K. Endo, F. Iida, Y. Kawabata, N. Endo, K. Itoh, and A. Takanishi, "Whole body emotion expressions for KOBIAN humanoid robot - preliminary experiments with different Emotional patterns - " in *Robot and Human Interactive Communication, 2009. RO-MAN 2009. The 18th IEEE International Symposium on*, 2009, pp. 381-386.

[14] A. A. I. Shafie, A. Khan, M.R., "Visual tracking and servoing of human face for robotic head Amir-II," in *International Conference on Computer and Communication Engineering (ICCCE), 2010*, Kuala Lumpur, Malaysia, 2010, pp. 1-4.

[15] K. M. Shafie A., "Design and Development of Humanoid Head," in *Proceeding of International Conference of Man Machine System (ICOMMS)*, Langkawi, Malaysia, 2006.

[16] M. N. K. a. s. A. A. Shafie. , N. I. Taufik Y., "Humanoid Robot Head," in *3rd International Conference on Mechatronics (ICOM)*, Kuala Lumpur, Malaysia, 2008.

[17] C. R. Zhao Wenyi, "A Guided Tour of Face Processing," in *Face Processing: Advanced Modeling and Methods*, ed: Academic Press, 2006, pp. 3-53.

[18] T. Matthew and P. Alex, "Eigenfaces for recognition," *J. Cognitive Neuroscience,* vol. 3, pp. 71-86, 1991.

[19] W. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant analysis of principal components for face recognition," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, 1998, pp. 336-341.

[20] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 19, pp. 696-710, 1997.

[21] Y. Yang, S. Ge, T. Lee, and C. Wang, "Facial expression recognition and tracking for intelligent human-robot interaction," *Intelligent Service Robotics,* vol. 1, pp. 143-157, 2008.

[22] M. Nilsson, J. Nordberg, and I. Claesson, "Face Detection using Local SMQT Features and Split up Snow Classifier," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, 2007, pp. II-589-II-592.

[23] F. Chaumette and S. Hutchinson, "Visual Servoing and Visual Tracking," in *Springer Handbook of Robotics*, ed, 2008, pp. 563-583.

[24] A. Azarbayejani, T. Starner, B. Horowitz, and A. Pentland, "Visually controlled graphics," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 15, pp. 602-605, 1993.

[25] D. Terzopoulos and K. Waters, "Analysis and synthesis of facial image sequences using physical and anatomical models," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 15, pp. 569-579, 1993.