Eye Tracking for Human-robot Symbiosis

Md. Al-Amin Bhuiyan

Abstract— **This paper addresses an image interpretation method for the establishment of human-robot symbiotic relationship. The system is based on visual information of human face from image sequences depending on the similarity measure of the hue components of the images in the YIQ color space and is organized with the implementation of an eye tracking system employing intelligent task scheduling for an entertainment social robot. Experimental results validate that the system is efficient and fast enough to use it in real-time application.**

***Index Terms*—Human-robot symbiosis, face detection, eye tracking, gaze direction, human-robot interface, AIBO**

I. Introduction

SYMBIOSIS is a sustained relationship between two different animals or plants of different species for the benefit of each other [1]. Recently this biological term is employed in robotics research to establish a symbiotic relationship between robots and human beings for their coexistence and cooperative work. Since humans can indicate their intentions and interests over gaze direction, eye tracking has, therefore, occupied an important role for visual perception and non-verbal information to instruct the robots. Eye tracking is, therefore, immerging tremendous interest in the advancement of human-robot interaction since it provides a natural and efficient way of representing expressions.

Eye tracking, a method of interacting with the computer, is primarily used for research into various fields such as driving, reading and learning patterns. In [2] the author shows how the investigation of eye movements can lead to the inference of underlying cognitive processes. In [3], the authors used five near-infrared light as the source of corneal reflex and computed the location of the sight and a rough estimate of the center of the pupil.

For the last couple of decades, different approaches [4-9] have been cited in literature on eye tracking and gaze estimation. Most of the methods of eye tracking use infrared lighting to illuminate the eye employing high speed and high resolution camera, organized lighting source and hardware equipment. In some approaches to eye tracking it is necessary to identify the Region of Interest (ROI). Baluja et. al [5,6] have proposed a neural network based non-intrusive technique to estimate the eye-gaze. This neural network based system used backpropagation algorithm and needed 260 epochs for training 2000 images and took 30-40 minutes. Pastoor et al. have developed an eye tracking system using an infrared camera [7] which assesses the pupil position by manipulating the differences in infrared reflection between the pupil and the surrounding iris and sclera which is sensitive to the head movement. Chen and Su have developed an eye tracking system by winner-update template matching method [8]. But its major drawback is that the computational cost of the template matching is so enormous that its applicability is limited for real-time systems. Matsumoto and Zelinsky [9] have assumed the eyeball to be a sphere and constructed a 3D model of the eye to measure the head pose and gaze estimation which necessitates template matching for face.

This paper explores a vision based eye tracking system which integrates face detection in complex backgrounds and localization of eyes on it. Face detection is achieved by skin color extraction. The system is organized with visual information of the face from the image streams in YIQ color space and computation of the difference between the consecutive frames of the video streams and thresholding the resulting image. Eyes are detected and localized by the geometrical analysis of the facial skeleton.

The rest of the paper is organized as follows. Section II describes face detection process. Section III illustrates eye localization and tracking method. Section IV describes object detection and knowledge based frame. Section V highlights experimental methods and results. Finally, Section VI draws the overall conclusions.

II. Face Detection

Face detection is attained by using skin-color segmentation with the approximation of the skin region of the images depending on the similarity measure of the hue and luminance attributes in the YIQ color model. Therefore, images are being searched in YIQ space depending on the amount of color content of these perceptually relevant colors, that is, whether the skin color value is substantially present in an image or not. The flowchart of the face detection process is shown in Fig. 1.

In the YIQ color space a color is represented by three components: luminance, hue and saturation. The YIQ provides a linear transformation of RGB as expressed by the following equation [10,11]:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.320 \\ 0.212 & -0.523 & 0.311 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

where *R,G, B* represent the red, green, and blue components which exist in the range [0,255], Y implies luminance channel and I,Q imply two chrominance channels to convey color information.

The pixels representing skin colors are thresholded empirically with the following equation:

Md. Al-Amin Bhuiyan is with King Faisal University, Al Ahsa 31982, KSA (corresponding author phone: 966-055-1973929; fax: 303-555-5555; e-mail: mbhuiyanr@kfu.edu.sa).

ISBN: 978-988-19253-4-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

$$(60 < Y < 200) \ \ and \ \ (20 < I < 50) \qquad (2)$$

The detection of skin regions by such hue segmentation process is shown in Fig. 2. The precise position of the face is then localized from the image with largest connected component analysis [11].

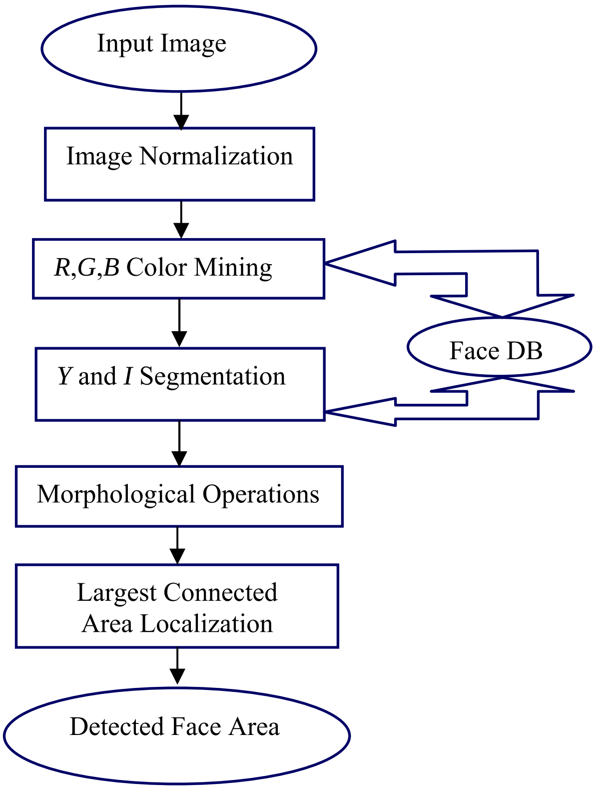

Fig. 1. Block diagram of the face detection method.

(a) Facial image.

(a) Skin color extraction.

(c) Largest component extraction.

Fig. 2. Detection of face region by skin color segmentation.

III. Eye Localization and Tracking

After localizing and segmenting the face area, the exact locations of the eye positions are searched by connected component analysis. For this, the original face images are exposed to some image processing operations, as shown in Fig. 3. After thresholding, the segmented image may come across by some holes in the face skin region. In order to remove these false regions, two types of morphological operations are performed: dilation and erosion [1].

The dilation of the binary image **f** by a structuring element **s**, represented by $\mathbf{f} \oplus \mathbf{s}$, is described as:

$$g(x,y) = \bigcup_{i=-\frac{n}{2}}^{\frac{n}{2}} \bigcup_{j=-\frac{m}{2}}^{\frac{m}{2}} f(x-i, y-j) \oplus s(i,j)$$

$$= \begin{cases} 1 & \text{if } s \text{ hits } f \\ 0 & \text{otherwise} \end{cases}$$

The erosion of a binary image **f** by a structuring element **s**, denoted by **f**Θ**s**, is described as:

$$g(x,y) = \bigcap_{i=-\frac{n}{2}}^{\frac{n}{2}} \bigcap_{j=-\frac{m}{2}}^{\frac{m}{2}} f(x-i, y-j) \Theta s(i,j)$$

$$= \begin{cases} 1 & \text{if } s \text{ fits } f \\ 0 & \text{otherwise} \end{cases}$$

Eyes are detected from the face profile, assigning a unique tag to each isolated candidate block by labeling the binarized image. Let $\mathbf{P}_i = (x_i, y_i), (i \in [1,2])$ be the two tag points of the image blocks of the left eye and right eye, respectively. Since in the binary image, all of the pixels have got the intensity value of 0 and 255 only, these tag points are localized with the following algorithm [11,12]:

[Step 1] Calculate the center point of the face:

$$\mathbf{P}(x,y) = (x_{\max}/2, \ y_{\max}/2)$$

[Step 2] Starting from $\mathbf{P}(x,y)$, search the first white pixel and assign it with the value of '100' (first tag value) as follows:

$$For \ \ x \to x_{\max}/2 \ \text{ to } 1 \text{ step } -1$$

$$For \ \ y \to y_{\max}/2 \ \text{ to } 1 \text{ step } -1$$

$$If \ \ \mathbf{Pixel}[x][y] == 255 \ then \ \mathbf{P}_i = 100$$

return

[Step 3] Search the entire window for the tag value '100' and assign all connected pixels with '100'.

Similarly, the 2nd tag point is determined from the top right quadrant of the searching window. Since the exact location of the pupil is very confusing, so instead of searching the pupil point, the position of the gravity center of the eyes are being determined from the black portions of the irises of the segmented image. If the gray level at each point (x, y) of the given area E of an eye is considered as the "mass", the center of gravity of eye can be determined. The pq-th moment of area E about the origin is expressed by:

ISBN: 978-988-19253-4-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

Fig. 3. Different image processing operations for eye localization.

$$m_{pq} = \sum_{(x,y) \in E}\sum x^p y^q \qquad (3)$$

where (x, y) are the coordinates of E. The center of gravity of E is the point $(\bar{x}, \bar{y})$ whose coordinates are expressed by:

$$\left.\begin{aligned} \bar{x} &= \frac{m_{10}}{m_{00}} = \frac{\sum\sum_{(x,y)\in E} xy^0}{\sum\sum_{(x,y)\in E} x^0 y^0} \\ \bar{y} &= \frac{m_{01}}{m_{00}} = \frac{\sum\sum_{(x,y)\in E} x^0 y}{\sum\sum_{(x,y)\in E} x^0 y^0} \end{aligned}\right\} \qquad (4)$$

After detecting and localizing the eye positions, the gaze directions are determined from the consecutive video frames.

M. Object Detection and Knowledge Based Conceptual Frame

The knowledge-based system is based on visual and geometrical information of the objects from the video streams and is commenced with acquiring, gathering, and disseminating the perceptual and semantic knowledge of the objects. The generalized frame-based knowledge engineering environment is employed as a software platform in implementing the human-robot symbiotic relation.

Objects are recognized from their partial information of color and shape attributes. Every candidate objects to be seen within the environment is pre-defined as a class frame in a frame hierarchy in the world model. First, the object regions are localized by means of color information such as “red”, ”green”, ”blue”, ”yellow”, ”cyan”, “black”, “pink” and so on. Then geometrical measurements are employed to obtain lines, circles, ellipses, etc. The knowledge stored in the shape model of the knowledge base, for example, the recognition of an ellipse and a couple of parallel lines in a red object, provides information of a red cup. An example of a frame description in the shape model to represent a Red-cup and a simplified image of identifying it by means of the knowledge about Cup_vessel within the shape model of the world model is shown in Fig. 4. Experiments are conducted on different shapes of objects, such as magnets (circular), balls (circular), cups (elliptical and linear), human faces (approximated as elliptical), and so on, as shown in Fig. 5.

Frame: Red-cup
A-kind-of: **Cup**
Color: **Red**
Has-parts: **Cup_vessel**
Ellipse-parameters: (x_0, y_0, a, b, θ)
Line-Parameters: $(x_i, y_i), i = [1,4]$

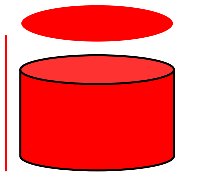

Fig. 4. Frame and image to describe a red cup.

M. Experimental Results and Performance

In order to justify the effectiveness of the proposed system, several experiments were carried out. The snapshot for the eye tracking image processing steps is shown in Fig. 5.

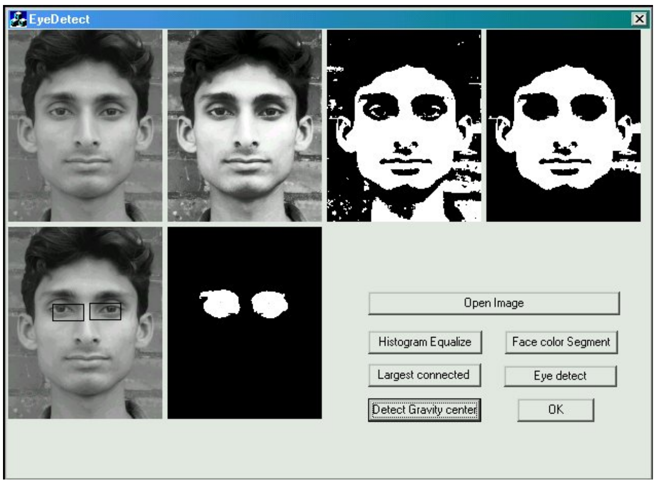

Fig. 5 Snapshot for the eye tracking system.

The accuracy of the eye tracking system has been evaluated on numerous video image sequences under varying lighting conditions. The success rate was computed by the ratio of the number of correct detection of eye in image sequences that of all the face image frames in the video. The detection rate is 98.5% for frontal view faces and 72.3% for non-frontal faces which is better than the conventional methods within certain restrictions. False detection happened during eye blinking, head movements, and morphological thresholding due to the failure of their geometrical

ISBN: 978-988-19253-4-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

information. The time-position graph for the horizontal and vertical actions of the right eye of a user is shown in Fig. 6.

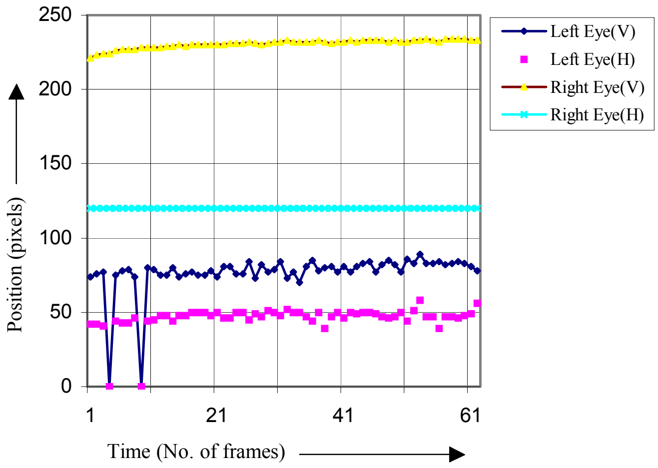

Fig. 6 The time-position graph for the horizontal and vertical actions of the right eye of a user.

It is apparent that eye blinks happened during the recording period. These are being suppressed by employing lowpass filters. The gaze detection system has been implemented for human-robot interaction. In order to implement the human-robot cooperative work, a social entertainment robot named 'Aibo' developed by Sony has been instructed through eye movements of the user in different directions. The snapshot for the human-robot interface is shown in Fig. 7.

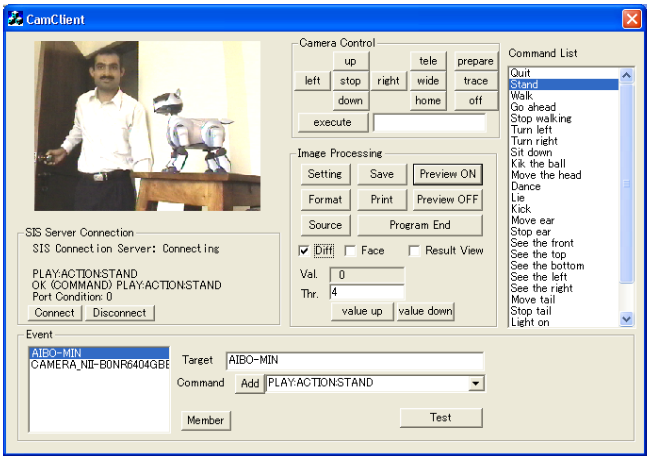

Fig. 7 Human-robot interface using face detection mode.

VI. Conclusion

A real time eye tracking system has been developed that analyses color video streams. The system uses a combination of color, edge, shape, and motion information to detect faces, and recognize and track the eye and gaze direction. Assuming a static head and the eye can only rotate in its socket, gaze points are estimated by tracking the gravity center of the user's eye. This is achieved by assuming the imaging geometry of the gravity center of the two eyes in the eye tracker co-ordinate mapping. The system has finally been implemented for instructing an entertainment robot, called AIBO.

The range of the robot's interaction with human beings is limited by many factors, including its visual perception. As new percepts are added, new dimensions of behavior are possible to include, for example face identification, facial expression, lip movement, head orientation, etc. Our main target is to make the system capable of detecting facial gestures and expressions and make the robots understand different kinds of objects and include more instructions to make them capable of grasping with natural behavior. Vision system requires prior knowledge of the images to be interpreted. Therefore, knowledge plays a significant role in image understanding. Our final goal is to provide more semantic and functional knowledge for image understanding and establish a human-robot symbiotic community.

References

[1] M.A. Bhuiyan, C.H Liu, and H. Ueno, "On Pose Estimation for Human-Robot Symbiosis", *International Journal of Advanced Robotic Systems, Vol. 5, No. 1, pp. 19-30 2008.*

[2] K. Rayner, "Eye movements in reading and information processing: 20 years of research", Psychological Bulletin (1998), 372-422.

[3] D. Gao, G. Yin, W. Cheng, and X. Feng, "Non-invasive eye tracking technology based on corneal reflex", International Workshop on Information and Electronics Engineering, Procedia Engineering, Vol. 29, pp. 3608- 3612. 2012.

[4] M.A. Bhuiyan, V. Ampornaramveth, S. Muto, and H. Ueno, "On Tracking of Eye For Human-Robot Interface", International Journal of Robotics and Automation, Vol. 19, No. 1, pp. 42-54, 2004.

[5] S. Baluja and D. Pomerleau, "Non-intrusive gaze tracking using artificial neural networks", Technical Report CMU-CS-94-102, Carnegie Mellon University, USA, 1994.

[6] H. Rowley, Henry, Baluja, Shumeet; Kanade, Takeo; "Neural network-based face detection, IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 20, No. 1, pp. 23-37, 1998.

[7] S. Pastoor, J. Liu and S. Renault, "An experimental multimedia system allowing 3D visualization and eye-controlled interaction without user-worn devices, IEEE Transactions on Multimedia, Vol. 1, No. 1, 1999, pp. 41-52.

[8] Y. S. Chen, C.H. Su and J.H. Chen, "Video-based real time eye tracking technique for auto-stereoscopic displays", Proceedings of the 5th conference on Artificial Intelligence and Applications, Taiwan, 2000, pp. 188-193.

[9] Y. Matsumoto and A. Zelinsky, "An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement", Proceedings of IEEE 4th International conference on face and gesture recognition FG'2000, Grenoble, France, 2000, pp. 499-505.

[10] M. Hasanuzzaman, T. Zhang, V. Ampornaramveth, M. A. Bhuiyan, Y. Shirai, H. Ueno, " Face and Gesture Recognition Using Subspace Method for Human-Robot Interaction", Advances in Multimedia Information Processing – PCM 2004, Lecture Notes in Computer Science, Volume 3331, pp 369-376, 2005.

[11] M.A. Bhuiyan, V. Ampornaramveth, S. Muto, and H. Ueno, "Face detection and facial feature localization for human-machine interface", NII journal, Vol. 5, No. 3, pp. 25-38, 2003.

[12] A.A. Bhuiyan, and C.H. Liu, "Intelligent Vision System for Human-Robot Interface", World Academy of Science, Engineering and Technology, V. 28, pp. 57-63, 2007.

ISBN: 978-988-19253-4-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)