ConBe Robot: The Development of Self-Perception and Expression in Face-to-Face Interaction

Sakmongkon Chumkamon and Eiji Hayashi
Department of Information Systems
Kyushu Institute of Technology
Iizuka, Fukuoka, Japan
m-san@mmcs.mse.kyutech.ac.jp, haya@mse.kyutech.ac.jp

***Abstract*— In social robot development of interaction system robot, it is necessary to develop the fundamental function such as the robot perception. Due to the robot should correctly interpret a behavior or mental expression of the human. If the robot has a good emotional insight of the human, it is the advantage for the robot perception. In this paper, we implement the significant technique that take an advantage to the robot such as the human detection, face detection and recognition. Basically, these techniques could further enable the robot capability of intelligent empathy from the expression of human. We intensively study the vision method for facial expression recognition (FER) to understanding the human emotion and interacting by the robot expression in particular case. The robot interaction is based on the interested person that the robot can recognize with their emotional expression. We also experiment the system in term of face-to-face between robot and user with demonstrate using the head robot along with the result, such as the performance of the perception and the behavior expression of the robot.**

Keywords—social robot; human-robot interaction; facial expression recognition

I. INTRODUCTION

A social robot utilizing in human life is recent pervasive, even in a house such as the automated vacuum cleaning robot based on Roomba vacuum cleaner [1]. This robot is the high growth field of robotic research and market. Due to it could entertain the human to encourage delighting a user by interactive communication to human that is crucial function of the robot. The human-robot interaction (HRI) is the interdisciplinary field that combines major study of human and robot. Many studies in HRI field are proposed from a various research areas, including artificial intelligence, human-computer interaction, pattern recognition, control system, electronic, mechanic, psychology, behavior expression system, social communication, neuroscience, etc. The point of HRI is as that the robot can be a friend of human and assisting. It is thus necessary to associate with artificial intelligence, especially intelligent agent (IA) which is a main role in HRI. The IA is an autonomous individual agent that has a perception observing an environmental state and has an action for robot behavior. The IA in HRI could be very complex because the robot needs to organize knowledge in many fields. In HRI, it is essentially capable of the empathy thereby it is crucial to the robot having the basis system as the recognition. For example, when the robot recognizes the object, the robot would autonomously comprehend the object and decide its responding action. According to the comprehension system, our previous research demonstrated the autonomous behavior producing by the motivation to imitate the animal [2]. The research was constructed with the arm robot and proposed the consciousness model of the robot depending on motivation level represented by the neurotransmitter dopamine. For looking forward and related by the proposed to provide the HRI system, in this paper, we study to implement the recognition and the social expression of the robot for interacting with a user.

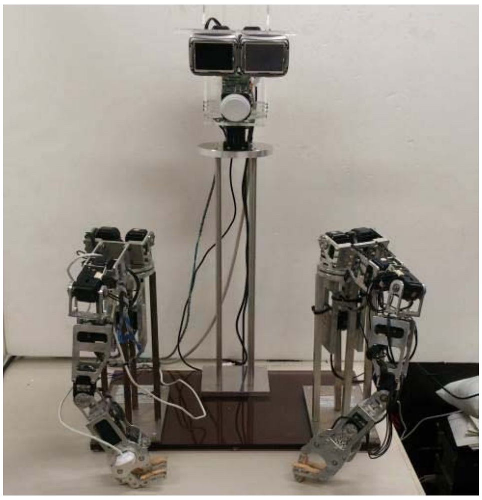

Fig. 1. ConBe robot which is comprised of two arm and one head.

In HRI, the recognition is significantly implemented with the robot to fulfil the robot perception. For example, human recognition, face detection, face recognition, facial expression recognition, can be implemented with the robot [3-5]. These recognitions are important in order to create the understanding the physical properties of humans.

In this paper, we aim to build the robot can socially interact with a human by interpreting the user's emotion from theirs face. This proposed organizes the recognition system that belonging to the emotion and motion of the human and implement robot behavior system for interaction between the robot's head and a user's facial expression. This system is operated on the robot that consists of two arms and one head.

978-1-4799-5955-6/14/$31.00 ©2014 IEEE

In the recognition system that we attempt to combine, there are human detection, face detection, face recognition. We use these techniques because the robot would highly weight for interesting with human and take more interesting when it can recognize the human. Later the robot would interact with them by the robot's expression using its eye correspondingly.

In the rested section of this paper, they are organized as that we describe our methodology of our proposed comprising of Conbe robot structure in section II, robot looking for interested person in section III, facial expression recognition in section IV, and the robot eye expression in section V. The experiment and results are presented in Section VI. We eventually conclude our work and suggest for future work in Section VII.

II. CONBE ROBOT STRUCTURE

This proposed is to develop the HRI system where it is based on the humanoid-like robot of the upper body. This robot is built to develop and study the robotic consciousness system, thereby enhancing the behavior of the robot for acting naturally like as an animal. The hardware of conscious behavior robot (Conbe) is composed of two manipulators and one head as shown in Fig. 1. The robot is also provided for vision system with a camera and expression system with the robot eye.

For each manipulator combined with six degrees of freedom arm and a DOF grippers where represents a hand-like consistency. Totally, seven attenuators are assembled for each manipulator. All attenuators are from Dynamixel DX-117 platform. Due to the determination of the angle for each joint of manipulator is difficult for moving to the target position, we then divided 7 DOF into 4 parts comparison by human arm, where each part represents to be a shoulder, an elbow, a wrist and a finger. The manipulator length is 450 millimeters.

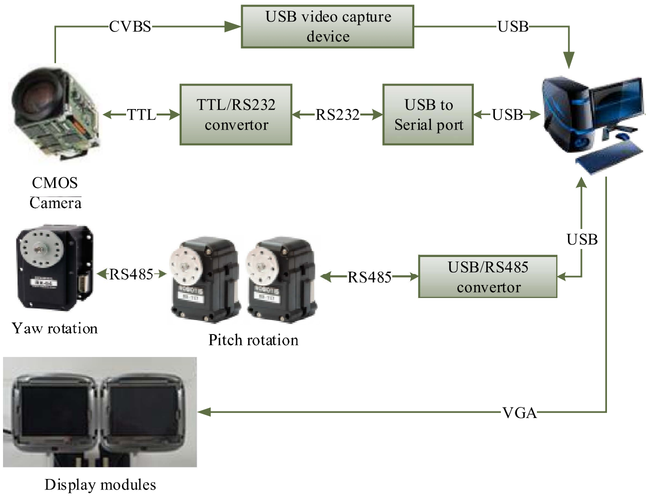

Fig. 2. Connection diagram of the head robot.

In this research, we build the head of the robot along with the robot vision system and expression system. The head dimension is 15×15×20 cm. For the head, we designed and constructed the structure from acrylic plastic material. The head also assembles with two DOF actuators to rotate in Cartesian coordinate where the first actuator performs as a neck; the second rotate for up down. These actuators were based on Dynamixel platform and controlled by packet command containing with an identity number of the actuator via serial 485-multidrop. The attenuator feedbacks provide the angle of rotation in degree, temperature, load input voltage, speed and so forth.

For the vision system, we embed the camera into the head to obtain the instant image from the embedded CMOS camera based on SONY FCB-H11. The camera is compatible full HD format with; NTSL and PAL. The camera model is equipped with the ×10 optical auto focus zoom lens and the ×12 digital zoom. The module can control the features of the camera with VISCA protocol via serial port. In this paper, we use this camera in PAL mode connecting to composite video channel and convert composite video channel to USB using video capture device. We use a computer to control all devices in the robot system. Fig. 2 shows the connection diagram of the proposed system.

According to eye expression, we use the 2.5 inch display to perform the virtual eyes where are implemented by C++ using OpenGL library. The eye model is composed of upper and lower eyelid, iris, cornea and scleral body. The robot can express the emotion and its gaze via the eye, depending on the environments. The virtual eye drawing and the eye movement are demonstrated in the experimental section.

III. ROBOT LOOKING FOR THE INTERESTED PERSON

In order to develop the interaction between human and robot, the robot is firstly necessary to answer what the object that the robot has the most interesting. In our research, we attempt to develop the robot that can interact with human among the various environments. As the reason, we implement the way that the robot can recognize the human, where we utilize the body and face recognition. For implementing the face and body detection, we equip with OpenCV library for coding the software. Firstly, for the vision system, the system would capture the image from the camera that is embedded in the robot head. Afterward, the algorithm would perform the face and body detection to identify the orientation of the user. The robot then tracks that user and attempt to recognize that user. If the robot can recognize it would interact with the person and convince user interacting. The procedure of the robot looking for the interested person is shown in Fig. 3.

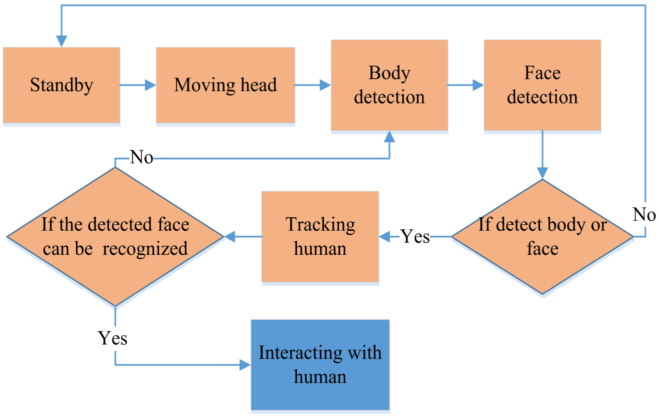

Fig. 3. Flow chart of the robot looking for the interested person

A. Face and body detection

For the perception of objects, this section would implement the function that the robot could locate the human thereby performing the face and body detection. Both object detection systems are implemented utilizing the Haar-based cascade classifier [4], [6]. This algorithm can detect the face properly and fast. The concept ideal is to search objects by the feature detector that similar to the original membered feature by comparison the physical characteristic of each region; for example, in the eye pair region, it should be darker than around the region such as forehead and cheek. This algorithm operates the integral images for the fast feature determination, using Adaboost to select the feature for learning algorithm, and cascading for fast rejection of non-face windows. The Haar-cascade classifier is constructed by the Haar-like features as a template for scanning the given image to find an object.

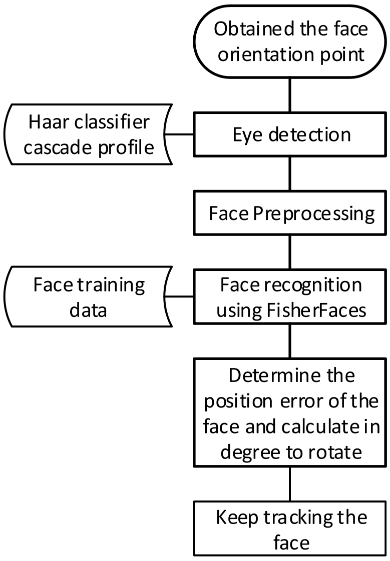
Obtained the face orientation point
Haar classifier cascade profile
Eye detection
Face Preprocessing
Face training data
Face recognition using FisherFaces
Determine the position error of the face and calculate in degree to rotate
Keep tracking the face

Fig. 4. The flow chart of face recognition algorithm

B. Face recognition

This implementation is to provide the recognition of the robot that the system is based on face-to-face condition. Therefore, the robot should recognize the person who the robot attends to interact. For the face recognition, we utilize the Fisherfaces algorithm in the software to label the given face. Presently, Fisherfaces method is also popular and reliable to implement. For the face classification, the pattern recognition of linear discriminant analysis (LDA) method is core to distinguish the face. Respect to LDA algorithm, we provide N sample images from c class or each face. Where $X = \{X_1, X_2, ..., X_c\}$, and each class representing to $X_i = \{x_1, x_2, ..., x_n\}$ as a vector in n-dimensional size of the image. The calculation of LDA is as follows in (1) (2) and (3). Where W_{LDA} is the LDA generalize Eigenvector that maximizes s_B which is the between-class scatter matrix, and s_w which is the within-class matrix. μ is the all-mean image of all classes and μ_i is the mean image of each class X_i. Fig. 4 is also shown the face recognition algorithm among the operation of our proposed.

$$w_{LDA} = \arg\max_w \frac{|w^T s_B w|}{|w^T s_w w|} \tag{1}$$

$$S_B = \sum_{i=1}^{c} N_i(\mu_i - \mu)(\mu_i - \mu)^T \tag{2}$$

$$S_w = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T \tag{3}$$

IV. Facial Expression Recognition

A. Constrained local model localizing facial features

The Constrained local model (CLM) is one of a powerful and ingenious method that can illustrate and indicate deformable objects or facial images, and many studies have implemented this attractive method [7]. The advantage of CLM is to stem from its use of the correlation among several small patches and an originative shape model, as well its robust and rapid tracking to unseen images. In addition to CLM, active appearance model (AAM), which is more precise, has also been very popular [8]. For comparison of the related methods, CLM methods can be used more efficiently for person-independent face alignment because CLM uses small local-region templates to local match in testing images, conversely AAM uses whole-face regions. CLM is necessary to provide images for training the models and defined shape that consists of the landmark points and the connections between the landmark points, which is shown in a 2D lattice. For example, the shape s of n landmark points represents by (4).

$$s = [x_1, y_1, ..., x_n, y_n]^T \tag{4}$$

For the shape relation, the region of each vertex corresponds to a source texture image. Equation (4) which we used x_n instead of $[x_n, y_n]$ thus we represented the equation by $s = [x_1, ..., x_n]$ where $x_i = [x_i, y_i]$ in 2D coordinates of an image. There T samples from data training from the images that we chose to train and assigned the landmark points in the local region. Other than this, we estimate the scale, rotation and translation from all samples that we use principal component analysis (PCA) for the approximated means. For the purposes of the present work, we implement this with non-rigid shape variation. A point distribution model (PDM) would be composed with the generalized rigid transformation, and locating the shape vertices with the given image [5].

$$x_i(p) = s\mathbf{PR}(\bar{x}_i + \mathbf{\Phi}_i\mathbf{q}) + \mathbf{t}; \quad (i = 1, ..., n) \tag{5}$$

Where $\mathrm{P} = \{s, \alpha, \beta, \mathrm{q}, \mathrm{t}\}$ represents the model parameters, which is composed of normalized scaling s, the rotation angles in 2D coordinates α and β, a translation of the shifting point **t** and non-rigid transformation parameter **q**. $\bar{x}_i$ represents a mean position of the i^{th} landmark and **P** denotes the projection matrix. We assume that the prior of the parameter can be normalized into a zero mean in a distribution and variance Λ at parameter vector **q**. Where x_i points in PCA provide $\bar{x}$ in (5) and Λ in (6).

$$p(\mathbf{p}) \propto N(\mathbf{q}; 0, \Lambda) \qquad (6)$$

The PCA of the PDM is to constrain the CLM and works with local or patch experts. For patch models, we use one classical probability method of 2D-Gaussian distribution to estimate the error landmark points. We then constructed the CLM model given by constructing a shape model and a trained patch model whose yields are considered independent and are multiplied.

$$J(p) = p(p)\prod_{i=1}^{n} p(l_i = 1 | x_i(p), I) \qquad (7)$$

Equation (7) where l_i denotes a random variable indicating whether the i^{th} landmark falls within its regional area, $p(l_i = 1 | x_i(p), I)$ is the probability of I image, and x_i indicates whether the i^{th} landmark is in its area. Additionally, another attractive detail of the CLM algorithm has previously been explained obviously which also presented the novel algorithm and compare the experiment with AAM by using human face images and magnetic resonance brain images [9], [10]. We build the facial models using CLM where Fig. 5 shows representative shape and patch models of the type used in the present study.

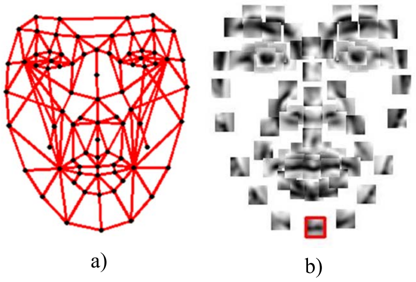

Fig. 5. CLM model a) Shape model and b) patch model

B. *Facial expression recognition of emotions*

After the facial parameters are obtained by the CLM algorithm, we use HMM to achieve the emotion classification using the face features location analyzing. The face features location is extracted by CLM and input to HMM for distinguish the expression. As algorithms for training the model, we use a Gaussian probability model. The objection of HMM is to the evaluation, decoding, and training where determined by Forward, Viterbi and Baum-Welch (BW) algorithms, respectively. BW re-estimates method is utilized to model the facial expression into HMM, where we use a left-right model [11]. For Considering HMM, $\lambda^{(k)} = (A^{(k)}, B^{(k)}, \pi^{(k)})$ which are the notations of HMM procedure that we use for constructing the expression models. S_i to S_j are the state transition probability that represented as $A^{(k)} = \{a_{ij}^{(k)}\}$. The observation probability $\boldsymbol{o}$ at state S_j represent by $B^{(k)} = \{b_j^{(k)}(o)\}$ and the initial state probability distributions represent by $\pi^{(k)} = \{\pi_j^{(k)}\}$. For continuous density HMM, we use a single component Gaussian distribution as the observation probability distribution given by (8) [11].

$$b_j(o) = \frac{1}{\sqrt{(2\pi)^n \left|\Sigma_j\right|}} \exp\left(-\frac{1}{2}(o-\mu_j)^t \Sigma_j^{-1}(o-\mu_j)\right) \qquad (8)$$

Where $\boldsymbol{\mu}$ is a mean vector and $\boldsymbol{\Sigma}$ is a covariance matrix that we solve by Viterbi training or the Baum-Welch method. For the purposes of the present study, we used a Baum-Welch algorithm. We construct and train the model for seven expressions to classify the facial expression. The concept is that the given face is similar to which class that means this class is the solution. Fig. 6 shows the overview of FER system that the system can recognize the emotion such as surprise, happiness, neutral, fear, sadness, disgust, and anger. Moreover, we still experiment the performance of this system and the corrected rate of these recognition system.

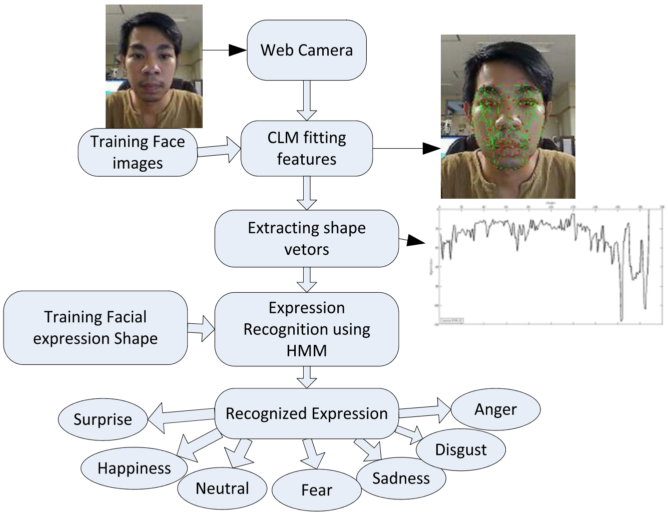

Fig. 6. The overview of FER system.

V. Robot Expression

For the implementation of the robot's expression, we apply the small display in the head to express the robot emotion. The robot would express its emotion via its eye, which is made by the 3D virtual eye software that would show in the experiment. With respect to the interested person, when the robot could recognize the person, the robot would attempt to play with that person and interacting with the user. Therefore, we simplify this task whereby imitating the user expression emotion to emulating for playing mutually between the user and robot. Fig. 7 shows the drawing eye structure and the orientation of the movement. For the robot eye's expression each emotion, we implement the appearance of robot's eye to mimic the human expression by study from the research that proposed recently [12, 13].

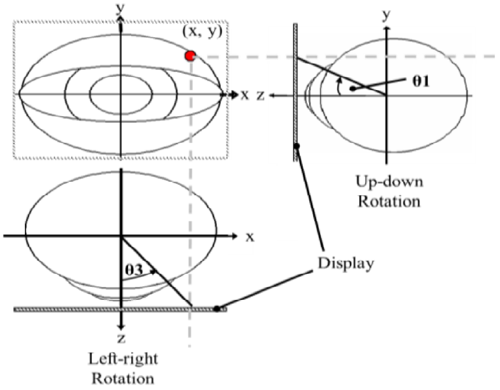

Fig. 7. The robot eye drawing and the movement

VI. Experiment

In this section, we demonstrate the robot to interact with the user in face-to-face. The experimental configurations are provided in our laboratory environment; and the robot head is set on the table belonging to an approximation of the human height, and the robot would look for a user. The given experimental user would emotionally express with his face in front of the robot head. Our proposed system was implemented consequently and combined together, there are body detection, face detection, along with recognition, facial expression recognition and robot eye expression. The machine configuration of these experiments was based on a laptop with CPU 2630QM.

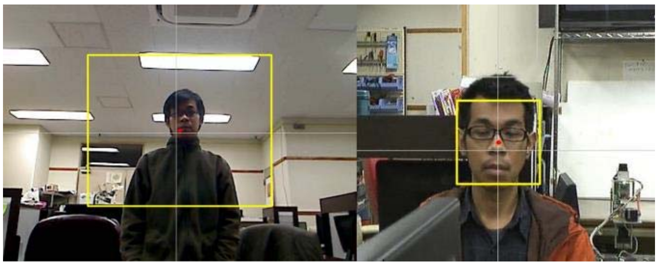

Fig. 8. The example results of body detection and face detection.

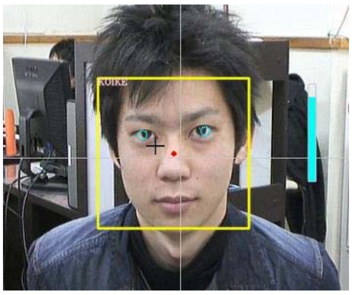

Fig. 9. The example image of face recognize tracking.

For the experiment of the face and body detection using Haar cascade classifier, we contribute the given user to test the human and face detection along with the robot tracking system because firstly the robot has to look for the interesting object or human. The example of this result is shown in Fig. 8. The robot can detect the object, whereas we restrict the size of the object image that is not bigger than 25×25. The performance frame rate of the detection is 16 frames per second and the velocity object while moving could not be over 180 pixels per second. In face recognition, when the robot would detect the face and recognize respectively, the robot will label that face and tracking. The example images while the face recognition system performing is shown in Fig. 9 and the operating frame rate is 13 frames per second. Fig 10 shows the performance of the recognized face tracking that shows the error in pixel of vertical and horizontal axis, while the robot tracking the given face. The result of the tracking of the recognized face is slightly slower than the tracking of face or body detection because that process has a longer execution time of face recognition process. With respect to the concept of the robot interaction, when the robot could recognize the given user, the robot will activate its behavior and interacting with the user if the robot were pleased with that user.

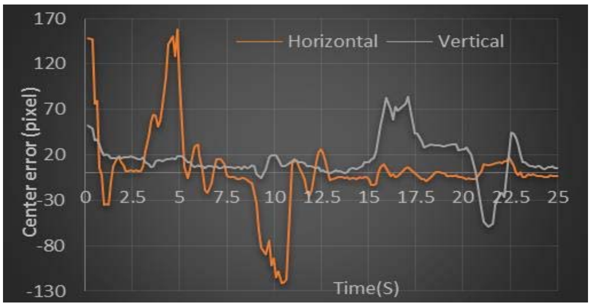

Fig. 10. The tracking errors for recognized face while the user moving left, right, up and down.

The interaction system is composed of two main sections, which are the facial expression recognition and robot eye expression. Both simultaneously operate when it can recognize the interested user. Firstly, for the FER, the robot uses the lattice shape model of human face and localizes the shape with the features on face using CLM. We then extract the facial parameter that is from the shape model. The facial parameter is composed of all connectivity in a shape model as an array. We then utilize continuous density HMM to classify the facial expression emotion from analysis of facial parameters. Our FER software in this paper was programmed by C++ language and utilizing OpenCV library. In this section, we tested extracting features using CLM, the emotion classification using HMM, and dynamic recognition of emotion.

For the experiment of feature extraction using CLM, the algorithm is necessary to be trained using facial images for constructing the model. We provide ten facial images with our laboratory environment as a background and train the CLM. Each image we defined 80 points of landmark and 187 connections due to these parameters are necessary for constructing the model. Then the CLM was trained to construct shape model and patch model for features fitting in testing images. Later, we test the localizing facial feature and extract the data to be a vector graph. Fig. 11 shows the example of extracting facial parameter into the graph and shows the

978-1-4799-5955-6/14/$31.00 ©2014 IEEE

characteristic graph of the happy expression where the x-axis means the sequent number of each connection in shape model and y-axis mean the length of the vector. The facial parameters would be the main data to be analyzed for emotional classification. The results of execution time in the additional experiment are shown in Fig 12. In this experiment, CLM has an average execution time as 27.97 milliseconds that means frame rate would be around 35.75 frames per second. As we can see in the results of the extracting graph each emotion, there is more complexity to distinguish that the facial parameter should be what the emotion.

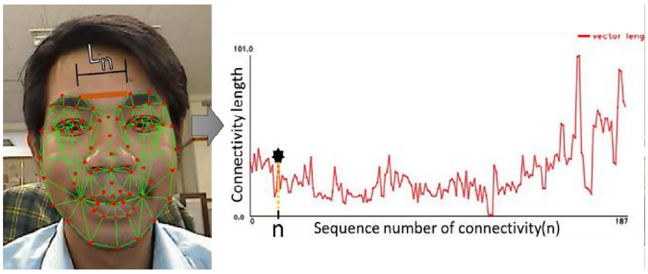

Fig. 11. The extractation of the facial parameters.

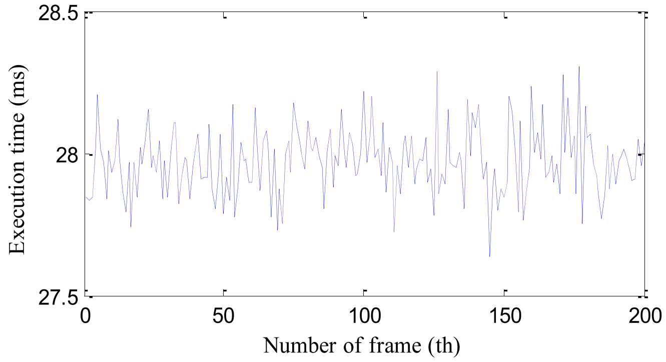

Fig. 12. The CLM execution time.

For the experiment of the emotional classification using HMM, the set of the facial image without spectacles includes 434 images divided to 84 training images and 350 testing images that means there are 12 training images per each emotion and 50 testing images per each emotion. The testing images are unseen images that exclude from training image. There are seven types of emotion, such as neutral, happiness, sadness, surprise, fear, anger and disgust. We then construct the models using these training data in different states model for comparing the performance. There are 80 states and 100 states of HMM. We then set up the system and test the given images. The results are given in a confusion matrix of recognition percentage as shown in Fig. 13 and Fig. 14. For average calculation time of each state, is shown in Fig. 15. In this system, we consider to use the HMM of 100 states because this model is optimize and suitable for our system by the experiment.

When the robot could recognize face and can perceive facial emotion, the robot would try to interact with the user. In this experiment, the robot would express its emotion via its eyes for imitating the human emotion. For the experiment, the given user would test with the robot head with his expression.

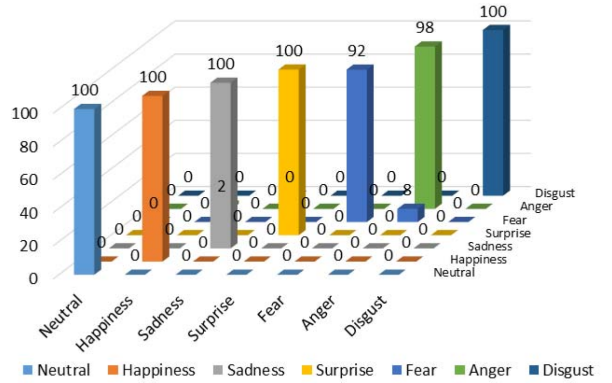

Fig. 13. Confusion matrix of the 80 states model.

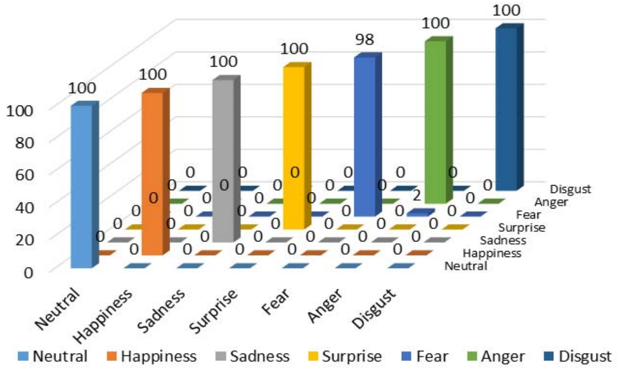

Fig. 14. confusion matrix of recognition correctness

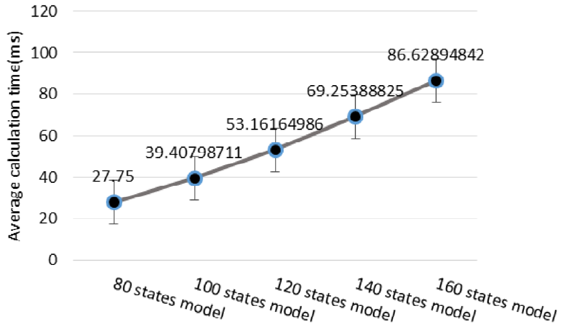

Fig. 15. Average calculation time of HMM for each state.

The user face images with the label name are provided in the face database of the robot software therefore the robot would recognize this user. The experiment successfully tested with the robot along with capture images that are shown in Fig. 16. In this experiment, the user performs his facial emotion and face with the robot based on basic emotion, such as neutral, happiness, sadness, surprise, fear, anger and disgust. For robot vision, the operation frame rate of this implementation is around eight frames per second and there is a little delay for the robot reaction because the calculation process of software.

VII. CONCLUSIONS

This paper proposes the system of the human-robot interaction in condition of face-to-face when the robot has the main motivation to interact with human. We then pay an effort to design our scenario and make the human-robot interaction system that combining our ConBe robot, the cognitive or recognition system, and the robot interaction or expression system. The system is successfully demonstrated by the head of the ConBe robot with given user, and present the results. With the results, the robot can interact with user in the condition of face-to-face whereby combining implementations of the face and body detection, face recognition, facial expression recognition, and the robot eyes expression. Additionally, we are moving forward for this proposed. Firstly, we would like to improve the FER system to be independent-person FER to assure the robot can cooperate with the human in the real world. The second, we would to combine system with the arms to increase the robot behavior that should embody the system operating properly.

Acknowledgment

We would like to appreciate to Jason Saragih who gives us some assistance in the part of CLM software, and our laboratory members for their contributed image.

References

[1] Forlizzi, Jodi, and Carl DiSalvo, “Service robots in the domestic environment: a study of the roomba vacuum in the home.” In Proceedings of the 1st ACM SIGCHI/SIGART conf. on Human-robot interaction, pp. 258-265, 2006.

[2] E. Hayashi, K. Ueyama, and M. Yoshida, “Autonomous motion selection via consciousness-based architecture,” IEEE Int. Conference on Ubiquitous Robots and Ambient Intelligence, pp. 401-402, 2011.

[3] Belhumeur P. N., Hespanha J. P., and Kriegman D. J., “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection.” In Pattern Analysis and Machine Intelligence, IEEE Transactions on, pp. 711-720, 1997.

[4] Viola P., Jones M., “Rapid object detection using a boosted cascade of simple features. In Computer Vision and Pattern Recognition,” Proceedings of IEEE Computer Society Conference on Vol. 1, pp. 511-518, 2001.

[5] S. Chumkamon and E. Hayashi, “Facial Expression Recognition using Constrained Local Models and Hidden Markov Models with Consciousness-Based Architecture,” IEEE/SICE Int. Symposium on System Integration, pp. 382-387, 2013.

[6] Lienhart Rainer and Jochen Maydt, “An extended set of haar-like features for rapid object detection.” IEEE Int. Conf. on In Image Processing. Proceedings, vol. 1, pp. 900. , 2002.

[7] Y.Wang, S.Lucey, J.Cohn, and J.M.Saragih, “Non-rigid face tracking with local appearance consistency constraint,” in IEEE Int. Conf. on Automatic Face and Gesture Recognition, October 2008.

[8] G. Edwards, C. Taylor, and T. Cootes. “Interpreting face images using active appearance models,” in Third IEEE Int. Conf. on Automatic Face and Gesture Recognition 1998 Proceedings, pp. 300-305, 1998.

[9] D. Cristinacce and T.F.Cootes, “Feature detection and tracking with constrained local models,” in Proc. of the British Machine Vision Conference, Vol. 3, pp. 929-938, 2006.

[10] D. Cristinacce and T. Cootes, “Automatic Feature Localisation with Constrained Local Models,” in Pattern Recognition, vol. 41, no. 10, pp.3054 -3067, 2008.

[11] Rabiner, Lawrence R., “A tutorial on hidden Markov models and selected applications in speech recognition,” Proc. of the IEEE vol. 77, no. 2, pp. 257-286, 1989.

[12] Hess Eckhard H., “Attittde and pupil size,” Scientific American, pp.46-54, 1965.

[13] Ekman Paul ed. Darwin and facial expression: A century of research in review. 2006.

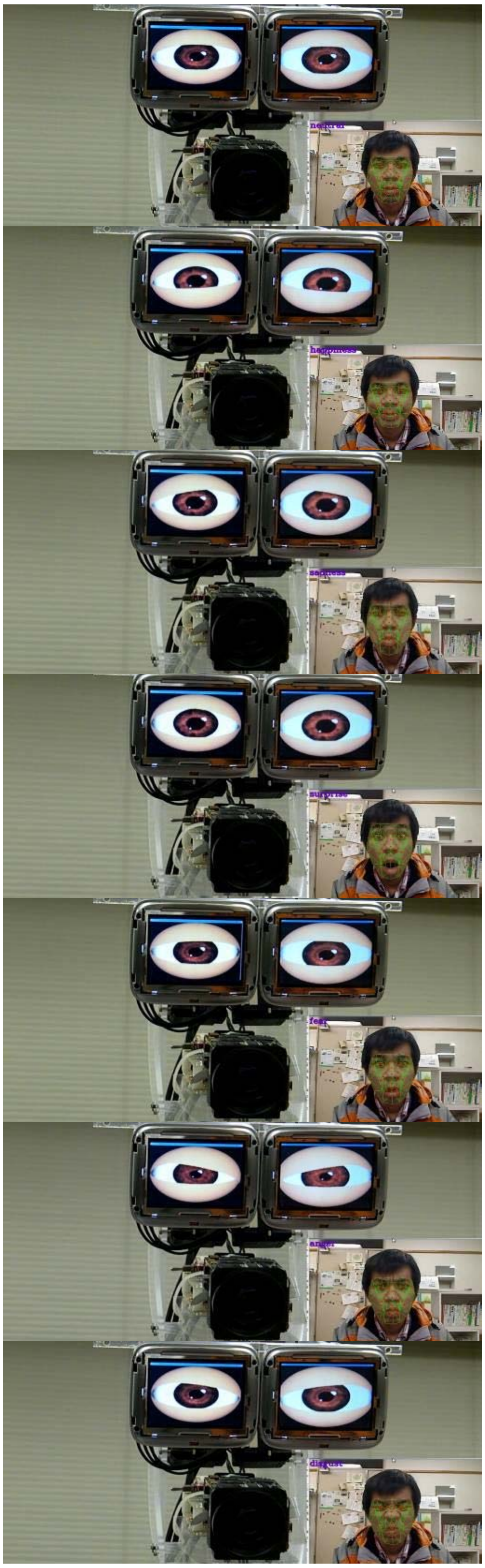

Fig. 16. The capture images while the robot interacting with user that express the facial emotion such as neutral, happiness, sadness, surprise, fear, anger and disgust respectively.