The Fifteenth International Symposium on Artificial Life and Robotics 2010 (AROB 15th '10),
B-Con Plaza, Beppu,Oita, Japan, February 4-6, 2010

A System for Synchronizing Nods of a CG Character and a Human Using Thermal Image Processing and Moving Average Model

R. Kato[1], Y. Yoshitomi[1], T. Asada[2], and M. Tabuse[1]

1: Div. of Environmental Sciences Graduate School of Life and Environmental Sciences Kyoto Prefectural University
2: Dept. of Environmental Information Graduate School of Human Environment Science Kyoto Prefectural University
1&2: 1-5 Nakaragi-cho Shimogamo Sakyo-ku Kyoto 606-8522 JAPAN
r_kato@mei.kpu.ac.jp, yoshitomi@kpu.ac.jp, t_asada@aoi.kpu.ac.jp, tabuse@kpu.ac.jp

Abstract: The purpose of the present study is to develop a system for communication between a human and a CG character for making him or her more peaceful and cheerful. The nonverbal communication using such as a facial expression, a nod, and a hand gesture is very important for reciprocal communication between humans. In the present study, an image registered by infrared rays which describes the thermal distribution of the face and neck has been used for developing a system for communication between a human and a CG character. In the present system, the CG character can synchronize its nod with a person's nod by predicting his or her nod angle. The measured feature parameter is inputted to a fuzzy algorithm system to obtain the nod angle of a person in front of an infrared ray camera, and then the Moving Average model is used for predicting the nod angle of the person. The average error of the nod angle obtained by the present system has been estimated as about $5°$. The CG character nods its head, not only when a human nods his or her head, but also when a human shakes his or her head left or right.

Keywords: Thermal image processing, nod, Fuzzy algorithm, AIC and Moving Average model.

I. INTRODUCTION

The nonverbal communication using such as a facial expression, a nod, and a hand gesture is very important for reciprocal communication between humans. These nonverbal communications are effective for expressing mutual consciousness. The realization of suitable communication between a human and a computer will help the developments of a computer itself, networking, a robot, and a computer system which can encourage a person living alone. In aging society, for example, Japan, something to talk to would be helpful for a person living alone and advanced in years to live peacefully.

The embodied communication technologies have been developed using the relationship between the speech and the nod [1]–[4]. On the other hands, several investigations for head gesture recognition have been reported [5]-[9]. In our daily lives, it is relatively difficult to take visual information on a person's head movement into account for synchronizing the nod of a CG character, compared to audio information on it. The main reason of the difficulty is that nuances of shade, reflection, and local darkness influence considerably the accuracy of head gesture recognition through the inevitable change of gray levels.

In the present study, in order to avoid the difficulty, an image registered by infrared rays (IR) which describes the thermal distribution of a face and neck has been used for developing a method for synchronizing nods of a human and a CG character.

II. THERMAL IMAGE ACQUISITION

The principle of thermal image generation comes from the well-known law by Stefan and Boltzmann, which is expressed by $W = \varepsilon \sigma T^4$, where W is radiant emittance (W/cm^2), ε is emissivity, σ is Stefan-Boltzmann constant (=5.6705×10^{-12}W/cm^{2}K^4), T is Temperature (K).

For human skin, ε is estimated at 0.98 to 0.99 [10], [11]. In the present study, however, the approximate value of 1 was used as ε for human skin. The values of ε for almost all substances are lower than that for human skin [10]. Accordingly, the human face and neck are easily extracted in the scene when the range of skin temperature is selected for producing the thermal image, using the value of 1 for ε [12], [13]. Fig.1 shows examples of male face and neck images by visible ray and IR. We can get a thermal face and neck image without lighting even at night. In principle, the temperature measurement by IR does not depend on skin color [11], darkness, or lighting condition, resulting in that the face, neck and their characteristics are easily extracted from the surrounding area on the thermal image.

III. BASIC PROCESSING FLOW

Fig.2 shows the flow chart of our method. We beforehand make the fuzzy rules between the ratio (hereinafter referred as the *nod*) of horizontal length to

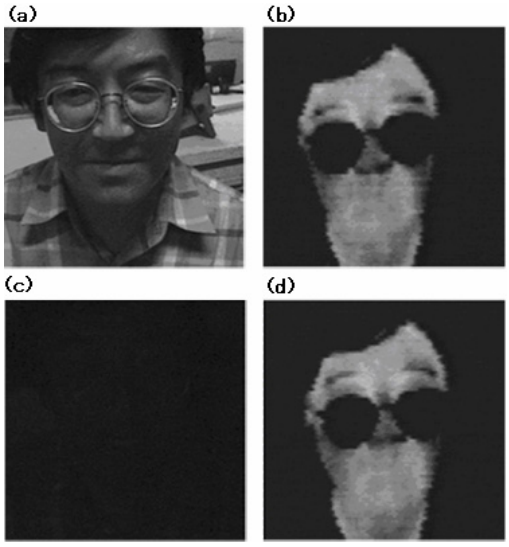

Fig. 1. Examples of face and neck image at night ; (a) visible ray with lighting, (b) IR with lighting, (c) visible ray without lighting, (d) IR without lighting [14]

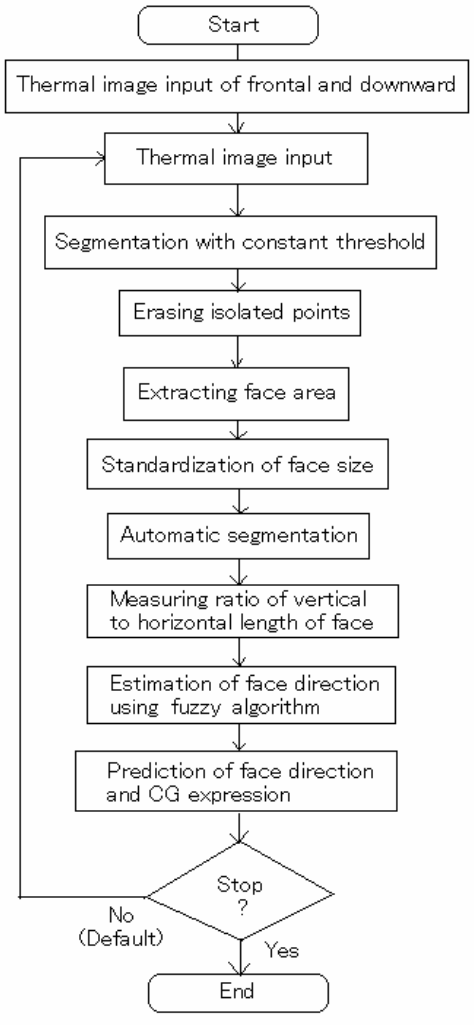

Fig. 2. Flow chart of our method

vertical length of face region on the thermal image and the face direction vector (hereinafter referred as $\boldsymbol{n}$) having the norm of 1. For making the rules, we prepare a thermal image for measuring the *nod*, while we use a motion capture system for measuring $\boldsymbol{n}$. Using the fuzzy algorithm and an input of the *nod*, $\boldsymbol{n}$ is outputted as three estimated values of elements of the vector. $\boldsymbol{n}$ is used for expressing the nod of a CG character. We explain the fuzzy algorithm and the process for measuring $\boldsymbol{n}$ in the Section IV.

At first, we get a pair of thermal images for the frontal and downward face directions for a subject to normalize the value of the *nod* as 1 in the case of frontal direction, and to transform the value of the *nod* of the subject to that of a typical subject. We explain the normalization and the transformation for the value of the *nod* in the Subsections IV(2) and VII(5), respectively. As a preprocessing for measuring the *nod*, a segmentation with a constant threshold, an erase of the isolated points, an extraction of the face area, a standardization of the size of face, and an automatic segmentation [15] are performed for a thermal image. Then, the *nod* is measured. For predicting the nod of a human, we use the Moving Average (MA) model. We explain the process for the prediction of the nod of the human in the Section V.

IV. FUZZY ALGORITHM FOR FACE DIRECTION ESTIMATION

1. Parameters used for fuzzy rules

Fig.3 shows a three-dimensional coordinate system used in the present study. $\boldsymbol{n}=(0,0,1)$ corresponds to the face direction when the face is frontal. Fig.4 demonstrates how to measure $\boldsymbol{n}$ using a motion capture system. For making the fuzzy rules and evaluating the accuracy of nod angle obtained by our method, three markers are attached on a subject's jaw, and both temples as shown in Fig. 4 (left figure) demonstrating an experiment scene where we place an IR camera and a computer monitor in front of the subject. We also place separately two cameras, which are not seen in Fig. 4 (left figure), of the motion capture system on the two diagonal positions from the frontal direction of the subject. With use of coordinates of three markers, $\boldsymbol{n}=(\alpha,\beta,\gamma)$ is determined by geometric calculation. The three elements of $\boldsymbol{n}$ are used as parameters of consequent in the fuzzy rules.

Fig.5 demonstrates face's downward rotation. The *nod* used as a parameter of antecedent in the fuzzy rules is measured as the ratio of vertical length to horizontal length of face. For eliminating the influence of hair style and throat region on the binary image to the utmost, the area having the horizontal pixels less than a

threshold which is experimentally decided is ignored in measuring the vertical length of face, and the area having the vertical pixels less than another threshold which is also experimentally decided is ignored in measuring the horizontal length of face.

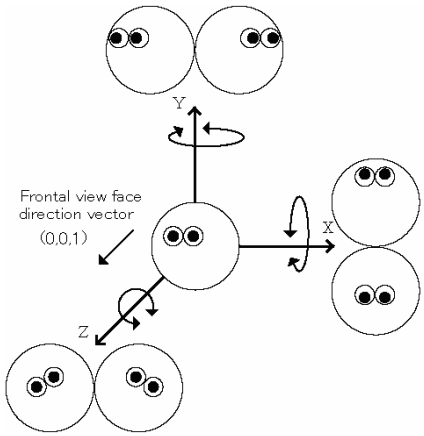

Fig. 3. Three-dimensional coordinate system

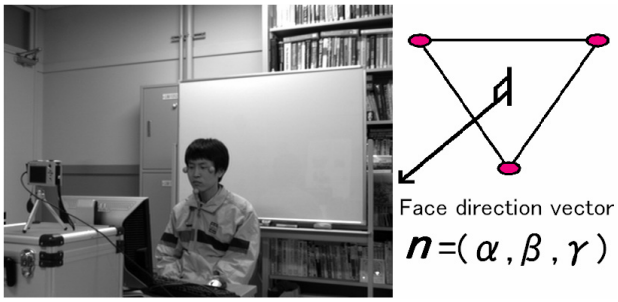

Fig. 4. Face direction vector $\boldsymbol{n}$
Left: face image with three markers to be measured their coordinates,
Right: way for making a face direction vector $\boldsymbol{n}$

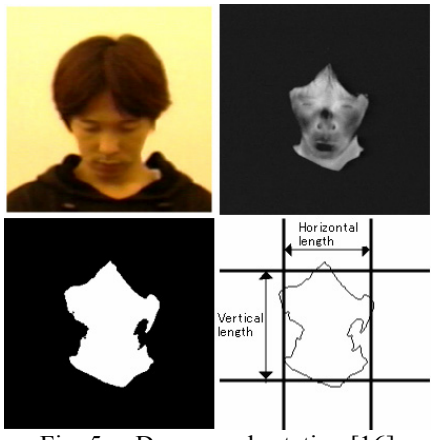

Fig. 5. Downward rotation [16]
Upper & left: visible image, Upper & right: thermal image, Lower & left: binary image, Lower & right: schematic diagram for explaining parameter *nod*

2. Fuzzy rules and algorithm

The fuzzy rules are described in (1). The fuzzy algorithm is described in (2)-(5).

$$\text{If } X = A_i \text{ then } Y_1 = B_{i1}, Y_2 = B_{i2}, Y_3 = B_{i3}, \quad (i = 1, 2, 3) \tag{1}$$

$$\mu^*_{B_{ij}}(Y_j) = W_i \mu_{B_{ij}}(Y_j), (i = 1, 2, 3, \ \ j = 1, 2, 3) \tag{2}$$

$$W_i = \mu_{A_i}(X^*), \ (i = 1, 2, 3) \tag{3}$$

$$B^0_j = \bigcup_{i=1}^{3} B^*_{ij}, \ (j = 1, 2, 3) \tag{4}$$

$$Y^*_j = \frac{\int \mu_{B^0_j}(Y_j) Y_j dY_j}{\int \mu_{B^0_j}(Y_j) dY_j}, (j = 1, 2, 3), \tag{5}$$

where X denotes a parameter of antecedent, Y_1, Y_2, Y_3 denote parameters of consequent, A_i, B_{i1}, B_{i2}, B_{i3} denote the corresponding fuzzy labels for X, Y_1, Y_2, Y_3, respectively, W_i denotes a fitness value of i th rule to the input of X^*. Y^*_j denotes an output for Y_j. The integral in (5) is performed in the whole range of Y_j. In the present study, X=*nod*, $Y_1 = \alpha$, $Y_2 = \beta$, $Y_3 = \gamma$ .We use three fuzzy rules described as 'frontal', 'rather downward', and 'downward'.

Fig.6 shows a schematic diagram of membership functions used for both antecedent and consequent. C_i denotes a fuzzy label of antecedent or consequent, for example, A_3. Z corresponds to a variable of antecedent or consequent, for example, X. For determining each actual shape of membership function, we use the measured values for the corresponding parameter. In the case of C_2, the corresponding measured value is neither a maximum nor minimum value among three measured values. In the case of C_1 and C_3, the corresponding measured values are minimum and maximum among three measured values, respectively. In the case of C_2, whose shape of membership function is a triangle, its membership function has a maximum value of 1 at the corresponding measured value and a minimum value of 0 at the corresponding measured values for C_1 and C_3. In the case of C_1, its membership function has a maximum value of 1 at its corresponding measured value or less and decreases linearly to a minimum value of 0 at the measured value for C_2. In the case of C_3, its membership function has a maximum value of 1 at the corresponding measured value or more and decreases linearly to a minimum value of 0 at the measured value for C_2. As each measured value, we use an average for six persons.

In addition, the value of *nod* is normalized by dividing by the value in the case of 'frontal'. We use a fuzzy algorithm for outputting $\boldsymbol{n}$ with three fuzzy rules and the value of the *nod* as an input. For integrating outputs from all rules, we use (4) and (5).

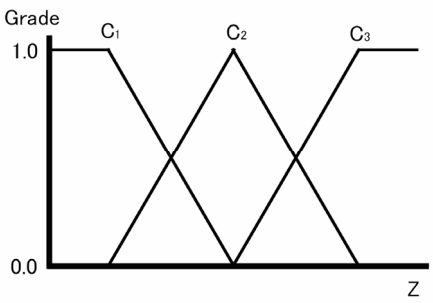

Fig. 6. Membership functions

3.Nod angle calculation

(6) defines a nod angle θ (°) which is calculated with γ.

$$\theta = 57.3\cos^{-1}\gamma \quad (6)$$

In the present study, θ is positive in case of downward rotation.

V. PREDICTION OF NOD ANGLE

For predicting the nod angle of a subject, we use the MA model expressed by (7).

$$\hat{\theta}(i) = \sum_{j=1}^{J} a(j)\theta(i-j), \quad (7)$$

where $\hat{\theta}(i)$ is a predicted value for $\theta(i)$, $a(j)$ is a coefficient for prediction, and J is a dimension for prediction. J is decided with Akaike's Information Criterion (AIC) expressed by (8).

$$AIC = n\ln S + 2p, \quad (8)$$

where n is the number of data, S is the summation of squared error, and p is the number of parameters. J is decided as p which gives the minimum value of AIC.

VI. FACE DIRECTION EXPRESSION USING CG

We use Virtual Reality Modeling Language (VRML) and Java for expressing a CG character. Fig.7 shows examples of CG expression of nod.

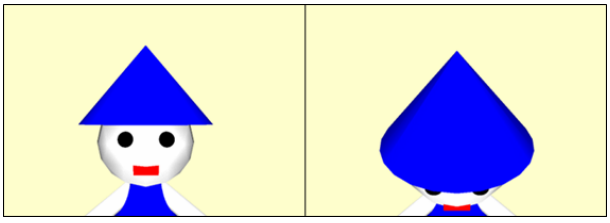

Fig. 7. CG expression of nod. Left: frontal view, Right: nod by 30 degree

VII.EXPERIMENT AND DISCUSSION

1. Experimental environment

The thermal image produced by a Thermal Video System (Thermo Shot F30) (NEC Avio Infrared Technologies Co., Ltd.) was transformed into a digital signal by an A/D converter (ADVC-300)(Thomson Canopus) and were input into a computer with an IEEE1394 interface board (1394-PCI3/DV6)(I・O Data Device). We used a computer (DELL Optiplex 745) (CPU: Intel Core 2 Duo 6600 2.4 GHz, a main memory: 2 GB, and OS: Windows XP (Microsoft)). We used Visual C++6.0 (Microsoft) and Java 2 SDK Standard Edition version 1.3, VRML2.0 for programming. A thermal image to be processed in a computer had a spatial resolution of 240×160 pixels and 8 bits gray levels at each pixel. We used Radish (Library) as a motion capture system for measuring the position of markers attached on the subject' face. We also used a computer, Dell Dimension 8300 (CPU: Pentium IV 3.2 GHz, main memory: 2.0 GB, OS: Windows XP), for performing the motion capture by Radish. Fig.8 shows a monitor screen of computer at experiment.

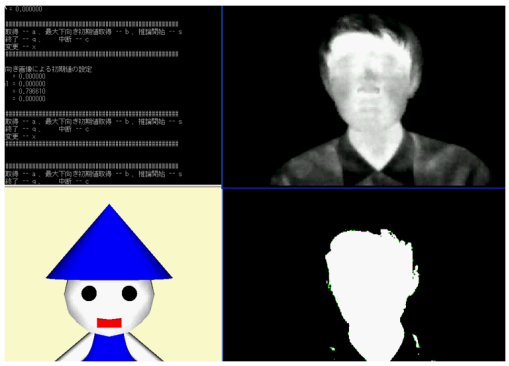

Fig. 8. Monitor screen of computer at experiment. Upper & left: expression on command prompt, Upper & right: thermal image, Lower & left: CG, Lower & right: binary image

2. Measurement of parameters used for fuzzy labels and treatment for input out of range decided by fuzzy labels

We obtained each fuzzy label as an average value for four male (A, B, C, D) and two female (E, F) subjects. Table 1 shows the fuzzy labels for each rule.

Table 1. Fuzzy labels for each rule

Variable / Rule	*nod*	α	β	γ
1 (frontal)	1.000	0.000	0.000	1.000
2 (rather downward)	0.908	0.027	0.147	0.989
3 (downward)	0.750	0.042	0.379	0.925

We got the nod angle θ using the output of γ, obtained by fuzzy algorithm, and (6). Because we use membership functions whose shapes are described in

Fig. 6, the output of γ is within the range of maximum and minimum obtained from three rules. When we got *nod* of 1.000 or above, the output by fuzzy algorithm was $\gamma = 1.000$. When we got *nod* of 0.750 or under, the output by fuzzy algorithm was $\gamma = 0.925$. For overcome the above problem, when the input of *nod* was under 0.750, we used a linear equation decided experimentally for getting the nod angle θ from *nod* having the value of 0.750 or under.

3. Approximation to nod angle θ obtained by motion capture system

Since we used the fuzzy labels shown in Table 1, the nod angle θ inevitably had errors for an individual subject. For predicting a nod angle of subject as precisely as possible, we approximated the nod angle θ to that obtained by the motion capture system, using a linear graph approximation method. For one male subject (G), we got the linear graph described in Fig.9 and used the relation for predicting his nod angle.

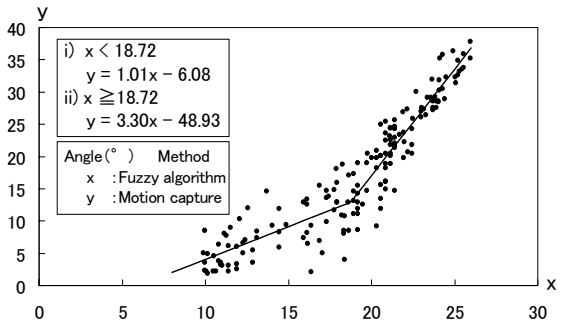

Fig. 9. Line graph approximation of nod angle from output using fuzzy algorithm

4. Prediction of nod angle θ of subject using MA model

According to (7) and (8), we got for the subject G the value of 23 as J which gave the minimum value of AIC when J is used for p, as shown in Fig.10. Then, we got (9) for predicting the nod of subject G.

$$\begin{aligned}\hat{\theta}(i) = {} & 1.39\hat{\theta}(i-1) - 0.19\theta(i-2) - 0.24\theta(i-3) \\ & - 0.11\theta(i-4) + 0.053\theta(i-5) + 0.032\theta(i-6) \\ & - 0.043\theta(i-7) - 0.028\theta(i-8) + 0.024\theta(i-11) \\ & + 0.03\theta(i-10) + 0.12\theta(i-11) - 0.15\theta(i-12) \\ & - 0.005\theta(i-13) + 0.0073\theta(i-14) + 0.11\theta(i-15) \\ & + 0.031\theta(i-16) - 0.14\theta(i-17) - 0.041\theta(i-18) \\ & + 0.18\theta(i-19) - 0.056\theta(i-20) + 0.052\theta(i-21) \\ & - 0.18\theta(i-22) - 0.14\theta(i-23)\end{aligned} \tag{9}$$

We did not have the value of $\theta(i-1)$ at the timing position of $i-1$ because it took about 0.13 s for getting $\theta(i-1)$. Accordingly, we used $\hat{\theta}(i-1)$ instead of $\theta(i-1)$ as described in (9). Fig.11 shows the accuracy of predicted nod angle for the subject G using (9) made beforehand with data for subject G at the previous experiment. The average error on predicted nod angle was $5.1°$. For a CG expression, the nod angle is transformed to 0 when $\hat{\theta}(i)$ is negative.

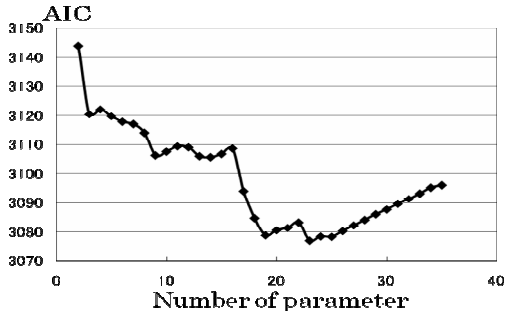

Fig. 10. AIC

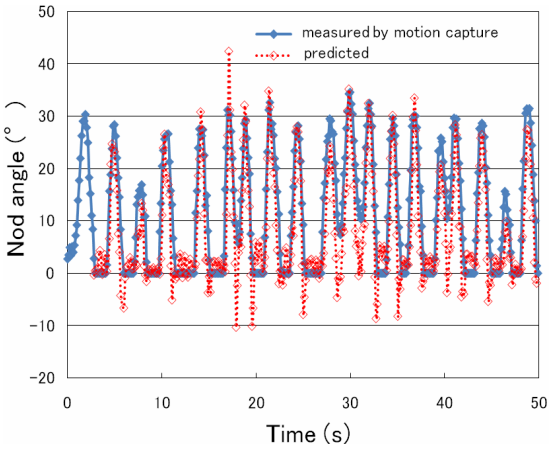

Fig. 11. Accuracy of predicted nod angle

5. Application to other subjects

We use the line graph approximation, as shown in Fig.9, for getting nod angle from the output of fuzzy algorithm. The line graph obtained for the subject G may not be appropriate for other subjects. However, it is time-consuming to get the line graph for each subject using a motion capture system. For overcoming this problem, when we use our system for another subject, we approximately transform the value of nod for another subject to that for the subject G, using a linear equation decided experimentally with two pairs of nod values of the subject G and another subject in cases of frontal and downward, followed by the same way as that for the subject G for getting the predicted nod angle.

For two male subjects (H and I), we used the above method for getting the predicted nod angle. The average errors for the subjects H and I on predicted nod angle were $4.4°$ and $5.5°$ respectively. The difference between the errors for the subject G and for the subjects H and I was within $1.0°$. This result shows that the

method using the line graph approximation of nod angle made for a typical subject (G) and the transformation of nod value from another subject to the typical subject is useful.

6. Future targets

The appearance of CG character gave a great influence on the impression of a user. It is necessary to improve the appearance of CG character so that it could give good impression to a user. In addition, a target of not only appearance but also movement of a CG character depends on the usage of the system. As the usage of future system, we assume an agent for an answer-phone using a computer and a virtual friend for a person living alone. In the present system, the movement of CG character is just nod. We will add other movements besides nodding, according to the usage.

We adopt the fuzzy algorithm for estimating n from the *nod*, because it is easy to improve its accuracy by adjusting the rules, the membership functions, and the way for integrating the outputs from rules for an input. In the present study, we propose a method for synthesizing nods of a CG character and a human. At the next stage, we will investigate the suitable embodied communication between a human and a CG character. The flexibility of fuzzy algorithm will help us to extend the present method to the next version.

Furthermore, we will compare the present method with the embodied communication technologies using the relationship between the speech and the nod [1]–[4]. Then, we will investigate a new technology for communicating a human and a CG character or a robot with use of human's speech besides his or her head gesture.

VIII. CONCLUSIONS

We have developed a method for communication between a human and a CG character by exploiting thermal image processing. The CG character can synchronize its nod with a human's nod by predicting his or her nod angle. The measured feature parameter is inputted to fuzzy algorithm system to estimate the nod angle of the person in front of the infrared ray camera, then we use the MA model for predicting the nod of person. The average error of the nod angle obtained by the present system was estimated at about $5°$. The CG character itself gave a great influence on the impression of user. Therefore we need to improve the CG character considering the usage of the system and the individual user.

REFERENCES

[1] Watanabe T, and Yuuki N (1989), A voice reaction system with a visualized response equivalent nodding. Advances in Human Factors/Ergonomics, 12A:396-403

[2] Watanabe T, Okubo M, and Ogawa H (2000), An embodied interaction robots system based on speech. J. Robotics and Mechatronics, 12(2):126-134

[3] Watanabe T, Danbara R, and Okubo M (2002), InterActor: speech-driven embodied interactive actor. Poc. of 11th IEEE Int. Workshop on Robot and Human Interactive Communication, 430-435

[4] Watanabe T, Okubo M, Nakashige M, and Danbara R (2004), InterActor: speech driven embodied interactive actor. Int. J. Human-Computer Interaction, 17(1):43-58

[5] Zelinsky A, and Heinzmann J (1998), A novel visual interface for human-robot communication. Advanced Robotics, 11(8):827-852

[6] Choi HI, and Rhee PK (1999), Head gesture recognition using HMMs. Expert Systems with Applications, 17(3):213-221

[7] Nonaka H (2003), Communication interface with eye-gaze and head gesture using successive DP matching and fuzzy inference. J. Intelligent Information Systems, 21(2):105-112

[8] Lin CS, Ho CW, Chang KC, Hung SS, Shei HJ, and Yehd MS (2006), A novel device for head gesture measurement system in combination with eye-controlled human-machine interface. Optics and Lasers in Engineering, 44(6):597-614

[9] Jia P, Hu HH, Lu T, and Yuan K (2007), Head gesture recognition for hands-free control of an intelligent wheelchair. Industrial Robot -An Int. Journal, 34(1):60-68

[10] Kuno H (1994), Infrared Rays Engineering (in Japanese). Tokyo, IEICE, 22

[11] Kuno H (1994), Infrared Rays Engineering(in Japanese). Tokyo, IEICE, 45

[12] Yoshitomi Y, Kimura S, Hira E, and Tomita S (1996), Facial expression recognition using infrared rays image processing. Proc. of the Annual Convention IPS Japan, 2:339-340

[13] Yoshitomi Y, Kimura S, Hira E, and Tomita S (1997), Facial expression recognition using thermal image processing. IPSJ SIG Notes, CVIM103-3, 17-24

[14] Yoshitomi Y, Miyawaki N, Tomita S, and Kimura S (1997), Facial expression recognition using thermal image processing and neural network. Proc. of 6th IEEE Int. Workshop on Robot and Human Communication, 380-385

[15] Yoshitomi Y, and Tanijiri T (2002), A machine for automatic segmentation of image and its method (in Japanese). Japan patent application 2002-344185

[16] Yoshitomi Y, Tsuchiya A, and Tomita S (1998), Face recognition using dynamic thermal image processing. Proc. of 7th IEEE Int. Workshop on Robot and Human Communication, 443-448