Gesture Recognition Using Combination of Acceleration Sensor and Images for Casual Communication between Robots and Humans

Y. Yamazaki[2], H. A. Vu[1], P. Q. Le[1], Z. Liu[1], C. Fatichah[1], M. Dai[1], H. Oikawa[1], D. Masano[1], O. Thet[1], Y. Tang[1], N. Nagashima[1], M. L. Tangel[1], F. Dong[1], K. Hirota[1], *Member, IEEE*

***Abstract*— A fuzzy logic based multi-modal gesture recognition method is proposed, where both camera images and hand motion data (given by a 3D acceleration sensor put on human wrists) are used to notify the emotion of humans to robots in real time. To demonstrate the validity, it is applied to a part of the home party scenario enjoyed by five eye robots and four human participants, where 8 types of human emotional-gestures are successfully recognized by robots. The proposed method aims to realize the casual communication between robots and humans in the mascot robot system that has been developed by authors' group.**

I. Introduction

IN order to control relationships between robots and humans in life environment, robotic systems may need to recognize human intentions and emotions. The casual communication robotic system, mascot robot system [1], has been proposed based on verbal information to recognize human intention and emotion. As for nonverbal approaches, gesture recognition has recently become attractive for spontaneous interaction with consumer electronics and mobile devices in the context of pervasive computing [2-6]. Most of previous works on the gesture recognition have been done based on visual information [7-9]. Using accelerometer data in gesture recognition is an emerging technique to improve recognition performance. Accelerometer-equipped devices, such as Apple iPhone, Nintendo Wii [10], provide new possibilities for interacting with a wide range of applications, such as home appliances and mixed reality.

A fuzzy logic based gesture recognition method is proposed and embedded in the mascot robot system to realize casual communication between robots and humans. It utilized both video image data from web camera and hand motion data given by a wearable 3D acceleration sensor on human wrists to identify the intentions and emotions. In the recognition algorithm, there are two main procedures, i.e., sensor data and image analysis. The processing of sensor data is based on fuzzy matching, where five fuzzy features with each axis (x, y, and z) are created from the values for each type of gestures, and then the candidate gestures are identified. In the case that fuzzy matching results are not decisive enough to identify the gesture, the outcomes from the image analysis procedure are used to make the final decision. The video processing algorithm is based on tracking of the movement of hands and relative distances between one's head and hands between frames. To demonstrate the validity, the proposed method is applied to a part of a mascot robot system with wearable 3D acceleration sensors (manufactured by Microstone Inc.) and Logicool Qcam(R)Connect cameras, where a home party scenario performed by five eye robots and four human participants. From the scenario, eight types of typical human emotional-gestures, i.e., toast, guide, goodbye, victory pose, banzai pose, putting hand on one's head and squatting; and putting hand on one's face, are recognized in the system. In experiment, each of six objects performs each of eight gestures for five times (30 gestures in all). With the fusion of both image and acceleration sensor data analysis, the recognition rate is compared to using sensor data or video processing individually. The system is implemented on a notebook PC (Let's Note, with Intel Core 1.06Hz, 1.5 GB Ram, Windows XP.) Results show that the proposed method can be used to improve the interaction between human and robotic systems.

In II the outline of the mascot robot system is summarized. A gesture recognition system is presented based on acceleration sensor and web camera in III. The IV presents the experimental environment/results with evaluation.

II. Mascot robot system

The mascot robot system is a networked robotic environment, where robots support humans in casual and friendly fusion. The systems have been composed of four fixed type eye robots, one mobile type eye robot, the speech recognition module (SRM), an information recommendation server, and a system server that is responsible for overall management. Robot Technology middleware (RT Middleware) is used to connect among the system's components. With RT middleware, each robot can be viewed as a networked component and the whole system can be managed from the view point of service level.

A system interface needs to express emotional reactions to user actions for casual communication. The eye robot that can express mentality and intentions is developed based on

Manuscript received January 31, This work was supported by JSPS KAKENHI 21300080

1. Department of Computational Intelligence and Systems Science, Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, G3-49, 4259 Nagatsuta, Midori-ku, Yokohama 226-8502, Japan. (Phone: +81-45-924-5686/5682; fax: +81-45-924-5676; e-mail: hrtjsps@ hrt.dis.titech.ac.jp).

2. Department of Electrical, Electronic & Information Engineering, Faculty of Engineering, Kanto Gakuin University, 1-50-1 Mutsuura-Higashi, Kanazawa-ku, Yokohama 236-8501, Japan. (email: yamazaki@ kanto-gakuin.ac.jp).

978-1-4244-8126-2/10/$26.00 ©2010 IEEE

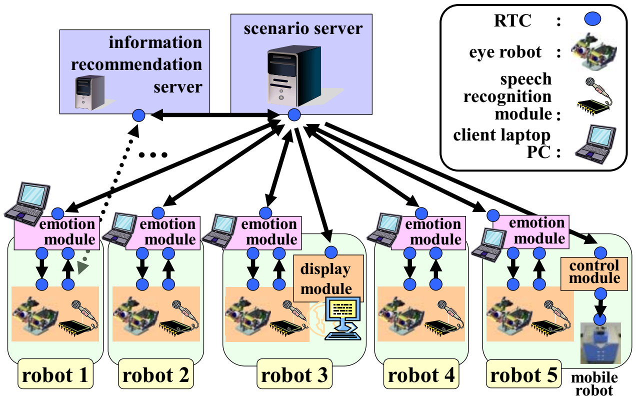

Fig. 1. the configuration of the mascot robot system

Fig. 2. the appearance of eye robots

Fig. 3. the appearance of Mobile eye robot

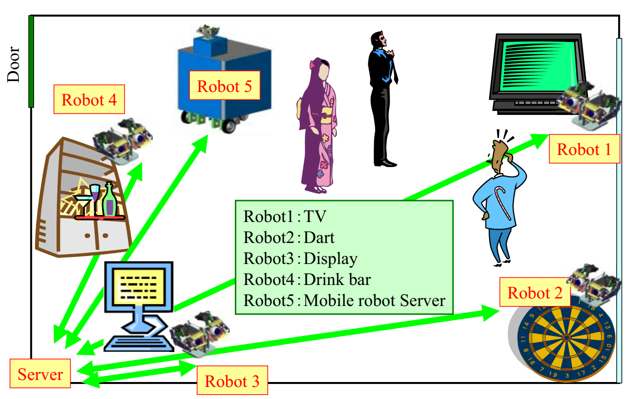

Fig. 4. the arrangement of the robots in the room

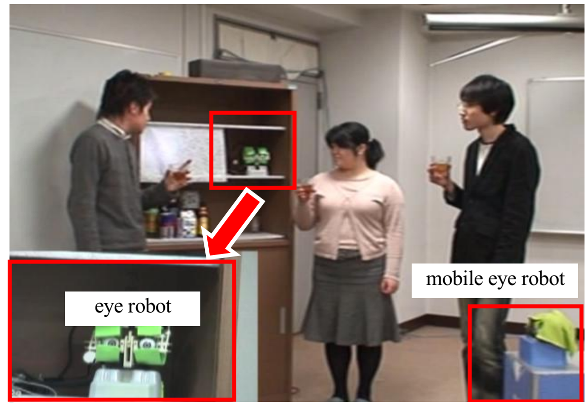

Fig. 5. The home party situation in the mascot robot system

the mechanisms of the human eye as a friendly robotic interface. The figures of the eye robot and the mobile eye robot are shown in Fig.1and Fig.2. The eye robot can express its mentalities and intentions using the motions based on extended pleasure-arousal space.

This system is tested in a living space simulating from a typical a home environment. The fixed type robots are placed on furniture and appliances such as the TV, a PC and a cabinet as a drink bar. The mobile eye robot moves along with users to help the speech recognition performance and support users. The arrangement of the system in a room is shown in Fig.4. A home party scenario is performed with human interlocutors in the space, where a host and two guests are supposed to enjoy the home party with five eye robots. The home party setting is shown in Fig. 5. In the scenario, eight types of typical human emotional-gestures, i.e., toast, guide, goodbye, victory pose, banzai pose, putting hand on one's head and squatting; and putting hand on one's face, are performed by humans.

Casual information recommendation based on speech information has already been confirmed in the system [1]. To embed the recognition function of unspoken message from humans in the mascot robot system, gesture recognition method based on multi-modal sensing is proposed, and shown in III.

III. Gesture Recognition Based on Multi-Modal Sensing

The gesture recognition system based on multi-modal sensing, i.e., the combination of an acceleration sensor recognition module and an images recognition module is presented. The inputs of the recognition system are data obtained from a three-axis acceleration sensor wrapped on wrist, and a web camera set in the robot system. The two modules processes input data simultaneously. The output of acceleration sensor module is obtained first. In case the acceleration sensor recognition module cannot identify the gesture, the video image recognition module is used based on the suggestion in the output of acceleration sensor module. The flow of the gesture recognition system is illustrated in Fig. 6. The detail of an acceleration sensor recognition module is shown in A, and an image recognition module is shown in B.

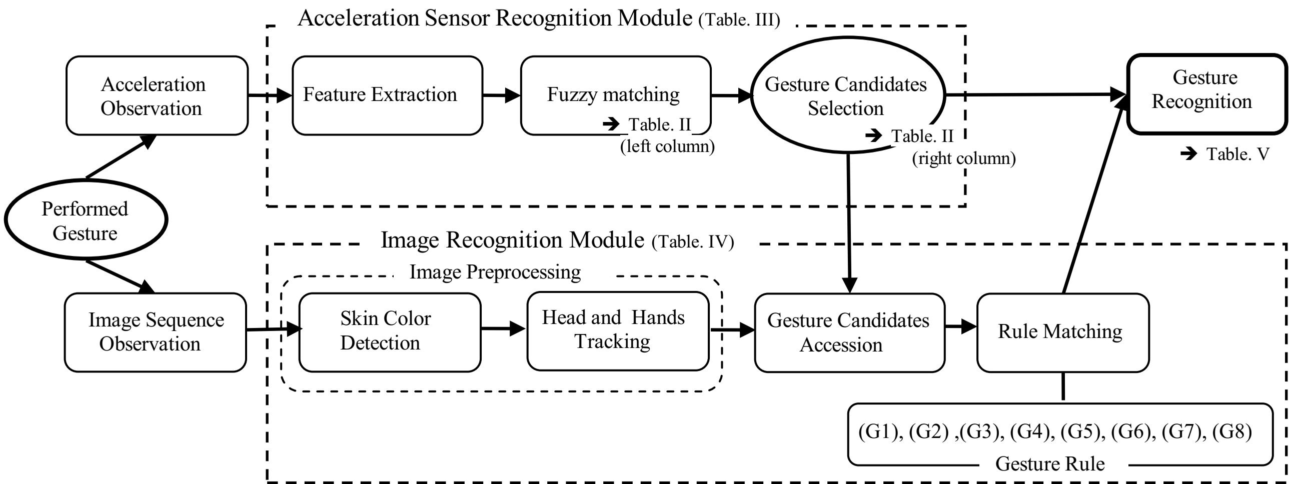

Fig. 6. The flow of the proposed gesture recognition system

Eight types of typical emotional-gestures in a home party scenario are recognized by the system. These gestures are presented as follows:

(G1) **Toast**: Raising a hand with a glass upwards from the waist level. (Fig. 7)

(G2) **Throw dart**: Moving a hand with a dart forward speedily from the head level.

(G3) **Victory pose**: From the relaxed position besides legs, the right hand is raised up to a level between the chest and the head. (This motion expresses happiness when the dart hits the target.)

(G4) **Banzai pose**: Both hands are first put down besides legs and then both are held right up. (This motion expresses happiness when the dart hits the target.) (Fig. 8)

(G5) **Squatting with hands over the head:** From the relaxed position the besides legs, both hands are put on the head and human subject squats. (This expresses disappointment when the dart loses the target.)

(G6) **Face covering**: Both hands are first put down besides legs and then right hand is held at face. (This expresses sadness when the dart loses the target.)

(G7) **Guiding**: From their relaxed position the besides legs, the right hand is swung to point towards the right direction.

(G8) **Farewell**: Both hands are first put down besides legs and then right hand is waved iteratively at right side.

The proposed method is applied to a part of a mascot robot system with a wearable 3D acceleration sensor manufactured by Microstone (Fig. 9) and Logicool Qcam(R) Connect web cameras.

Frame 12 Frame 20 Frame 28

Fig. 7. “Toast” gesture

Frame 5 Frame 26 Frame 42

Fig. 8. “Banzai pose” gesture

Fig. 9. 3D portable wireless accelerometer and how it is put on

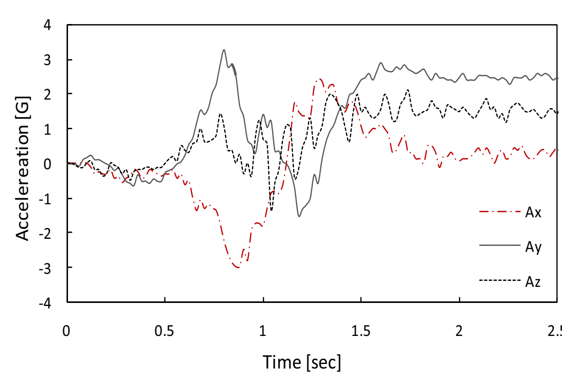

Fig. 10. Sample acceleration data of “Toast” gesture

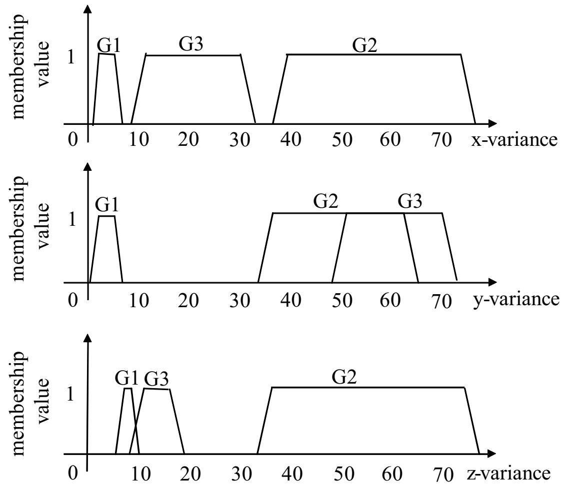

Fig. 11. The examples of membership functions for three-axis variance

Acceleration Sensor Recognition Module

To recognize gesture robustly and accurately, a fuzzy recognition system is developed. There are three steps in the proposed system. The first is getting data from an acceleration sensor. The second is data pre-processing. In this step, the challenge is how to calculate the starting point and the ending point of a gesture on the timeline. Then, the features are extracted from the acceleration sensor data. There are five features, i.e., maximum, minimum, mean, variance value, and mean crossing rate. In the third step, fuzzy matching is used to identify the gesture. The Fig. 10 shows the sample acceleration data of "Toast" gesture.

1) Feature Extraction from Acceleration Data

With each gesture time series data, the maximum and minimum are calculated at first. Then the parameters of mean, variance, and mean crossing rate are calculated by:

$$mean = \sum_{k-1}^{N} sample_k / N \quad (1)$$

$$\mathrm{var}\,iance = \sum_{k-1}^{N} (sample_k - mean)^2 / N \quad (2)$$

$$mcr = \sum_{k}^{N-1} \mathrm{sgn}(sample_k - mean) - \mathrm{sgn}(sample_{k+1} - mean) \quad (3)$$

where ***N*** is the number of total sample of a gesture in the time series data. All of the parameters have be calculated on axis x, y, and z.

In order to indentify the gesture from the eight types of gestures, a fuzzy measure of each parameter for each axis is proposed, where 15 membership functions are prepared for each gesture. The example variances of membership functions for three gestures (G1, G2, and G3) are shown in Fig. 11, where the membership functions are decided by three experimental samples. When a series date of gesture is put in the system, the 15 fuzzy measures which are calculated from 15 membership functions is used as a criterion to select the set of suitable gestures. This set of gestures is used in the image recognition module.

B. Image Recognition Module

Gestures are also recognized by video image processing along with acceleration sensor recognition to help making decision when the system cannot recognize the gestures from acceleration sensor which is the main part of algorithm. Video image sequences are captured while a web camera is put in front of human within home environment background. The gesture is extracted from each frame in the stream by three main steps: skin-color detection, head and hands tracking and rules matching. "Toast" and "Throw Dart" are shot from the side and the rest from front. Video image processing algorithm is shown in Fig. 6.

1) Skin Color Detection and Tracking

The skin detection algorithm is based on the HSV (Hue-Saturation-Value) color space with its robustness to the way that humans tend to perceive colors. Generally, a pixel's color can be considered as skin color if $h_{min} < H < h_{max}$ and $s_{min} < S < s_{max}$,where H,S, and V are the pixel's color components in the HSV color space and, h_{min}, h_{max}, s_{min}, s_{max} are the threshold values. From this operation, three regions including head and two hands can be found as groups of skin color areas.

2) Rule Matching for Image Recognition

From the output of the acceleration sensor recognition module, few gestures are selected. So, the simple rules are used to match the correct gesture. The frames recognized in the section above are matched to gestures by the following rules that are dependent on the moving directions and are tolerant to individual variations.

(G1) **Toast**: One hand is held up at 30 to 90 degrees tilt position from the horizontal line.

(G2) **Throw dart**: One hand is moved almost straight forward quickly.

(G3) **Victory pose**: One hand is held up over shoulders.

(G4) **Banzai pose**: Both hands are held up over head.

(G5) **Put both hands over head and squat**: Both hands are close to head and head squats.

(G6) **Put one hand on face**: Hand approaches close to head.

(G7) **Guiding**: Hand is swung to the side and is stopped at a distance longer than half of the distance between head and hand.

(G8) **Goodbye**: Hand is waved at the side from right to left and left to right over twice.

To evaluate the recognition system, it is embedded to a mascot robot system – a robot system in home environment. One acceleration sensor is set for each person, and one web-camera is set for each robot. The mascot robot system inputs the data of human gesture from acceleration sensor and web-camera, and outputs the recognition result.

TABLE II GESTURE RECOGNITION PROCESS BY ACCELERATION SENSOR AND NUMBER OF MATCHED GESTURES

	The number of count of gesture category (times)								The number of candidate gestures (n = 1-8, N/A)								
	G1	G2	G3	G4	G5	G6	G7	G8	1	2	3	4	5	6	7	8	N/A
G1: Toast	**30**	1	7	7	19	18	10	1	**10**	3	4	6	3	3	0	1	0
G2: Throw dart	1	**29**	3	5	5	2	2	5	**23**	1	3	0	1	1	1	0	0
G3: Victory pose	0	0	**27**	24	25	21	2	15	0	7	3	**10**	9	1	0	0	0
G4: Banzai pose	0	1	17	**29**	**29**	19	4	5	1	3	9	**16**	0	1	0	0	0
G5: Put both hands over head and squat	0	0	10	8	**30**	24	0	15	0	**18**	2	5	5	0	0	0	0
G6: Put one hand on face	0	0	1	3	**30**	**30**	11	5	0	**18**	8	1	2	1	0	0	0
G7: Guiding	0	0	0	0	7	27	**30**	0	3	**20**	7	0	0	0	0	0	0
G8: Goodbye	0	0	9	3	**30**	15	3	28	0	8	**17**	4	1	0	0	0	0

(N/A : NOT APPLICABLE) Output the results <- | -> To image recognition module

TABLE I SPECIFICATIONS OF ACCELEROMETER USED

Item			Specifications	Unit
Internal Sensor	Accelerometer	Axis	3(Ax, Ay, Az)	axis
		Acceleration	±20/±60 (switchable)	m/sec^2
		Response Frequency	0～100	Hz
	Angular Rate Sensor	Axis	3(ωx, ωy, ωz)	axis
		Anglular Rate	±300	deg/sec
		Response Frequency	0～50	Hz
External sensor		No. of Channels	8 or less	ch
		Service Voltage	3	VDC
		Input Voltage Range	0～3	V
A/D resolution			1024 (10 bit)	LSB
Sampling cycle			1～200	msec
Wireless Connection Standard			Bluetooth	class1
Wireless Connection Distance (reference value)			30	M
Power			rechargeable lithium-ion battery	
Operation Time (estimation)			10	Hour
Operation Temperature Range			0～40	℃
Weight			about 60	g

TABLE III GESTURE RECOGNITION RESULTS FOR MULTI-MODAL SENSOR

Performed gesture	Gesture category								
	G1	G2	G3	G4	G5	G6	G7	G8	N/A
G1: Toast	**27**	0	2	0	0	0	1	0	0
G2: Throw dart	0	**29**	0	0	0	0	1	0	0
G3: Victory pose	0	0	**24**	2	0	4	0	0	0
G4: Banzai pose	0	0	0	**28**	2	0	0	0	0
G5: Put both hands over head and squat	0	0	0	0	**29**	0	0	1	0
G6: Put one hand on face	0	0	0	0	3	**27**	0	0	0
G7: Guiding	0	0	0	0	0	2	**28**	0	0
G8: Goodbye	0	0	0	0	3	0	0	**27**	0

(N/A : NOT APPLICABLE)

TABLE IV GESTURE RECOGNITION RESULTS FOR ACCELERATION SENSOR

Performed gesture	Gesture category								
	G1	G2	G3	G4	G5	G6	G7	G8	N/A
G1: Toast	**23**	0	0	0	2	5	0	0	0
G2: Throw dart	0	**29**	0	0	1	0	0	0	0
G3: Victory pose	0	0	8	6	**9**	5	0	2	0
G4: Banzai pose	0	0	4	**15**	10	1	0	0	0
G5: Put both hands over head and squat	0	0	0	1	**17**	10	0	2	0
G6: Put one hand on face	0	0	0	0	6	**24**	0	0	0
G7: Guiding	0	0	0	0	0	3	**27**	0	0
G8: Goodbye	0	0	0	0	14	0	*0*	**16**	0

(N/A : NOT APPLICABLE)

TABLE V GESTURE RECOGNITION RESULTS FOR VIDEO IMAGE PROCESSING

Performed gesture	Gesture category								
	G1	G2	G3	G4	G5	G6	G7	G8	N/A
G1: Toast	**17**	0	4	0	0	4	5	0	0
G2: Throw dart	4	**18**	1	0	0	0	6	1	0
G3: Victory pose	6	0	**24**	0	0	0	0	0	0
G4: Banzai pose	2	2	0	**26**	0	0	0	0	0
G5: Put both hands over head and squat	0	0	0	0	**29**	0	0	1	0
G6: Put one hand on face	6	1	0	0	0	**23**	0	0	0
G7: Guiding	6	2	0	0	0	0	**22**	0	0
G8: Goodbye	0	1	1	0	0	0	0	**28**	0

(N/A : NOT APPLICABLE)

IV. GESTURE RECOGNITION EXPERIMENTS BASED ON MULTI-MODAL SENSING

A. Experimental Environment for Gesture Recognition System

In order to measure gesture recognition rate using an acceleration sensor, video image processing, and multi-modal sensors, eight gestures mentioned in III.B are selected from the party scenario designed for a mascot robot system. Gesture recognition experiments are carried out with a three dimensional acceleration sensor and video acquired from a camera used to track hand movements. The proposed systems are implemented in Visual C++ 2005 language and OpenCV library for video processing, and are run on a notebook computer with Intel® Core2Duo 1.06 GHz processor, Windows XP operating system, and 1.5 GB of RAM.

A wearable acceleration sensor, (W=45×D=45×H=20mm) manufactured by Microstone Inc., is used to get three dimensional acceleration data for movements on axis x, y, and z, and the system receives 3D movement data via a Bluetooth connection. Each human subject wears an acceleration sensor on his/her wrist as shown in Fig. 9. Six subjects with different height, who are widely selected from various organizations, are invited to perform the eight gestures for five times (240 gestures in all) in the party scenario of Mascot Robot System. Table I gives specifications of the used acceleration sensor. Each video sequence containing around 50 frames is captured by a Logicool Qcam(R) Connect camera mounted on a notebook with 30 fps rate and 640x480 pixels resolution. Human subjects perform

TABLE VI GESTURE RECOGNITION RATES (%) *FOR ACCELERATION SENSOR, VIDEO IMAGE PROCESSING, AND* MULTI-MODAL SENSORS

	Accelerometer	Camera	Multi-modal sensing: Accelerometer + Camera
G1: Toast	76.7	56.7	**90.0**
G2: Throw dart	**96.7**	60.0	**96.7**
G3: Victory pose	26.7	**80.0**	**80.0**
G4: Banzai pose	50.0	86.7	**93.3**
G5: Put both hands over head and squat	56.7	**96.7**	**96.7**
G6: Put one hand on face	80.0	76.7	**90.0**
G7: Guiding	90.0	73.3	**93.3**
G8: Goodbye	53.3	**93.3**	90.0
Average	66.3	77.9	**91.3**

(N/A : NOT APPLICABLE)

gestures under a constant background, lighting conditions and a constant 2-3 meter distance from the camera.

B. Experimental Results on Gesture Recognition

Gesture recognition system based on multi-modal sensing are examined, and compared with the recognition system by an acceleration sensor and the system only by video image processing.

The result of the process in the system multi-modal sensing system is given in Table. II, when each of the gestures is supplied as input to the system. Left eight columns (G1-G8) give count of recognitions for each typical gesture, while right left columns (n=1-8, N/A, i.e. Not Applicable) give counts for each number of matched gestures. The number of recognized gestures obtained from the system of the multi-modal sensing, the acceleration sensor, and video image processing are given in Table III, IV, and V respectively. In Table III, IV, and V, each row corresponds to one of eight categories of gestures and each column gives the number of recognized gestures for each gesture category performed by all subjects. The average recognition rates for each recognition system are calculated for each gesture in Table VI.

Overall, 91.3%, 66.3%, and 77.9% of gestures are recognized in average with the multi-modal sensing, the acceleration sensor, and video image processing, respectively. As a result, the proposed multi-modal sensing method gets the highest recognition rate in average of all gestures. As for individual gesture, the proposed method gets the highest recognition rate in the seven types of gestures out of eight (the 3 gestures are two-way tie for highest). In the recognition for the one gesture (G8: Goodbye), the proposed method gets 90.0%, and is 3.3% lower than the highest score.

The recognition method based on different sensors has a particular inclination. With acceleration sensor, the “Throw dart” gesture has the best recognition rate of 96.7%, while the “Victory pose” gesture has the lowest recognition rate of 26.7%. With video image processing, the “Put both hands over head and squat” gesture has the best recognition rate of 96.7%, while the “Toast” gesture has the lowest recognition rate of 56.7%. With multi-modal sensing, more than 80% of the gestures are recognized for all types of gestures. This result confirms the advantage of the proposed multi-modal sensing method compared with the recognitions based on single sensor.

V. CONCLUSION

A fuzzy logic based gesture recognition method is proposed, where the eight types of typical emotional-gestures are recognized based on multi-modal sensing, i.e., the combination of the 3D acceleration sensor and web cameras. The proposed method includes two modules, i.e. the acceleration sensor recognition module and the image recognition module. The acceleration sensor module precedes the image recognition module. In the acceleration sensor module, five fuzzy features with each axis (x, y, and z) are first generated by the values for each type of typical emotional-gestures (15 features in all), and then candidate gestures are decided by fuzzy matching. In the case that fuzzy matching results are not decisive enough to identify one gesture, the outcomes from the image recognition module are applied to make the final decision. In the image recognition modules, the recognition result is decided using rule matching based on the skin color detection and the tracking data of the head and hands.

To demonstrate the validity, the proposed method is applied to a part of the mascot robot system with wearable 3D acceleration sensors (manufactured by Microstone) and Logicool Qcam(R)Connect cameras, where a home party scenario is performed by five eye robots and six human participants. From the scenario, eight types of typical human emotional-gestures, i.e., “Toast”, “Throw dart”, “Victory pose”, “Banzai pose”, “Put both hands on one’s head and squat”, “Put hand on one’s face”, “Guiding”, and “Goodbye”, are recognized in the proposed system. In experiment, each of six subjects performs each of eight gestures for five times (30 gestures in all for each gesture). The system is implemented on a notebook PC (Let’s Note with Intel Core 1.06Hz, 1.5 GB Ram, Windows XP), where gesture recognition system based on multi-modal sensing are tested and compared with the competitors, i.e., the recognition system by using an acceleration sensor and the system only by video image processing.

The proposed multi-modal gesture recognition method gets 91.3% of recognition rate in average of all types of gestures. The proposed method gets the highest recognition rate in the seven types of gestures out of eight (the 3 gestures are two-way tie for highest). More than 80% of gestures are recognized as for all types of gestures, which shows that the proposed multi-modal sensing method advances to the recognition methods based on single sensor. As a result, the validity of proposed multi-modal gesture recognition is confirmed.

The proposed fuzzy logic based multi-modal gesture recognition method provides a casual communication robotic environment for mascot robot system, where robots assist humans in view of the home party atmosphere based on recognitions of spoken and unspoken messages from humans.

REFERENCES

[1] Y. Yamazaki , H. A. Vu, Q. P. Le, K. Fukuda, Y. Matsuura, M.S. Hannachi, F. Dong, Y. Takama, and K. Hirota, "Mascot Robot System by integrating Eye Robot and Speech Recognition using RT Middleware and its Casual Information Recommendation," *3rd International Symposium on Computational Intelligence and Industrial Applications (ISCIIA2008)*, Nov.2008, pp. 375-384.

[2] L. Bao, et.al., "Activity Recognition from User-Annotated Acceleration Data," *Pervasive Computing* , 2004, pp. 1-17.

[3] J.Y. Yang, J.S. Wang, Y.P. Chen, "Using acceleration measurements for activity recognition: An effective learning algorithm for constructing neural classifiers," *Pattern Recognition Letters*, vol. 29, 2008, pp. 2213-2220.

[4] T. Pylvanainen, "Accelerometer Based Gesture Recognition Using Continuous HMMs," *IbPRIA, LNCS 3522*, 2005, pp. 639-646.

[5] J. Wu, G. Pan, D. Zhang, G. Qi, S. Li, "Gesture Recognition with a 3-D Accelerometer," *UIC, LNCS 5585*, 2009, pp. 25-38.

[6] J. Liu, L. Zhong, J. Wickramasuriya, V. Vasudevan, "uWave: Accelerometer-based personalized gesture recognition and its applications," *Pervasive and Mobile Computing*, vol. 5, 2009, pp. 657-675.

[7] H. S. Park, E. Y. Kim, S. S. Jang, S. H. Park, M. H. Park, H. J. Kim, "HMM-based gesture recognition for robot control," *IlbPRIA, LNCS 3522*, 2005, pp. 607-614.

[8] C. Shan, T. Tan, Y. Wei, "Real-time hand tracking using a mean shift embedded particle filter," *Pattern Recognition Letters*, vol. 40, 2007, pp. 1958-1970.

[9] J. Varona, A. Jaume-i-Capo, J. Gonzalez, F.J. Perales, "Toward natural interaction through visual recognition of body gesture in real time," *Interacting with Computers*, vol. 21, 2009, pp. 3-10.

[10] Nintendo, Nintendo Wii, http://www.nintendo.com/wii/