The Realization of Robot Theater: Humanoid Robots and Theatric Performance

Chyi-Yeu Lin, Chang-Kuo Tseng, Wei-Chung Teng, Wei-Chen Lee, Chung-Hsien Kuo, Hung-Yan Gu, Kuo-Liang Chung, Chin-Shyurng Fahn

***Abstract*—This paper aims to report on the technical development and the realization of the world first robot theatre performance with a cast comprising two biped androids and two twin-wheeled two-armed humanoid robots. Each of the biped androids has a head with human-face, capable of showing multiple facial expressions, and a pair of 7 DOF arms with a multi-jointed hand capable of grabbing objects of various shapes and performing a number of trained talent shows. With a torso and two jointed legs, the android is capable of static walking. Each of two dual wheeled humanoid robots has a pair of 7 DOF arms and multi-jointed hands. These intelligent robots have advanced artificial intelligence including computer vision, voice recognition, and voice synthesis, and are capable of performing a number of entertaining programs on the stage.**

I. Introduction

The well-developed industrial robot technology has replaced human power in many high-speed, high-accuracy and repeating works at various advanced factory production lines. Since the beginning of 21st century, rapid development in related technologies for intelligent robots has built a solid foundation to new advantageous robotic applications in the future. Different from industrial robotic technology, intelligent robotics technology involves many new technologies in various categories. By uniting more powerful artificial intelligence, many faster and smaller multi-functional sensors, more efficient communication technology, and faster, economical computer calculation, intelligent robots in new age can provide numerous new potential utilities beyond our imagination. Bill Gates' declaration, "By 2025, there will be robots in every family," backed-up the great develop potential of intelligent robots.

There are many types of intelligent robot applications, including rescue, service, cleaning, health care, surveillance, educational, entertainment, companion, military, and many other special utilities. While modern technology is bringing huge stress to human life, it is worthy to build entertaining robots for providing laughter to human. Humanoid robot has become a hot research category; many countries have developed various humanoid robot systems. For example, in Japan, Sony has developed many innovative entertaining robots, such as four-foot robot AIBO [1], dual-foot robot QRIO [2]; besides, Honda's dual-foot robot ASIMO [3], Hitachi's Emiew-2 [4], Mitsubishi's Wakamaru [5], and Korean KAIST's Albert HUBO [6] are all able to provide entertainment and interaction functions with human. However, their major designing purpose was not to perform in a robot theater. Other countries have created several robots that may be applied in robot theater development, such as ROMAN [7], David Hanson Robot [8], KASPAR [9], and Doldori [10]. Most of these android robots have various facial expressions, but lack of moving ability and thereby fail to move on the stage by themselves. Their high price and costly maintenance prevents common people to own a robot. The concept of "Robot Theater" is to offer an easy opportunity for people to watch the entertaining performance of theatric robots and become more attached to intelligent robots.

One robot with the highest level musical performing function is Toyota Partner Robot [11], which can play a number of instruments in ways highly resemblance to human. In 2005 Aichi Expo, Japan, hundreds of thousand visitors lined up for hours each day to watch the Partner Robots' history-making performance. When a human watches humanoid robots with appearance like himself/herself performing mankind skills, his/her mind would be filled with special passion and satisfaction. Therefore, a robot theater featured by humanoid robots and androids is very attractive to human beings in every age. As one form of robot theaters, University of Osaka held a robot drama event with Wakamaru robots [5] in November 25, 2008. However, only wheeled robots played the leading roles. The drama content mainly emphasizes in story plot.

The main purpose of this study is to create robots with high-creative and entertaining values. In this study, we have built four intelligent humanoid robots (Two biped androids and two twin-wheeled robots) which can perform in robot theaters, and have actually played entertaining programs to the public. The biped androids have mechanical heads with 23 degrees of freedom to show various facial expressions and their bodies similar to human appearance with 38 degrees of freedom. Besides, they have artificial intelligence including vision, hearing and speech synthesis. The two twin-wheeled

Chyi-Yeu Lin is currently a Professor at the Department of Mechanical Engineering and the Director of Center for Intelligent Robots in National Taiwan University of Science and Technology. (e-mail: jerrylin@mail.ntust.edu.tw)

Chang-Kuo Tseng is currently working toward the Ph.D. degree in the Mechanical Engineering at National Taiwan University of Science and Technology.

Wei-Chen Lee is in the Department of Mechanical Engineering in National Taiwan University of Science and Technology.

Chung-Hsien Kuo is in the Department of Electrical Engineering in National Taiwan University of Science and Technology.

Wei-Chung Teng, Hung-Yan Gu, Kuo-Liang Chung, and Chin-Shyurng Fahn are in the Department of Computer Science and Information Engineering at National Taiwan University of Science and Technology.

robots each have two mechanical arms with 7 degrees of freedom, and are able to play drums and perform real-time drawing skills on the stage.

The first robot theater public performance (December 27, 2008, Taipei) presented by a male and a female android and two twin-wheeled robots in the world is named “Robot Fantasia”. The play is about a child’s imagination for robots. With guidance of a female robot (woman plays) which comes from the distant future to the Earth, modern robots’ development and functions are introduced. Pica, one of the wheeled robots, real time sings on musical notations and lyrics, and draws face portraits, showing that robots have started to learn human skills. Janet and Thomas, the two androids, perform a musical section from “The Phantom of the Opera” with hand-manipulated puppets and themselves, showing that robots may express love and with emotions. The culminating music and dance are combined with Ringo’s, the second wheeled robot, drumming and Janet and Thomas’ dancing, illustrating joy in the robot world. At the end, Janet and Thomas made a historical first kiss between two androids to close this robot theater performance. In the following chapters, the theatric robots, robot’s skills, motion planning technology used on the robots, and the actual performance in the robot theater are discussed.

II. The Robot Protagonist

A. Androids

Janet and Thomas, the protagonists of the robot theater, are both biped humanoid robots, as shown in Fig. 1. Their body structures are based on Wabian-2 [12], which was designed by Takanishi Lab of Waseda University, Japan. The detailed specification is listed in TABLE 1. The body and head are controlled separately by two systems. The body part has 38 degrees of freedom, including 7 degrees of freedom on each leg, 1 on the hip, 7 on each arm, 3 to control finger motion on each hand, and 2 on the waist. Robots can walk with self-balance.

TABLE 1
The Humanoid Robot Specifications

Parameters	Values/details
Height	160 cm
Weight	72 kg
DOF	23(Head) + 38(Body)
Vision Sensor	2 CCD cameras
Sensor	6-axis force sensor * 4, Photo sensor, Gyro
Actuator	Server Motor (Head), DC motor (Body)
Battery	Ni-MH 48V 4200 mAh*2
CPU	Pentium 4 M 2.0 GHz *2(Head & Body)
OS	Windows XP Professional (Head), QNX (Body)

The system planning of the android head developed in this work can be seen in Fig. 2. The rotation angle information for each servo motor corresponding to a specific facial expression is stored in the PC. When a specific facial expression is required for display on the face robot, pre-recorded information is sent to the servo motor controller so as to create the expression. Facial skin that can deform and show a number of facial expressions is made from a type of silicon with proper color dye. The facial skin is designed to have various thicknesses in different regions of the face to increase similarity to human face skin.

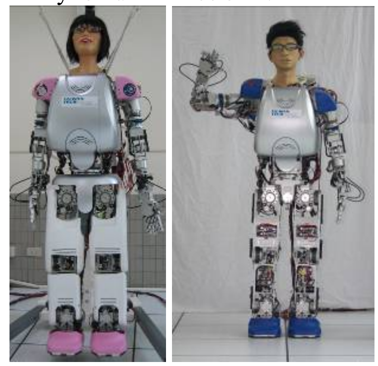

Fig. 1 humanoid robots: Janet (left) and Thomas (right)

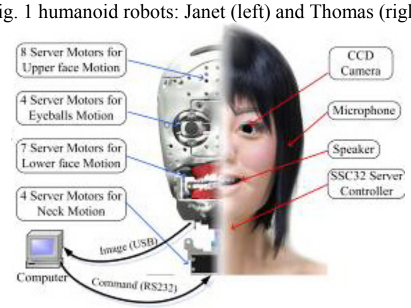

Fig. 2 System architecture of the android head

Since the silicon face skin is soft and stretchable, a support frame is needed to keep it in position, as shown in Fig. 3. The surface of the support frame is complicated and therefore is produced using a rapid prototyping technique. Many holes are made on the frame, through which wires attached to the servo motors can be connected to the interior surface of the face skin. For easy installment and maintenance, the frame is composed of three parts, the forehead module, the eye module, and the mouth module.

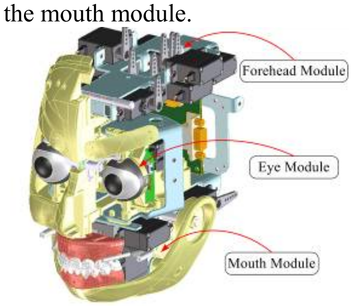

Fig. 3 RP face skin support frame.

B. Twin-wheeled robots

Pica and Ringo are two twin-wheeled self-balancing robots, as shown in Fig. 4, Pica the left and Ringo the right. Two robots have similar structures. Pica’s major function is drawing, and Ringo’s is drumming. They can balance, move forward, backward, and turn around on two wheels. Their foundations are similar to twin-wheeled man-carrying electrical car of Segway HT [13]. They have robotic arms similar to Janet and Thomas. The specification of two dual-wheeled robots can be seen in TABLE 2. The control system of their bodies is divided into two parts; one controls foundation balance by PIC18 chip, and the arm motion control is driven by a computer as in androids.

Fig. 4 Twin-wheeled robot: Pica (left) and Ringo (right)

TABLE 2

THE TWIN-WHEELED ROBOT SPECIFICATIONS

Parameters	Values/details
Height	124 cm
Weight	50 kg
DOF	20(Arms)+ (Wheels)
Vision Sensor	CCD cameras
Sensor	Gyro
Actuator	DC motor
Battery	Ni-MH
CPU	Pentium 4 M 1.6 GHz
OS	Linux

Two-wheel balancing platform system was designed based on two-wheel differential concepts [14], mainly composed with two symmetrical sets of driving units, transmission units and framework. The off-centered driving unit delivers kinetic energy to wheels through the kinetic transmission system. Because the dual-wheel balancing platform is an independent system, all its control system must be enclosed in it. The experiment result shows that these robots can balance, move forward and backward, and turn around. Fig. 5 shows the angle and displacement histories during the robots movement from toppled to standing in balance. The robot can achieve balance phase in 1.05 seconds.

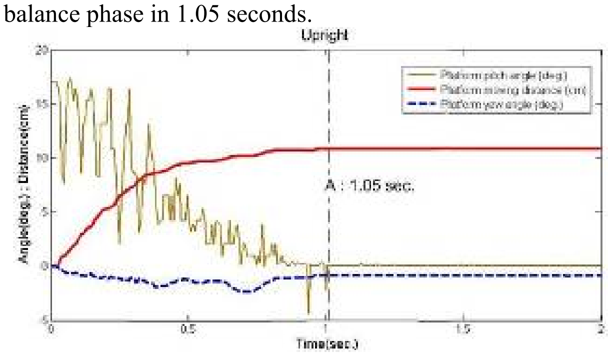

Fig. 5 The angle and displacement histories from toppled to standing in balance

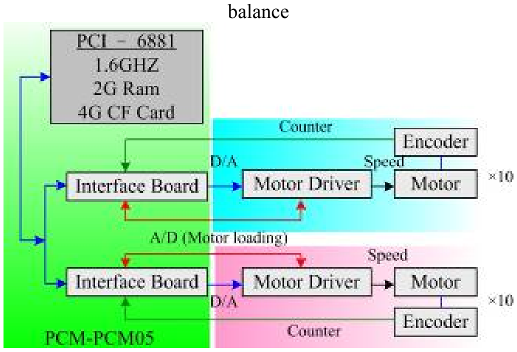

Fig. 6 The arm control system of the two-wheeled robot

To control the arm system, two interface boards are used, each controlling ten motors, include the 7 DOFs on each arm, and 3 DOFs of each hand. The hardware system is depicted in the Fig. 6.

III. AUTONOMOUS SIMPLIFIED MUSICAL NOTATION READING AND SINGING

A. *Simplified Musical Notation Recognition*

The experimental webcam installed in the eyeball of the wheeled robot is the Logitech product QuickCam® Sphere MP. Its video capture can be up to 640x480 pixels, still image capture 1280x960 pixels, and frame rate 30 frames per second. Fig. 7 illustrates the flowchart of our image capturing and recognition system.

Once the captured image of simplified musical notation completely presents on the screen by properly locating the mark, the system will start to detect the fringes of the paper from the neighbourhood of the mark in a clockwise tracking sequence. Here, the boundary of the paper is identified by the Sobel edge detector [15].

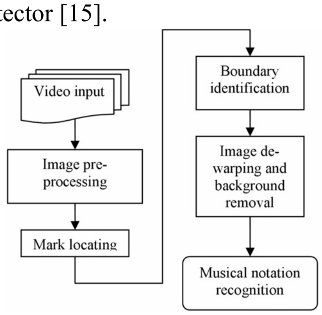

Fig. 7 Flowchart of our image capturing and recognition system

According to the extent of the simplified musical notation illustrated on the screen, those pixels inside the extent are preserved and the system continue to exploit four corner points of the paper to constitute a skewed quadrilateral image used for the calculation of de-warping parameters [16]. An image interpolation method is implemented to increase its resolution.

The recognition process consists of four steps. In Step 1, the dewarped image without background is binarized by using Otsu's thresholding method [17]. In Step 2, the binarized image is partitioned into non-overlapped blocks where each block contains a simplified musical notation or an English character by using the projection techniques corresponding to x-axis and y-axis, respectively. In Step 3, for each block, the affine-invariant matched value between the block containing the simplified musical notation and the model block stored in the database in advance is determined; and then the matched value between the simplified musical notation and the model block is determined. The matched value is used to interpret the meaning of simplified musical notation in the block as shown in Fig. 8. Finally, English characters are merged into meaningful words according to the distance between two adjacent English characters. The recognition output, which includes English words and simplified musical notation, will be used as the input of the Mandarin singing voice synthesis system.

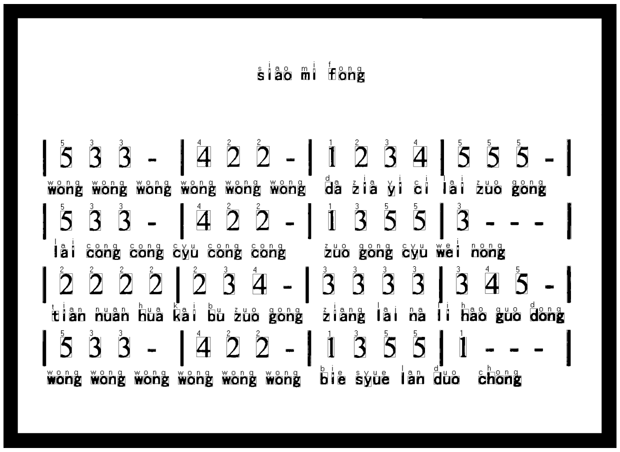

Fig. 8 The interpretation result of the numbered musical notation.

B. *Singing Voice Synthesis*

To synthesize music signals, additive synthesis, subtractive synthesis, and FM (Frequency Modulation) synthesis are notable techniques [18]. In this paper, however, the technique of HNM (harmonic plus noise model) originally proposed by Y. Stylianou [19] is used as a foundation and then extended. HNM is selected so that the synthetic Mandarin singing voice would have a higher signal clarity and naturalness level. HNM splits the spectrum of a signal frame into two halves of unequal widths to better model the spectrum. The lower frequency half is modelled as harmonic partials, while the higher frequency half is modelled as noise signal components.

Here, the syllable is selected as the unit for synthesis processing. This is because Mandarin is a syllable-prominent language, and each syllable is of the structure CxVCn. The Cx of a syllable may be null, a voiced consonant, or an unvoiced consonant, while the Cn may be null, or nasal, as in /n/ or /ng/. Additionally, the V may be a vowel, diphthong, or triphthong. If the Cx is a long unvoiced consonant (e.g., /s, p/), its synthetic signal will be generated as a noise signal with HNM. If the Cx part is a short unvoiced consonant (e.g., /b, d/), its synthetic signal will be directly copied from the corresponding part in the recorded syllable. Otherwise, the Cx is a voiced consonant (e.g., /m, r/) and is considered together with the remaining phonemes. Then, their synthetic signals are generated as harmonic partials plus noise signal with HNM.

IV. Human Portrait Generation System

This function aims to develop a human portrait generation system that enables the two-armed humanoid robot, Pica, to autonomously draw the face portrait of the person sitting in front of Pica. This portrait generation system converts a face image captured by the CCD camera installed on the head of Pica, to line segments that constitute a portrait of a good artist quality and are suitable for the robot arm to draw within a short period of time. A selected reduced number of pixel points on the line segments of the portrait are used to control the motion of the robot arm. The control points on the portrait plane are then automatically transformed into the robot's coordinates. A PD controller drives the motors of the robot arm to complete the real-time portrait drawing and signature.

The image algorithm will first use the face detection function [20] to locate the human face, then use the center-off techniques [21][22] to attain the representative curves and line segments on the face, and next use the median algorithm to remove unwanted noises. After that, the facial feature processing techniques [23] are implemented to locate eyebrows, eyes, nose, and mouth. The hair regions are located by using the segmentation techniques. The final line segments representing the face portrait are recorded and the coordinates of the key points on these line segments are sent to the arm motion control algorithms to prepare the portrait painting.

At the end of the image processing, a large number of point coordinates representing the human face portrait are recorded as in the 2D drawing paper coordinate system. These coordinates will be transformed to the global 3D world coordinate (x,y,z) system for the robot. Based on the 3D coordinates, the joint angle trajectory of Pica's drawing arm is computed.

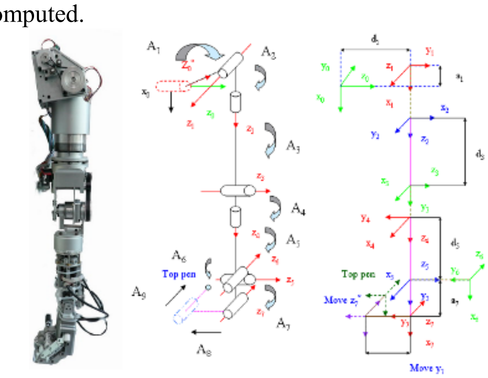

Fig. 9 Pica's humanoid robot arm with 7 DOFs

Three main methods are often used to obtain the inverse kinematics solutions, the closed form, the geometric and the numeric solutions. The closed form (algebraic form) and the geometric inverse kinematics are significantly more efficient than the numeric solutions. But the geometric form inverse kinematics solutions become complicated for Pica's arm because it has the angle (15 degree) at the first joint. Therefore, the closed form inverse kinematics is used in this case. Fig. 9 shows the configuration of Pica's 7 DOF arm and its drawing pen, and also the representation of the motor distribution, link lengths, and the local coordinate systems.

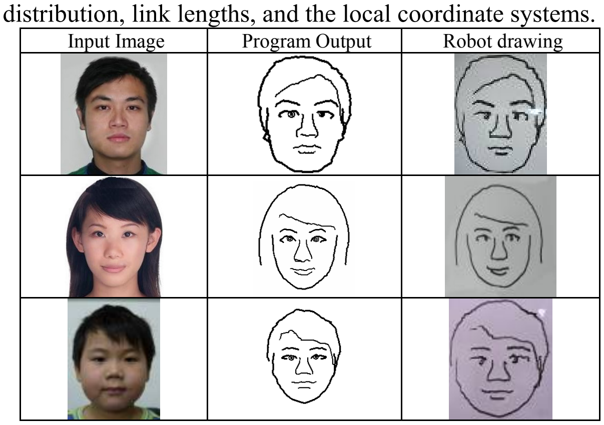

Fig. 10 Robot portrait drawing results

A number of people have been used on the real-time robot portrait drawing experiments. As long as the illumination conditions and the face restrictions are met, the real-time portrait drawing success rate is quite high. Fig. 10 shows the CCD captured image, the image process algorithm generated image and the robot drawn portrait for 3 people. It took 4-6 minutes to finish the portrait drawings.

V. REALIZATION OF ROBOT THEATER

Dec. 27, 2008, the first robot theater show was performed publicly in National Taiwan University of Science and Technology, Taipei, Taiwan. The show was divided into 5 acts. Act 1 was Pica's real time singing with musical notes, as shown in Fig. 11. After Pica read the lyric extemporaneously composed by the audience, she sang the simplified musical notation through image recognition and voice synthesis.

Fig. 11 Pica performs real time read and sing

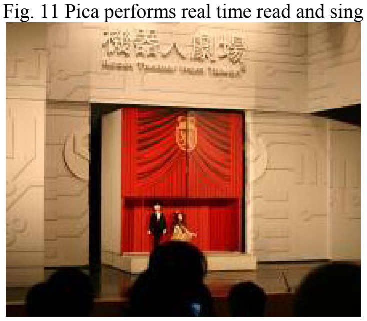

Fig. 12 Janet and Thomas playing puppets

Act 2 was Janet and Thomas manipulating puppets behind a small stage (see Fig. 12), to perform a section from "The Phantom of the Opera". In the following Act 3, Janet and Thomas opened the door on the small stage and came out. They became the hero and heroine in "The Phantom of the Opera" (see Fig. 13).

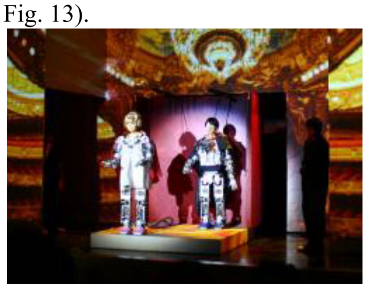
Fig. 13 Janet and Thomas are singing "The Phantom of the Opera"

In Act 4, Pica, as a painter, eased up the audience's emotion. We invited a child to sit in front of Pica. Pica caught the child's image with CCD and immediately sketched the child's portrait on a drawing board. The entire section took approximately 5 minutes (see Fig. 14).

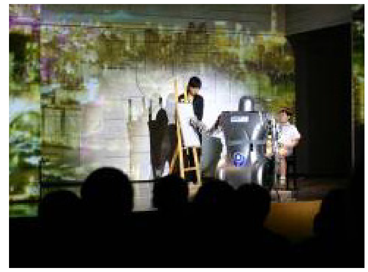
Fig. 14 Real time human face portrait by Pica

The last Act was Ringo drumming and Janet and Thomas dancing, thus bringing the show to culminating point (see Fig. 15). As seen in Fig. 16, at the end of the performance, two robots kissed each other in applause, to commemorate the first moment male and female androids kissing each other in history.


Fig. 15 Robots playing drums and dancing

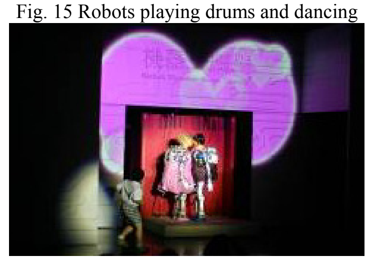
Fig. 16 Historical robot kiss

The first performance of the robot theatre was almost smooth. However, during the rehearsal, we found a couple of problems needed to be solved as the following:

1) Communication problem between the robots and the control sides: All the robots were controlled through wireless network. The four robots required totally 15 computers connected to the wireless network to control their synchronizing performance motion. The network needed to deliver image, sound, and control commands, thus it must be provided with broad bandwidth and stable AP (Access Point). In this performance, it was found that the wireless network often lost signals. After testing APs from four brands, the wireless network was abandoned, and wire network replaced to solve the communication problem. In the future, theater performance studies must consider the wireless network's problem. Solutions may include decreasing image and sound delivery, reducing broad bandwidth occupation, increasing signal strength of wireless network, and preventing signal interference.
2) Robot battery problem: The entire performance time in the robot theatre was approximately 50 minutes. For humanoid robots, 4200 mAh battery can be only used for 20 minutes. The batteries need to be replaced three to

four times during the show, increasing risk to the performance. While replacing batteries, outside power source must be connected to the robots, or the robots shall be completely shutdown. If the power source was unstable during replacing, the computers might be crashed. To restart the computer systems would take 6-8 minutes, thus performance progress might be delayed. Therefore, the robots needed outside power source. However, it's not adorable to expose wires on the stage. The practical power charging system on the stage is a valuable research topic and one goal we will be working on.

3) Heat emission of robots: Environment on the stage and in the lab was completely different. Intensive lighting on the stage would raise ambient temperature. During theatric operation, the heat-emitting devices, such as robots' computers, motors and motor drivers, need higher cooling efficiency.

VI. Concluding Remarks

Two androids and two wheeled robots are made for the robot theater, performing various entertaining shows. Androids are able to simulate many human face expressions and thus suitable for many characters on the stage. Their mouths can make proper shapes to harmony with Chinese and English languages. The wheeled robots can make self-balancing on two wheels. Their arms can complete performance motions including sketching and drumming with support from application programs. We will continuously devote our research efforts in making more robot characters for robot theaters, and develop more performance techniques. We believe that in the future robot theaters will become one of important intelligent robot applications and great human entertaining activities.

Acknowledgment

This research was financially funded by the National Science Council of the Republic of China (Taiwan) under the grant number NSC 96-2218-E-011-002. Their support made this research and the outcome possible.

References

[1] M. Fujita, and H. Kimo, "Development of a quadruped robot for robot entertainment," *Autonomous Robots*, Vol. 5, pp. 7-18, Kluwer Academic, 1998.

[2] M. Fujita, K. Sabe, Y. Kumki, T. Ishida, and T. Doi, "SDR-4X II: A small humanoid as an entertainer in home environment," *International Symposium of Robotics Research*, Siena, Italy, Oct. 2003.

[3] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, and K. Fujimura, "The intelligent ASIMO: system overview and integration," *IEEE/RSJ International Conference on Intelligent Robots and System,* Vol. 3, pp.2478-2483, 2002.

[4] Y. Hosoda, S. Egawa, J. Tamamoto, K. Yamamoto, R. Nakamura, and M. Togami, "Basic Design of Human-Symbiotic Robot EMIEW," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 5079-5084, 2006.

[5] R. HIURA, KEN ONISHI, H. OKAZAKI, and S. ASANO, "Development and demonstrate of reseptionist robot "wakamaru"," Nippon Kikai Gakkai Robotikusu, Vol.2006, pp.1A1-E02, 2006.

[6] J. Oh, D. Hanson, W. Kim, I. Han, J. Kim, and I. Park, "Design of Android type Humanoid Robot Albert HUBO," *IEEE/RSJ International Conference on Intelligent Robots and Systems*, Beijin, China, 2006.

[7] J. Hirth, N. Schmitz, and K. Berns, "Emotional architecture for the humanoid robot head roman," *IEEE International Conference on Robotics and Automation (ICRA)*, Rome, Italy, April 11-13 2007, pp. 2150–2155.

[8] D. Hanson, "Expanding the Aesthetics Possibilities for Humanlike Robots." *IEEE Humanoid Robotics Conference*, special session on the Uncanny Valley, Tskuba, Japan, December 2005.

[9] M. Blow, K. Dautenhahn, A. Appleby, C. L. Nehaniv, and D. Lee, "The Art of Designing Robot Faces - Dimensions for Human-Robot Interaction," *Proceedings of HRI06*, pp. 331-332, 2006.

[10] H.S. Lee, J.W. Park, and M.J. Chung, "A Linear Affect–Expression Space Model and Control Points for Mascot-Type Facial Robots," IEEE Transactions on Robotics, vol. 23, no. 5, pp. 863-873, 2007.

[11] Toyota Partner Robots. (2008, December). http://www.toyota.co.jp/en/special/robot.

[12] Yu Ogura, Kazushi Shimomura, Hideki Kondo, Akitoshi Morishima, Tatsu Okubo, Shimpei Momoki, Hun-ok Lim, and Atsuo Takanishi, "Human-like Walking with Knee Stretched, Heel-contact and Toe-off Motion by a Humanoid Robot," *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3976, 2006.

[13] Segway. (2008, December). Available: http://www.segway.com.

[14] Siegwart, Roland and Nourbakhsh, Illah R, *Introduction to Autonomous mobile Robots*, Massachusetts Institute of Technology, MIT Press, Cambridge, 2004.

[15] R. C. Gonzalez, and R. E. Woods, *Digital Image Processing*, 2nd Ed., Prentice-Hall, Upper Saddle River, NJ, 2002.

[16] Z.C. Li, T.D. Bui, Y.Y. Tang, and C.Y. Suen, *Computer Transformation of Digital Images and Patterns*, World Scientific, Hackensack, NJ, 1989.

[17] N. Otsu, "A threshold selection method from gray level histogram," *IEEE Trans. on Systems, Man, and Cybernetics*, SMC–8, pp. 62–66, 1978.

[18] C. Dodge, and T. A. Jerse, *Computer Music: Synthesis, Composition, and Performance*, second edition, Schirmer Books, New York, 1997.

[19] Y. Stylianou, "Modeling Speech Based on Harmonic plus Noise Models", *Nonlinear Speech Modeling and Applications*, Springer-Berlin, pp.244-260, 2005.

[20] A.L. Yuille, P. Hallinan and D.S. Cohen, "Feature Extraction from Faces Using Deformable Templates," *International Journal of Computer Vision*, Vol. 8, No.2, pp. 99-111,1992

[21] Chyi-Yeu Lin, and Chien-Wei Li, "Human Portrait Generation System for Robot Arm Drawing," *The 13th Conference on Artificial Intelligence and Applications*, Taipei, Taiwan, 2008.

[22] Xiaoyang Mao, Yoshiyasu Nagasaka and Atsumi Imamiya, "Automatic Generation of Pencil Drawing From 2D Images Using Line Integral Convolution," *CAD/Graphics*, 2001.

[23] Shigefumi Yamamoto, Xiaoyang Mao and Atsumi Imamiya, "Enhanced LIC Pencil Filter," *IEEE Computer Graphics, Imaging and Visualization Proceedings*, 2004.