Proceedings of the 2002 IEEE/RSJ
Intl. Conference on Intelligent Robots and Systems
EPFL, Lausanne, Switzerland • October 2002

The intelligent ASIMO: System overview and integration

Yoshiaki Sakagami Ryujin Watanabe Chiaki Aoyama
Shinichi Matsunaga Nobuo Higaki Kikuo Fujimura*

Honda R&D Co.Ltd.
1-4-1 Chuo Wako-shi Saitama, Japan
*Honda R&D America Inc.
800 California Street Mountain View, CA 94041

Abstract

We present the system overview and integration of the ASIMO autonomous robot that can function successfully in indoor environments. The first model of ASIMO is already being leased to companies for receptionist work.

In this paper, we describe the new capabilities that we have added to ASIMO. We explain the structure of the robot system for intelligence, integrated subsystems on its body, and their new functions.

We describe the behavior-based planning architecture on ASIMO and its vision and auditory system. We describe its gesture recognition system, human interaction and task performance. We also discuss the external online database system that can be accessed using internet to retrieve desired information, the management system for receptionist work, and various function demonstrations.

1.Introduction

Autonomous robotics has been an active research area for long time. Humanoid robots are especially desirable in human society as they can work well in indoor environments that have been designed for humans. Having a human form makes it easier for people to identify as compared to other forms. Humanoid robots like ASIMO can potentially assist humans in their daily tasks, bringing additional value to the human society.

Currently, however, humanoid robots are not skilled enough to perform many tasks that humans routinely do. This is because their sensory system is poor compared to humans. Robots have to process several types of raw sensory data such as orientation obtained by a gyro sensor, forces obtained by force sensors, image pixel data obtained by cameras and sound signals obtained by microphones. Then there is filtered intermediate level data such as velocity, acceleration, optical flow, edges and color. Finally there is high-level identifiable data like human face, gesture, posture, staircase, door, and natural language sentences.

We have developed a sensory system that processes multiple types and levels of data on ASIMO. We have also developed functions to understand human intentions to interact with them and to perform various tasks.

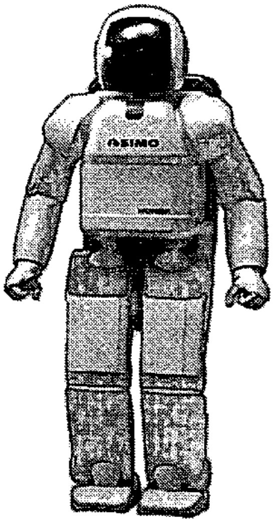

Figure 1: ASIMO

2.Related work

We have developed biped walking robots, humanoid robots P1, P2, P3 and ASIMO (Figure 1) over the past 16 years. Our initial goal was to realize a biped walk control to walk and turn in any direction, as well as for going up or down stairs in ordinary indoor human environments [1]. In this paper, we focus on the vision and auditory system for human interaction on ASIMO.

Various vision and auditory systems have been developed for robot intelligence. Applications vary from robust navigation in uncertain environments to identification of humans and interaction with them using gestures and voice. The humanoid robot Cog [2] realized humanlike intelligence by distributed computer systems outside the robot body. SDR-3X [3] is a small stand-alone humanoid entertainment robot. This robot has a vision system so as to detect colored balls and keep balance on slanted floor. Flo [4] is wheel type robot to serve elderly people, providing healthcare and other information related to activities of daily living. In this system, map based navigation as well as human

0-7803-7398-7/02/$17.00 ©2002 IEEE

interaction using voice, face detection and tracking were used. Robovie [5] is a wheel type robot for interacting with people by generating speech based on joint attention. This system is able to draw the person's attention to the same sensor information as the robot and omit words that are clear from the context.

There are various approaches for robot intelligence. Artificial intelligence techniques for simulating human thinking include symbolic processing, cognitive modeling based on brain mechanism, and emerging new logics such as in artificial life. Our approach is to model the relationship between sensory information and behavior directly.

In section 3, we describe the system hardware and software structure. In section 4, we describe the vision and auditory sensing system, navigation and human interaction. In section 5, we explain the planning behavior based architecture. Sections 6, 7, and 8 contain discussions on external database system, demonstrations and conclusions, respectively.

3.System structure

The current version of ASIMO is a highly autonomous system compared to our first version that used external computation for planning and action selection by GUI. The ASIMO hardware specification is described in Table 1.

Table 1: *ASIMO hardware specification*

Weight	52kg
Height	120cm
Moving velocity	0-1.6km/h
Biped cycle	Variable cycle/step
Grasping force	0.5kg
Actuator	Servo+Harmonic
Leg Force Sensor	6 axis force sensor
Body Sensor	Gyro, acceleration
Power supply	38.4V/10Ah(Ni-MH)
Head	2DOF
Shoulder	3DOF x 2
Elbow	1DOF x 2
Wrist	1DOF x 2
Finger(Grasping)	1DOF x 2
Crotch	3DOF x 2
Knee	1DOF x 2
Leg	2DOF x 2

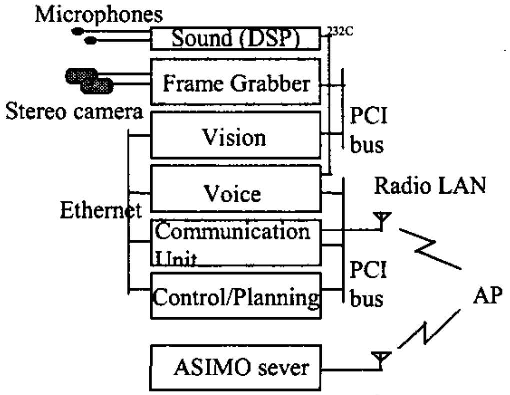

Figure 2: *ASIMO computational structure*

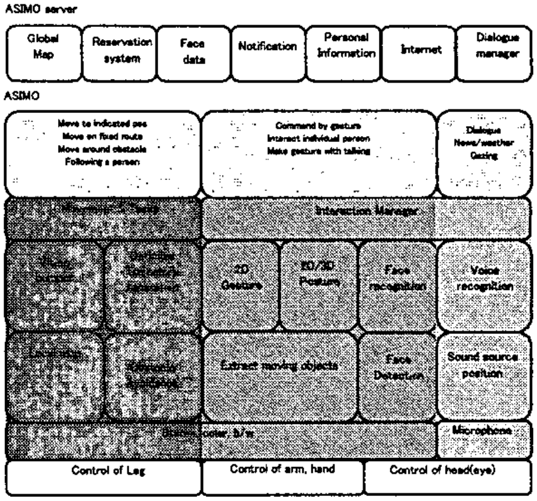

Figure 3: *Function and software*

A vision and auditory system has been installed on ASIMO for navigation in ordinary environments and for human interaction. The intelligence system consists of a frame grabber, a PC for image processing, a PC for speech recognition and synthesis, a processor for control and planning, a radio communication network controller unit for communication with the external system and DSP board for detecting sound sources (Figure 2).

Two board color cameras are installed on head unit to obtain stereo images which are processed to compute depth. A frame grabber connects the vision computer with the PCI bus for high speed data transfer. The vision system is separated from the processor group of control and planning. Two microphones for sound detection system are installed on the front side of head. Ethernet is used to communicate among processes with different speeds (e.g. slow vision process, fast auditory system).

In the external system of ASIMO, we have developed a map management system for navigation and specification of tasks in man-made environments such as offices, museums, and hospitals. This system is able to send commands to the robot and allows selection of tasks for execution at specified locations (such as recognition and speech dialogue). There is also a database of face images of different people used to identify people and recognize them at subsequent meetings.

Current ASIMO uses several different operating systems, and a message board for asynchronous internal communication to achieve tasks and motions. The control system needs fast processing to perform actions. Vision processing is slower compared to control processing. The planning system has to handle both fast (e.g. obstacle avoidance) and slow situations (e.g. route trace moving).

A communication server manages socket port numbers for communication among vision and planning processes. The planning system is an agent-based distributed architecture system. The planning system is event driven with no central control to handle unforeseen situations. The construction of various functions and softwares of ASIMO is shown in figure 3.

4.Sensory system

4.1 Camera system and image capture

The vision system of ASIMO runs on a PC. The stereo method is based on SAD (Sum of Absolute values of Differences). It computes a depth map from two CCD cameras with b/w images, and does calibration for lens distortion and rectification (Table 2). The cameras in the head are located on top of the robot body with many degrees of freedom. To calculate camera pose, image capture is synchronized with all joint angles. The frame grabber has several types of I/O signals and captured images are synchronized with body motion (Table 3).

The vision system for navigation and interaction takes images from the frame grabber and processes them to extract 3D objects and moving objects (Figure 4).

Table 2: *Specification of stereo system*

Base line length	74mm
Imaging sensor	1/3"Color CCD x 2
Picture elements	768(H)x480(V)
Focal length	4mm
CPU	Mobile Pentium III-M 1.2GHz
Disparity image size	320(H) x 240(V)
Disparity range	32 pixels
Stereo frame rate	20fps
Software processing	Correction of lens distortion and rectification

Table 3: *Specification of frame grabber*

bus interface	CompactPCI
input signal	S-Video/RS-170A
input channel	2ch(S-VIDEO). 2ch(b/w)
Sampling	NTSC 4fsc(14.31818MHz)
sync.	VS, HD, VD
LUT	Y 8bit→16bit, CrCb16bit→16bit
frame memory	256Mbyte(90frame ring buffer)
sync. Pulse	0~16msec/field. Delay;5VTTL
async. Pulse	ON/OFF range;0~16msec
time code	sync pulse stamp on frame
Shutter	SONY XC-ES alternate
Transfer	64bit DMA
host OS	WindowsNT 4.0

4.2 Speech system and sound localization

For speech recognition and synthesis, we use a commercial voice recognition engine and speech synthesis product. However, the audio quality and intonation of voice need more work and they are not yet satisfactory for use on the robot. We control sentence and words tags for smooth speech. Sound source detection is able to identify human voice tones and step sounds. The shape of envelope of sound signal and the direction of the sound is computed from the volume and time difference of the signals at two microphones. The sound detection resolution is one degree of separation over 5 meters of distance. When someone calls ASIMO, it turns its head to face the person. When something falls on floor, it turns its head to gaze what happened.

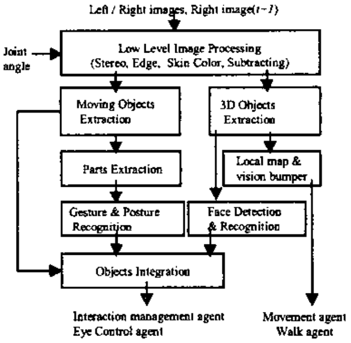
Left / Right images, Right image($t-1$)
Joint angle
Low Level Image Processing (Stereo, Edge, Skin Color, Subtracting)
Moving Objects Extraction
3D Objects Extraction
Parts Extraction
Local map & vision bumper
Gesture & Posture Recognition
Face Detection & Recognition
Objects Integration
Interaction management agent
Eye Control agent
Movement agent
Walk agent

Figure 4: *Vision system software configuration*

4.3. Obstacle detection for navigation

Many autonomous robots perform navigation under an ordinary environment using vision, sonar, infrared sensor and range finders. ASIMO makes an in memory local map using vision. It reconstructs a map of the neighborhood surrounding the robot and uses it to move to the point of interest on a pre-defined route. The vision bumper handles obstacle surrounding the robot with simple 8 bit pattern from the local map data (Figure 5). Obstacles detected from stereo depth map data are updated frame by frame in time sequence. Obstacles are modeled using their bounding box. ASIMO can turn its head horizontally from +/-83 degrees, and can therefore make a wider local map covering +/-113 degrees.

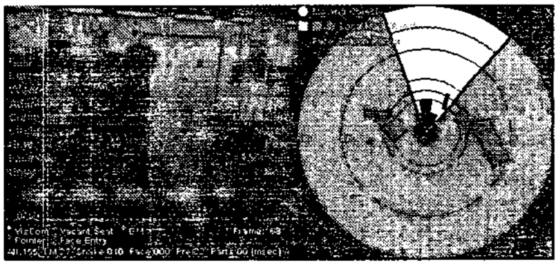

Figure 5: *Image and top view of surroundings. White region is robot's view. Purple rectangles are obstacle. Blue cone patterns are range of vision bumper.*

4.4.Human and gesture recognition for interaction

Interaction is important for any kind of robot to perform tasks in human society like carrying luggage, pushing a cart, serving drinks, taking tools from the table and so on. The robot has to understand human's high-level task requirements by using its low-level sensory modules such as eyes, ears and tactile sensors .

Human tracking algorithm is able to track humans and their actions. An optical flow based algorithm extracts the foreground from the image even when robot head and body are in motion. Snake algorithm extracts contours of human shapes and can separate multiple people in the scene. Human head position is estimated at the top of the contour.

Face detection makes use of a model of skin color to extract face contours. Face recognition is based on the Eigenvector Method [6] (Figure 6).

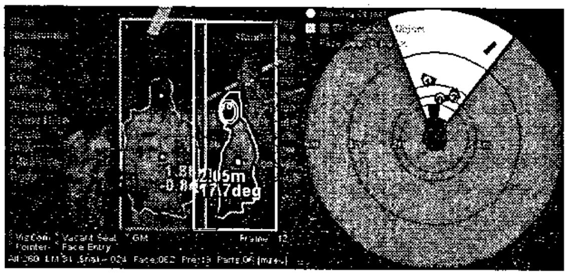

Figure.6: Extraction of humans in the scene. Each color region is human position, point in contour is the human contour center. Numbers describe the direction and length of human from robot. Oval on face signifies recognized person.

Our 2D gesture recognition algorithm detects the position of the hand and estimates the action using a Bayes statistical model. It is able to identify hand, face and side and front profiles of body. Recognized gestures include handshake, hand circling, bye-bye, hand swing, high-hand and come here call. Human gesture computation is done using the robot front view image (Figure 7).

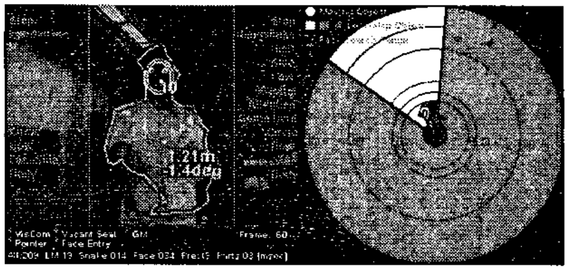

Figure 7: Handshake gesture recognition.

Our 3D gesture recognition algorithm can recognize pointed hand gesture based on the head and hand position relationship using depth map data from stereo (Figure 8).

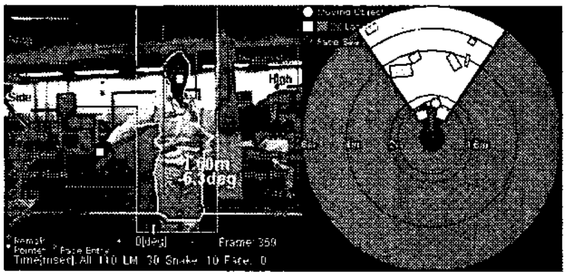

Figure 8: Pointing gesture recognition (blue bar)..

5.Planning system

5.1 Software architecture

Many models of autonomous robot architecture for navigation and interaction have been proposed. RHINO [7] gave tours in museum for visitors. The system was organized in hierarchical modules to control a robot.

The ASIMO planning architecture is behavior-based and has deliberative and reactive layers. Behavior-based architecture [8] combines deliberative planning like high-level human commands and reactive behavior control for navigation in a rapidly changing environment. Behavior agents use distributed processing, are event driven, use asynchronous communication, and have no supervisor in each layer. The behavior agents in our system are listed in table 4.

***Table 4:** Functions of planning agents*

1) Agents in Deliberative Layer

Movement Agent	Getting route data from Global map. Calculating direction towards sub goal and approach there. Obstacle avoidance by potential method. Issue command for task.
Interaction management agent	Switching scenario of task. Selecting a dialogue depending on task. Selecting an action using posture/gesture recognition result and voice recognition.
Dialogue Agent	Switching dialogue based on voice recognition. Switching a dialogue to speech with action Retrieving and making dialogue form Internet.

2) Agents in Reactive Layer

Walk Agent	Making a step command for walk depends on surrounding objects. Obstacle avoidance by vision bumper.
Eye control Agent	Attention a moving object. Attention and gazing sound. Gaze by command. Sleeping./ look around as no any input.
Sound source detection Agent	Detecting a position of sound source. Evaluation of human voice tone.
Robot control interface agent	Walk command, action command accepted from other agent. Broadcast a latest state of robot.

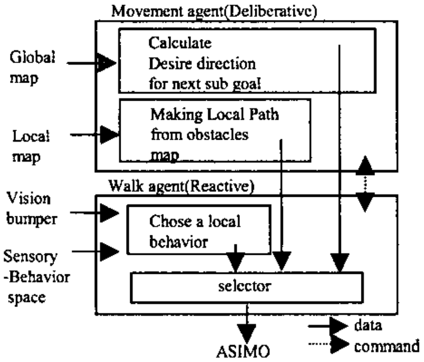

Figure 9: Obstacle avoidance mechanism

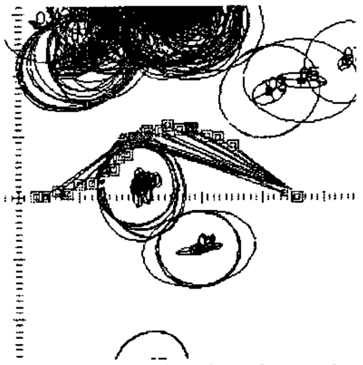

Figure 10: New route for obstacle avoidance.Red rectangles show generated local path. Blue oval are obsatcles.

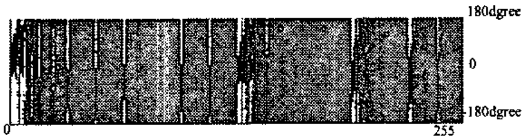

Figure 11: Sensory-behavior space. Vertical dots; a case of vision bumper pattern (256). Horizontal dot; safe direction to move (5 degrees/dot). Black; not safe to move, White; safe to move. Gray; no data.

Each agent handles data from the vision system, auditory system and other agents. The cycle time of deliberative agents is larger than that of reactive agents. For example, *Movement agent* makes a route every 500 ms. to avoid static obstacles. Static obstacles are detected by a potential method using a local map data that is updated by the vision system every 120 ms. Object avoidance in *Walk agent* has a 33 ms. cycle. *Walk agent* can stop and turn the robot when an obstacle suddenly appears. It uses the sensory behavior space generated by reward and penalty based learning on a simulator to decide motion direction (Figures 9,10,11).

5.2 Integration

All agents run simultaneously and handle minimally necessary input data. Once the condition for agent invocation is satisfied, the agents become active and work on achieving their tasks. In case of conflicting behaviors, agents negotiate with each other by sending commands.

Agents receiving commands inhibit or modify their behaviors by updating parameters for acting behavior. For example, when *Interaction management agent* gets a command from a known person, it forwards the same command, frame number of image, and other information to *Dialogue agent*, *Move agent* and *Eye control agent*. *Dialogue agent* selects and executes appropriate speech synthesis. *Move agent* inhibits motion behavior toward sub goal. *Eye control agent* receives data of a moving object with frame number at every cycle and selects the object to gaze.

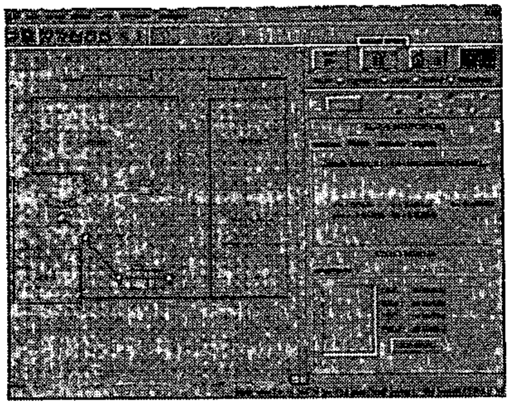

Figure 12: A GUI of Global Map system.

For multi processor communication, we use a message board based communication in real time operating system and a socket based communication between real time system and non real time system. Vision and auditory systems are part of non real time system. A name server mechanism is used to manage port numbers to establish communication among agents. Agents can start or terminate a process any time.

6.External system

6.1 Database of individual information

To recognize known people, the face recognition algorithm uses eigenvector algorithms on face and eye image data. Known people's face image database is kept on the ASIMO server. When the robot system initializes, the algorithm takes the stored eigenvector data from server. Other information like names, company names, address and hobbies is also available.

Processes in vision and planning system can refer to the external database via the network.

6.2 Reservation system

The task of ASIMO receptionist is to guide a visitor to a meeting room or a waiting place and to call a company representative to notify him about the visitor. The reservation system allows company people expecting visitors to choose meeting place and time and obtain visitor details in a web GUI. When a visitor approaches ASIMO, the moving human shape is extracted, and face is detected and recognized. Then the interaction manager confirms the schedule of meeting with the reservation system and sends a notification command to client program on user PC. Once the company user replies to the notification, ASIMO guides the visitor to the meeting place specified in the web GUI.

6.3 Global map system

The Global map system GUI is shown in figure 12. It supports the following functions:

1) Loading, saving, updating and scaling a map data.

2) Playing an object at any position on map, changing its characteristics and manipulating it.

3) Making a route and sub-goal.
4) Selecting a route with minimum cost.
5) Switching a route from planning agents.
6) Setting a task (e.g. talking, action, image processing) at sub-goal.
7) Deciding a region for walking.
8) Sending a command (start, stop, home, pause) to planning.
9) Monitoring robot behavior state.

7. Demonstrations

The intelligent ASIMO can perform reception work autonomously in the following basic order (Figure 13, Video 1)
1) Detect an approaching visitor.
2) Find and track a visitor's face.
3) Verbally greet the visitor with a gesture.
4) Recognize the individual and confirm.
5) Check appointment meeting time, place and people.
6) Notify the visitor's arrival.
7) Guide visitor to meeting place.

The functions for navigation are as follows (Video 2):
1) Walking on a route specified in the GUI of Global map system, and performing pre-specified tasks at sub-goal.
2) Detecting obstacles.
3) Stopping if an obstacle is encountered.

The other interaction functions are as follows:
1) It is possible to call the robot. This is useful when calling person is far, and/or the voice command has too much noise. The robot first finds the direction of calling person by sound source detection and tries to find a call sign gesture. If a call sign is identified, the robot approaches the calling person and follows him/her. The robot can also be stopped by gesture. (Video 3)
2) The robot knows handshake gesture of people. This is a good and basic action at first meeting.
3) Sometime, we want ASIMO to wait at the some location. 3D hand gesture can be used to point a location to the robot where it should go and wait. (Video 4)

8. Conclusion

We have integrated autonomous functions for navigation and interaction in ASIMO. A stereo camera, frame grabber and a computer image processing have been added to the current rental model. Equipped with the vision system, the robot can not only navigate in an ordinary environment, but also understand human requirements. Our auditory system is useful for interacting with people and understanding commands from a distant location. The deliberative and reactive architecture can perform high level planning and rapid response. The vision, auditory, and planning systems also use information from the external database system.

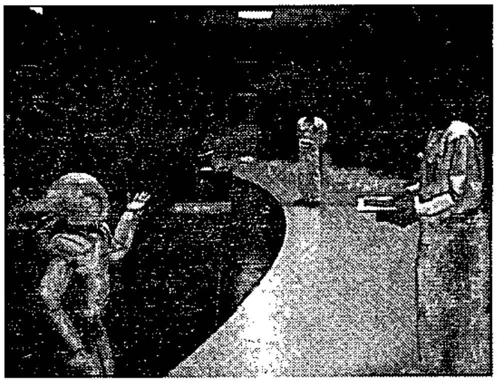

***Figure 13:** Receptionist ASIMO.*

In the future we plan to develop a more robust sensory and planning system to realize highly autonomous functions on ASIMO. Our goal is to demonstrate a personal robot that can perform helpful tasks to support human daily activities.

References

[1] Hirai, K.,Hirose, M., Haikawa, Y., Takenaka, T., ìThe Development of Honda Humanoid Robotî, Proc. of the 1998 IEEE International Conference on Robotics & Automation, pp.1321-1326, 1998.

[2] Brooks, R., Breazeal, C., Marjanovic, M., Scassellati, B., Williamson, M., ìThe Cog Project: Building a Humanoid Robot, , Computation for Metaphors, Analogy and Agents, Vol. 1562 of Springer Lecture Notes in Artificial Intelligence, Springer-Verlag, 1998.

[3] Ishida, T., Kuroki, Y., Yamaguchi, J., Fujita, M., Doi, T., ìMotion Entertainment by a Small Humanoid Robot Based on OPEN-Rî, Proc. of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.1079-1086, 2001.

[4] Galtus, G., Fox, D., Gemperle, F., Goetz, J., Hirsch, T., Magaritis, D., Montemerolo, M., Pineau, J., Roy, N., Schulte, J., Thrun, S., ìTowards Personal Service Robots for the Elderly.î, Computer Science and Robotics, Carnegie Mellon University, 1998.

[5] Imai, M., Ono, T., Ishiguro, H., ìPhysical Relation and Expression: Joint Attention for Human-Robot Interactionî, Proc. of 10th IEEE International Workshop on Robot and Human Communication, 2001.

[6] Turk, M., Pentland, A., ìEigenfaces for Recognitionî, Journal of Cognitive Neuroscience, Vol.3, No.1, pp.71-86, 1991.

[7] Burgard, W., Cremers, B. A., Fox, D., Hahnel, D., Lakemeyer, G., Schulz, D., Steiner, W., Thrun, S., ìExperiences with an interactive museum tour-guide robot.î, Artificial Intelligence 114, pp.3-55, 1999.

[8] Arkin, R.C., Behavior-based Robotics, MIT press, 1998.