Embodied Communication Between Human and Robot in Route Guidance

Guillermo Enriquez, Yoshifumi Buyo, and Shuji Hashimoto

Science and Engineering, Waseda University
3-4-1 Okubo, Shijuku-ku, Tokyo, Japan 169-8555
{enriquez,buyo,shuji}@shalab.phys.waseda.ac.jp

Abstract. Walking with someone or guiding them to a destination is a simple task that humans do everyday. However, while almost any mobile robot can navigate to a given point while a human walks behind it, these robots do not take into consideration whether the human is following along properly. Our research involves the use of a wireless sensor network for navigation of a guidance robot and a vision-based tracking system for human awareness. The network creates a virtual directional field that provides a directional imperative to the robot, and the vision system allows it to detect if a following human has strayed. We also present a demonstration of our system in a real world application.

Keywords: Human-aware, guidance, tracking, wireless sensor network.

1 Introduction and Objectives

The act of guidance is a simple task that we humans do without much thought. However, in reality, there are many variables to be considered that our brains handle subconsciously. Factors such as path planning, distance from person being guided, and obstacles are all taken into consideration as we move. Robots however are incapable of making these decisions without human programming. While there have been robots that guide the humans in various places such as museum and railway station[1]-[4], these robots did not take into consideration whether or not the humans we actually following. Our system not only moves the robot from its present location to a desired destination, but it also takes into account the present location of the human being guided. Should the human stray, the robot takes measures to collect the human before continuing on to the final destination as described in the flowchart of the robot behavior Fig.1. In such way there exists an embodied communication between human and robot even in simple route guidance. While previous robots utilized vision and maps to navigate the environment our system uses a wireless sensing network to navigate to the destination. In addition, it makes use of a vision tracking system to determine whether or not the human is following the robot properly. In order to test our system, we created a mock-environment to simulate a restaurant.

M.J. Smith, G. Salvendy (Eds.): Human Interface, Part I, HCII 2007, LNCS 4557, pp. 824–829, 2007.

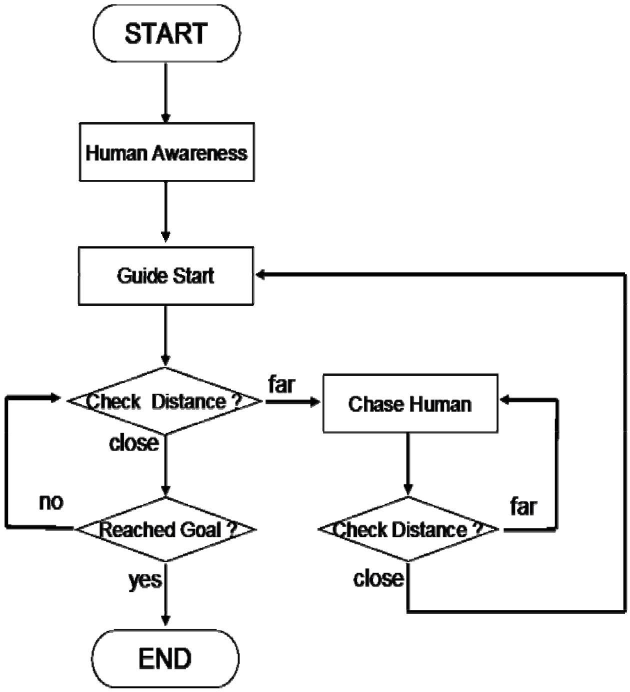

Fig. 1. The flowchart of the guide robot system

2 Navigation Through a Directional Field

Our navigation system has two facets, depending on the state of the robot; the wireless network, and the vision-based tracking system. The wireless sensor network comprised of a number of MICAZ motes and a mobile mote that is mounted on the robot. The network provides general localization and is the basis of the directional field. These motes are pre-programmed with their position within the room. As major static obstacles in restaurants do not often change, pre-programming the map is not costly.

2.1 Wireless Sensor Network

In recent years, interest in wireless sensor networks has grown increasingly due to their wide array of applications. They are ad-hoc, small in size, low in cost, and very scalable. For our system we used MICAZ (MPR2400J) motes and an MIB520 programming board. A total of eight motes were dispersed around the experiment room and pre-programmed with their location and other information. The pre-programming of the motes is not costly in our experiment as restaurants rarely change the position of major obstacles such as walls, partitions, or tables. A ninth mote was used in conjunction with a MIB520 board as an interface between the robot and the sensor network. The mote on the robot, or mobile mote, runs an altered version of the network program allowing localization to be processed.

2.2 General Localization

In order to perform general localization the mobile mote reads a series of signals from the neighboring motes. The signals emitted by the motes, like all radio signals, decay and weaken in strength as they travel. In [5] a localized mobile robot was used to determine the positions of un-localized motes. Our system implements the reverse in order to localize the mobile robot. Each time the approximate position of the robot is calculated, a distance to each readable mote is taken. These distances are used to create bounding boxes around each mote. Essentially an imaginary box around each mote is created with the mote at the center and a distance of *d* to each side, where *d* is the approximated distance from the current mote to the robot. The boxes are represented by their bottom-left and upper-right corners (1). An initial probability box of the robot is the entire environment. The intersection of all the bounding boxes is performed iteratively over all the motes, the final iteration representing the probable area in which the robot is currently located (2). For further calculation purposes, the approximate position of the robot is taken as the center of the probability box (3).

$$B.box = \{(x, y, d) : (blX, blY, trX, trY)\} \tag{1}$$

$$P.box = \{(B.box_i \cap P.box')\} \tag{2}$$

$$approx = \{(P.box) : (x, y)\} \tag{3}$$

2.3 Directional Field

The wireless sensing network performs two duties in our system. Not only is it used in localization, but it provides a directional field that provides basic directional information to the robot. This directional field can be imagined as a set of gradients

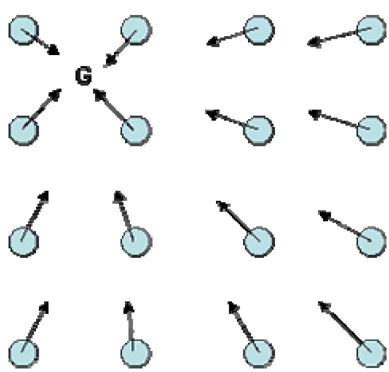

Fig. 2. A visual representation of the Directional Field concept

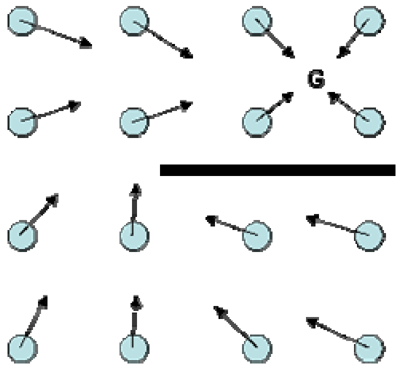

Fig. 3. A visual representation of the Directional Field concept with an obstacle

pointing downwards toward the goal as shown in Fig. 2. The robot examines these gradients and attempts to maintain a downward path. Essentially the robot is rolling downhill on a virtual slope created by the directional field. The directional field can also be used to avoid major obstacles as the directional vectors can be aimed around them as shown in Fig. 3.

3 Human Awareness

Our system provides human awareness through a visual tracking system. This tracking is performed by utilizing a certain difference in luminance value between the background region and the human region of the image. The system extracts human region by focusing on the color feature of the human's clothes.

3.1 Frame Entrance Recognition

The system is started by recognizing when a human appears behind the robot. This is done by frame subtraction method. The robot remains stationary and keeps capturing images until the human enters frame comes within a predefined distance. $I(x, y, t)$ represents the pixel value at (x, y) and at t of the input images. We get a binarized subtraction image $BSI(x, y, t)$ which is computed as

$$BSI(x, y, t) = \begin{cases} 255 & I(x, y, t) - I(x, y, t-1) \geq th \\ 0 & otherwise \end{cases} \quad (4)$$

Then, we apply erosion and dilation several times to the BSI in order to eliminate noise or small connected components. After that, we calculate the smallest rectangle which includes all pixels of value 255. When the height of the rectangle is over 90% of the input image, we assume that the human has come sufficiently close to the robot and the system can get the information of the human from the image such as color of clothes and other features. We call this rectangle the initial human region.

3.2 Feature Extraction

We apply the Histogram Back-Projection method to determine the probability distribution image of the human clothes region as follows:

1. Convert the input images to HSV color system.
2. Set the region of interest (ROI) in the initial human region.
3. Calculate H value histogram of the ROI. $\{\hat{q}\}_{u=0\ldots m}$ represents the frequency of the value u of H.
4. Transform pixel value based on the histogram.

$$\left\{ \hat{p}_u = \frac{\hat{q}_u}{\max(\hat{q})} \times \hat{q}_u \right\}_{u=1\ldots m} \tag{5}$$

That is, the histogram values are rescaled from $[0, \max(q)]$ to the new range $[0, 255]$, where pixels with the highest probability of being in the histogram will map as visible intensities in the 2D histogram back-projection image.[6]

Next, we binarize the histogram back-projection image with a threshold which is determined statistically. Then, we apply erosion and dilation to the binarized image and compare the area size of the connected component to determine the human clothes area. In this research, we calculate the approximate distance between the human and the robot by the ratio of extracted human clothes area size to initial human clothes area size.

4 Experiment and Future Works

While this approach has many applications, we simulated a generic restaurant in order to test our system. In many restaurants, customers must wait for a host or hostess to seat them at their table. This is the aspect we chose to replace as it is a common, everyday example of guidance. There are many factors that could affect the guidance and cause it to be interrupted midway.

Using a hallway and several rooms, we dispersed several motes in an array. These motes form the directional field. We placed obstacles such as tables and chairs in order to more closely simulate an actual restaurant.

Under normal conditions, i.e. the human follows the robot properly without straying; we expected our robot to perform properly. In order to test its capability of handling interruptions, we simulated the human stopping, straying, and a combination of both. These cases represent the human stopping perhaps to answer their cell phone or perhaps wandering off to find the bathroom or greet a friend they have spotted in the restaurant.

The details of the experimental results will be shown in our web site with some movie clips soon at http://www.shalab.phys.waseda.ac.jp/index-j.html.

As a future work we are planning to improve the ability of the robot so that it can guide a group of humans. Although we employed wheeled mobile base in our robot at present, we are also considering to use a biped walking system developed in our institute[7] to make the more practical robot that can move on a floor with a difference in level.

Acknowledgments

This research was funded in part by a grant from the Japan Society for the Promotion of Science 21st Century COE Program "Coexistence of Humans and Robotic Technology in the Super-Aging Society" and the Waseda University WABOT-HOUSE Laboratory.

References

1. http://www.cs.cmu.edu/~minerva/
2. http://www.informatik.uni-bonn.de/~rhino/
3. http://robotics.naist.jp/research/ikochan/index.html
4. Suzuki, K., Camurri, A., Ferrentino, P., Hashimoto, S.: Intelligent Agent System for Human-Robot Interaction through Artificial Emotion. Proceedings. of SMC, pp. 1055–1060 (1998)
5. Shenoy, S., Tan, J.: Intelligent Robots and Systems, Proceedimgs of IROS, pp. 1636–1641 (2005)
6. Allen, J.G, Xu, R.Y.D, Jin, J.S: Object Tracking Using CamShift Algorithm and Multiple Quantized Feature Spaces. Proceeding of the Pan-Sydney area workshop on Visual information processing, pp. 3–7 (2003)
7. Ogura, Y., Aikawa, H., Shimomura, K., Kondo, H., Morishima, A., Lim, H.-o., Takanishi, A.: Development of a New Humanoid Robot to Realize Various Walking Pattern Using Waist Motions. Proceedings of ROMANSY16 (2006) 16, 279–286 (2006)