Development of the robotic travel aid "HITOMI"

Shinji Kotani*, Hideo Mori, Noriaki Kiyohiro

Department of E.E. and C.S., Yamanshi University, Takeda 4, Kofu 400, Japan

Abstract

The Robotic Travel Aid (RoTA) is a motorized wheelchair equipped with vision, sonar, and tactile sensors, in addition to a map database system. The visually impaired can get orientation, mobility and obstacle information from RoTA, and can inquire about their present location, landmarks and the future part of the route. The concept is implemented on a RoTA called "HITOMI". It can guide the impaired to avoid vehicles along a road with lane marks or along a sidewalk marked with Braille.

1. Introduction

Various kinds of Electronic Travel Aid (ETA) have been developed for the visually impaired. Among them the Moat sensor and Sonic guide have been available for 20 years, but have not yet become widespread. Why are they not commonly used by blind people? Jansson has suggested that it is because most ETAs give information about the environment only a few meters ahead, that can easily be obtained by using a long cane [4]. Another reason, we think, is that the sound by which ETAs give him information disturbs his echolocation, and vibration is too poorer to inform him of the environment.

The guide dog is the best travel aid, but it is difficult to train enough guide dogs as are needed.

The number of guide dogs in the world (e.g. 8,000–10,000 in USA, 4,000 in UK 730 in Japan) shows this difficulty. In Japan, shortage of training staff and budget (about $25,000 per dog) make it difficult for the guide dog to become widespread.

The Mechanical Engineering Laboratory of Japan started project MELDOG to develop a robot guide dog in 1977 [9]. By the use of CCD sensor the robot detected barcode-like landmarks fastened on the road. This research project came to an end after seven years and the robot was never used by the blind.

Recently, vision-guided vehicles have been investigated by many researchers. Some of them follow a lane mark and avoid obstacles. We have proposed a behavior-based locomotion strategy called "sign-pattern-based stereotyped motion" and have applied it to mobile robots named "HARUNOBU" [6]. Recently, car-navigation
systems have been produced commercially by the electronics and car industries. Combining these technologies of the mobile robot and the car-navigation system, we have been developing the Robotic Travel Aid (RoTA) since 1990. In this paper we give an outline of RoTA.

2. RoTA requirements

The RoTA is not a substitute for the guide dog but it is an advancement of ETA. Required functions of RoTA are listed as follows.

2.1. Target of RoTA

The required functions of RoTA are different depending on the level of visual disability, the age of losing sight and whether the blind person has auditory impairment or not. In general, the higher the age of losing sight, the more difficult is the training in echolocation. Furthermore, for an older person, it is very difficult to memorize and recall the route from the start to the goal. Our RoTA is designed for those who lost their sight in an old age and have difficulty in memorizing in the route.

2.2. Size of RoTA

To utilize a video camera as a sensor, RoTA should be large enough to stabilize the video image. The larger the rota is in size, the smaller is the vibration of a video camera attached to it. A motorized wheelchair was used as the undercarriage of rota. It is 1170 (depth), 700 (width), 1500 (height) mm in size and 80 kg in weight. The wheelchair is big and heavy enough for the blind person to walk holding the handle bar.

2.3. Sensors of RoTA

RoTA requires information about orientation and mobility. Orientation information is required to get the goal, or the sub-goal is obtained by the use of vision to detect openings for passage or space without any obstacles. Mobility information means what is required by the blind person to control walking continuously and to keep balance while walking. RoTA makes proper use of sensors to get information of orientation and mobility. Table 1 shows multiple sensors of RoTA.

Information about orientation can be obtained by sensing the passage ahead within 5 m. Only vision can provided such a range of sensing. We believe monocular and monochromatic vision is sufficient for getting orientation. Vision is also useful to detect obstacles such as vehicles and pedestrians. However, it cannot detect a wall-like obstacle that the a homogeneous surface.

The sonar is useful to detect a wall-like obstacle, but its effective range is limited to 3 m in an outdoor environment.

2.4. Sign pattern based stereotyped motion

We have previously proposed a behavior-based action strategy, stereotyped motion (STM) [5]. It is defined as a fixed action pattern that makes the robot perform a primitive skillful action. We assume that five STMs, “moving along”, “moving toward”, “moving for sighting”, “following a person” and “moving along wall” are enough to perform any locomotion from start to finish. All complex actions such as obstacle avoidance can be defined as a chain of these STMs. Cooperative camera pan/tilt control is included in STM.

A pattern or feature of the environment that is utilized to initiate STM or modify STM to fit the environment is called a sign pattern (SP). The STM is different from subsumption architecture [1] which does not use SP. The STM is a goal-oriented action and can perform a mission or task, but subsumption architecture is a reflective action and cannot perform a mission.
The advantages of using SP-based STMs are:
(1) The robot can move even when the information about the future part of the passage is incomplete. For instance, what is round a corner is invisible, the robot can turn the corner by a chain of STMs that includes collision avoidance for an obstacle that might suddenly appear.
(2) The reaction time to an SP is very short, as it does not need motion planning.

2.5. Moving along and moving toward

Jansson in his psychological study defined two kinds of perceptual guiding action, “walking along” and “walking toward”. Applying these to RoTA we define two STMs: “moving along” and “moving toward”.

Moving along is defined as an STM that consists of two cooperative actions of the under carriage and video camera systems. It moves along an elongated SP keeping the distance from SP constant, changing the camera direction to keep the SP in the center of the video image. The elongated SP may include a lane mark of a road, the edge of a sidewalk, fences and a boundary of a campus path.

Moving toward is also defined as an STM that consists of the cooperative action of the two systems to move toward a sub-goal. The RoTA has to search for the SP of sub-goal and watch through its video image. An SP of a sub-goal includes only a zebra-crossing mark at present, but in the future the entrances of building and stairs.

2.6. Obstacle avoidance

When RoTA moves along a road or sidewalk, most critical obstacles are vehicles and bicycles. The RoTA’s obstacle avoidance is carried out through four tasks: “moving along” in which an obstacles is found, “moving for sighting” finds the right-hand side of the obstacle, “moving along a wall” which passes by the obstacle, and again “moving along”, as shown in Fig. 1.

2.7. Crossing an intersection

One of the problems for the visually impaired in walking outdoors is to cross a road intersection. The RoTA can find the intersection if there is zebra-crossing mark on the road by detecting it visually.

2.8. Map-based guidance

The digital map system of RoTA gives two kinds of map information, one is for RoTA itself and another for the impaired person.
The map information for RoTA is almost the same road information as that for car navigation, and it includes multiple SPs for each road and sidewalk. The multiple SPs are necessary to increase the robustness of the guidance.

The map information for the impaired is used to let him know the landmarks of the present location landmark and the route represented by a command list such as “go straight” or “turn right”.

The map database system of RoTA is made up of two sub-systems: the basic system and a portable system. We think the basic system should be managed by the welfare office and be updated every month. The portable system is part of the original one and is created by the impaired person or a volunteer before every journey. We have implemented the two systems on a well-known Japanese commercial digital map system, because that system covers almost all areas of Japan. The two systems are represented by a network of nodes and branches. A branch represents a passage. Its attributes include multiple SPs for RoTA and landmarks for the impaired. A node represents an intersection.

When the impaired person does not know where he is or loses his way to the destination, he can interrogate the portable system.

2.9. semi-automatic navigation

While an SP is detected visually, the impaired person may follow RoTA which keeps “moving along” or “moving toward”. However, when the SP disappears, RoTA stops, refers to the map system and makes an inference as to why the SP has disappeared. The inference of RoTA may be “we must reach the end of the SP” or “we will probably meet an obstacle”. He will test the environment through his residual senses, including auditory, tactile senses, and understand the inference through his knowledge of environment, traffic, weather and time. The RoTA makes a list of possible STMs, he selects one from the list and RoTA performs it.

2.10. The blind-oriented interface

The RoTA has to inform the blind person of four kinds of information: mobility information, orientation information, obstacle/intersection information and map-based information, as shown in Fig. 2. A command bar with a Braille key is fixed on the rear part of the robot. By holding this bar the user can get the mobility and orientation information.

![Fig. 2. Blind-oriented interface.](image-url)
Obstacle/intersection information detected by RoTA is issued as warning and alarm messages through the voice interface. The impaired person lets RoTA know the destination by the Braille key.

3. Implementation and results

The above specification are implemented on a RoTA called HITOMI ("pupil" in Japanese). Its system configuration and a photograph are shown in Figs. 3 and 4. We will briefly describe the main points.

3.1. Undercarriage

One of the most serious problems of the mobile RoTA is that it cannot go up or down stairs. The motorized wheelchair cannot go over a step more than 3 cm in height. We made the wheelchair back-to-front, because the rear drive wheels are larger in diameter than the free front wheels and are able to go over higher steps.

3.2. Camera platform

The problems of the camera platform are how to reduce the electric power consumption and how to decrease the vibration of the video image during the locomotion. At first we used servo systems for factory automation. They are high in electric power consumption and very large in size. We use commercially available servo systems developed for model aircraft. They are one-fifth both in size and in power consumption compared to industrial servo systems. Its direction accuracy is 0.5° in pan

---

**Fig. 3. System configuration of HITOMI.**
3.5. Zebra-crossing mark detection

When RoTA comes near to a road intersection, it knows this fact by map matching, and inserts a zebra-crossing mark searching process in every ten cycle of the lane mark detection processes [11]. When the mark is found, it is followed until RoTA is 3 m in front of it. In its searching and following process, the road image is binarized and horizontal and vertical projections are performed on the image, as shown in Fig. 5. By analyzing the two projections the mark is identified. Since some obstacles momentarily show almost the same projection as the mark, RoTA sometimes mistakes the obstacle for the mark, but this error can be corrected by checking the projection in the successive frames.

3.6. Vehicle detection

The vehicle is one of the most troublesome obstacles, whether it is moving or stationary. We have proposed a simple useful vehicle detection algorithm, which is based on the fact that the sunny part of the road shines in the sun and sky light, while the underneath part of the vehicle is dark because of the faint environment light through the gap between the vehicle and the ground [2]. The underneath and shaded parts seem to be the same in luminance, but this is an illusion of brightness constancy. Our algorithm is easy to implement by software: right and left edges of the vehicle are also used for vehicle identification. An example of intensity curve in a window which is located in the underside of a vehicle is shown in Fig. 6. We tested this method in four traffic scenes of (1) partly shaded road, (2) entirely shaded road, (3) nonshaded road, (4) a road in cloudy; more than 97% of vehicles were successfully detected (Table 2).

3.7. Pedestrian detection

Pedestrian detection is very important when RoTA moves along a sidewalk. We have proposed a rhythm model to detect and follow a pedestrian [7]. The model is based on the following fact.
Rhythm constancy: when a person walks, their volume changes rhythmically in area, width and height. The volume is specified by the men and standard deviation of walking frequency $F_p$ and $O_p$.

Advantages of the rhythm model are as follows.

1. Total image-processing time is shorter than that of any other model because few non-structured features are used in processing.
2. Rhythm does not vary with for distance change between person and observer.
3. Rhythm is independent of illumination and therefore robust against time and weather changes.
4. Rhythm is easy to detect when a person is wearing clothes.

A disadvantage of the rhythm model is that someone with ill-intent can deceive the robot by a simple trick.

The process of pedestrian detection by the rhythm model is decomposed into four sub-processes:

---

**Table 2**

<table>
<thead>
<tr>
<th>Weather</th>
<th>Shadow on the road</th>
<th>Number</th>
<th>Success rate (%)</th>
<th>Under-detection</th>
<th>Over-detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fine</td>
<td>Partly</td>
<td>294</td>
<td>92</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Fine</td>
<td>Entirely</td>
<td>191</td>
<td>97</td>
<td>3.2</td>
<td>0</td>
</tr>
<tr>
<td>Fine</td>
<td>None</td>
<td>272</td>
<td>98</td>
<td>2.2</td>
<td>0</td>
</tr>
<tr>
<td>Cloudy</td>
<td>None</td>
<td>405</td>
<td>98</td>
<td>1.3</td>
<td>0.3</td>
</tr>
</tbody>
</table>
motion detection by frame subtraction and binarization, motion segmentation by head-to-feet window, feet tracking, attribute acquisition, and rhythm matching.

In motion segmentation, horizontal and vertical projections are performed, as shown in Fig. 7, to set up a head-to-feet window for pedestrian detection. Right and left-foot windows, which are 1/5 in height and 1/2 the width of the head-to-feet window, are set up for feet tracking in successive images, and the location and area of blobs in the two windows are measured.

In rhythm matching, a power spectrum estimate method is operated on the two time series of areas of the right and left foot windows. When the first components of the power spectrum of the two are matched and are within two standard deviations of the mean rhythm of walking, the objects in the two windows are judged to be the feet of a pedestrian.

In our experiments, video tapes of pedestrians on the road environment in cloudy conditions were used. Image subtraction is performed for every two frames (66 ms). Fig. 8 shows a subtracted image enclosed by a 1/5 size window. The
subtracted image of the window is transformed into the binary image by a threshold operation. The area of a binary image is used for rhythm detection. The power spectrum is obtained by a maximum entropy method from 64 area measurements (66 ms × 64 = 4.2 s). Fig. 9 shows the time series of the area and the obtained frequency components. It is found from Fig. 9(b) that the rhythm of walking is 1 s in frequency. This method works well when the robot stops and looks at pedestrians at a distance of 7-30 m through the video camera. The success rate was 94% for one pedestrian, and 95% for non pedestrian. Detection errors are mainly caused by setting window incorrectly.

4. Concluding remarks

We have been developing the robotic travel aid (RoTA) "HITOMI" which guides the visually impaired in the road environment. The HITOMI is a small mobile robot which utilizes a motorized wheelchair as its undercarriage. A vision system is equipped to detect the road, vehicles and pedestrians. A multiple-sonar system is fitted to detect walls and other obstacles which the vision system cannot detect. A portable digital map system is equipped to give the undercarriage a sequence of commands by which it follows the route from start to destination. It also gives the vision system detection parameters of sign patterns and landmarks of the route. The digital map includes the names of intersections and buildings. The user can ask where he is and the map system replies through a synthesized voice. A command bar is attached at the rear part of HITOMI. The user stands behind the RoTA and follows by grasping the command bar. He can get the mobility and orientation through the motion of HITOMI.

The HITOMI cannot recognize obstacles as well as man does. The success rate of vehicle and pedestrian detection is between 92 and 98%. When the visually impaired person walks guided by HITOMI, he may meet a traffic accident. To avoid this the semi-automatic navigation is applied. When HITOMI senses an environmental change, it infers its cause and tell the user its inference and the next plan of motion by the synthesized voice. The user confirms the inference.

---

![Fig. 8. A subtracted image of a pedestrian and a bicycle.](image)

![Fig. 9. Rhythm pattern and power spectrum of the pedestrian in Fig. 8: (a) Time series of the pedestrian's blob. (b) Power spectrum for rhythm of the pedestrian.](image)
by his residual senses and permits HITOMI to perform the plan or makes it wait for his permission. Generally speaking the impaired person does not want to have to obey completely what the robot commands. By the active use of his residual senses, his independence of life will be promoted.

Acknowledgements

This paper has been partially supported by Grant-Aid for scientific Research from the Ministry of Education No. 03650343.

References