Robotic travel aid for the blind: HARUNOBU-6

Hideo Mori and Shinji Kotani

Department of Electrical Engineering, Yamanashi University,
Takeda-4, Kofu 400-8511, JAPAN

forest@es.yamanashi.ac.jp, kotani@es.yamanashi.ac.jp

ABSTRACT

We have been developing Robotic Travel Aid (RoTA) "HARUNOBU" to guide the visually impaired in the sidewalk or campus. RoTA is a motor wheel chair equipped with vision system, sonar, differential GPS system, dead reckoning system and a portable GIS. We estimate the performance of RoTA in two viewpoints, the viewpoint of guidance and the viewpoint of safety. RoTA is superior to the guide dog in the navigation function, and is inferior to the guide dog in the mobility. It can show the route from the current location to the destination but cannot walk up and down stairs. RoTA is superior to the portable navigation system in the orientation, obstacle avoidance and physical support to keep balance of walking, but is inferior in portability.

1. INTRODUCTION

Among 307,000 visually impaired in Japan 65,000 are the complete blinds. Most of them lost their sight in the elderly age. It is very difficult for the aged to learn to walk with the long cane or the guide dog, because they are not so rich in the auditory and haptic sensing and have not good memory for the cognitive map.

We have been developing Robotic Travel Aid (RoTA) "HARUNOBU" since 1990 to guide the visually impaired in the sidewalk or campus (Kotani S., Mori H. & Kiyohiro N.,1996). RoTA is a motor wheel chair equipped with vision system, sonar, differential GPS system, dead reckoning system and a portable GIS(Geographic Information System). In designing the RoTA, we add a guidance function and a safety function to the conventional mobile robot functions.

MoBIC Project (the mobility of blind and elderly people interacting with computers) was carried out from 1994 to 1996 with support of the TIDE program of the Commission of the European Union. It developed MoBIC travel Aid (MoTA) which consists of MoBIC Pre-Journey System (MoPS) and MoBIC Outdoor System(MoODS) (Pertie H., et al.,1996). MoPS is a simulator that helps the exploration of a previously unknown area and the selection and preparation of a route before an actual walk. MoODS is a portable system that gives assistance during the walk. It consists of a small wearable PC kernel of 16 x 11 x 7 cm in the size, a GPS, an electronic campus and a pair of special earphones that prevent masking the ambient sound essential for echo location. The system provides on-route information about the current position. The system informs the traveler automatically when they are leaving the chosen route or if the accuracy of the system has degraded. A prototype of MoTA was developed and estimated through a field test and found the design philosophy was useful in the human navigation.

RoTA is superior to the guide dog in the navigation function, and is inferior to the guide dog in the mobility. It can show the route from the current location to the destination but cannot walk up and down stairs. RoTA is superior to MoTA in the orientation, obstacle avoidance and physical support to keep balance of walking, but is inferior in portability. The functional comparison of RoTA, the guide dog and MoTA is shown in Table 1.

In the road environment the most important objects may be the car and the pedestrian. Conventional methods for the car and pedestrian detection seem to simulate the perception of the human beings. We get the idea of objects discrimination from the study of ethologist (Tinbergen N.,1969). He shows that the animal behavior is represented by a chain of fixed action patterns even if the behavior is an advanced and complex one. To explain the mechanism of the behavior Tinbergen proposes three concepts: sign stimulus, Central Excitatory Mechanism (CEM) and Innate Releasing Mechanism (IRM). The animal dose not recognize objects as human being does, it makes a response not to the whole of the object but to the part inherent in the object. The part of the object that activates the fixed action pattern is called sign stimulus. CEM is similar to

 ISBN 0 7049 1141 8

the modern multi-tasking system in the modern computer. All the fixed action patterns are in the dormant state, and when a sign stimulus appears the IRM activates one of the fixed action patterns corresponding to the sign stimulus. We think that Tinbergen's concepts are useful to configurate the vision based mobile robot. We use sign pattern(SP) as the technical term instead of the sign stimulus. Sign pattern is different from the landmark in three factors as shown in Table 2. The purpose of the landmark is to verify the current location, on the other the sign pattern is used to activate and guide the fixed action of the robot. We think the basic fixed action patterns are Moving-along SP, Moving-toward SP, Following-a-person, Turning-corner, Avoiding-obstacle, Moving-for-sighting SP.

Table 1. *RoTA, MoTA and guide dog robot*

	Obstacle avoidance & orientation	Mobility	Portability	Navigation
RoTA	O	O	×	O
MoTA	×	×	O	O
Guide dog	O	O	O	×

Table 2. *Comparison of landmark & sign pattern*

	Sign pattern	Landmark
Purpose	To guide the fixed action pattern	To verify the cuuent location
Object	Permanent and temporal objects	Permanent objects
Representation	Simple feature; edge, rhythm, shadow	2-D & 3-D model

2. GUIDANCE

A Geographic Information System (GIS) is required as the base of the navigation system of RoTA. The GIS of RoTA has to include the robot guide information and the human guide information. The robot guide information should give the sensor system of the robot the information about the environment.

2.1 Sign pattern

The robot does not recognize the total environment but it recognizes only two kinds of signals required to guide the robot in the environment. One is a sign pattern and the other is a landmark. For instance SP in Moving-along means a signal used to correct the location and heading errors of the dead reckoning system. RoTA uses as the SP of Moving-along an elongated feature on the road such as road boundary, lane mark, fence, tactile block and so on. We define the rhythm of walking as the SP of Following-a-person. As the SP of Avoiding-car we define the shadow underneath the car.

2.2 Robot guide information

To keep safe and to follow the Japanese traffic regulation RoTA should move on the sidewalk and zebra crossing marks. For this reason we define the path on which RoTA and the blind can move safely. When the road has the sidewalk of enough width for RoTA, the path is specified on it. When the road has not a sidewalk, the path is specified right or left roadside which is free from falling into creek, downstairs and depression. The digital map of the GIS includes a road network, a path network, sign patterns and land marks. The road network includes road information such as the type, the distance, the direction and absolute location of the street and the junction. After route searching the GIS feeds to the locomotion control system the robot guide information along the route. Fig.1 shows a snapshot of the display of RoTA in Moving-along SP in Yamanashi University campus. The upper middle part shows a video image in which a SP searching window is described by a large square and SP tracing windows are described by small squares. The right upper part shows the robot coordinate system in which a detected sign pattern is described by a line segment. The left upper part shows the heading of RoTA. The lower right part shows the digital map of the campus, the center of which shows the current location of RoTA. The left lower part shows a differential GPS sky map in which solid circles show received satellites and open circles show non-received satellites.

2.3 Map learning by practice

To make the digital map to guide the robot, one should select landmarks sign patterns in the course and measure the distance and orientation between intersections. This measurement requires much efforts, moreover visual sign patterns change with the time (morning/daytime/evening), weather (sunny/cloudy) and the season. For these reasons RoTA has a function of the map learning by the practice.

Before the practice the operator gives RoTA a rough map represented by a list sections. The section is defined as the part between intersections and is specified by an approximate distance and a direction. In the first practice RoTA moves along the course based on the rough map and detects a SP and corrects the lateral location error based on the SP. In the SP detection RoTA works in two modes; searching mode and tracing mode. In searching mode the vision module detects SP candidates with a wide view angle and selects one which matches with the section of the rough map in the direction. Then the vision mode begins, it traces the SP obtained in searching mode with a narrow visual angle in the predicted position. It takes less processing time in tracing mode than in searching mode. The sign pattern information obtained by the first practice can improve the performance of traveling not only in the traveling time but also in the safety of locomotion.

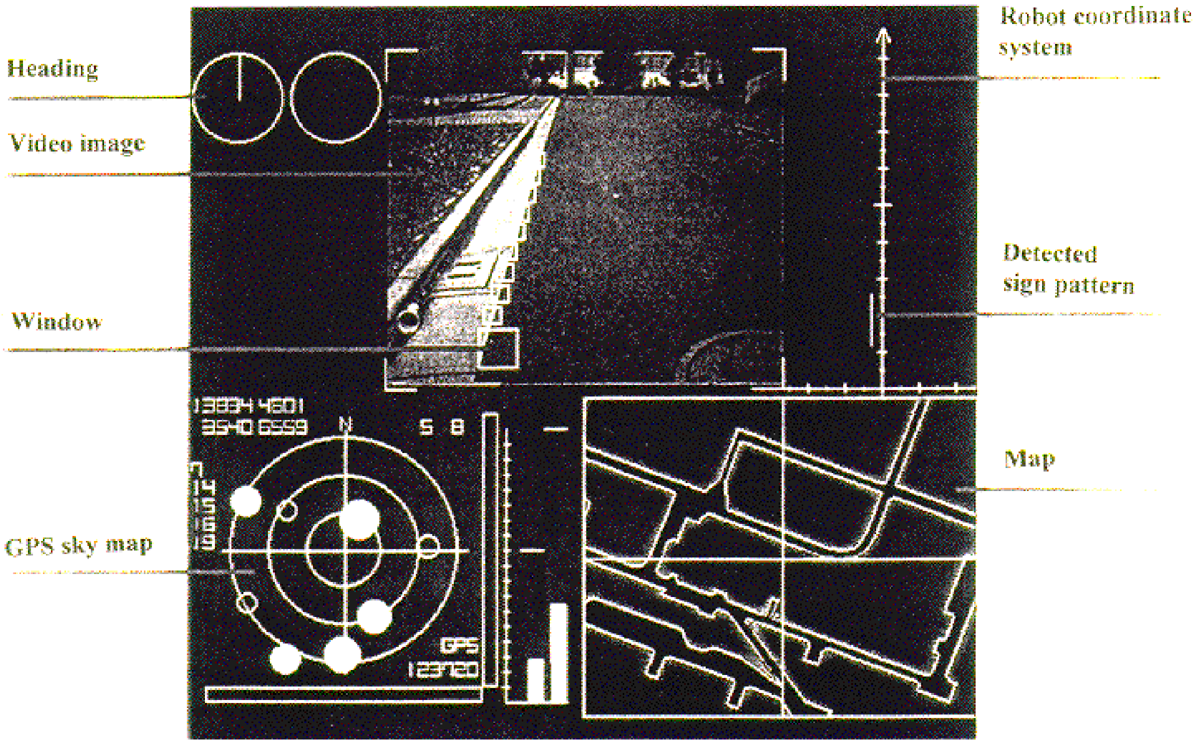


Fig.1. *A snapshot of the display of RoTA in Moving-along SP in the campus*

During the first practice the vision module stores the trajectory and SPs with their location and direction. After the first practice the learning process omits the noisy SPs and then fills gaps between neighboring SPs. The new SPs are used to update the rough map. The new map includes SP information about its location and direction. The second practice can improve its performance by using the new map.

Fig.2 (a) and (b) show the first and the second practice of HARUNOBU-4 in our campus. . A broken line shows the trajectory, and a line segment shows a SP candidate. A small closed circle indicates a searching point. In the first practice, four closed circles at corner N_2 show that the vision module repeat searching until it gets the real SP of the direction N_2N_4. At T-shape intersection N_3 the vision module misses SP and after three searching processes it detects SPs at the opposite(right) side of the passage, and tracing one of them. HARUNOBU-4 reaches at point B and finds the traveled distance is over the specified approximate distance and makes a U-turn immediately. After two searchings it finds the real SP of direction N_3N_4. In the second practice as shown in Fig.2(b) the searching point drastically decrease from fifteen to six.

2.4 Human guide information

We are developing a human guide information system. Its basic concept is almost the same as MoPS] When the blind is unsure the current location, he/she push the button, then the system tell the current location through a synthesized voice maker. When the blind wants to know future path to the goal, the system answers the time, distance and the number of turning to the goal.

3. PEDESTRIAN DETECTION BY RHYTHM

Conventional human motion tracking method includes the modeling of human body and the matching the model with the real data. The stick figure model is a well-known model of human body, but it should be modified by the distance, the clothes that will be changed by man or woman, summer or winter.

When one walks in the sidewalk, the rhythm of the walk is almost constant. The rhythm can be seen in the swing motion of feet and hands and in the up down motion of the head and shoulder. Among these motion the feet motion is the most detectable by the computer vision, because their rhythm is clearer than those of the head and the shoulder, the background of the feet is simpler than those of head and shoulder, and the clothes and another part of the body do not cover the feet in the image. The rhythm of the feet is a good sign pattern as it is easy to detect by the computer vision.

The difficult process of scaling to fit the object image to the model is not needed. It is free from the distance, the clothes and the weather. The implemented method is as follows(Yasutomi S, Mori H. & Kotani S.,1996).

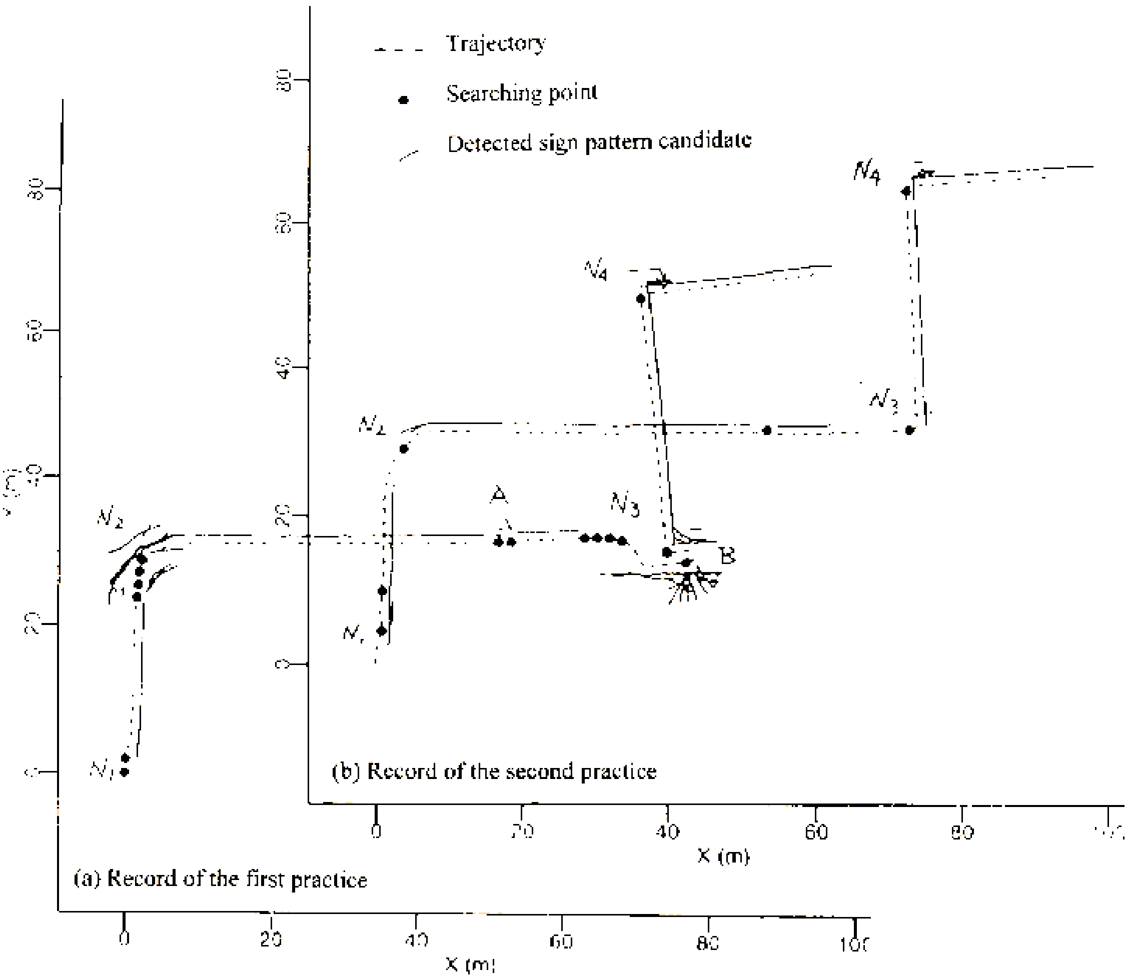


Fig. 2. *An example of SP learning*

3.1 Motion segmentation

The frame subtraction is applied to detect moving objects as shown in Fig.3. So this method is effective when the video camera is in the stationary state. A horizontal projection is operated after binarizing the subtracted image. The horizontal projection is sliced by a threshold to obtain *H* segment that may represent the height of a person. A vertical projection is operated and sliced by another threshold to obtain *V* segment that may represent the width of the person. If *V* segment satisfies the threshold of width, window W of *HV* in size is assumed to be the head to feet window of the person. Then the right and the left foot window, W_R and W_L, each of which is 1/5 of the *H* segment and 1/2 of *V* segment are set up in the lowest part of W. This process is called finding process, and is followed by tracking process as follows. Window W_R and W_L of the last frame are a little enlarged in length and width to trace the feet in the next frame, and the horizontal and vertical projections are operated on the new binarized subtracted image. New W_R and W_L are obtained by the same slicing operation as the finding process.

Proc. 2nd Euro. Conf. Disability, Virtual Reality & Assoc. Tech., Skövde, Sweden, 1998

3.2 Rhythm matching

The most significant features of W_R and W_L are (1)the ordinate of the bottom of the windows and (2)the area of the binarized subtracted image in the windows, as the ordinate shows the distance of the person, and the periodic change of the area of the sliced image depends on the rhythm of walking.

Auto-correlation function is operated on time series of the area of W_R and W_L. When the primary components of the power spectrum of the two time series are satisfied with 2σ of the mean rhythm of walking, W_R and W_L are judged as the feet of a person. An example of time series of the area in W_L and its power spectral are in Fig 4.

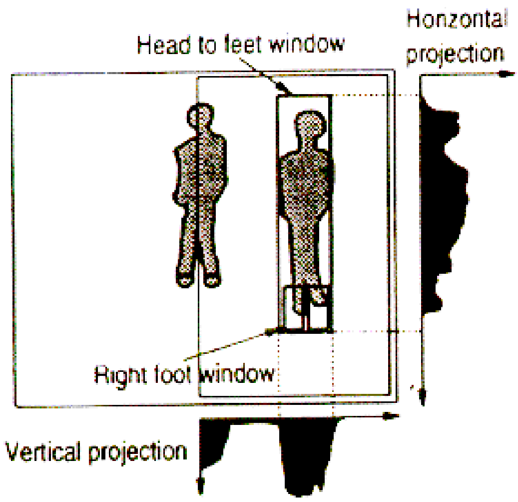


Fig. 3. *Setting of three windows*

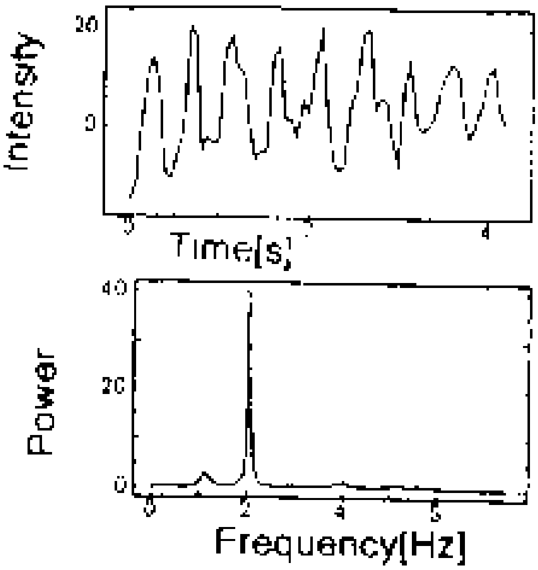


Fig. 4. *An Example of the time series of the area of WL and its power spectral*

	Samples	Correct	False
Pedestrian	407 (334:pants,45:short pants & short skirt, 28:long skirt)	94.9%	5.1%
Non-pedestrian	106	96.2%	3.8%
Total	513	95.0%	5.0%

Table 3. *Results of pedestrian detection*

3.3 Results of pedestrian detection

Pedestrian detection algorithm for stationary camera is implemented on a monochromatic image processing system (HITACHI Co.Ltd.,IP-2000). It samples a moving object every 67ms and takes 64 samples (4.3sec) to judge the object by the rhythm. We fixed a video camera in our campus 1m in the height and 15in the depression angle, and recorded 407 pedestrians and 106 non-pedestrian including bicycles and dogs on a videotape in a cloudy day. Among the pedestrians 82% of them wear pants, 11% of them wear short pants or short skirt and 7% of them wear long skirt. The experimental results on the videotape are shown in Table 3. Five % of errors are caused by (1) noise of video signal that makes jitters on the image, (2) swaying of trees and grass that make the same rhythm as that of pedestrian, (3) the same color of shoes as that of the asphalt paved road.

4. DANGER ESTIMATION AT AN INTERSECTION

When the driver and the blind keep the traffic regulation perfectly, they will not meet with any accident. However as they often pay less attention to the right and the left sides of an intersection, they will occasionally have an accident. According to the statistics of the traffic accidents in Japan, about 50% of them occur at or near intersections. To avoid collision we should estimate danger level of vehicles at or near the intersection (Kotani S.,Mori H. & Charkari N.M.,1996).

4.1 Car detection by shadow

The sunlight and the diffused sky light do not reach the underneath a car. The image intensity of the underneath part is almost noise level in the video image. Its intensity is lower than any other part such as a wet part or a patched part repaired by new asphalt even in the cloudy day as shown in Fig.5.

These phenomena are applied to the SP definition of the car. A window is set up in the lane, and .the vertical projection of the window is obtained. When the projection curve shows a flat bottom of a certain width with cliffs at the right and left side as shown in Fig.5, we define the bottom as the sign pattern of a car. Three levels of danger are defined in this work, 0: *Safe* 1: *Warning* 2: *risky*. The robot detects the location s_i and moving direction r_j of the car by its sign pattern and predicts its future path based on the Japanese traffic regulation. The danger coefficient d_{ij} for a vehicle at (s_i,r_j) is defined as follows. When the future paths of the vehicle and the robot do not cross (d_{ij}=0), possibly cross (d_{ij}=1), surely cross (d_{ij}=2).

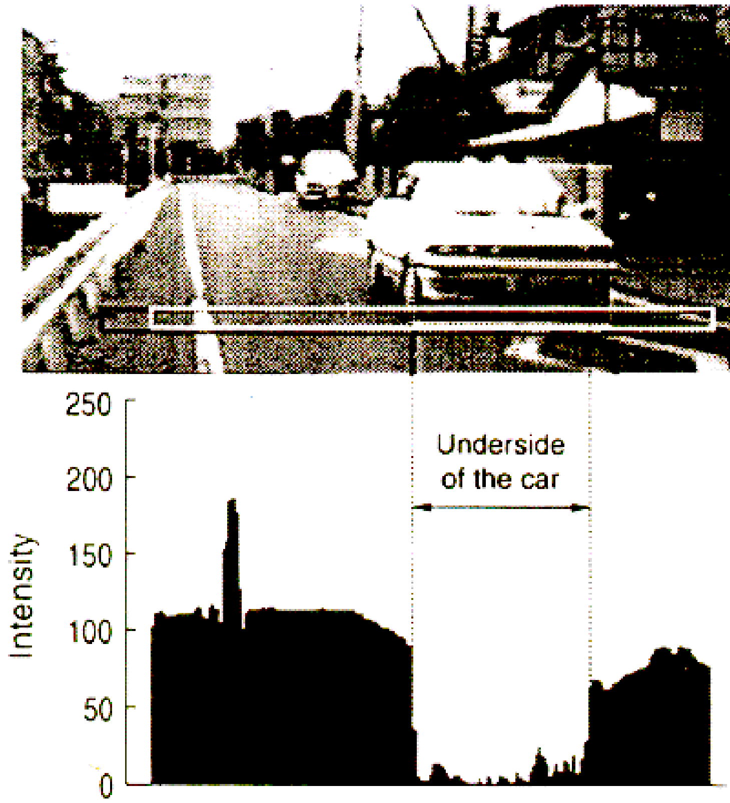


Fig. 5. *An intensity curve in a window which is set up underneath the car*

4.2 Japanese traffic regulation

We formulate the traffic regulation including the behavior of the careless driver as follows. (J1) Vehicles move along the left lane mark. (J2) Vehicles follow typical path. (J3) When the driver moves straight, he will only pay attention to the front. When he turns left, he will pay attention to the front and the left. When he turns right, he will wait until all the straight moving and right turning cars pass by. (J4) When the blind starts moving across the intersection, the car must not obstruct his/her way.

4.3 Robot's traffic regulation

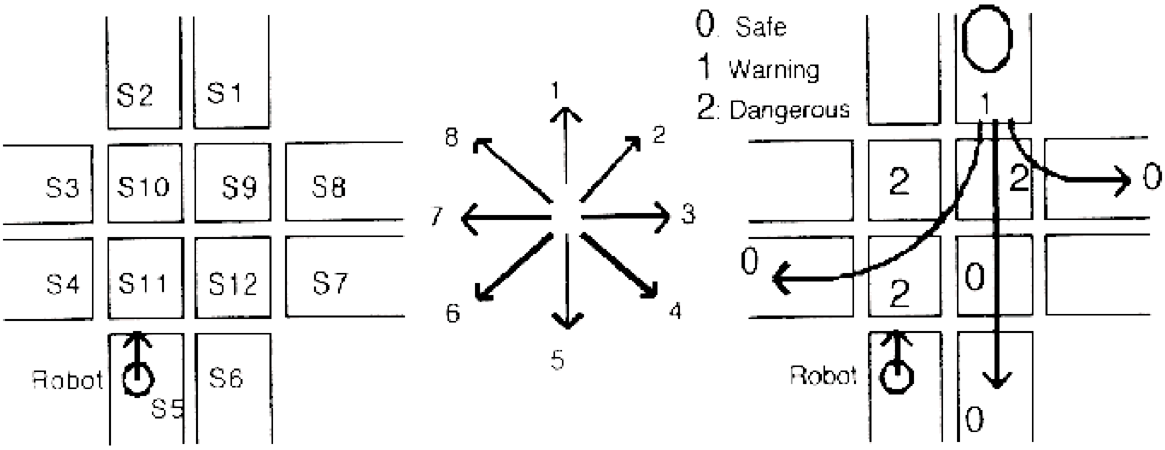
We consider that the robot follows the same traffic regulation as the guide dog. (R1) The robot moves along the left side of the road. (R2) When the danger estimate value is safe, the robot sends the blind the permission message to start crossing. (R3) After the blind receives the permission message, he gives the robot a start command. (R4) The robot has an intelligent disobedience function. The robot does not follow the blind's command before the danger estimate value becomes safe. Based on the traffic regulation of the car and the robot, danger matrices d_{ij} are given as shown in Fig.6.

4.4 Results of danger estimation

The car detection algorithm is implemented on a personal computer of CPU486 (100Mhz) with an image

Proc. 2nd Euro. Conf. Disability, Virtual Reality & Assoc. Tech., Skövde, Sweden, 1998

processing board (HITACHI Co.Ltd., IP-2000). The car tracking is performed every five frames. The danger level is estimated while a car passes an intersection within 3 - 4 sec. A video camera was fixed 1m in the height and 2.5m apart from a T shape intersection as shown in Fig.7. It recorded 105 cars that pass the intersection on a videotape. We looked B at the display and judged the estimation made by the computer. The performance of the computer for 105 cars shows 90% of success as shown in Table 4. Among 10 misjudgments, eight misjudgments were caused by two or three cars in successive running less than 20m apart. The vision system is successful in tracking the first car, but often fails in detecting the second car. Remaining one was caused by the trajectory of an ill-mannered driver. The last one was caused by the mis-tracking of a car that was too large to process at video rate. Fig.8 shows examples of three cases. In the right side of the display six parameters are shown; "TIME" indicates the quantized time, "Moving car" shows the result of the process, "DIST" shows the obtained distance of the car from the video camera in meters, "Speed" shows the estimated velocity in KM/h, "AP-TIME" indicates the predicted arrival time at the intersection. The Trajectory of the vehicle is shown at the bottom of the right side part. Finally the danger coefficient of the car is represented *safe*, *warning* and *risky*.



(a) Sections at the cross *(b) Quantized directions* *(c) Danger estimation for a front vehicle*

Fig.6. *Danger estimates of a vehicle at an intersection*

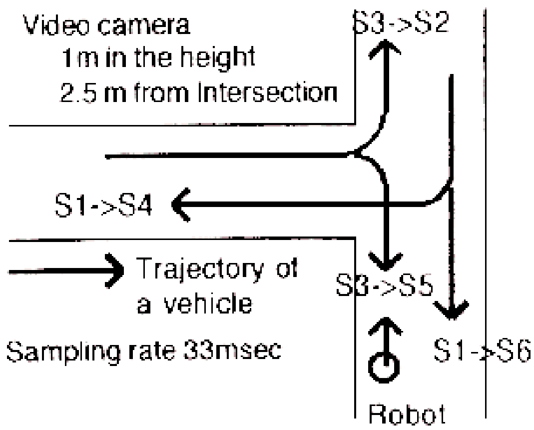


Fig.7. *Experimental set up at a T shape intersection*

Course	No. of cars	Correct	False
S1 → S6	36 (100%)	30 (83%)	6 (17%)
S1 → S4	16 (100%)	15 (93%)	1 (7%)
S3 → S2	23 (100%)	22 (95%)	1 (5%)
S3 → S5	30 (100%)	28 (93%)	2 (7%)
Total	105 (100%)	95 (90%)	10 (10%)

Table 4. *Results danger estimation at the T shape intersection*

5. EXPERIMENTAL RESULTS

We have implement the concept of RoTA on a color vision-based mobile robot HARUNOBU-6 as shown in Fig.9. It has a motor wheel chair (SUZUKI Co.Ltd., MC14) as the undercarriage part, a color video camera with pan/tilt control(Sony EVI-G20) and a real time image processing board (HITACHI I Co.Ltd., IP-2000) as the vision module, two sonar range sensors(IZUMI I Co.Ltd., SA6A-L2K4S, 130kHz), an optical obstacle sensor(SUNX Co.Ltd., PX24ES), a dead reckoning system with an optical gyroscope (HITACHI WIRE I Co.Ltd., OFG-3) and a differential GPS system (MATSUSHITA DENKO I Co.Ltd., GS-5). The performance of these sensors are shown in Table 5. The vision module is used to get the information of orientation and navigation. The sonar range sensor is used to get mobility information. The optical obstacle sensor is used for reflective obstacle avoidance. A horizontal bar is attached the rear of HARUNOBU-6. By touching the bar the blind can keep his balance in walking and can feel the surface of the ground through its vibration. He/she can get the mobility and orientation information through the motion of HARAUNOBU-6.

The performance of RoTA "HARUNOBU-6" was tested by three test courses.

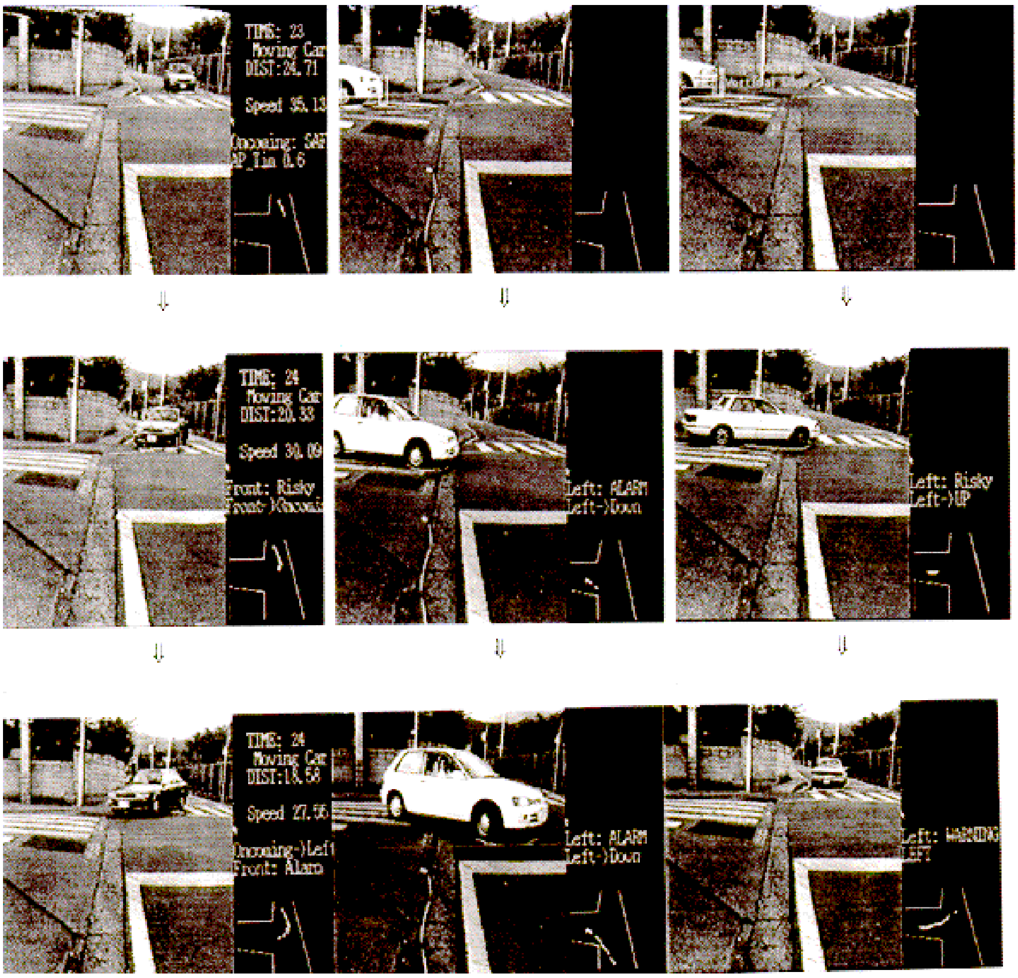
The first test course is set up in a small zone of our university campus of 50m by 50m. In this course HARUNOBU-6 changes 360 degrees in its heading. The illumination of sunlight changes from back light to counter light. From the technical point of view this experiment give us the problem of iris control. A blind who lost his sight by *retinosis pigmentosa* tested HARUNOBU-6. He said the robot was useful for him to move from building to building. He suggested us that a step attached the rear of the robot would be useful to rest himself during the locomotion. He can escape from the accidents by getting off the step.

The second test course is set up in an open space of Kofu stadium, In such open field the bind feels difficulty in orientation because he cannot use the echo location. Although the position error(3σ) of the differential GPS is 2 meters, it is useful in only open space. The open space is a good place to guide RoTA by the differential GPS.

The third test course is set up in the hospital of YAMANASHI MEDICAL UNIVERSITY. To guide a patient of ophthalmology from the doctor's office to his/her ward a nurse is required. Instead of the nurse our RoTA is expected. The illumination of the corridor is not homogeneous, therefore it is difficult to detect SP and obstacles by the vision. The sonar range sensor and the optical sensor are used in the hospital.

Table 5. *Performance of sensors of HARUNOBU-6*

Sensor	Detected objects	Range
Vision module	Road edge, car, Pedestrian	2 – 30 [m]
sonar range sensor	Right and left side wall	0.2 – 2 [m]
Optical obstacle sensor	Suddenly appearing obstacle	0.1 – 1.5 [m]



(a) Enter from ahead, turn left   *(b) Enter from left, come here*   *(c) Enter from left, go ahead*

Fig.8. *Some results of the danger estimation system*

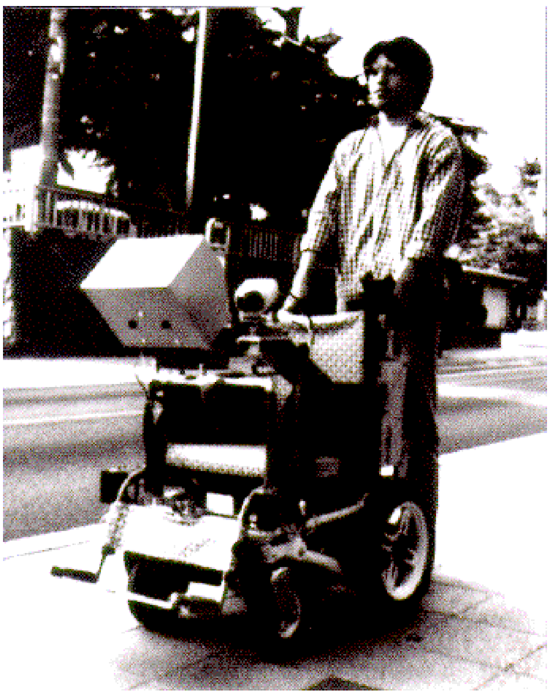


Fig. 9. *HARUNOBU-6*

6. CONCLUDING REMARKS

We have a plan to develop several RoTAs in corporation with Japanese companies to make field tests by two kinds of the blind. The first is the blind who can walk with the guide dog and the second is the diabetic who loses his sight recently and cannot walk without a helper.

The guide dog user will want to walk in the crowded streets for visiting and shopping. We think he can use PC with voice interface to communicate with RoTA. The difficult problem in this case is how to make the map data base.

The blind of diabetes will want to learn walking in a safe place such as the campus of a hospital or a park. The difficult problem is the human interface because the diabetic loses not only the vision but also auditory sense and the haptic sense. He will not be able to use PC.

This work is supported by Grant-in-aid for Scientific Research on Priority Areas No.07245105, Grant-in-Aid for Scientific Research(B)(2) No.07555428 by the Ministry of Education, Science, Sports and Culture, The Mechanical Industry Development & Assistance Foundation (MIDAF).

7. REFERENCES

Kotani S., Mori H. & Kiyohiro N.(1996), Development of the robotic travel aid “HITOMI, *Robotics and Autonomous Systems*, *17*, pp.119-128

Kotani S., Mori H. & Charkari N.M.(1996), Danger estimation of the Robotic Travel Aid(RoTA) at intersection, *Robotics and Autonomous Systems*, *18*, pp.235-242

Mori H., Charkari N.M. and Matushita T.(1994), On-line Vehicle and Pedestrian Detection Based on Sign Pattern, *IEEE Trans. On Industrial Electronics*, **41**, *4*, pp384-391

Pertrie H., Johnson V., Strothotte T., Rabb A., Fritz S. And Michel R.(1996), MOBIC: Designing a Travel Aid for Blind and Elderly People, *Journal of Navigation*, **49**, *1*, pp.45-49

Tinbergen N.(1969), *The study of instinct*, The Claredon Press Oxford

Yasutomi S., Mori H. & Kotani S.(1996), Finding pedestrians by estimating temporal-frequency and Spatial-period of the moving objects, *Robotics and Autonomous Systems*, *17*, pp.25-34