6-4

Measurement of Ground Surface Displacement using Stereo Vision and Mechanical Sensors on Humanoid Robots

Takashi Yoshimi
3-D Vision Systems Research Group
Intelligent Systems Institute, AIST
Central 2, 1-1-1 Umezono
Tsukuba-shi, Ibaraki 305-8568 Japan
tak-yoshimi@aist.go.jp

Yoshiaki Kawai
3-D Vision Systems Research Group
Intelligent Systems Institute, AIST
Central 2, 1-1-1 Umezono
Tsukuba-shi, Ibaraki 305-8568 Japan
y.kawai@aist.go.jp

Yutaro Fukase
Space and Robot System Group
Institute of Technology
Shimizu Corporation
4-17, Etchujima 3-chome, Koto-ku
Tokyo 135-8530, Japan
fukase@shimz.co.jp

Hidekazu Araki
Technology Fusion Laboratory,
Matsushita Electric Works, Ltd.
1048, Kadoma
Osaka 571-8686, Japan
araki@prod.mew.co.jp

Fumiaki Tomita
3-D Vision Systems Research Group
Intelligent Systems Institute, AIST
Central 2, 1-1-1 Umezono
Tsukuba-shi, Ibaraki 305-8568 Japan
f.tomita@aist.go.jp

Abstract

This paper presents a system for measuring ground surface displacement using stereo vision and mechanical sensors on a humanoid robot. The stereo vision system generates a set of range data for each view by a correlation method as ground surfaces can be regarded as smoothed surfaces with random texture patterns. The mechanical sensors are encoders, gyro sensors and acceleration sensors that measure the angles of joints, angular velocity and acceleration of the robot. With those angles and the link geometry, the transfer matrices between different views are generated and used to register sets of range data.

1. Introduction

When a humanoid robot, one which has two legs for locomotion, travels over rough ground, it requires information on the ground conditions to plan an appropriate route. Such information can be obtained by measuring the surface displacement of the ground. Although static environments, which may include fixed ground shapes (e.g. a floor) or obstacles (trees, houses, walls, etc.), can be measured and given in advance as map data, dynamic environments, such as soft ground shapes (e.g. sand) and removable obstacles(e.g. boxes), must be measured before the robot moves.

Because the information needed for action is limited in the vicinity of the robot, it is an appropriate solution to have the robot carry a measurement system, in other words a vision system for navigation.

Much research has been done on vision systems for mobile robots([1]), but few studies are specific to unstructured environments. Krotkov et al. used a stereo vision system for their RATLER (Robotic All-Terrain Lunar Explorer Rover)[6]. A laser range finder was used for hazard detection on the Mars Pathfinder [4].

The above studies were for wheel-based robots. Although wheel-based robots use terrain information to decide routes or to control speed, legged robots (including two-legged robots) need to use more detailed ground shape information to control their walking procedure, because legged robots are more prone than wheel-based robots to be unstable in compensate for their superior DOF.

This paper introduces the vision system used on HRP-2, a product of the Humanoid Project in AIST. The robot is designed to work cooperatively with humans in outdoor environments such as construction sites, and requires the capability to decide route of travel on unstructured ground surfaces.

For the vision system on the humanoid robot, range data registration of multiple views is necessary for two reasons. One reason is that a wide angle lens cannot be used because the vision system is also used for manipulation. The Second reason is that, because of the weight and power consumption limits of the vision system, it cannot have a zoom lens or a congestion control mechanism.

Range data registration is achieved by the transfer matrix calculated from the mechanical sensors that read the angles of the joints of the robot. There are 30 joints on the robot, and 16 of them participate in the transfer matrix. Range data of the different views are registered by the matrix. There are also acceleration sensors and gyro sensors that give the transfer matrix between the global coordinates and the robot's coordinates.

In section 2, the mechanical configuration of the humanoid robot is described. In section 3, the vision system based on a stereo vision using correlation is described. In section 4, the coordinate systems and transfer matrix are described. In section 5, results of our experiment on unstructured ground are presented and the analyzed.

2. Hardware of HRP-2

This section describes the mechanical configuration of the HRP-2 humanoid robot and the sensors for measuring the geometrical alignments of the joints of the robot.

Figure 1 shows an overview of the HRP-2 humanoid robot. The robot is designed based on a prototype model HRP-2P[2]. It is 1,539[mm] in height, 621 [mm] in width, 355 [mm] in thickness and 58 [kg] in weight. The robot's maximum walking speed is 2.0 [km/h].

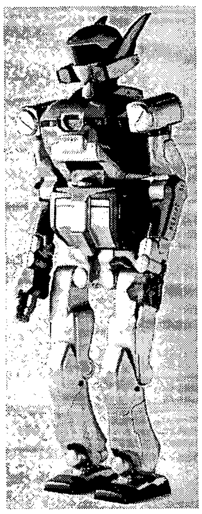


Figure 1. HRP-2.

HRP-2 has 30 D.O.F in total, 2 in the neck, 2 in the waist, 14 in the arms and 12 in the legs. (Fig.2) The mechanical sensors includes 30 encoders, one for each joint, 3-D gyro sensors and 3-D accelerometers.

The robot has two on-board CPU's for locomotion and the vision system. It has also a radio LAN system to communicate with off-board computers.

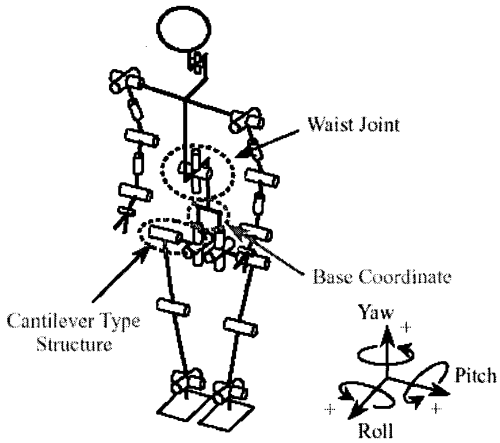


Figure 2. Joint Configuration.

3. Vision System

HRP-2 has a vision system for 3-D measurement and object recognition [3].

Figure 3 shows the trinocular stereo camera system. The camera distance of the right and the left camera is 140[mm] and the center camera is located 70[mm] above. The view angle of each camera is 31.8° horizontal and 24.5° vertical. The cameras do not have zoom systems or congestion control mechanisms, but they do have electrical shutters to adapt to the brightness of the scene.

HRP-2 has two computers, one is for the vision system and the other is for the control system. The vision system's CPU is a PentiumIII 1.26[GHz], with 256 [MB] memory, and the capture card is a Cybertek CT-3001RGB. The image processing and 3D range data calculation is processed by the VVV (Versatile Volumetric Vision) software system[8]. The correlation method is applied for measuring ground displacement and the segment-based stereo method is applied for object recognition[7]. The acceleration technique proposed by Okada et al. [5] is also used.

The area which can be measured in one measurement of the ground is a trapezoid where the distance is from 880[mm] to 1950[mm], the closest edge is 750 [mm] and the farthest edge is 1050 [mm].

The ground displacement measurement error due to the discrete value of the disparity is approximately 20 [mm],

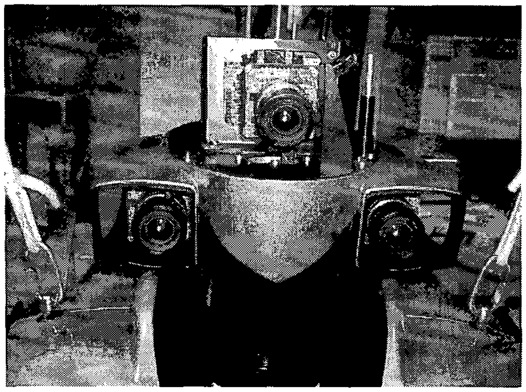


Figure 3. HRP-2 vision system.

but with interpolation it can be reduced to approximately 15 [mm]. Figure 4 shows histograms of z (vertical direction) value for a flat floor. Two curves are histograms with and without interpolation, which shows improvement by interpolation clearly. Because the origin of the foot coordinate is shifted 100 [mm] from the floor, it has a peak at –100 [mm].

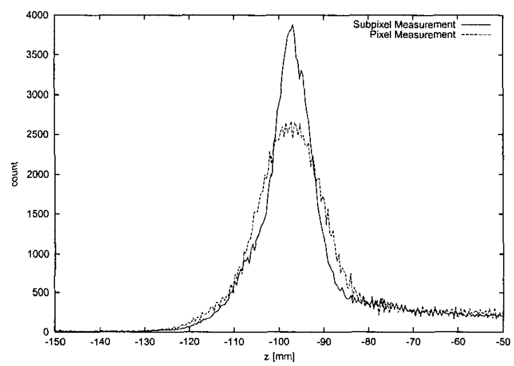


Figure 4. Histogram of floor z level.

4. Transformation

Range data obtained by the vision system has values in the coordinates where the vision system is calibrated. To register the range data of multiple views, they must be transformed into a world coordinate that is fixed on the ground.

The vision system is in the head of the robot, and it is supported by the neck joints (2 DOF), waist joints (2 DOF) and leg joints (12 DOF). The angles of all joints can be measured simultaneously, and from them the transform matrix between the vision coordinates and the world coordinates is calculated.

The following is the calibration method for the transform matrix between the vision coordinates and the robot coordinates.

A marker is put on the robot finger and the robot moves the hand to several places in the field of the vision (Fig. 5). The center position of the marker is detected using the vision system at each time. Using the detected marker positions in the vision coordinates and the finger positions in the robot coordinates, the transform matrix between the two coordinate systems is calculated.

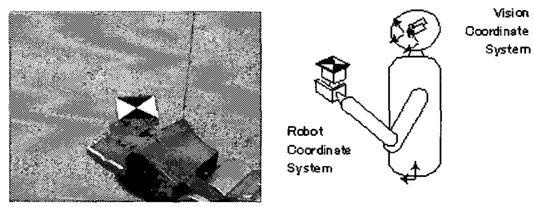


Figure 5. Calibrating transform matrix.

The following is the relation between the range data in the vision coordinates and in the world coordinates.

$$\boldsymbol{X}_w = T_{vw}\boldsymbol{X}_v, \quad (1)$$

where $\boldsymbol{X}_v$ and $\boldsymbol{X}_w$ denote homogeneous vectors in the vision and world coordinates, and T_{vw} is a homogeneous transform matrix. Actually, T_{vw} is a product of multiple transform matrix T_i, where each denotes the matrix related to one of the joints.

$$T_{vw} = \prod_i T_i \quad (2)$$

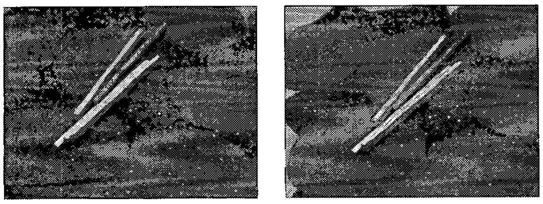
We use two coordinates which can be used as the world coordinates. They are the robot coordinates and the foot coordinates. The origin of the robot coordinate is located at the center of the body. As the z axis of the coordinates is automatically balanced to be vertical by means of gyro-sensors, it is regarded as a world coordinate. Another is the foot coordinate located at the right foot.

As the leg joints does not contribute to T_{vr}, the transfer matrix between the robot coordinates and the vision coordinates, T_{vr} may be more free from measurement error caused by the encoders of the joints than T_{vf}, the transfer matrix between the foot coordinate and the vision coordinate. Another assumption is that T_{vr} may be influenced by errors of the gyro-sensors when controlling the z-axis vertically.

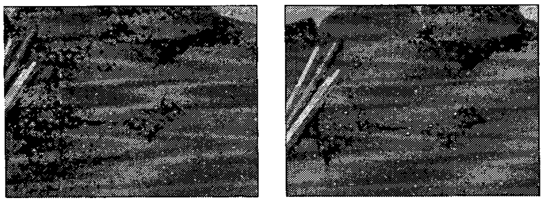
5. Experiments

An outdoor ground scene with some obstacles was set up in our laboratory. Figure 6 shows images taken from the left and the right camera for three views, 20° left view, center view and 20° right view. These views were taken by just rotating around the yaw joint of the waist. Other joints are not moved intentionally, but can be moved by the automatic balancing control system of the robot.

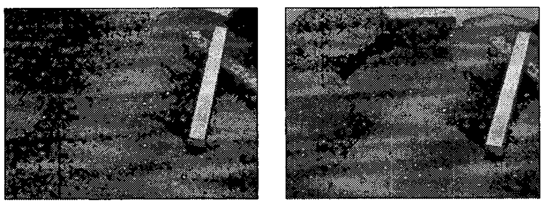
Figure 7 shows the overlapping area of the three views. Almost one third of the data is overlapped both between the



(a) Left 20°



(b) Center



(c) Right 20°

Figure 6. Left and right camera images for three view directions.

left and the center views, and between the center and the right views.

Figure 7 8 shows actual range data from the top view and the front view. The circle patterns appearing in the texture are caused by artificial mounds, which are approximately 200 [mm] in diameter. The mound locates in the center is the tallest at 20 [mm] in height. Other mounds are shorter(5[mm]). There are some piles of timber in the left and right view area. The left pile of timber is 60 [mm] in height and the right one 150[mm].

Figure 9 shows the error in the range data in the overlapped area shown in Fig.7. The horizontal axis shows the error in the z axis and the vertical axis is a count of points. The data sets are (1) the left and the center views and (2) the center and the right views, and for each data set, transformation with the robot coordinates and the foot coordinates is performed.

As the difference caused by the transform matrix is very small, the curves are almost the same. The left peak at –6.6 [mm] is for data set 1, and the right peak at –5.1 [mm] is for data set 2. Almost all data have error between –15 [mm] and 5 [mm]. Total errors include errors in calculating range data (Fig. 4) and errors in transformation, but the latter are relatively small and can be considered negligible. The peak drift from zero was likely caused by some calibration error in the yaw rotation axis of the waist.

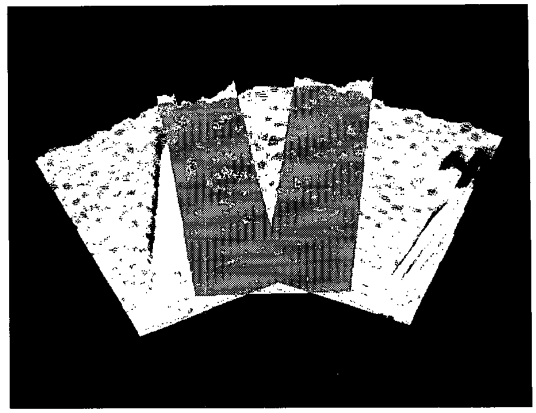


Figure 7. Three views mapped on $z = 0$ plane.

In the experiments described in this paper, the robot stood on a flat floor, which means the z-axis of the foot coordinates is assured to be vertical. In an actual outdoor environment, this is not assured so the foot coordinates may be unreliable in rotation. On the other hand, though the robot coordinates are controlled to be vertical and are reliable in rotation, the exact displacement from the ground is unsure so the elevation is not reliable. These coordinates can be combined to produce a more robust transform matrix in an outdoor environment.

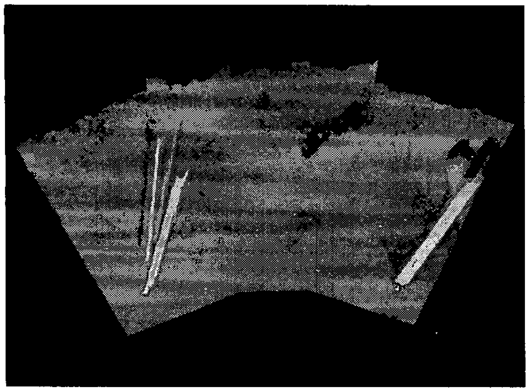
As a preliminary experiment, the range data is processed to decide a clear route for the robot. Figure 10 shows sliced range data at $z = z_G + 10$ [mm], where z_G is the z value of the ground level obtained from Fig. 4. Where the z value is smaller than $z_G + 10$ [mm] are shown in white. The range data indicates that to the right of the area, a wide route under $z_G + 10$ [mm] exists.

Another result of data processing (Fig. 11) shows the sliced image by the angle between the normal vector and z axis. The image is colored in four levels of the angle, from lighter to darker, under 10 deg., under 20 deg., under 30 deg., and otherwise. Although the quality of the range data is improved by interpolation, it is still not free from step-like noise. Such noise affects the calculation of the normal vector and makes the errors large. Nevertheless, some clear area can be found in the right part of the figure.

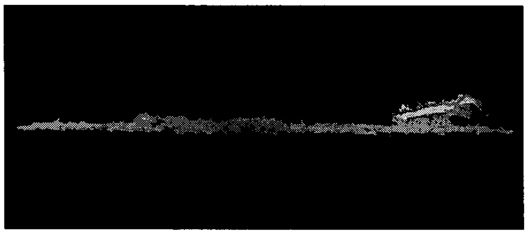
Calculation errors in the range data can be improved by using zooming and congestion mechanisms. With such devices, the normal vector difference map can also be used for route decision making.

6. Conclusions

This paper presented how the HRP-2 humanoid robot gets ground surface displacement information by a registration of multiple view range data. First the robot configuration and the vision system for measuring range data was presented. Next, the transformation between the vision and the world coordinates was described. Finally, our experimental results using



(a) Top view.



(b) Front view.

Figure 8. Registered range data.

an outdoor scene were described. Our results show that errors caused by transformation are relatively small compared to errors in range data calculation. And our preliminary results on route decision in the registered range data are promising.

Future study includes increasing the variation of scenes, improving range data accuracy with the additional optical system, and performing walking experiments with the robot in the route found in the scene.

Acknowledgements

We express our thanks to Kenji Kaneko, Hirohisa Hirukawa and the researchers in the Humanoid Robot Group in AIST and to Takakatsu Isozumi and others in Kawada Industries, Inc. for developing a beautiful humanoid robot. They also pleasantly accepted our requests for changes in the vision system. Our co-workers in the VVV group in AIST gave helpful advice especially on correlation method improvements.

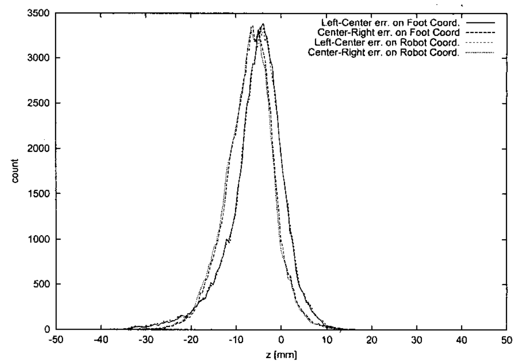


Figure 9. Histogram of error magnitude between (1) range data for left and center views and (2) range data for center and right views, for both foot and robot coordinates.

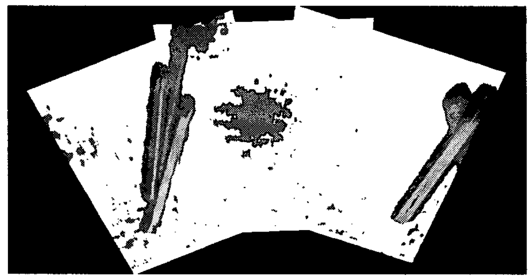


Figure 10. Clear zone where $z < z_G + 10[mm]$. z_G is the ground level.

References

[1] G. DeSouza and A. Kak. "Vision for Mobile Robot Navigation: A Survey". *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(2):237–267, Feb. 2002.

[2] K. Kaneko, F. Kanehiro, S. Kajita, K. Yokoyama, K. Akachi, T. Kawasaki, S. Ota, and T. Isozumi. "Design of Prototype Humanoid Robotics Platform for HRP". In *Proc. IEEE/RSJ Int'l Conf. Intelligent Robots and Systems*, pages 2431–2436, Oct. 2002.

[3] Y. Kawai, Y. Fukase, R. Ikeno, Y. Ishiyama, , and F. Tomita. "Development of the Vision System for the HRP-2P Humanoid Robot". In *Proc. of MVA 2002*, pages 257–260, Dec. 2002.

[4] L. Matthies, T. Balch, and B. Wilcox. "Fast Optical Hazard Detection for Planetary Rovers Using Multiple Spot Laser Triangulation". In *Proc. IEEE Int'l Conf. Robotics and Automation*, volume 1, pages 859–866, Apr. 1997.

[5] K. Okada, S. Kagami, M. Inaba, and H. Inoue. "Onbody Realtime Disparity Image and Flow Generation System based on Highspeed Correspondence Matching using PC". *Journal of the Robotics Society of Japan*, 18(6):896–901, Sept. 2000. (in Japanese).

[6] R. Simmons, E. Krotkov, L. Chrisman, F. Cozman, R. Goodwin, M. Hebert, L. Katragadda, S. Koenig, G. Krishnaswamy, Y. Shinoda, W. Whittaker, and P. Klader. "Experience with Rover Navigation for Lunar-Like Terrains". In *Proc. 1995 IEEE Int'l Conf. Intelligent Robots and Systems*, volume 1, pages 441–446, Aug. 1995.

[7] Y. Sumi, Y. Kawai, T. Yoshimi, and F. Tomita. "3D Object Recognition in Cluttered Environments by Segment-Based Stereo Vision". *Int'l J. of Computer Vision*, 46(1):5–23, Jan. 2002.

Figure 11. Clear zone extracted by normal vectors.

[8] F. Tomita, T. Yoshimi, T. Ueshiba, Y. Kawai, Y. Sumi, T. Matsushita, N. Ichimura, K. Sugimoto, and Y. Ishiyama. "R&D of Versatile 3D Vision System VVV". In *Proc. IEEE Int'l Conf. on SMC*, pages 4510–4516, Oct. 1998.