

Integration of Real-time Binocular Stereo Vision and Whole Body Information for Dynamic Walking Navigation of Humanoid Robot

Kei Okada, Masayuki Inaba, Hirochika Inoue
Department of Mechano-Informatics
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku
Tokyo 133-8656 JAPAN
k-okada@jsk.t.u-tokyo.ac.jp

Abstract

A method to segment a floor and an obstacle region from images, is a fundamental function for robots in real-world. This paper describes a floor detection method by integrating binocular stereo vision and whole body information, for walking direction control of a humanoid robot.

We developed the humanoid robot navigation system using vision based local floor map. The developed system consists of a map building stage and a walking direction control stage. In the map building stage, the system builds a local floor map around a robot by integrating floor region information from visual input and whole body posture information. Plane Segment Finder (PSF) algorithm, which is able to extract planner surface from 3D vision input, is utilized to segment a floor and an obstacle regions. Floor region segmentation from input images is represented in view coordinates, then whole body posture information is utilized to transform from view coordinates to body coordinate to build a local floor map. In another stage, the system search for open space direction on the local floor map and control walking direction toward open space to avoid obstacles.

Finally, walking navigation experiments based on floor detection using a life-size humanoid robot are shown.

1. Introduction

Recently many humanoid robots have been developed, specially in the research field of dynamic walking control[1, 2, 3, 4, 5]. However these humanoid robots is not able to autonomously walking based on environment recognition. To realize a humanoid robot in real-world and unknown environment, a sensor based navigation function is required.

The three dimensional recognition of environments is important for a robot that behaves in a real world. Recently many remarkable real-time depth map gener-

ation systems are developed and some of them are sold commercially[6, 7, 8, 9].

Although the development of these real-time depth map generation systems are a remarkable achievement toward understanding real environments, there are few robots which utilize depth map information[10, 11].

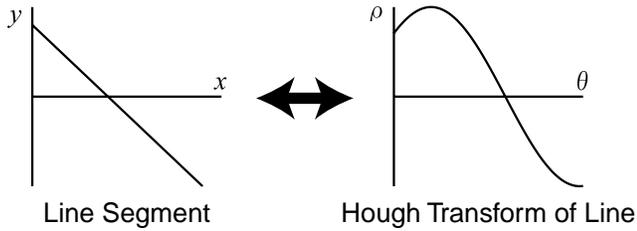
In this paper we present a method to segment floor region and obstacle region by integrating binocular stereo vision and whole body information, for walking direction control of a humanoid robot.

We developed the humanoid robot navigation system using vision based local floor map. The developed system consists of a map building stage and a walking navigation stage. In the map building stage, the system builds a local floor map around a robot by integrating floor region information from visual input and whole body posture information. In order to segment floor and obstacle regions, Plane Segment Finder (PSF) algorithm is utilized, which is able to extract planner surface from 3D vision input.

In this paper, we describes an algorithm of PSF for extract planner surface detection from input images that includes 1) Planner surface segment candidate extraction using 3D Hough Transformation from depth map information, 2) fitting the planner surface segment candidates to the depth map images to detect the partial planner surface region, since the extracted planner surface segment candidates are general planes, with no boundary. To achieve real-time Plane Segment Finder system, we apply 1) Recursive correlation method for depth map generation, 2) Partial planner surface segment extraction by using Randomized Hough Transformation method for PSF.

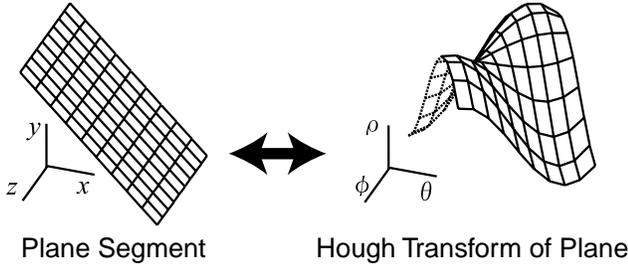
In the another stage, the system search for open space direction on the local floor map and control walking direction toward open space to avoid obstacles.

Finally, walking navigation experiments based on floor detection using a life-size humanoid robot are shown.



Line Segment

Hough Transform of Line



Plane Segment

Hough Transform of Plane

Figure 1. Hough Transform of 2D line vs. Hough Transform of a planner surface

2. Plane Segment Finder(PSF)

In order to segment floor region and obstacle region from binocular stereo vision inputs, we utilize 3D Hough Transformation based 3D planner surface detection method, which we call Plane Segment Finder (PSF)[12].

In previous research of computer vision area, many researchers have been investigating Range Image Segmentation method which is to label input distance image into planner surface patches with Region growing method[13], Split-and-merge method[14], Clustering method[15, 16], Scan line structure method[17], and so on.

These computer vision researches utilized laser range finder which is able to precise distance information, however these system usually too large to embedded in a mobile robot. Planner surface detection system for robot application requires small equipment, such as binocular vision system. However distance information from stereo vision system is less precise. Therefore, a planner surface detection method that is robust against noise is required. Thus, Hough Transformation method[18, 19] which is a well-known method which is robust to noise and occlusions, is utilized.

The algorithm of PSF is as follows. 1) Planner surface segment candidate extraction using 3D Hough Transformation from depth map information. Since, the extracted planner surface segment candidates are general planes with no boundary, 2) fitting the planner surface segment candidates to the depth map images to detect the partial planner surface region is applied.

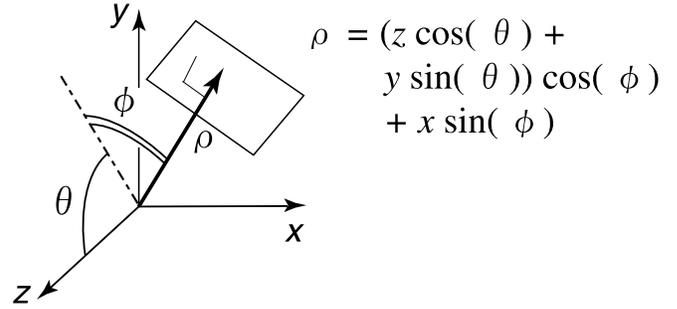


Figure 2. Parametric representation of a planner surface

2.1. Binocular Stereo Vision based Planner Surface Extraction using 3D Hough Transformation

We utilized the Hough transform method for extracting planner surface segment candidates. The Hough Transformation method is a well-known method which is robust to noise and occlusions, and generally used for extracting lines, circles or ellipses.

Figure 1 shows difference between Hough Transform of a 2D line and Hough Transform of 3D planner surface. For a straight 2D line segment extraction using the Hough Transform method, one point on a 2D line is transformed to a curved line in a Hough space (parameter space). In the case of our Hough transform for extracting planner surface extraction, a point on a 3D plane is transformed to a curved surface.

To apply the Hough Transform method for a planner surface extraction, we adopt the following parametric representation of a plane. Figure 2 shows this representation, where ρ is the distance between a plane and the origin, ϕ is angle against the x axis, θ is angle against y the axis. (x_0, y_0, z_0) is the point on the plane.

$$\rho = (x_0 \cos(\phi) + y_0 \sin(\phi)) \cos(\theta) + z_0 \sin(\theta)$$

This equation has the following features.

1. A *point* in 3D space becomes a *curved surface* in a Hough space.
2. A *plane* in 3D space becomes a *point* in a Hough space.

Hence, to extract plane segment candidates, 1) we transform (“vote”) all 3D points into the Hough space, 2) Detect peak points in the Hough space, which correspond to planner surface candidates in 3D space.

2.2. Planner Surface Region Detection by Fitting to Depth-map Image

The planner surface segment candidates which are extracted from distance information using the Hough

Table 1. Accuracy of plane segments detection

	ρ [cm]	θ [degree]	ϕ [degree]
Plane 1	10.11	98.8	56.3
Plane 2	12.32	98.8	56.3
Plane 3	14.95	98.8	56.3

ρ is distance between a plane and the origin.

ϕ is angle against x axis.

θ is angle against y axis.

Table 2. Calculation time of each functions on PentiumIII-750MHz (Dual CPU machine))

Function	Time
LoG filter *1	5.92
Shift and rotate a input image	10.14
Disparity image generation *2	36.04
Distance Information Calculation	5.49
Smooth Distance Information *3	33.87
Plane Candidate Extraction *4	141.28
Plane Segment Detection	0.34
Total time of processing	233.08

(unit:msec): The size of images are 128×128 pixel.

*1 7x7 filter for 2 images.

*2 search area is 0 to 24.

*3 5x5 average filter.

*4 parameter space is $60 \times 72 \times 72$ for ρ, θ, ϕ .

Transform, are general plane, i.e. there is no boundary. Thus a method to detect partial plane segments from plane segment candidates is required. We utilize the following simple method.

For each plane segment candidate:

1. Generate the virtual depth map image of the extracted plane segment candidate using parameters of the plane.
2. Calculate distance between the virtual depth map image and the real depth map image (the input disparity image) at each pixel.
3. If the distance is lower than the threshold, this point categorized to the partial plane segment.

3. Design and Development of Real-Time Plane Segment Finder System

For robotics applications, real-time and robustness are important features. We developed PC-based real-time

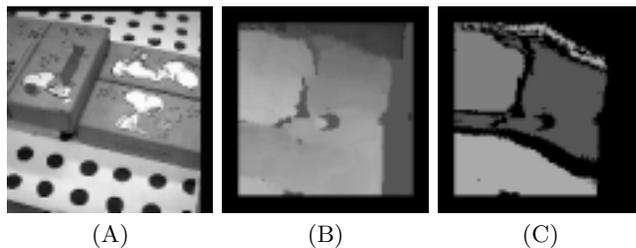


Figure 3. A) Input image. B) Depth map (Brighter is closer). C) Result of plane segment detection.

Plane Segment Finder system. The advantage of using PC as a vision system is 1) Current CPU is as fast as hardware system due to high frequency clock cycle and SIMD-based multimedia instruction set suit visual processing. 2) Software approach enables us to implement complex algorithms such as reliability evaluation.

Our system consists of two parts, one is depth map generation system [22] and the other is plane segment detection system.

3.1. Real-Time and Reliable Depth Map Generation Sub-System

To achieve the real-time depth map generation system, we utilized following three key issues. (1) Recursive Correlation Method [23, 7], (2) Algorithmic Optimization for 2nd Level Cache, (3) Multimedia instruction set (MMX) implementation.

The Recursive Correlation Method reduces the computation time of finding corresponding points between two images from $O(N^2W^2D)$ to $O(N^2D)$, where N is the input image size, W is the window size of correlation and D is the max disparity size.

We apply the ‘‘Consistency Checking Method’’ [24, 25] to generate the reliable depth map. Since stereo matching suffers from occlusions or mismatches fundamentally.

3.2. Randomized Hough Transform for Real-Time Plane Segment Finder

The Hough Transform method has the advantage that it is robust to noise. However, the computational cost and required memory size are very high.

These disadvantages are very significant problems to develop real-time system for robotics applications. To cope with this problem, a method called Randomized or Probabilistic Hough Transform [20, 21] has been introduced. We apply this idea to our Hough Transformation for extracting planes.

In the original Hough Transform method, each point on the input image is transformed to a curved surface, therefore we need to vote for $\Theta - \Phi$ points in Hough array for each point in the distance image.

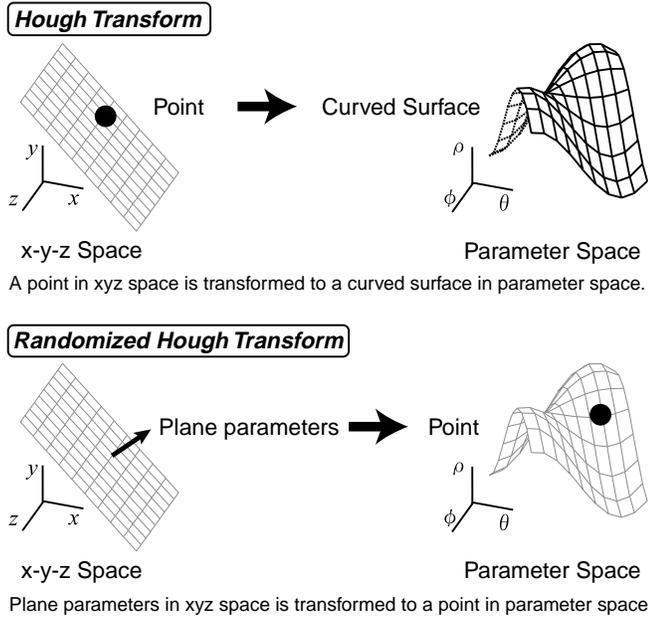


Figure 4. Randomized Hough Transform

By introducing Randomized Hough Transform method, we need to vote for S points in Hough array for each point in the distance image (S is constant).

This method reduces the computational cost of Hough Transform. The points to be voted in Hough array for a pixel in distance image is $S(\text{constraint})$ points, where $\Theta - \Phi(\text{resolution of each axis})$ points without Randomized Hough Transform method.

3.3. Evaluation of Plane Segment Finder System

We placed blocks on a desk. *Figure 3(A)* is the input image, and the depth image is shown in *Figure 3(B)*. Then, we detected plane segments from the depth image using PSF, *Figure 3(C)* shows the results of PSF. Pixels which have the same gray level are pixels on the same plane segment. These results show there are three planes in the input image.

Figure 3(D) is three-dimensional representation of the result of PSF. Each detected plane segment is rendered with a texture-map. The images shows the same 3D data from different view points, and the arrows in the images are $x - y - z$ axis (5[cm]).

Table 1 shows the accuracy of the detected plane segments for each 3 blocks. From the results, we can see that the distances between each plane are 2.21[cm], 2.62[cm] respectively, where the real height of block is 2.5[cm].

Table 2 shows the calculation time of the developed system. The system is able to detect planes at a rate of up to 4[Hz]. The size of input image is 128x128 pixels, and window size of correlation is 21, disparity size is 32. The size of Hough space is 100 for ρ and 72 for θ, ϕ .

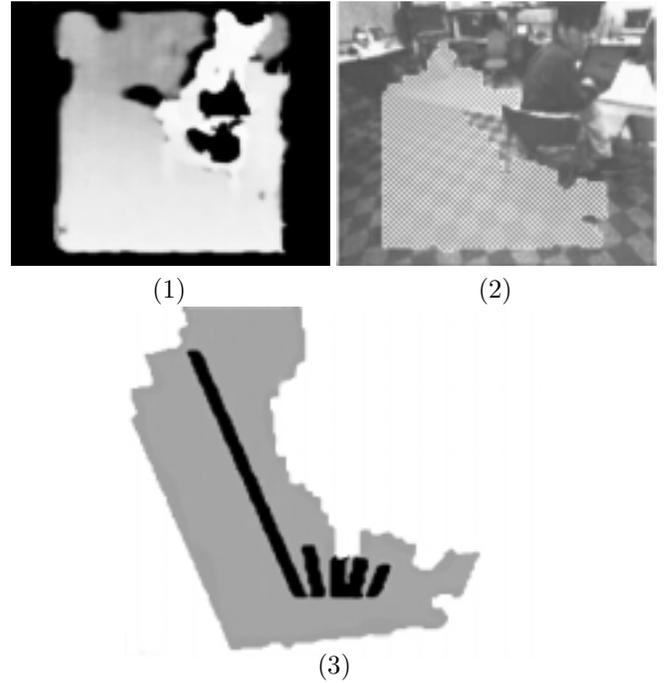


Figure 5. (1) Depth map (2) Result of floor recognition (3) Generated local map and open space search

4. Local Floor Map Building by Integrating PSF Results and Whole Body Posture Information

Though local floor map is required for walking navigation of the robot, the floor region segmentation results of PSF is represented in view coordinates, then the whole body posture information is used to perform the transformation from view coordinates to body coordinate to build a local floor map.

Figure 5 shows the results of floor region segmentation of PSF and local floor map by integrating the results of PSF and whole body posture information.

Figure 5(1) is the depth map image, *(2)* is the result of map building by integrating detected floor region information, which is represented in view coordinates. *Figure 5(3)* shows the local floor map, which is represented in body coordinates of the robot. Each pixel of this map has a binary information whether floor region or non-floor region, which includes both obstacle regions and regions out of the field of view.

5. Walking Navigation Experiments of Humanoid 'H7' using Vision-based Local Floor Map

Walking navigation experiments based on floor detection by integrating binocular stereo vision based Plane

Segment Finder and whole body information with the life-size humanoid robot H7[26] were carried out[27].

The robot control it's walking direction based on the result of open space search on the 2D local floor map as shown in previous section and *Figure 5(3)*. The open space search method examines which direction an obstacle exists between -15 to +15 [degree] every +5 [degree]. Black lines in the image show each direction of the open space search. *Figure 6* show a result local floor map based walking direction control.

6. Conclusion

This paper describes walking direction control system of a humanoid robot with map building based on integration of vision based floor region recognition and whole body information, and open space direction search method.

The system builds a local floor map around a robot by integrating floor region information from visual input and whole body posture information. Plane Segment Finder (PSF) algorithm is utilized to segment floor and obstacle regions. To control walking direction to avoid obstacles, the system search for open space direction on the local floor map.

Finally, walking navigation experiments based on floor detection using a life-size humanoid robot are shown.

References

- [1] Koichi NISHIWAKI, Tomomichi SUGIHARA, Satoshi KAGAMI, Fumio KANEHIRO, Masayuki INABA and Hirochika INOUE: Design and Development of Research Platform for Perception-Action Integration in Humanoid Robot : H6, Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'00), pp. 1559–1564, (2000).
- [2] Kazuo Hirai: Current and future perspective of honda humanoid robot. In Proc. IEEE/RSJ International Conference Intelligent Robots and Systems (IROS'97), pp. 500–508, (1997).
- [3] Sony Corporation: SDR-4X, <http://www.sony.co.jp/SonyInfo/News/Press/200203/02-0319>, (2002).
- [4] Shuji Hashimoto, Hideaki Takanobu et al.: Humanoid Robots in Waseda University-Hadaly-2 and WABIAN, IEEE-RAS International Conference on Humanoid Robots (Humanoids2000), (2000).
- [5] H. Inoue, S. Tachi et al.: HRP: Humanoid robotics project of MITI, Proceedings of the First IEEE-RAS International Conference on Humanoid Robots, Cambridge, MA, September 2000.
- [6] K. Konolige. Small Vision Systems: Hardware and Implementation. In Y. Shirai and S. Hirose, editors, *Robotics Research: The Eighth International Symposium*, pp. 203–212. Springer, 1997.
- [7] T. Kanade, A. Yoshida, K. Oda, H. Kano, and M. Tanaka. A Stereo Machine for Video-rate Dense Depth Mapping and Its New Applications. In *Proc. of the 1996 International Conference on Computer Vision and Pattern Recognition*, pp. 196–202, Jun 1996.
- [8] Point Grey Research. <http://www.ptgrey.com>.
- [9] P. Corke, P. Dunn, and J. Banks. Frame-rate stereopsis using non-parametric transforms and programmable logic. In *Proc. IEEE International Conference on Robotics and Automation*, pp. 1928–1933, 1999.
- [10] K. Okada, S. Kagami, M. Inaba, and H. Inoue. Vision-based action control of quadruped legged robot jrob-1. In *Proc. of 9th International Conference on Advanced Robotics(ICAR'99)*, pp. 451–456, 1999.
- [11] T. Wiliamson and C. Thorpe. A trinocular stereo system for highway obstale detection. In *Proc. IEEE International Conference on Robotics and Automation*, pp. 2267–2273, 1999.
- [12] K. Okada, S. Kagami, M. Inaba and H. Inoue: Plane Segment Finder : Algorithm, Implementation and Applications, Proceedings of International Conference on Robotics and Automation (ICRA'01) pp 2120-2125 (2001).
- [13] A.E. Johnson, S.B. King, and R. Szeliski. Extraction of concise and realistic 3-d models from real data. Crl 95/7, Digital Equipment Corporation, Cambridge Research Lab, October 1995.
- [14] B. Parvin and G. Medioni. Segmentation of range images into planar surfaces by split and merge. In *In Proc of International Conference on Computer Vision and Pattern Recognition(CVPR '86)*, pp. 415–417, 1986.
- [15] Richard L. Hoffman and Anil K. Jain. Segmentation and classification of range images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 9, No. 5, pp. 608–620, September 1987.
- [16] P.J Flynn and A.K. Jain. BONSAI: 3D Object Recognition Using Constrained Search. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 10, pp. 1066–1075, October 1991.
- [17] X. Y. Jiang and H. Bunke. Fast segmentation of range images into planner regions by scan line grouping. *Machine Visoin and Applicatoins*, Vol. 7, No. 2, pp. 115–122, 1994.
- [18] D.H. Ballard. Generalizing the Hough Transform to Detect Arbitrary Shapes. *Pattern Recognition*, Vol. 13, No. 2, pp. 111–122, 1981.
- [19] R.O. Duda and P.E. Hart. Use of the Hough Transform to Detect Lines and Curves in Pictures. *Communicatoins of the ACM*, Vol. 15, No. 1, pp. 11–15, 1972.
- [20] N. Kiryati, Y. Eldar and A.M. Bruckstein: A Probabilistic Hough Transform, *Pattern Recognition*, vol. 24, no. 4, pp 303-316 (1991).

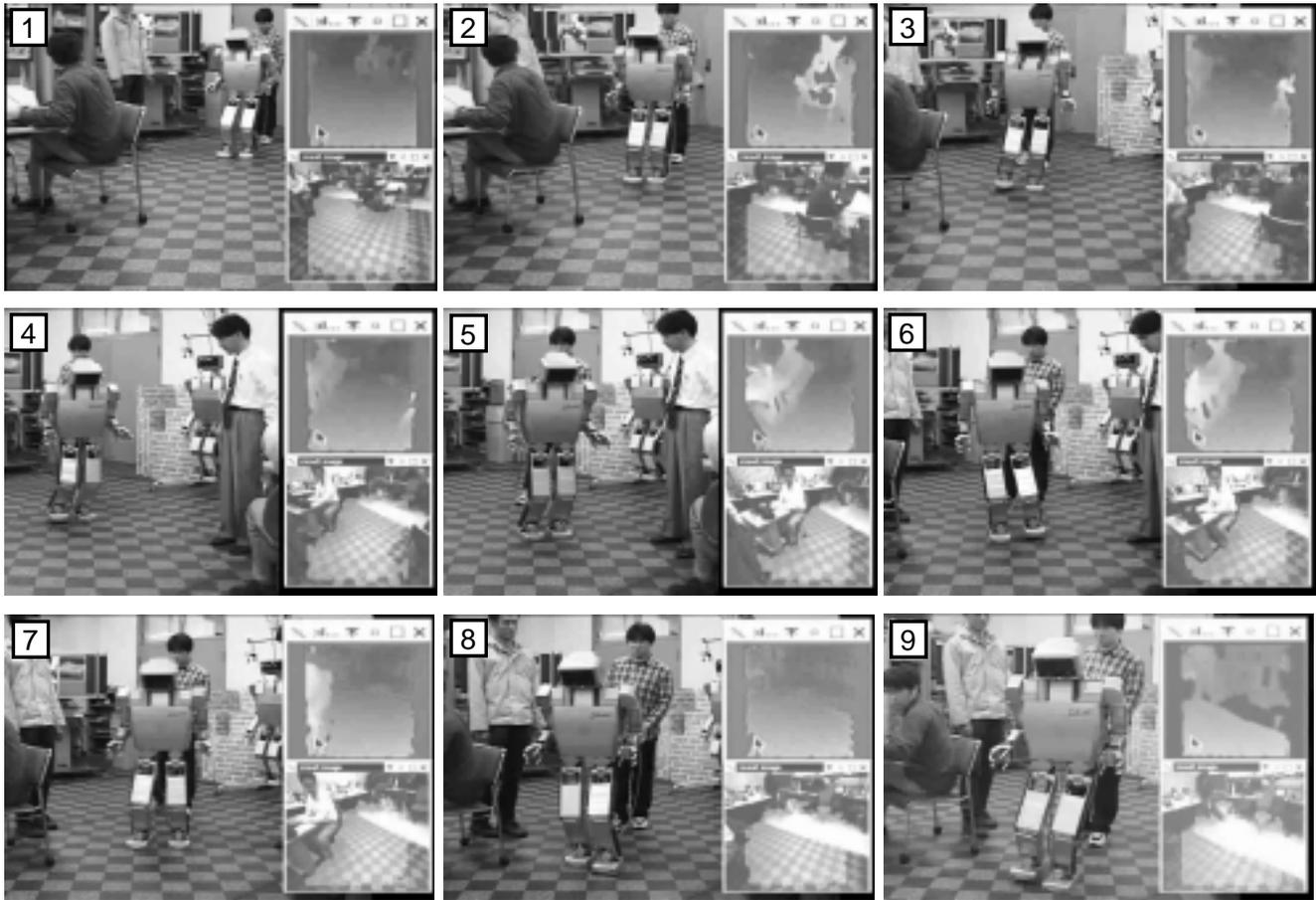


Figure 6. The humanoid robot based on walking direction control using open space search with local floor map

- [21] Heikki Kälviäinen, Petri Hirvonen, Lei Xu and Erkki Oja: Comparisons of Probabilistic and Non-probabilistic Hough Transforms, Proceedings 3rd European Conference on Computer Vision, pp 351-360, (1994).
- [22] S. Kagami, K. Okada, M. Inaba, and H. Inoue. Design and implementation of onbody real-time depthmap generation system. In *Proc. of IEEE International Conference on Robotics and Automation (ICRA'00)*, pp. 1441-1446, 2000.
- [23] O. Faugeras, B. Hots, H. Mathieu, T. Viéville, Z. Zhang, P. Fua, E. Théron, L. Moll, G. Berry, J. Vuillemin, P. Bertin, and C. Proy. Real Time Correlation-Based Stereo: Algorithm, Implementations and Applications. Technical Report N°2013, INRIA, 1993.
- [24] P. Fua. A Parallel Stereo Algorithm that Produces Dense Depth Maps And Preserves Images Features. In *Machine Vision and Applications*, pp. 35-49, 1991.
- [25] R. Bolles and J. Woodfill. Spatiotemporal Consistency Checking of Passive Range Data. In T. Kanade and R. Paul, editors, *Robotics Research: The Sixth International Symposium*, pp. 165-183. International Foundation for Robotics Research, 1993.
- [26] Koichi Nishiwaki, Satoshi Kagami, Yasuo Kuniyoshi, Masayuki Inaba, and Hirochika Inoue. Online generation of humanoid walking motion based on a fast generation method of motion pattern that follows desired zmp. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2684-2689, 2002.
- [27] K. Nishiwaki, S.Kagami, J.J. Kuffner, K. Okada, Y. Kuniyoshi, M. Inaba, and H. Inoue. Online humanoid locomotion control using 3d vision information. In *Proc. IEEE/RSJ Int. Symposium on Experimental Robotics (ISER'02)*, 2002.