

Navigation of a Mobile Robot using Mono-Vision and Mono-Audition

W. Choi*, C. Ryu**, and H. Kim***

Department of Automation Engineering, Inha University

Inchon, 402-751, Korea

*g1991096@inhavision.inha.ac.kr

**g1982580@inhavision.inha.ac.kr

***hikim@inha.ac.kr

ABSTRACT

Incorporation of various types of sensors increases the degrees of autonomy and intelligence of mobile robots (mobots) in perceiving surroundings, which at the same time imposes a large computational burden on data processing. The purpose of the research presented in this paper is to develop a low-cost multi-sensor system and incidental algorithms for autonomous navigation of a mobot.

This paper proposes digital image processing schemes for map-building and localization of a mobot using a monocular vision system and a single ultrasonic sensor in indoor environments. In localization, the camera calibration is preceded so that the depth information can be acquired from the image obtained by a single camera. For map-building, fast and effective image processing techniques based on morphology are applied to reduce computational complexities.

For preliminary experiments, we have integrated a mobot system whose main components are a mono-vision system, a single ultrasonic sensor, and a notebook PC, mounted on a mobile base. The proposed algorithms were implemented, and the mobot was able to localize itself in an allowed position error range and to locate dynamic obstacles moving reasonably fast inside a building. The overall results demonstrate the suitability of the proposed methods for developing autonomous service mobots in indoor environments.

1. INTRODUCTION

One of the most fundamental requirements for intelligence of autonomous mobots is the automated navigation skill of traveling from a current position to a goal position without unplanned collision. The intelligent navigation consists of three major components: (i) *path planning* where a mobot searches for optimal paths between the two positions based on global information provided by a map, (ii) *collision avoidance* where a mobot keeps itself away from unexpected obstacles using sensor measurements while traveling, and (iii) *localization* in which a mobot finds its position in the global coordinate system of the map[1].

In particular, the exact and accurate map is essential as a model of environment for path planning and localization. Although the map can be constructed by human operators, it is very cumbersome to update the map when the map requires constant modification for frequently changing environment. The map-building is an ancillary function of a mobot constructing a digital map based on its sensor measurements.

Meanwhile, the purpose of perception of autonomous mobots in navigation is two-fold, localization and obstacle detection. While active localization requires pre-known artificial landmarks[2], passive localization detects structural features of environment using various sensors. Although vision is not necessary for most of mobot tasks, it is desirable for unstructured environments due to its long-range, high resolution, and passive sensing ability[3]. This paper presents a map-building scheme using single ultrasonic sensor, where an occupancy grid map is converted to a grayscale image and then various digital image processing techniques are applied to the image to extract spatial features and possible routes in indoor environment. It also describes simple but effective procedures of camera calibration and image processing techniques for localization using a mono-vision system. The localization includes a camera calibration process which enables the mobot to measure the range in a single image frame, and utilizes linear structural features in indoor environment such as doors and corridor lines.

2. PATH PLANNING USING IMAGE PROCESSING

There are two representation models of map for autonomous mobots, grid-based map and topological map. While grid-based map[4] models mobot's environment by equi-spaced grid cells each of which contains a degree of occupancy indicating the presence of obstacles in the corresponding region of the environment, topological map[1] represents environment as a graph where nodes correspond to distinct places such as corners and edges, and arcs are clear routes between pairs of nodes. Since both models possess orthogonal strengths and weaknesses, one can be supplementary to the other. This section presents a series of morphological image processing procedures to find an optimal path from an occupancy grid-based map.

2.1. Building an occupancy grid-based image

In order to acquire an occupancy grid-based map, the mobot repeats stop-and-go at every 1m interval following the wall of indoor environment as shown in Fig. 1, and fires the single ultrasonic sensor around with the interval of 18° as depicted in Fig. 2.

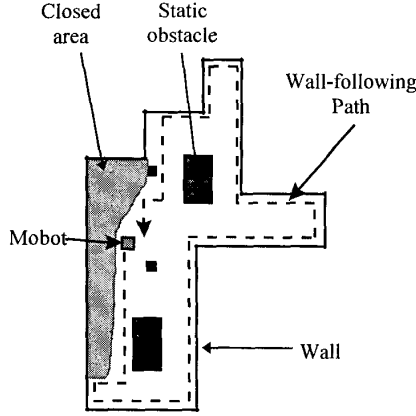


Fig. 1. Wall-following path for map-building.

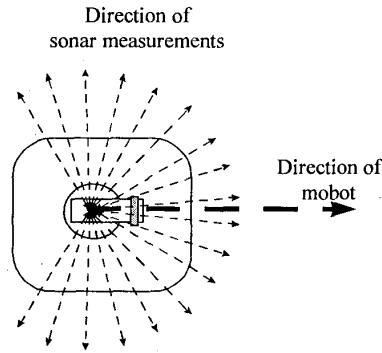


Fig. 2. Directions of ultrasonic sensor measurements.

At each range measurement d by the sonar sensor, the occupancy value at a grid cell (r, θ) , where r is the distance from the sensor and θ is the angle from the beam direction, is given by a joint probability density function as following:

$$p(d|r, \theta) = \begin{cases} \frac{(1 - p_{min})}{\sqrt{2\pi}\sigma_d} \exp\left[-\frac{(r-d)^2}{2\sigma_d^2}\right] + p_{min} & (r-d) \leq 3\sigma_d \\ 0.5 & (r-d) > 3\sigma_d \end{cases}$$

The above density function is jointly formed by a Gaussian distribution in r direction and a uniform distribution in θ direction, which is similar to the one proposed by Elfes[5]. The modification is made so that the occupancy values of the cells whose distance from the sensor is less than d cannot be lower than p_{min} , and those of the cells far behind d are assigned 0.5 to

represent uncertainty. This non-zero minimum value assignment prevents erroneous interpretation of measurements caused by abrupt changes of probabilities in later updating. And, the uniform distribution along θ direction is assumed to account for the fact that the cells inside the beam angle of the sonar sensor cannot be discriminated. Fig. 3 is a 3D plot of the proposed probability density function for $d=2m$, $p_{min}=0.4$, and $\sigma_d=0.2m$.

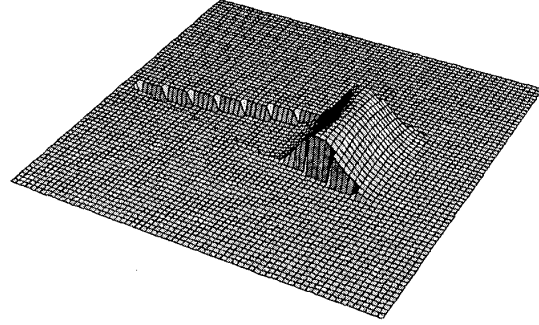


Fig. 3. 3D plot of occupancy values in grid-based map.

Initial degrees of occupancy for cells are set to 0.5, and as the mobot travels and makes sonar measurements, the occupancy values of each cell is updated by using the Bayesian probability updating rule as

$$P[(r, \theta)\{d\}_{t+1}] = \frac{p[d_{t+1}|(r, \theta)] \cdot P[(r, \theta)\{d\}_t]}{\sum_{(r, \theta)} p[d_{t+1}|(r, \theta)] \cdot P[(r, \theta)\{d\}_t]}$$

where $\{d\}_t$ represents the set of entire measurements by the sonar sensor until t . The above updating rule reinforces the degree of occupancy at a certain grid cell from the past value based on the current measurement.

After completing a wall-following path, the degrees of occupancy ranging from 0 to 1 are inverse-linearly converted to integers between 0 to 255 in order to visualize the occupancy grid map by 8bit-per-pixel gray scale image. Fig. 4 is the resulted image for the environment in Fig. 1, where each pixel has the spatial resolution of 0.25m by 0.25m and the darker pixels show more probably occupied region. Once the map is transformed to a 2D gray scale image (it is called *occupancy image*), various image processing methods can be applied to extract spatial information useful for mobot's autonomous navigation. In this research, some of basic functions in morphological image processing[6] have been utilized in order to carry out path planning in the occupancy image.

2.2. Morphological image processing

Morphological image processing is a technique used to manipulate the shape of regions of an image for aesthetic purposes, whose basic idea is to exploit prior knowledge of the shape of image distortions in order to remove these distortions. In the context of morphological image processing, the so-called

structuring element and the basic operators *erosion* and *dilation* are the focus of attention. Erosion removes pixels from region borders while dilation adds pixels to a border. A structuring element which is a mask operator of a given shape determines the removal or addition of pixels by being handled like an ordinary local operator in digital image processing.

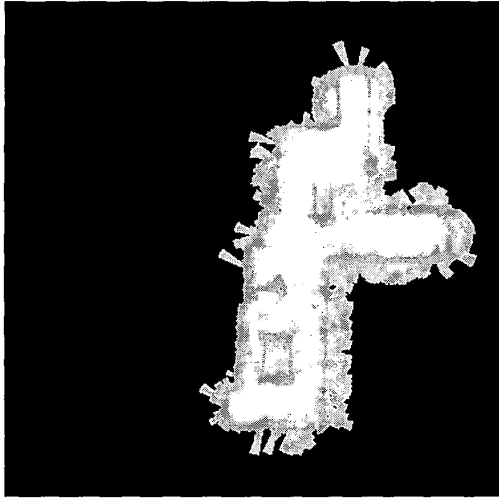


Fig. 4. Occupancy image.

Fig. 5 shows the proposed procedure of image processing to extract an optimal path of the mobot from the occupancy image of the environment. The operations in morphological image processing using the structuring element are based on the following rules :

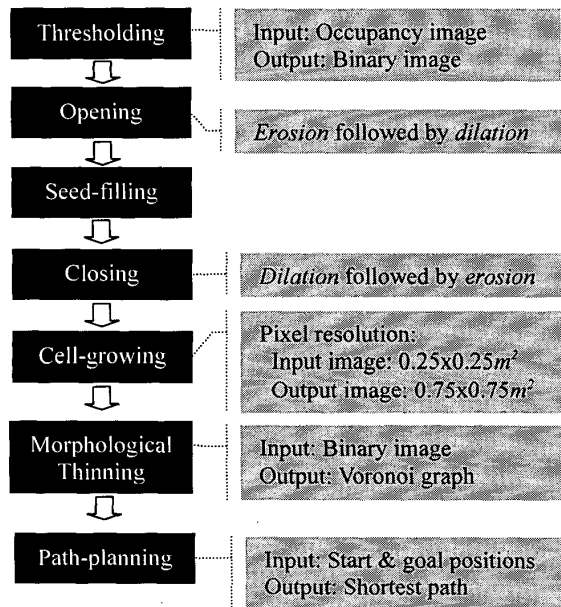


Fig. 5. Procedure of image processing to extract an optimal path.

- Erosion : if the *whole* structuring element lies inside a region, then set the current pixel in the output image to 1.
- Dilation : if *at least one* pixel of the structuring element lies inside a region, then set the current pixel in the out image to 1.

The function of each step is as following:

- Thresholding* dichotomizes multi-level occupancy into two states, *clear* or *occupied*. Fig. 6 is the result of thresholding Fig. 4 with the threshold value of 0.5. It is easily noticeable that a sonar sensor gives quite amount of inconsistent readings.

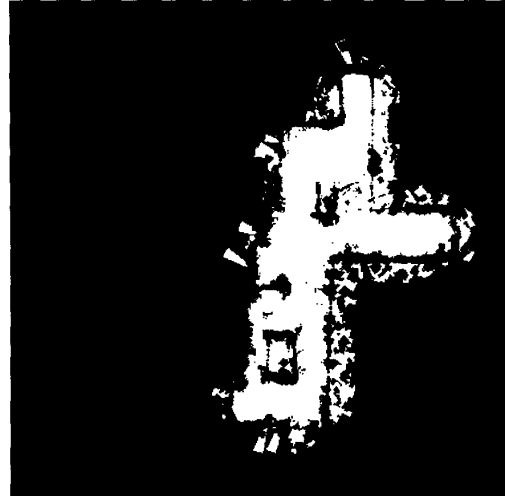


Fig. 6. Binary image after thresholding.

- Opening* removes the borders of frayed regions due to the incorrect sonar measurements, whose result is shown in Fig. 7. The borders become smooth and clear, and the regions corresponding to the incorrect measurements are separated from the target region.

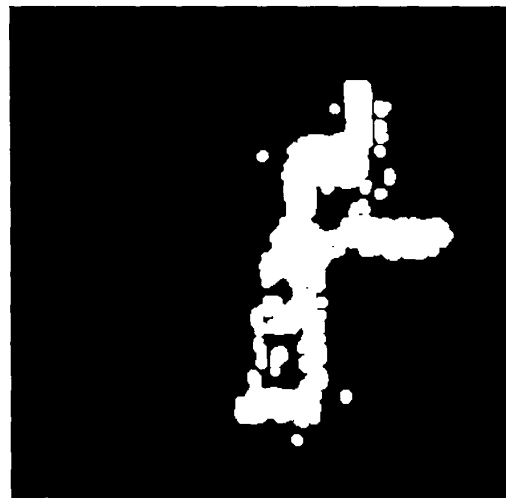


Fig. 7. Result image after opening.

- iii) *Seed-filling* eliminates the small blobs isolated from the target region.
- iv) *Closing* fills the gaps inside the region and along the border to make sure that there is no occupied cell inside the clear region. Fig. 8 is the result of seed-filling followed by closing.
- v) *Cell-growing* combines a block of 3x3 pixels into a single cell. The resulted image is more condensed and the spatial resolution is degraded but still high enough for positioning of the mobot whose base area is $0.75 \times 0.75 m^2$.
- vi) *Morphological thinning* extracts the Voronoi graph of the target region as shown in Fig. 9.

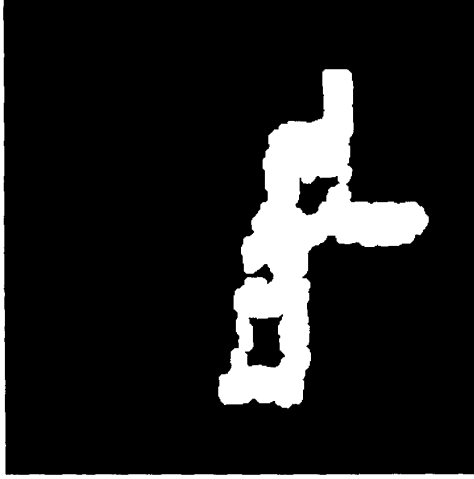


Fig. 8. Result image after seed-filling then closing.

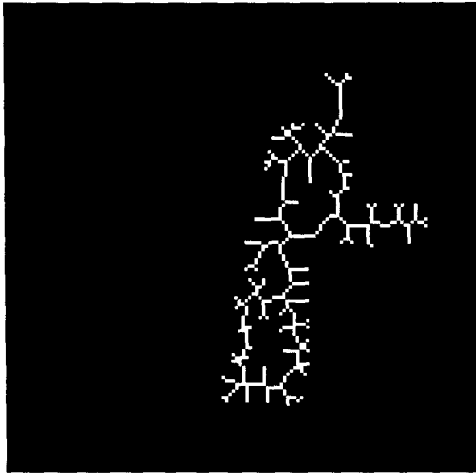


Fig. 9. Voronoi graph after morphological thinning.

In the Voronoi graph, the pixels having more than three branches become the nodes, then an arc is established between a pair of nodes if there is no occupied cell in their connected path. Each arc is assigned a weight corresponding to its distance.

3. LOCALIZATION WITH MONO-VISION

Perception of environment using a mono-vision system requires a mathematical modeling for image projection and coordinate transformation. In practice, the exact mathematical modeling cannot be obtained because of the lack of necessary parameters. Furthermore, as shown in Fig. 10, 2D image does not provide depth information due to the projective transformation from 3D to 2D. Consequently, extraction of the depth from a 2D image can be accomplished only with a certain amount of constraints about environment. In this work, it is assumed that spatial features of environment lie on or near the floor and that the mono-vision system is looking at the floor with a fixed slant angle. This section focuses on camera calibration and image processing for extracting linear features for localization of the mobot based on a mono-vision system.

3.1. Modeling and calibration of vision system

In Fig. 10, a point vector $w = (X, Y, Z)^T$ in the world coordinate system is mapped to a pixel $c = (x, y, z)^T$ in the image coordinate system as following:

$$c_h = PCR_\beta R_\alpha R_\theta Gw_h$$

where c_h is in the homogeneous image coordinate, w_h is in the homogeneous world coordinate, P is the projection matrix of the camera lens, C is the translation matrix by r , R_β , R_α and R_θ are the rotation matrices according to Z , X , and Y -axis by the angles β , α , and θ respectively, and G is another translation matrix by $w_0 = (X_0, Y_0, Z_0)^T$. This mapping relation can be simplified as

$$\begin{bmatrix} kx \\ ky \\ kz \\ k \end{bmatrix} = c_h = A \cdot w_h = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

where k is an arbitrary constant.

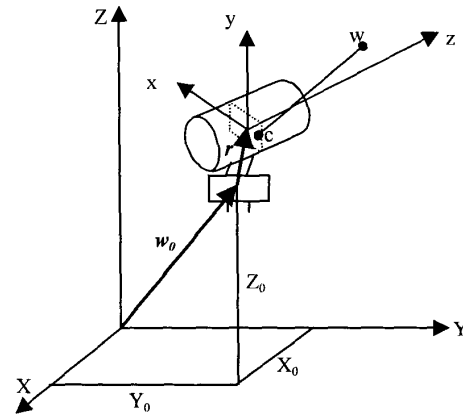


Fig. 10. Geometry of a camera system.

For the inverse-mapping from the image coordinate to the world

coordinate, z in c_h is not given. And, according to the above assumption, (X,Y) which represents 2D positions in the floor is computed while the height Z in w_h is ignored. Then, the inverse-mapping can be written as

$$\begin{bmatrix} kX \\ kY \\ k \end{bmatrix} = \hat{w}_h = \hat{A}^{-1} \cdot \hat{c}_h = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}$$

In order to find the elements of $\hat{A}^{-1}$, the inverse-mapping relation can be also expressed in the form of

$$\begin{aligned} gxX + hyX + iX - ax - by - c &= 0 \\ gxY + hyY + iY - dx - ey - f &= 0 \end{aligned}$$

Since there are two equations with nine unknowns, at least five pairs of (x,y) and (X,Y) can determine a unique $\hat{A}^{-1}$. Once $\hat{A}^{-1}$ is found, (X,Y) in the world coordinate is obtained from the corresponding (x,y) as

$$X = \frac{ax + by + c}{gx + hy + i} \quad \text{and} \quad Y = \frac{dx + ey + f}{gx + hy + i}$$

By differentiating X and Y with respect to x and y respectively, the spatial resolutions along both directions in the world coordinate are expressed as

$$\begin{aligned} \Delta X &= \frac{(ah - bg)y + (ai - cg)}{(gx + hy + i)(gx + hy + i + g)} \\ \Delta Y &= \frac{(eg - hd)x + (ei - hf)}{(gx + hy + i)(gx + hy + i + h)} \end{aligned}$$

The above equations predict the actual spatial resolutions along horizontal and vertical directions over the floor.

3.2. Image processing for heading and position correction

Mobots use odometry to acquire the relative position and heading from the initial position and heading. The odometry is an open-loop control which can not cope with the slip between wheels and floor. Even a slight difference in the diameters of wheels causes an error in heading. The position and heading errors by odometry accumulate as the mobot travels. Hence, periodical localization and correction are necessary to nullify the accumulated errors[7].

Fig. 11 summarizes the procedure of localization performed in this research. A single frame image from the mono-vision system is processed to extract linear features in indoor environment such as edges along the floor and doors. Canny's operator[8] and Hough transform[9] are modified and utilized for edge detection and line detection, respectively. The positions of the detected linear features are inversely mapped from the image coordinate to the world coordinate by $\hat{A}^{-1}$. Then, the mobot's actual position is computed by referencing the positions of door corners over the floor edge, and the heading direction by the direction of the floor edge lines.

The proposed method for localization is not only implemented but also computationally optimized to overcome the computational burden that is often imposed by most of digital image processing techniques.

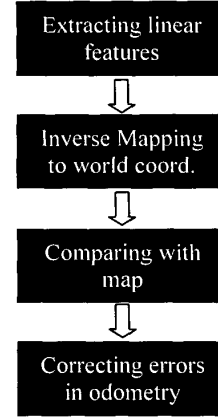


Fig. 11. Procedure of localization using mono-vision system.

4. EXPERIMENTAL RESULTS

Fig. 12 shows the mobot used in the experiment. a single ultrasonic sensor and a mono-vision system along with a notebook PC are mounted on a mobile base.

The Voronoi graph can be considered as a weighted graph where if there is a direct path between a pair of nodes, the weight of the corresponding arc is its distance, otherwise the weight is set to infinity. Given a start and a goal positions, a shortest path between them can be found by applying the Dijkstra's algorithm to the weighted graph. Fig. 13 is an example of path planning, where thin lines are direct paths among nodes. Some of them are passing through the obstacles. This is caused by incorrect reconstruction of the obstacles in the proposed image processing method, which needs to be corrected and fixed as a future work.

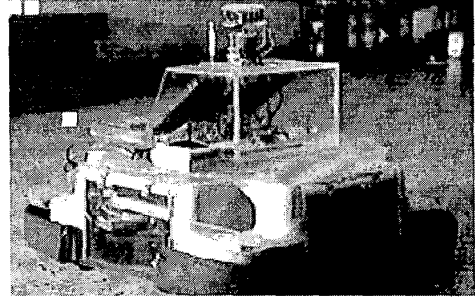


Fig. 12. A mobot system in experiment.

For calibrating the mono-vision system, the camera is mounted to look at the region of $3.2m(W)$ by $4.5m(L)$ on the floor $2m$ ahead of the mobot. The size of the corresponding image is $480(x)$ by $220(y)$. 26 pairs of known positions in the region are found and applied to the two linear equations for $\hat{A}^{-1}$. Using the least square estimation, it is uniquely determined as

$$\hat{A}^{-1} = \begin{bmatrix} -0.1403 & -0.0021 & 34.6207 \\ -0.0089 & 0.0248 & -68.5416 \\ -0.0000 & -0.0013 & 0.1131 \end{bmatrix}$$

The position error in the inverse mapping by the above $\hat{A}^{-1}$ was 1.4cm in average with the maximum of 3.6cm, which is quite accurate for positioning of the mobot. It also implies that the mono-vision system has been properly calibrated.

In order to verify the correctness of the predicted spatial resolution ΔX and ΔY derived above, they are compared with the actual resolutions obtained by real measurements of distance divided by the number of pixels. As shown in Fig 14, their difference was within 0.3mm.

Experimental navigation has carried out, and the proposed method for localization demonstrated its usefulness by allowing position errors only within 4.7cm.

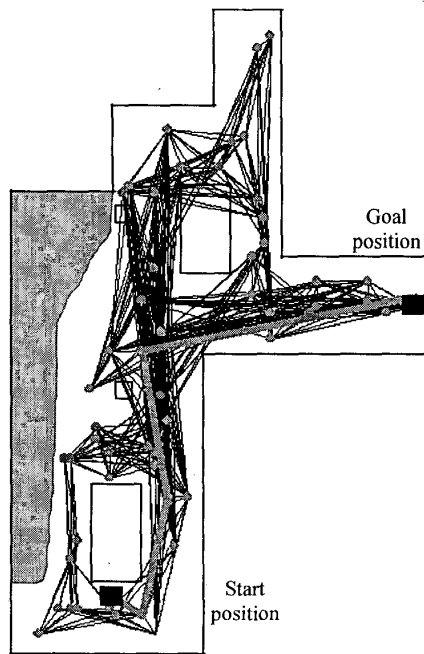


Fig. 13. Example of path planning.

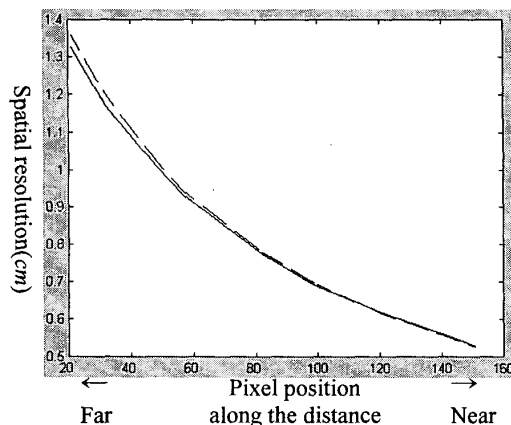


Fig. 14. Comparison of the predicted spatial resolution with the actual spatial resolution.

5. CONCLUSION

The purpose of this paper is to develop a low-cost multi-sensor system of a mono-vision sensor and a single ultrasonic sensor for autonomous navigation of a mobot in indoor environment. In this paper, measurements from the single sonar sensor and the Bayesian updating rule produce the occupancy grid-based map in a wall-following round trip, and then a morphological image processing scheme converts the occupancy image into a Voronoi graph of the environment. This paper also implements a camera calibration process by which the mobot is provided distance information with a mono-vision sensor.

The effectiveness of the proposed methods for autonomous navigation of a service mobot is preliminarily demonstrated by experiments with a mobot in indoor environment. Future works include extending the image processing methods for a multi-sensor system with a stereo-vision sensor and multiple sonar sensors and integrating them to a service mobot.

REFERENCES

- [1] D. Kortenkamp and T. Weymouth, "Topological Mapping for Mobile Robots using a Combination of Sonar and Vision sensing," *Proc. of 12th National Conference on AI*, 1994, pp. 979-984.
- [2] M. Betke and L. Gurvits, "Mobile Robot Localization Using Landmarks," *IEEE Trans. on Robotics and Automation*, Vol. 13, No. 2, 1997, pp. 251-263.
- [3] A. Fujimori, P.N. Nikiforuk, and M. Gupta, "Adaptive Navigation of Mobile Robots with Obstacle Avoidance," *IEEE Trans. on Robotics and Automation*, Vol. 13, No. 4, 1997, pp. 596-602.
- [4] H. Moravec and A. Elfes, "High resolution maps from wide angle sonar," *Proc. of IEEE International Conference on Robotics and Automation*, 1985, pp. 116-121.
- [5] A. Elfes, "Using occupancy grids for mobile robot perception and navigation," *IEEE Computer*, Vol. 22, No. 6, 1989, pp. 46-57.
- [6] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, Addison-Wesley Co., 1993.
- [7] C. Chang and K. Song, "Environment Prediction for a Mobile Robot in a Dynamic Environment," *IEEE Trans. on Robotics and Automation*, Vol. 13, No. 6, 1997, pp. 462-472.
- [8] J. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 8, No. 6, 1986, pp. 679-697.
- [9] B. Burns, et. al., "Extracting Straight Lines," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 8, No. 4, 1986, pp. 174-188.