

# Localization and Approaching to the Human by Mobile Home Robot

Saeed Shiry Ghidary  
Computer System, Eng. Faculty,  
Kobe University,  
Kobe, Nada Ku, 1-1,657

Yasushi Nakata, Toshi Takamori, Motofumi Hattori  
Computer System, Eng. Faculty,  
Kobe University,  
Kobe, Nada Ku, 1-1,657

## Abstract

This paper introduces a robotic system that uses a PTZ camera to find the human in the room and locate its position. Then the robot is guided to go near to the human. This includes human detection and localization, robot localization and robot navigation. Human's head and face is detected by using a combination of motion detection, shape analysis and color information. The 3-D position of human's head and face is estimated by using the depth from focus method. For robot localization we use a very fast measuring system that utilizes ultrasonic and infrared signals simultaneously. For robot navigation we use a map of environment consisting of doors, walls and static objects. After locating human and robot a visibility graph is searched for shortest path and obstacle avoidance is done in a reflexive method using ultrasonic sensors.

## 1 Introduction

The ability to recognize human inside the room and localize its 3D position is subject of research in many different applications. In home robot and human interaction it would be necessary for robot to know where is the location of person in the environment.

The purpose of robotic system introduced in this paper is to find the human in the room and guide the robot to go near to him or her. This may be used for example in delivering some thing to the human or other kind of interaction and support. In this paper we will address some basic problems for human support by home robot. This includes human recognition and localization, robot positioning in indoor environment and robot navigation. We describe an application where a mobile robot uses a camera to detect the human and estimate its position.

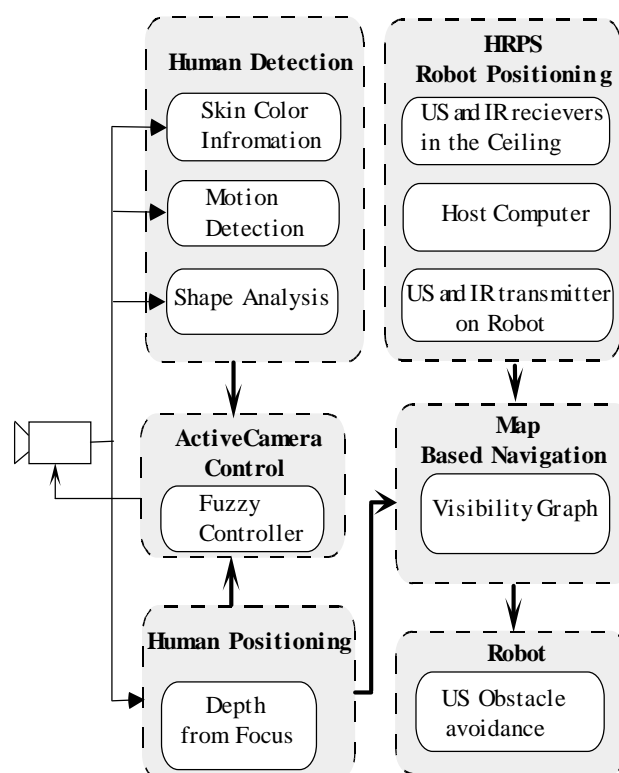


Fig. 1. System overview showing relationship of each module.

## 2 System structure

This system is implemented on the Ymabico mobile robot. Our robot is equipped with a Sony EVI D30 PTZ CCD camera, which is used for human detection and localization. There are 8 ultrasonic transceivers around the robot for obstacle avoidance. The robot communicates with host computer with wireless LAN. Figure 1 shows the overview of system and relationship between its component.

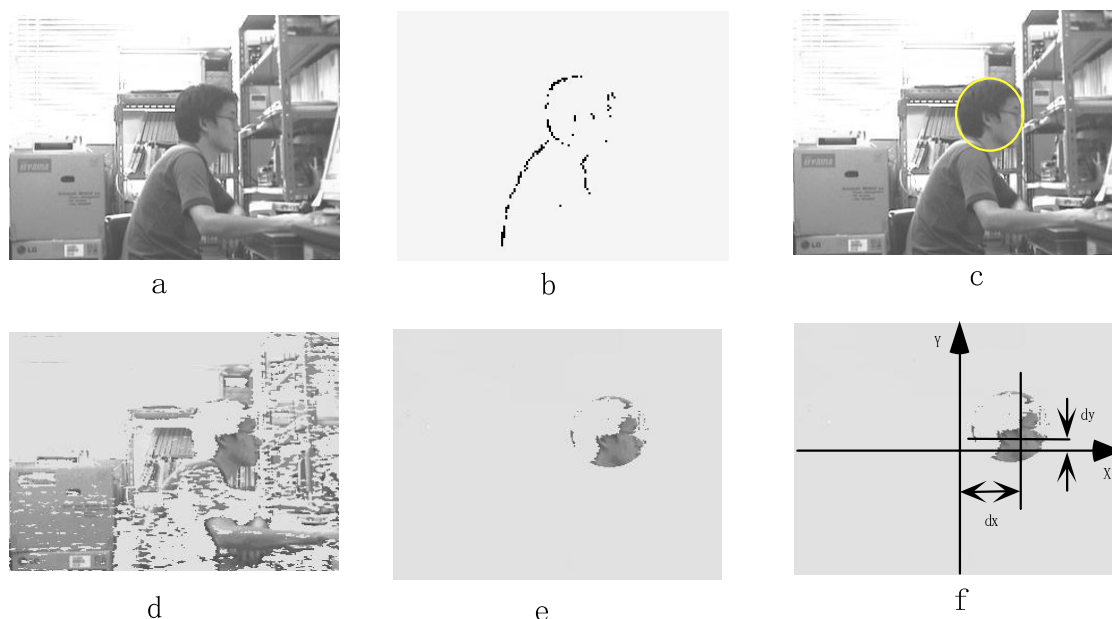


Fig. 2. a) Person working with computer. b) Difference binary image. c) Located head. d) Color detection in whole scene. e) Color detection in head area. f) Face position relative to image center.

Host computer is a controlling computer used for global robot localization and path planning. It is responsible for finding the robot in the room and guiding it toward the human. There is a master slave relationship between host computer and mobile robot. Host computer consists of a powerful PC with digital I/O interface and wireless LAN connection. It collects data from sensors on the mobile robot or other sensors in the environment and after processing the data, orders the mobile robot to do the necessary function. Commands such as "move", "turn", "capture" and "measure" are sent by host computer to the robot and the robot provides different data for the host.

We find the position of the robot in the room using HRPS [1] system which consists of ultrasonic and infrared sensors. We have installed ultrasonic and infrared receivers in the environment for robot localization.

Robot uses CCD camera to search for the human using skin color information, motion detection and shape analysis. Then by active control of camera parameters robot makes the camera to automatically focus on the face and uses information from focusing ring of camera to measure distance between human face and camera. By having information about the absolute position of robot and camera and its pan and tilt angles it is possible to find absolute position of the person in the room.

We use an absolute world model for automatic path planning and execution. After locating the robot and human, the whole path of the mobile robot is computed by searching in the visibility graph. Then the robot is guided to traverse this path and avoid obstacles in a reactive

method. In the following section details of implementation of each module is explained.

### 3 Human detection and localization

There is a large amount of research in human head and face detection using motion [2], color information [3], shape [4] or combination of them [5].

Our system locates a human's head and face by using motion detection, Hough transform and a statistical color model. Face color provides a simple and fast tracking cue for active camera control, which tries to locate face inside a window by changing pan, tilt and zoom of camera based on fuzzy rules. This will put the face in the center of the camera's field of view.

By calculating the temporal derivative of an image and thresholding at a suitable level to filter out noise we segment the image into regions of motion and stationary objects (fig. 2.b). We use the difference image from motion analysis to localize the head. Head can be supposed to be a nearly rigid object and its shape in an image can be approximated by a rigid two-dimensional model. We use a circular model for representing the head. This model may not completely fit the human head, but it is simple and provides a fast head locating method. We search in motion data for the top point of the head and use a simplified Hough transform to locate circle passing from this point (fig. 2.c).

Then we search for facial area inside the head circle (fig. 2.e) using a statistical color model that consists of

two-dimensional Gaussian distribution of normalized face colors. Although different people have different color appearances, several studies have shown that such a difference can be reduced by intensity normalization and skin colors of different races fall into small cluster in normalized RGB or HSV color spaces [6].

We trained the statistical classifier from a sampling of peoples with a variety of races and skin colors. The parameters of the distribution are obtained by maximum likelihood estimation. A pixel is identified to have skin color if corresponding probability is greater than a threshold. The details of head and face detection are reported in our previous work [8]. Figure 2 shows steps taken to detect head and face by proposed method.

#### 4 Robot localization

In order to robot to be useful it must be able to localize itself in the environment. We use two mechanisms to find position of the robot; HRPS [1] and dead reckoning. Dead reckoning induces cumulative errors caused by wheel slippage or rough surfaces that can quickly become extremely severe. Therefore external sensors should be used for adjustment of the robot position. In order to solve this problem we make use of HRPS to help robot to estimate its position. HRPS is a very fast robot localization system using ultrasonic and infrared sensors. When robot is moving it uses dead reckoning to measure how much it has been moving. But for correcting the final position it uses the HRPS.

HRPS system basically consists of one transmitter module, which is, mounted on the mobile robot and some (6 in our system) receiver modules, which are installed on the fixed points in the ceiling of the room.

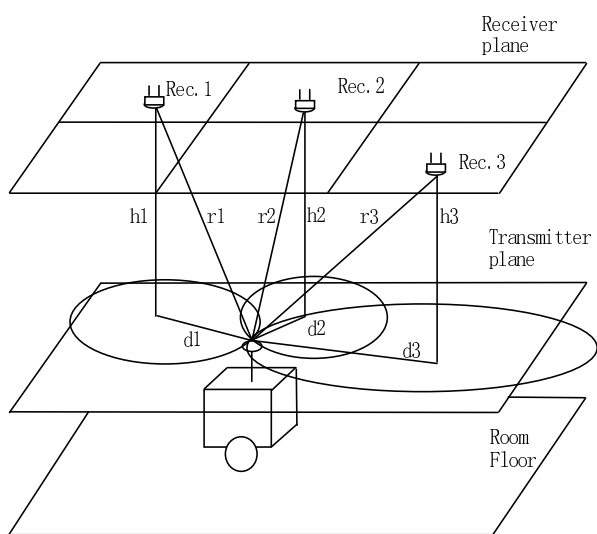


Fig.3. Robot positioning principle.

The transmitter module consists of ultrasonic and infrared transmitter arrays and can produce a burst of both signals at the same time. The receiver module also consists of ultrasonic and infrared sensors and necessary circuit for detection and amplification of these signals.

When the transmitter module sends a couple of ultrasonic and infrared pulses, the receiver module receives the infrared signal almost at the same time (the time of flight of light is negligible and the delay of detecting circuit is considered as a fix value in calculations). This signal starts a counter at the receiving circuit. This counter will continue to count until the reception of ultrasonic signal which, will stop the counter. The value of this counter is proportional to the time of flight of ultrasonic pulse and can be used to calculate the distance between receiver and transmitter.

For measuring the position of the robot, the host computer sends the "measure" command via wireless LAN to the robot and robot produces a couple of ultrasonic and infrared signals. Then the host computer gets the data of receivers and uses them to calculate the position of robot.

The location and the number of receivers are designed so that at least 3 receivers will catch these signals from robot. As the receivers are fixed at a plane with known heights and also the transmitter is fixed in a known point on the robot, the vertical distance between transmitter and receivers are known in a priori. So the problem of localization of robot can be done in two dimensions. As shown in figure 3, the range (r) to any given receiver is projected to the plane of transmitter (d), which maps out a circular locus of possible positions.

$$d = (r^2 - h^2)^{1/2} \quad (1)$$

The interception of two or more of these loci gives the position of the robot.

The resolution of receiver circuit in measuring the range is 1.5 cm and it can detect signals coming from a distance up to 4 m. The positioning error is less than 5 cm.

#### 5 Human localization using depth from focus

This system estimates the 3D position of human's head and face by using the depth from focus method. Traditionally there are two classes of approaches for depth computation: active and passive. Active approaches use infrared, ultrasonic or laser range finder for measuring the distance while passive approaches place only a sensor in the scene to acquire visual information. Examples for passive approaches are depth from stereo, motion, shading, texture, focus or defocus. Focus interpretation is a valuable alternative to stereo vision because it doesn't require solving correspondence for depth recovery [9].

Most imaging systems such as standard cameras can only be in focus for one distance at a time. Surface regions at this focal distance will produce sharp images; surface at other distances will produce blurred images. Surfaces at different distances can be brought into focus one at a time by adjusting the power (focal length) of the optical system. In the human visual system this is accomplished by changing the shape of the lens. In camera systems, they change the position of one or more camera lenses.

Several researchers have proposed depth cues based on focusing information [9,10]. A common technique involves focusing the object on the image detector in the camera. The distance is then determined from camera setting. This technique is called depth from focusing [11]. The other common technique is depth from defocusing, which calculates depth from degree of image blur [13].

In general, autofocus techniques for video camera maximize the high frequency components of an image by adjusting the focusing lens. Focused images have more high frequency components than defocused images of a scene, and thus defocused images can be described as the result of convolving the focused image with the blurring function that plays the role of a low-pass filter.

In this system we use information from focusing ring of an autofocus camera to measure depth information. This ring provides depth information about the object in the center of camera's field of view. By active control of camera parameters we make the camera to autofocus on the face. In this way we can compute the distance between human face and camera using the data from focus ring encoder. By having information about the absolute position of robot and pan and tilt angles of camera we can compute absolute position of the person in the room.

This method is quite simple and needs only calibration of one camera parameter to obtain the relationship between distance to object in focus and the autofocus lens position. Details of implementation is reported in [7].

## 6 Tracking and Camera Control

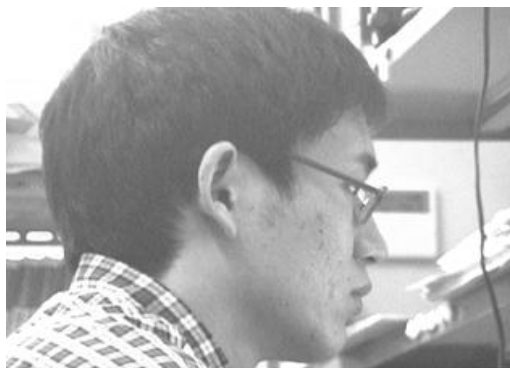
In our system tracking keeps the object of interest centered in the image by adequately moving the camera.

In order to use autofocus ability of camera, the camera must have an appropriate zoom setting to make the target large enough in the image. During depth measurement the camera control policy is to change the zoom so that the width of the face occupy approximately more than 50% of the width of the camera image.

To obtain a close-up view of a particular subject, three parameters need to be determined: the pan angle, the tilt angle, and the zoom factor. Also for achieving high speed tracking it is important to use both position and velocity control.



a



b

Fig.4. a) Head is detected using head detection algorithm. b) Head is brought inside the camera's field of view by changing pan, tilt and zoom factor.

In this system we use Fuzzy logic rules and reasoning for control of speed and direction of camera. Fuzzy controller has this advantage that we can do tracking without need for camera calibration or modeling.

Input values are  $dx$ ,  $dy$  (Fig.2.f) which determine the difference between current position of face in the image and image center point.

Controller has four outputs: Value of pan angle, value of tilt angle, pan speed and tilt speed. Triangular membership function is applied to the outputs. The inference is of Mamdani inference engine type and center of gravity defuzzification method is implemented. The fuzzy controller consists of a rule base for speed and direction control of camera. It contains rules for change of tilt angle, change of pan angle and values of speed for pan and tilt change. The rules are formulated in classical logic form.

In order to detect features such as head motion and face color in the image, the camera must have an appropriate zoom setting. But for measuring distance, system cannot use wide zoom and has to change the zoom to provide closer view. Therefore the zooming is a

dynamic function in which system selects close view when measuring the distance and select wide view when tracking.

While tracking, the camera is moved in a saccadic way in which no motion processing occurs during the camera movement.

In the case of person with small motions, for example one working with computer or watching TV or washing dishes, robot can locate person's position in the room with less than 10 cm error. The measurable distance between robot and human is from 90 cm up to 340 cm.

## 7 Navigation

The robot uses an internal grid-based map and its sensory system to navigate in the environment. The environment is considered to be structured with known static objects and randomly placed furniture. The map consists of doors, walls and static objects. As the robot knows its initial position and final goal position and has a map of the world, the path planning layer constructs a path to the goal and begins to traverse that path. Whenever the robot detects an obstacle in the path by its obstacle avoidance layer, it executes an obstacle exploration algorithm to find out boundaries of obstacle and replans to find alternative path when it can not proceed in the intended path.

Navigation is almost like subsumption method. It is accomplished by several layers: a low level module that avoids objects, and a higher level that moves the robot in a particular direction, ignorant of obstacles. The combination provides a simple way of getting from current state to the goal.

It is supposed that human doesn't move during motion of the robot and dynamics are mainly caused by refurbishings.

Obstacle avoidance is based on ultrasonic sensor information. There are 8 such sensors located around the robot. When obstacles are found we update the map and a new path may be computed from robot's current position to the goal. Therefore the planned paths may take into account the placement of dynamic objects as soon as they become relevant.

### 7.1 Path planning

Path planning is a fundamental issue in robotics. The purpose of a path planner is to compute a path, i.e. a continuous sequence of configurations that leads the robot to its goal. The existence of an absolute world model allows for automatic path planning and execution, and for subsequent route revisions in the event a new obstacle is encountered.

The whole path of the mobile robot is expressed by an assembly of straight line segments and turning angles. For

planning a path between robot and human we use visibility graph approach [12]. A visibility graph is a graph,  $G(V, E)$ , where  $V$  is the set of all vertices of obstacles within an environment including the coordinates of the start and the destination, and  $E$  is the set of all edges connecting any two vertices in  $V$  that do not pass through any obstacles within the environment. Once a visibility graph has been constructed for a given environment, a search algorithm is used to find the shortest path between starting and destination points.

The configuration of the robot position is specified by a 3D vector, whose elements are the 2D position of the front wheel and the orientation of the robot. The size of robot is considered by setting a margin around each obstacle. The size of the margin is set to half of the robot width.

## 8 Experiment results

Our experiment was aimed to put all different modules of system together in leading the robot to go near the human with considering all uncertainty in robot position, human position and obstacle perception. The experiment has been conducted in a laboratory of  $6 \times 4 \text{ m}^2$  with robot moving with speed of  $v=0.2\text{m/sec}$ .

Figure 5 displays an example of a plan generated by planner and executed path. Robot uses dead reckoning to traverse each partial path and then at each stop point it uses HRPS to measure its absolute position. Kalman filter can be used for fusing data from two sensors. Due to the crosstalk effect in ultrasonic sensors we have limited use of HRPS to prevent confusion between HRPS and obstacle avoidance system. This can be prevented if we use other sensory system for obstacle avoidance.

HRPS provides very fast and exact information about position of the robot but the error in robot's direction is considerable.

Positional information provided by human localization system have enough accuracy for semi static human but due to speed limitation of autofocus system it loses accuracy for moving human but it is still applicable if we can use low accuracy data for rough estimation.

## 9 Conclusion

We have demonstrated a system, which can search for the human inside room and guide a mobile robot to go near to it. This is accomplished through the integration of three key modules: Robot positioning by HRPS system, Human detection by using color and motion information and Human localization by depth estimation from focus information.

Our system has application in robot and human interaction but the method can be used for searching other landmarks in the environment also. Currently we do not apply human

identification but we are going to use it to identify and track a specific user in a multi person environment.

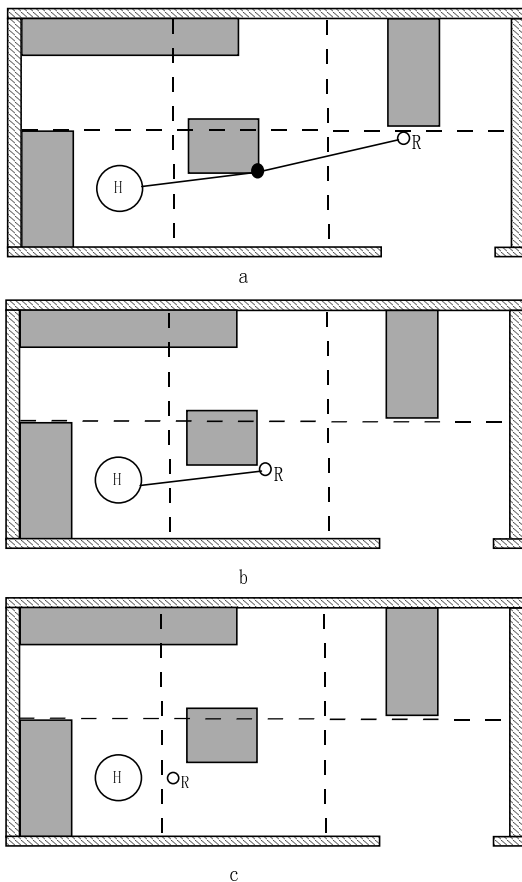


Fig. 5. a) A path is computed between robot and human using visibility graph. b) Robot is guided to the first part of path using dead reckoning, then robot is localized using HRPS and a new path is computed. c) Robot is stopped near the human using distance criteria.

## References

- [1] S.S. Ghidary, T. Tani, T. Takamori, M. Hattori, "A new Home Robot Positioning System (HRPS) using IR switched multi ultrasonic sensors", IEEE SMC Conf., Tokyo, Japan, Oct.1999.
- [2] S.J. Mckenna, S. Gong, H. Liddell, "Real-time tracking for an integrated face recognition system", 2nd Workshop on Parallel Modeling of Neural Operators, Faro, Portugal, Nov. 1995.
- [3] J. Yang, A. Waibel, "A real Time Face Tracker", In Proc. Of WACV, Sarasota Florida USA, 1996.
- [4] S. Birchfiel, "An Elliptical Head Tracker", 31st Asilomar Conf. On Signals, Systems, and Computers, Nov. 1997.
- [5] H. P. Graf, E.Cosatto, D.Gibbon, M. Kocheisen, "Multi-Modal System for Locating Heads and Faces", In Proc. 2nd Int. Conf. Automatic Face and Gesture Recognition, IEEE Computer Soc. Press, pp. 88 – 93, 1996.
- [6] J.C. Terrillon, H. Fukamachi, "Modeling the Chrominance of Human Skin for Face Detection in Natural Color Images", The fifth ATR Symposium on Face and Object Recognition, Japan, Apr. 1998.
- [7] S. S. Ghidary, Y. Nakata, T. Takamori, M. Hattori, "Human Detection and Localization at Indoor Environment by Home Robot", IEEE SMC Conference, Nashville, USA, Oct. 2000.
- [8] S. S. Ghidary, Y. Nakata, T. Takamori, M.Hattori, "Head and Face Detection at Indoor Environment by Home Robot", in Proc. of ICEE2000, Iran, May 2000.
- [9] Y. Xiong , S. Shafer, "Depth from Focusing and Defocusing", Tech. report CMU-RI-TR-93-07, Robotics Institute, Carnegie Mellon University, March 1993.
- [10] M. Subbarao, T. S. Choi, A. Nikzad, "Focusing Techniques," Journal of Optical Engineering, pp. 2824-2836 Nov. 1993.
- [11] E.Krotkov,"Focusing", Inte. Journal of Computer Vision, Vol. 1, No. 3, pp 223-238, Oct. 1987.
- [12] J.C. Latombe, "Robot Motion Planning", Kluwer Academic Publishers, second edition, 1991.
- [13] A.P. Pentland, "A New Sense For Depth Of Field", IEEE Trans. Patt. And Machine Intell. PAMI(9), pp. 522-531,1987.