We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

3,800
Open access books available

116,000
International authors and editors

120M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
Vision based Systems for Localization in Service Robots

Paulraj M.P. and Hema C.R.
Mechatronic Program,
School of Mechatronic Engineering,
Universiti Malaysia Perlis
Malaysia

1. Introduction

Localization is one of the fundamental problems of service robots. The knowledge about its position allows the robot to efficiently perform a service task in office, at a facility or at home. In the past, variety of approaches for mobile robot localization has been developed. These techniques mainly differ in ascertaining the robot’s current position and according to the type of sensor that is used for localization. Compared to proximity sensors, used in a variety of successful robot systems, digital cameras have several desirable properties. They are low-cost sensors that provide a huge amount of information and they are passive so that vision-based navigation systems do not suffer from the interferences often observed when using active sound or light based proximity sensors. Moreover, if robots are deployed in populated environments, it makes sense to base the perceptional skills used for localization on vision like humans do.

In recent years there has been an increased interest in visual based systems for localization and it is accepted as being more robust and reliable than other sensor based localization systems. The computations involved in vision-based localization can be divided into the following four steps [Borenstein et al, 1996]:

(i) Acquire sensory information: For vision-based navigation, this means acquiring and digitizing camera images.

(ii) Detect landmarks: Usually this means extracting edges, smoothing, filtering, and segmenting regions on the basis of differences in gray levels, colour, depth, or motion.

(iii) Establish matches between observation and expectation: In this step, the system tries to identify the observed landmarks by searching in the database for possible matches according to some measurement criteria.

(iv) Calculate position: Once a match (or a set of matches) is obtained, the system needs to calculate its position as a function of the observed landmarks and their positions in the database.
2. Taxonomy of Vision Systems

There is a large difference between indoor and outdoor vision systems for robots. In this chapter, we focus only on vision systems for indoor localization. Taxonomy of indoor based vision systems can be broadly grouped as [DeSouza and Kak, 2002]:

i. Map-Based: These are systems that depend on user-created geometric models or topological maps of the environment.

ii. Map-Building-Based: These are systems that use sensors to construct their own geometric or topological models of the environment and then use these models for localization.

iii. Map-less: These are systems that use no explicit representation at all about the space in which localization is to take place, but rather resort to recognizing objects found in the environment or to tracking those objects by generating motions based on visual observations.

In among the three groups, vision systems find greater potential in the map-less based localization. The map-less navigation technique and developed methodologies resemble human behaviors more than other approaches, and it is proposed to use a reliable vision system to detect landmarks in the target environment and employ a visual memory unit, in which the learning processes will be achieved using artificial intelligence. Humans are not capable of positioning themselves in an absolute way, yet are able to reach a goal position with remarkable accuracy by repeating a look at the target and move type of strategy. They are apt at actively extracting relevant features of the environment through a somewhat inaccurate vision process and relating these to necessary movement commands, using a mode of operation called visual servoing [DeSouza and Kak, 2002].

Map-less navigation include systems in which navigation and localization is realized without any prior description of the environment. The localization parameters are estimated by observing and extracting relevant information about the elements in the environment. These elements can be walls, objects such as desks, doorways, etc. It is not necessary that absolute (or even relative) positions of these elements of the environment be known. However, navigation and localization can only be carried out with respect to these elements.

Vision based localization techniques can be further grouped based on the type of vision used namely, passive stereo vision, active stereo vision and monocular vision. Examples of these three techniques are discussed in detail in this chapter.


Making a robot see obstacles in its environment is one of the most important tasks in robot localization and navigation. A vision system to recognize and localize obstacles in its navigational path is considered in this section. To enable a robot to see involves at least two mechanisms: sensor detection to obtain data points of the obstacle, and shape representation of the obstacle for recognition and localization. A vision sensor is chosen for shape detection of obstacle because of its harmlessness and lower cost compared to other sensors such as...
laser range scanners. Localization can be achieved by computing the distance of the object from the robot’s point of view. Passive stereo vision is an attractive technique for distance measurement. Although it requires some structuring of the environment, this method is appealing because the tooling is simple and inexpensive, and in many cases already existing cameras can be used. An approach using passive stereo vision to localize objects in a controlled environment is presented.

3.1 Design of the Passive Stereo System
The passive stereo system is designed using two digital cameras which are placed on the same y-plane and separated by a base length of 7 cm in the x-plane. Ideal base lengths vary from 7 cm to 10 cm depicting the human stereo system. The height of the stereo sensors depends on the size of objects to be recognized in the environment, in the proposed design the stereo cameras are placed at a height of 20 cm. Fig. 1 shows the design of mobile robot with passive stereo sensors. It is important to note both cameras should have the same view of the object image frame to apply the stereo concepts. An important criterion of this design is to keep the blind zone to a minimal for effective recognition as shown in Fig.2.

![Fig. 1. A mobile robot design using passive stereo sensors](image)

![Fig. 2 Experimental setup for passive stereo vision](image)

3.2 Stereo Image Preprocessing
Color images acquired from the left and the right cameras are preprocessed to extract the object image from the background image. Preprocessing involves resizing, grayscale conversion and filtering to remove noise, these techniques are used to enhance, improve or
otherwise alter an image to prepare it for further analysis. The intention is to remove noise, trivial information or information that will not be useful for object recognition. Generally object images are corrupted by indoor lighting and reflections. Noise can be produced due to low lighting also. Image resizing is used to reduce the computational time, a size of 320 by 240 is chosen for the stereo images. Resized images are converted to gray level images to reduce the pixel intensities to a gray scale between 0 to 255; this further reduces the computations required for segmentation.

Acquired stereo images do not have the same intensity levels; there is considerable difference in the gray values of the objects in both left and right images due to the displacement between the two cameras. Hence it is essential to smooth out the intensity of both images to similar levels. One approach is to use a regional filter with a mask. This filter filters the data in the image with the 2-D linear Gaussian filter and a mask. The mask image is the same size as the original image. Hence for the left stereo image, the right stereo image can be chosen as the mask and vice versa. This filter returns an image that consists of filtered values for pixels in locations where the mask contains 1’s, and unfiltered values for pixels in locations where the mask contains 0’s. The intensity around the obstacle in the stereo images is smoothened by the above process.

A median filter is applied to remove the noise pixels; each output pixel contains the median value in the M-by-N neighborhood [M and N being the row and column pixels] around the corresponding pixel in the input image. The filter pads the image with zeros on the edges, so that the median values for the points within [M N]/2 of the edges may appear distorted [Rafael, 2002]. The M -by- N is chosen according to the dimensions of the obstacle. A 4 x 4 matrix was chosen to filter the stereo images. The pre-processed obstacle images are further subjected to segmentation techniques to extract the obstacle image from the background.

### 3.3 Segmentation

Segmentation involves identifying an obstacle in front of the robot and it involves the separation of the obstacle from the background. Segmentation algorithm can be formulated using the grey value obtained from the histogram of the stereo images. Finding the optimal threshold value is essential for efficient segmentation. For real-time applications, automatic determination of threshold value is an essential criterion. To determine this threshold value a weighted histogram based algorithm is proposed which uses the grey levels of the image from the histogram of both the stereo images to compute the threshold. The weighted histogram based segmentation algorithm is detailed as follows [Hema et al, 2006]:

**Step 1:** The histogram is computed from the left and right gray scale images for the gray scale values of 0 to 255.

Counts \( a(i) \), \( i=1,2,3,\ldots,256 \)

where \( a(i) \) represents the number of pixels with gray scale value of \( (i-1) \) for the left image.

Counts \( b(i) \), \( i=1,2,3,\ldots,256 \)

where \( b(i) \) represents the number of pixels with gray scale value \( (i-1) \) for the right image.
Step 2: Compute the logarithmic weighted gray scale value of the left and right image as

\[
\text{ta}(i) = \log(\text{count a}(i)) \times (i-1) \quad (1)
\]

\[
\text{tb}(i) = \log(\text{count b}(i)) \times (i-1) \quad (2)
\]

where \( i = 1, 2, 3, \ldots, 256 \)

Step 3: Compute the logarithmic weighted gray scale

\[
\text{tam} = \frac{1}{256} \sum_{i=1}^{256} \text{ta}(i) \quad (3)
\]

\[
\text{tbm} = \frac{1}{256} \sum_{i=1}^{256} \text{tb}(i) \quad (4)
\]

Step 4: Compute \( T = \min(\text{tam}, \text{tbm}) \). The minimum value of ‘tam’ and ‘tbm’ is assigned as the threshold.

Threshold of both the stereo images are computed separately and the min value of the two thresholds is applied as the threshold to both the images. Using the computed threshold value, the image is segmented and converted into a binary image.

### 3.4 Obstacle Localization

Localization of the obstacle can be achieved using the information from the stereo images. One of the images for example, the left image is considered as the reference image. Using the reference image, the \( x \) and \( y \) co-ordinates is computed by finding the centroid of the obstacle image. The \( z \) co-ordinate can be computed using the unified stereo imaging principle proposed in [Hema et al, 2007]. The unified stereo imaging principle uses a morphological ‘add’ operation to add the left and right images acquired at a given distance. The size and features of the obstacle in the added image varies in accordance with the distance information. Hence, the features of the added image provide sufficient information to compute the distance of the obstacle. These features can be used to train a neural network to compute the distance (\( z \)). Fig.3 shows images samples of the added images and the distance of the obstacle images with respect to the stereo sensors. The features extracted from the added images are found to be good candidates for distance computations using neural networks [Hema et al, 2007].
The x, y and z co-ordinate information determined from the stereo images can be effectively used to locate obstacles and signs which can aid in collision free navigation in an indoor environment.

### 4. Active Stereo Vision for Robot Orientation

Autonomous mobile robots must be designed to move freely in any complex environment. Due to the complexity and imperfections in the moving mechanisms, precise orientation and control of the robots are intricate. This requires the representation of the environment, the knowledge of how to navigate in the environment and suitable methods for determining the orientation of the robot. Determining the orientation of mobile robots is essential for robot path planning; overhead vision systems can be used to compute the orientation of a robot in a given environment. Precise orientation can be easily estimated, using active stereo vision concepts and neural networks [Paulraj et al, 2009]. One such active stereo vision system for determining the robot orientation features from the active stereo vision system in indoor environments is described in this section.
4.1 Image Acquisition
In active stereo vision two or more cameras are used, wherein the cameras can be positioned to focus on the same imaging area from different angles. Determination of the position and orientation of a mobile robot, using vision sensors, can be explained using a simple experimental setup as shown in Fig. 4. Two digital cameras using the active stereo concept are employed. The first camera (C1) is fixed at a height of 2.1 m above the floor level in the centre of the robot working environment. This camera covers a floor area of size 1.7m length ($L$) and 1.3m width ($W$). The second camera (C2) is fixed at the height ($H_2$) of 2.3 m above the ground level and 1.2 m from the Camera 1. The second camera is tilted at an angle ($\theta_2$) of 22.5°.

The mobile robot is kept at different positions and orientation and the corresponding images ($O_{a1}$ and $O_{b1}$) are acquired using the two cameras. The experiment is repeated for 180 different orientation and locations. For each mobile robot position, the angle of orientation is also measured manually. The images obtained during the $i^{th}$ orientation and position of the robot is denoted as ($O_{ai}$, $O_{bi}$). Sample of images obtained from the two cameras for different position and orientation of the mobile robot are shown in Fig. 5.

![Fig. 4. Experimental Setup for the Active Stereo Vision System](image-url)

Fig. 4. Experimental Setup for the Active Stereo Vision System
4.2 Feature Extraction

As the image resolution causes considerable delay while processing, the images are resized to 32 x 48 pixels and then converted into gray-scale images. The gray scale images are then converted into binary images. A simple image composition is made by multiplying the first image with the transpose of the second image and the resulting image \( I_u \) is obtained. Fig. 6 shows the sequence of steps involved for obtaining the composite image \( I_u \). The original images and the composite image are fitted into a rectangular mask and their respective local images are obtained. For each binary image, sum of pixel value along the rows and the columns are all computed. From the computed pixel values, the local region of interest is defined. Fig. 7 shows the method of extracting the local image. Features such as the global centroid, local centroid, and moments are extracted from the images and used as a feature to obtain their position and orientation. The following algorithm illustrates the method of extracting the features from the three images.

**Feature Extraction Algorithm:**

1. Resize the original images \( O_a, O_b \).
2. Convert the resized images into gray-scale images and then to binary images. The resized binary images are represented as \( I_a \) and \( I_b \).
3. Fit the original image \( I_a \) into a rectangular mask and obtain the four coordinates to localize the mobile robot. The four points of the rectangular mask are labeled and cropped. The cropped image is considered as a local image \( I_{al} \).
4. For the original image \( I_a \), determine the global centroid \((G_{ax}, G_{ay})\), area \((G_{aa})\), perimeter \((G_{ap})\). Also for the localized image \( I_{al} \), determine the centroid \((L_{ax}, L_{ay})\) row sum pixel values \((L_{ar})\), column sum pixel values \((L_{ac})\), row pixel moments \((L_{urm})\) column pixel moments \((L_{ucm})\). The above features are associated to the orientation of the mobile robot.
5. Repeat step 3 and 4 for the original image \( I_b \) and determine the parameters \( G_{bx}, G_{by}, G_{br}, G_{by}, L_{bx}, L_{by}, L_{br}, L_{bc}, L_{brm} \) and \( L_{bcm} \).
6. Perform stereo composition: \( I_u = I_a \times I_b^T \) (where \( T \) represents the transpose operator).
7. Fit the unified image into a rectangular mask and obtain the four coordinates to localize the mobile robot. The four points of the rectangular mask are labeled and cropped and labeled and cropped. The cropped image is considered as a local image.
8) From the composite global image, the global centroid \((G_{ux}, G_{uy})\), area \((G_{ua})\), perimeter \((G_{up})\) are computed.

9) From the composite local image, the local centroid \((L_{ux}, L_{uy})\), row sum pixel values \((L_{ur})\), column sum pixel values \((L_{uc})\), row pixel moments \((L_{urm})\), column pixel moments \((L_{ucm})\) are computed.

The above features are associated to the orientation of the mobile robot.

Fig. 6. (a) Resized image from first camera with 32 x 48 pixel, (b) Edge image from first camera, (c) Resized image from second camera with 48 x 32 pixel, (d) Edge image from second camera with transposed matrix (e) Multiplied images from first and second cameras with 32 x 32 pixel.

Fig. 7. Extraction of local image (a) Global image (b) Local or Crop image.
5. Hybrid Sensors for Object and Obstacle Localization in Housekeeping Robots

Service robots can be specially designed to help aged people and invalids to perform certain housekeeping tasks. This is more essential to our society where aged people live alone. Indoor service robots are being highlighted because of their potential in scientific, economic and social expectations [Chung et al, 2006; Do et al, 2007]. This is evident from the growth of service robots for specific service tasks around home and work places. The capabilities of the mobile service robot require more sensors for navigation and task performance in an unknown environment which requires sensor systems to analyze and recognize obstacles and objects to facilitate easy navigation around obstacles. Causes of the uncertainties include people moving around, objects brought to different positions, and changing conditions.

A home based robot, thus, needs high flexibility and intelligence. A vision sensor is particularly important in such working conditions because it provides rich information on surrounding space and people interacting with the robot. Conventional video cameras, however, have limited fields of view. Thus, a mobile robot with a conventional camera must look around continuously to see its whole surroundings [You, 2003]. This section highlights a monocular vision based design for a housekeeping robot prototype named ROOMBOT, which is designed using a hybrid sensor system to perform housekeeping tasks, which includes recognition and localization of objects. The functions of the hybrid vision system alone are highlighted in this section.

The hybrid sensor system combines the performance of two sensors namely a monocular vision sensor and an ultrasonic sensor. The vision sensor is used to recognize objects and obstacles in front of the robot. The ultrasonic sensor helps to avoid obstacles around the robot and to estimate the distance of a detected object. The output of the sensor system aids the mobile robot with a gripper system to pick and place the objects that are lying on the floor such as plastic bags, crushed trash paper and wrappers.

5.1 ROOMBOT Design

The ROOMBOT consists of a mobile platform which has an external four wheeled drive found to be suitable for housekeeping robots; the drive system uses two drive wheels and two castor wheels, which implement the differential drive principle. The left and right wheels at the rear side of the robot are controlled independently [Graf et al, 2001]. The robot turning angle is determined by the difference of linear velocity between the two drive wheels. The robot frame has the following dimensions 25cm (width) by 25cm (height) and 50cm (length). The robot frame is layered to accommodate the processor board and control boards. The hybrid sensor system is placed externally to optimize the area covered. The housekeeping robot is programmed to run along a planned path. The robot travels at an average speed of 0.15m/s. The navigation system of the robot is being tested in an indoor environment. The robot stops when there is an object in front of it at the distance of 25cm. It is able to perform 90°-turns when an obstacle is blocking its path. The prototype model of the robot is shown in Fig.8.
5.2 Hybrid Sensor System

The hybrid sensor system uses vision and ultrasonic sensors to facilitate navigation by recognizing obstacles and objects on the robot’s path. One digital camera is located on the front panel of the robot at a height of 17 cm from the ground level. Two ultrasonic sensors are also placed below the camera as shown in Fig.9 (a). The ultrasonic sensors below the camera is tilted at an angle of 10 degrees to facilitate the z co-ordinate computations of the objects as shown in Fig.9(b). Two ultrasonic sensors are placed on the sides of the robot for obstacle detection (Fig.9(c)). The two ultrasonic sensors in the front are used for detecting objects of various sizes and to estimate the y and z co-ordinates of objects. The ultrasonic system detects obstacles / objects and provides distance information to the gripper system. The maximum range of detection of the ultrasonic sensor is 3 m and the minimum detection range is 3 cm. Due to uneven propagation of the transmitted wave, the sensor is unable to detect in certain conditions [Shoval & Borenstein 2001]. In this study, irregular circular objects are chosen for height estimation. Therefore the reflected wave is not reflected from the top of the surface. This will contribute to small error which is taken into account by the gripper system.

Fig. 8. Prototype model of the housekeeping robot.

Fig. 8. Vision and ultrasonic sensor locations (a) vision and two ultrasonic sensors in the front panel of the robot, (b) ultrasonic sensor with 10 degree tilt in the front panel, (c) ultrasonic sensor located on the sides of the robot.
5.3 Object Recognition

Images of objects such as crushed paper and plastic bags are acquired using the digital camera. Walls, furniture and cardboard boxes are used for the obstacle images. An image database is created with objects and obstacles in different orientation and acquired at different distances. The images are dimensionally resized to 150 x 150 sizes to minimize memory and processing time. The resized images are processed to segment the object and suppress the background. Fig.9 shows the image processing technique employed for segmenting the object. A simple feature extraction algorithm is applied to extract the relevant features which can be fed to a classifier to recognize the objects and obstacles. The feature extraction algorithm uses the following procedure:

Step1. Acquired image is resized to 150 x 150 pixel sizes to minimize memory and processing time.
Step2. Resized images are converted to binary images using the algorithm detailed in section 3.3. This segments the object image from the background
Step3. Edge images are extracted from the binary images to further reduce the computational time.
Step4. The singular values are extracted from the edge images using singular value decomposition on the image matrix.

The singular values are used to train a simple feed forward neural network to recognize the objects and the obstacle images [Hong, 1991; Hema et al, 2006]. The trained network is used for real-time recognition during navigation. Details of the experiments can be found [Hema et al, 2009].

Fig. 9. Flow diagram for Image segmentation

5.4 Object Localization

Object localization is essential for pick and place operation to be performed by the gripper system of the robot. In the housekeeping robot, the hybrid sensor system is used to localize the objects. Objects are recognized by the object recognition module; using the segmented object image the x co-ordinate of the object is computed. The distance derived from the two ultrasonic sensors in the front panel is used to compute the z co-ordinate of the object as shown in Fig. 10. The distance measurement of the lowest ultrasonic sensors gives the y co-ordinate of the object. The object co-ordinate information is passed to the gripper system to
perform the pick and place operation. Accuracy of 98% was achievable in computing the \( z \) co-ordinate using the hybrid vision system.

![Fig. 10. Experimental setup to measure the \( z \) co-ordinate.](image)

The ROOMBOT has an overall performance of 99% for object recognition and localization. The hybrid sensor system proposed in this study can detect and locate objects like crushed paper, plastic and wrappers. Sample images of the experiment for object recognition and pick up are shown in Fig.11.

![Fig. 11. picking of trash paper based on computation of the object co-ordinates (a) location 1 (b) location 2.](image)

### 6. Conclusion

This chapter proposed three vision based methods to localize objects and to determine the orientation of robots. Active and passive stereo visions along with neural networks prove to be efficient techniques in localization of objects and robots. The unified stereo vision system discussed in this chapter depicts the human biological stereo vision system to recognise and localize objects. However a good database of typical situations is necessary to implement the system. Combining vision sensors with ultrasonic sensors is an effective method to combine the information from both sensors to localize objects. All the three vision systems were tested in real-time situations and their performance were found to be satisfactory. However the methods proposed are only suitable for controlled indoor environments, further investigation is required to extend the techniques to outdoor and challenging environments.
7. References


Localization and mapping are the essence of successful navigation in mobile platform technology. Localization is a fundamental task in order to achieve high levels of autonomy in robot navigation and robustness in vehicle positioning. Robot localization and mapping is commonly related to cartography, combining science, technique and computation to build a trajectory map that reality can be modelled in ways that communicate spatial information effectively. This book describes comprehensive introduction, theories and applications related to localization, positioning and map building in mobile robot and autonomous vehicle platforms. It is organized in twenty seven chapters. Each chapter is rich with different degrees of details and approaches, supported by unique and actual resources that make it possible for readers to explore and learn the up to date knowledge in robot navigation technology. Understanding the theory and principles described in this book requires a multidisciplinary background of robotics, nonlinear system, sensor network, network engineering, computer science, physics, etc.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:
