**Computer Science Department of The University of Auckland**
**CITR at Tamaki Campus** (http://www.citr.auckland.ac.nz)

CITR-TR-75 November 2000

# Robot Localisation Using an Omnidirectional Colour Image

David C.K. Yuen[1] and Bruce MacDonald[1]

## Abstract

We describe a vision-based indoor mobile robot localisation algorithm that does not require historical position estimates. The method assumes the presence of an *a priori* map and a reference omnidirectional view of the workspace. The current omnidirectional image of the environment is captured whenever the robot needs to relocalise. A modified hue profile is generated for each of the incoming images and compared with that of the reference image to find the correspondence. The current position of the robot can then be determined using triangulation as both the reference position and the map of the workspace are available. The method was tested by mounting the camera system at a number of random positions positions in a 11.0m x 8.5m room. The average localisation error was 0.45 m. No mismatch of features between the reference and incoming image was found amongst the testing cases.

[1] The Department of Electrical and Electronic Engineering, The University of Auckland, Private Bag 92019, Auckland, New Zealand.

# Robot Localisation Using an Omnidirectional Colour Image

David C.K. Yuen* and Bruce A. MacDonald

The Department of Electrical and Electronic Engineering,
The University of Auckland, Private Bag 92019, Auckland, New Zealand.
d.yuen@auckland.ac.nz,b.macdonald@auckland.ac.nz

**Abstract.** We describe a vision-based indoor mobile robot localisation algorithm that does not require historical position estimates. The method assumes the presence of an *a priori* map and a reference omnidirectional view of the workspace. The current omnidirectional image of the environment is captured whenever the robot needs to relocalise. A modified hue profile is generated for each of the incoming images and compared with that of the reference image to find the correspondence. The current position of the robot can then be determined using triangulation as both the reference position and the map of the workspace are available. The method was tested by mounting the camera system at a number of random positions positions in a 11.0m $\times$ 8.5 m room. The average localisation error was 0.45 m. No mismatch of features between the reference and incoming image was found amongst the testing cases.

## 1 Introduction

Under the traditional deliberative motion control architecture, a robot needs to know its own position in the environment before making a navigation plan. If the robot is first switched on or wants to re-position itself after getting lost, no reliable previous position estimates will be available for the localisation stage. Many common localisation methods, notably dead-reckoning using extended Kalman filtering [4], cannot cope with such a condition.

In this paper, we describe a passive, vision-based localisation technique that does not involve the use of historical position estimates, and takes advantage of the richer information in an image. An omnidirectional imaging system is introduced to provide colour and textual information to the system. The distinctive features from an incoming image are extracted using a region segmentation method. The extracted features are then matched with those from a reference image to generate matched landmarks. The placement of artificial landmarks in the environment is unnecessary.

In section 2, we review previous work in vision-based localisation methods that do not require historical position estimates. Section 3 outlines our localisation approach. It also describes the image segmentation and triangulation

* This work was supported in part by the Foundation for Research, Science and Technology, New Zealand, with a Top Achiever Doctoral Scholarship.

techniques adopted in the system. The test results are discussed in section 4 before summarising the paper in section 5.

## 2 Maps and Landmarks

Map matching can usually be carried out without the use of an image. A local map is first generated for the area around the robot, using the measurements from a laser or ultrasonic range finder [2, 7]. The local map is then matched against different regions of a global map, at different orientations. Since the map matching uses a local distance map, the localisation process can be confounded if objects with similar shapes are present in the environment. Also, the correlation operation requires considerable computation.

Many industrial robots are guided by bar codes [5], reflective tape [3, p313–317], ceiling light patterns [3, p472–477] or other artificial landmarks. A global positioning system (GPS) is a notable example of an artificial emitter in an outdoor navigation. While the landmark recognition step is usually quite simple, the cost of laying out and maintaining the well calibrated landmarks can be very expensive, and impractical in some environments.

Visual images usually have high spatial resolution and can provide details such as the colour and texture of the object being observed. With the extra information provided by visual sensors, the robot can have a better understanding of the complex surroundings. In many cases, natural landmarks can be extracted from the incoming images.

Using the concept of a "view field" [1], tiny visual features may be extracted from an image together with their relative spatial relations, to form a landmark. The memory requirement for the storage of typical indoor scenes is thus reduced to about 16000 bytes per $m^2$. Both Lin and Zhang [6, 8] process the sparsely sampled omnidirectional image with neural networks to extract landmarks for localisation, in which 120 and 1600 bytes were retained respectively for each image frame. While the storage of these landmarks requires only modest amount of memory, the image capturing stage involves a lot of preparative work and makes the localisation system quite inflexible.

## 3 Vision-Based Localisation System

Algorithm 1 shows the overall process of localisation. Our method assumes an *a priori* map for the environment. An omnidirectional image is used to simplify camera motion; panning control is not required.

To locate the robot, a vertically central strip of an omnidirectional image is segmented into regions by analysing the horizontal hue profile, then matched against region boundaries in a reference image, and triangulation is used to calculate the new robot position.

The imaging system comprises two Sony EVI-D31 cameras and two OMT SEQ-P1S frame grabber cards with a Pentium based controller, to be mounted on

**Algorithm 1** Localise

On first invocation, call Initialise()
CurrentImage = ObtainImage()
Create all tokens of 3 consecutive region MHI median values for ReferenceImage
Create all tokens of 3 consecutive region MHI median values for CurrentImage
Find longest token match between ReferenceImage and CurrentImage
**for** each of the first, middle and last matching boundary pairs: **do**
    Triangulate position from the map position of the boundary pair
**end for**
return the average of the three position estimates

**ObtainImage:**

Take 8 images at $45^o$ increments, link them together to one image
Extract the 30 pixel high central strip
Calculate the MHI for each pixel in the strip
**for** each 10-pixel wide band **do**
    Calculate the band MHI median
**end for**
Find region boundaries by differentiating the band median sequence
**for** each region between boundaries **do**
    calculate the region MHI median
**end for**
return the sequence of region MHI values

**Initialise:**

ReferenceImage = ObtainImage()
Load the environment map
Calculate the map positions of boundaries in ReferenceImage

our mobile robot as a multi-purpose flexible vision system. To ensure controllable images for testing the current development stage, a single camera is mounted on a tripod. The images captured for this study have a resolution of $320 \times 240$ pixels and a colour depth of 24 bits. To facilitate comparing results, the zoom control of the camera was adjusted for a view angle of $45^o$ (horizontal) $\times$ $34^o$ (vertical) at 84cm above the floor. At each location, 8 images were taken in $45^o$ increments. At present the camera head should face the same direction when taking the first image amongst each series; the purpose is to discover the robot position and later we expect to remove this constraint and also discover the orientation. The 8 images were linked together to form a panoramic view of the environment, shown in Figure 1. A horizontal strip of $2560 \times 30$ pixels is then cut from the center of the omnidirectional image and used for the rest of the processing.

The representation of the image may be further simplified by extracting the hue channel of an HSV model. For humans, colour discontinuity often represents separation between objects. While the hue channel is relatively immune to variations in illumination, some hue values have little meaning and are sensitive to minor changes, notable values near white, gray and black. The modified hue index (MHI) is then defined:

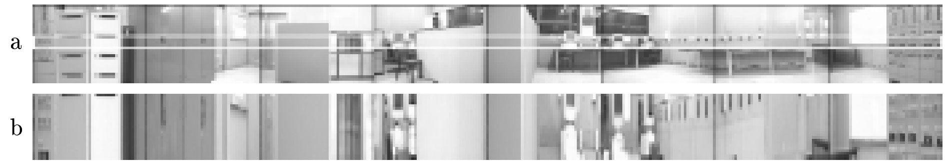


**Fig. 1.** Omnidirectional view of the workspace: a) the original panoramic image. b) The horizontal strip cut from original view, which is marked by the white box shown in image a). (The view shown in b) has been stretched vertically for better display.)

$$MHI = \begin{cases} -2/3 * \pi & S >= 0.15 \text{ and } V >= 10 \quad \text{(black)} \\ -1/4 * \pi & S < 0.15 \quad \text{and } V >= 90 \quad \text{(gray)} \\ -1/3 * \pi & S < 0.15 \quad \text{and } V > 90 \quad \text{(white)} \\ H & \text{otherwise} \quad \text{(other colours)} \end{cases} \quad (1)$$

where H,S,V represents the hue $[0, 2\pi)$, saturation $[0, 1]$ and value $[0, 100]$.

The image is divided into 10-pixel wide vertical bands and the median MHI is computed for each band. Most of the smaller uncharted objects, e.g. network cable ducts, electric switches etc, are removed by band median filtering.

When viewing a large object, we may find regions with relatively constant values in the MHI band median profile, as illustrated in Figure 2. The regional boundaries may represent object edges or distinctive changes in the surface features of objects. We can locate potential regional boundary lines by thresholding the differentiated MHI band median profile. To facilitate the later matching operation, a "region median" is calculated for each detected region by calculating the median MHI of all the bands within the region boundaries.

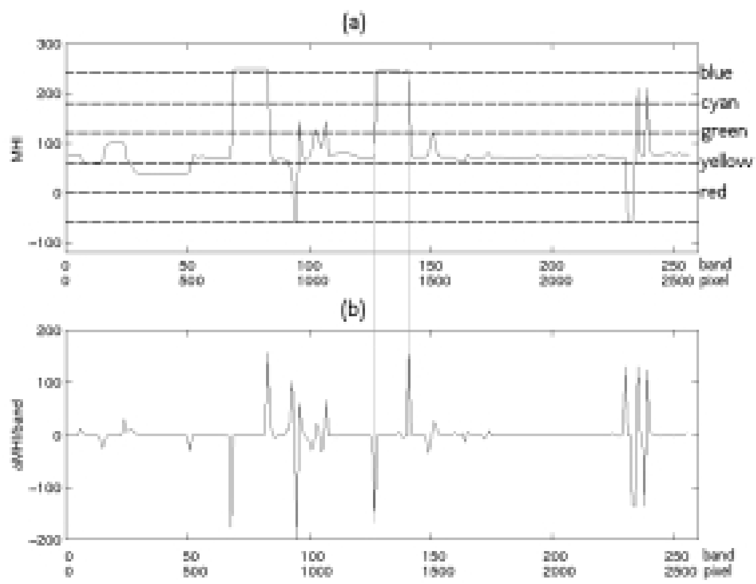


**Fig. 2.** (a) The modified hue index profile. (b) the differentiation of the MHI profile.

### 3.1 Preparations for the Map and Reference Image

Since the band median filtering method removes minor features, the level of detail required in the map is not high, and maps should not be difficult to maintain. The complexity of the environment determines the minimum number of reference images that needs to be taken. If the visibility of different parts of the workspace to the reference point is blocked, more reference points are required. In this study, a simpler environment was considered where only one reference point was sufficient. The exact position of the reference point was determined by surveying before taking the first image.

The viewing angle from the reference point to the edges of the large objects can be calculated from the coordinates of the regional boundaries on the omnidirectional image. The map position of these objects can then be estimated by extending the line-of-sight at the given viewing angle until an intersection is formed on the map, as depicted in Figure 3.

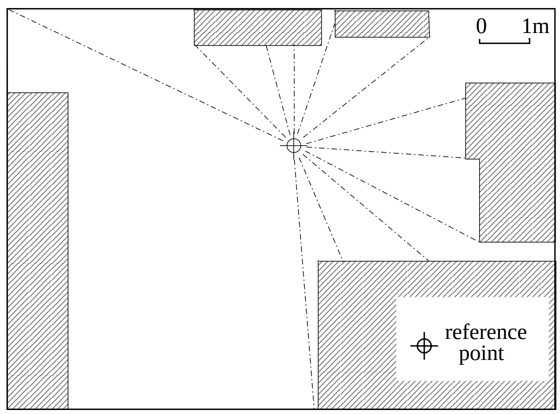


**Fig. 3.** Mapping of the observed feature for the reference image. The map position of an observed feature can be found by extending the line-of-sight at the given viewing angle until an intersection is formed.

### 3.2 Localisation System

An omnidirectional snapshot of the environment is taken whenever the robot needs to re-locate itself, and the MHI is calculated to identify regions. Since the positions of large objects are known, the current position of the robot can be identified using triangulation once enough matches have been established between boundary lines in the reference and current images, that represent features in the map.

The feature matching process is crucial to the performance of the localisation stage. When the robot moves to different parts of the room, the relative size of the regions on the MHI profile may change. Some features may become too small and be left unaccounted for. Due to the presence of uncharted objects, some unexpected features may appear while some expected ones may be occluded. Also changes in reflectance of object surfaces may appear as features after MHI processing. The proposed matching algorithm should be tolerant to these defects.

Omnidirectional images have the important property that the *sequence* of modified hue regions remains the same, providing all the objects are still visible to the observer. A sequence of triples is formed for the reference image by grouping the region median values of three consecutive regions (that is for regions $\{(1,2,3),(2,3,4),(3,4,5),\ldots\}$) into "tokens." The list of region median values for the current image is then searched to locate the possible matches for each of the reference tokens. A match is declared if the region medians for each of the three consecutive regions of current image are within a certain tolerance from the respected regions of the reference token. The tolerance level was set to $\frac{5}{36}\pi$ radians in this study. Ideally, we can obtain a token sequence match from the incoming image that contains as many regions as the reference. In practice the longest token is taken as the best match.

The location and orientation of the robot $(x, y, \phi)$ can be found by solving the following non-linear simultaneous equations:

$$\tan(2 * \pi - \phi - \theta_i) = \frac{y_i - y}{x_i - x} \tag{2}$$

where $x_i, y_i$, represent the $x, y$ coordinates of the $i^{th}$ object edge on the map, and $\theta_i$ represents the observed angle of the $i^{th}$ object edge from the robot. See Figure 4 for further explanation.

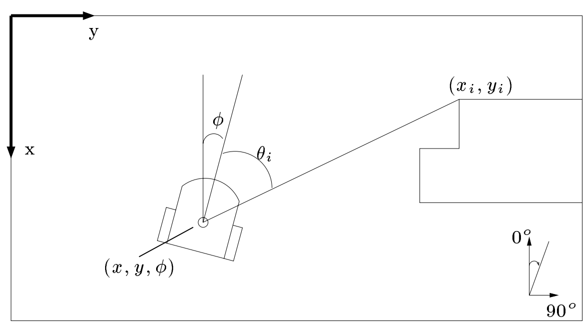


**Fig. 4.** Geometric conventions.

In this study, the camera head was aligned to a fixed direction before taking the first image. The localisation module thus needs to solve for only the two position variables $(x, y)$, So a minimum of two matched features are required.

As an initial investigation, the average is taken of three sets of position estimates, which are generated by taking the observed angles of the first, last and the middle regional boundaries of the longest token match from equation 2.

## 4 Results and Discussions

The vision-based localisation method was tested in an 11.0m × 8.5m laboratory. As shown in Figure 5, nine random testing positions were generated. The test results are shown in Table 1. The average localisation is 0.45 m with a standard deviation of 0.22 m. No mismatch was found between the reference and current image when examined the longest token match for each testing case.

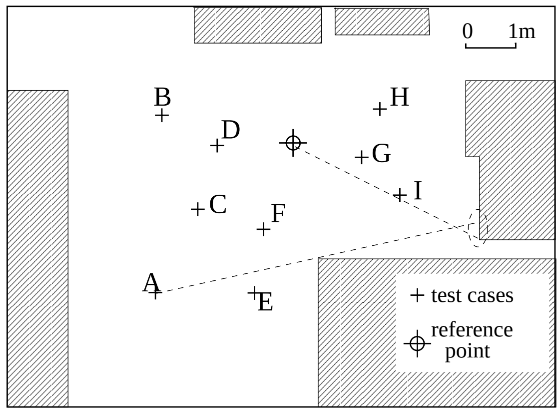


**Fig. 5.** The testing environment for the localisation algorithm. The influence of partial occlusion is demonstrated. The dotted lines from location A and the reference point represent their line-of-sight when supposedly viewing the same edge of an object. Due to partial occlusion, the robot at location A is not really the true edge and thus leads to a large localisation error.

**Table 1.** Localisation error of the testing cases

| Position | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| x-coordinate (m) | 5.57 | 2.11 | 3.93 | 2.70 | 5.58 | 4.53 | 2.95 | 1.98 | 3.69 |
| y-coordinate (m) | 2.98 | 3.10 | 3.82 | 4.14 | 4.91 | 5.13 | 7.02 | 7.43 | 7.79 |
| localisation error (m) | 0.66 | 0.91 | 0.50 | 0.38 | 0.25 | 0.45 | 0.23 | 0.27 | 0.44 |

Although the proposed method may not be accurate enough for the use in a standalone localisation system, that does not poise a serious problem. In this study, we intend to develop a vision-based localisation system that does not depend on the historical position estimates. In this way, the relative rough position estimates can be refined using more established localisation methods, such as extended Kalman filtering.

The test samples that give large localisation error are located far away from the reference point. The view can be quite different from that captured at the reference point. For example, only a fraction of the partition can be visualised at location A. As a result, the observed boundary at location A is not really the true edge of the partition (circled with dots in Figure 5) and thus leads to a large error.

In the current system, the robot position was calculated using only three of the matched features with the rest being discarded. These other matches could potentially be used to improve the accuracy and robustness of the technique. In addition, range sensors can be introduced to the system to reduce the ambiguities arisen during various stage of the operation.

## 5 Conclusion

A vision-based robot localisation system is proposed that does not involve the use of historical position estimates. A modified hue profile is generated for each of the incoming omnidirectional images. The extracted hue regions are matched with that of the reference image to find corresponding region boundaries. As the reference image, exact location of the reference point and the map of the workspace are available, the current position of the robot can be determined by triangulation.

The method was tested by placing the camera set-up at a number of different random positions in a 11.0m $\times$ 8.5m room. The average localisation error was 0.45 m. No mismatch of features between the reference and incoming image was found. While the proposed localisation method may not be sufficiently accurate if used alone, it provides a good initial position estimate for the use of other more established localisation methods, such as extended Kalman filtering.

## References

1. Christian Balkenius. Spatial learning with perceptually grounded representations. *Robotics and Autonomous Systems*, 25:165–175, 1998.
2. Alberto Elfes. Sonar-based real-world mapping and navigation. *IEEE Journal of Robotics and Automation*, RA-3(3):249–265, June 1987.
3. H.R. Everett. *Sensors for mobile robots: theory and application.* A.K. Peters Ltd., 1995.
4. Leopoldo Jetto, Sauro Longhi, and Giuseppe Venturini. Development and experimental validation of an adaptive extended kalman filter for the localization of mobile robots. *IEEE Transactions on Robotics and Automation*, 15(2):219–229, 1999.
5. John J. Leonard and Hugh F. Durrant-Whyte. Mobile robot localization by tracking geometric beacons. *IEEE Transactions on Robotics and Automation*, 7(3):376–382, 1991.
6. Long-Ji Lin, Thomas R. Hancock, and J. Stephen Judd. A robust landmark-based system for vehicle location using low-bandwidth vision. *Robotics and Autonomous Systems*, 25:19–32, 1998.
7. Brian Yamauchi. Mobile robot localization in dynamic environment using dead reckoning and evidence grids. In *Proceedings of the 1996 IEEE International Conference on Robotics and Automation*, pages 1401–1406, Minneapolis, Minnesota, Apr 1996. IEEE.
8. Jianwei Zhang, Alois Knoll, and Volkmar Schwert. Situated neuro-fuzzy control for vision-based robot localisation. *Robotics and Autonomous Systems*, 28:71–82, 1999.