

Department of Electrical  
and  
Computer Systems Engineering

Technical Report  
MECSE-8-2007

Robust Appearance Based Visual Route Following for  
Navigation in Large Scale Outdoor Environments

A. M. Zhang and L. Kleeman

**MONASH**  
UNIVERSITY

# Robust Appearance Based Visual Route Following for Navigation in Large Scale Outdoor Environments

Alan M. Zhang and Lindsay Kleeman

Intelligent Robotics Research Centre

Department of Electrical and Computer Systems Engineering

Monash University, Clayton, Victoria 3800, Australia

[alan.alanzhang@gmail.com](mailto:alan.alanzhang@gmail.com)

[Lindsay.Kleeman@eng.monash.edu.au](mailto:Lindsay.Kleeman@eng.monash.edu.au)

### Abstract

This paper presents a navigation algorithm that enables mobile robots to retrace routes previously taught under the control of human operators in outdoor environments. Possible applications include robot couriers, autonomous vehicles, tour guides and robotic patrols. The appearance based approach presented in the paper is provably convergent, computationally inexpensive compared to map based approaches and requires only odometry and a monocular panoramic vision sensor. A sequence of reference images are recorded during the human guided route teaching phase. During the autonomous phase, current image is compared against the reference image using image cross-correlation performed efficiently in Fourier domain to recover the difference in relative orientation. Route following is achieved by compensating for this orientation difference. Extensive experimental results in urban environments presented in the paper demonstrate robustness against lighting variations and occlusion to an extent not seen in other literature. The cumulative distance traveled by the robot in these experiments exceeds 20km.

### I. INTRODUCTION

Route following is defined as the ability to repeat the same route after having traversed it once under external control. This is an important navigation capability for a mobile robot and an active area of research (Tang and Yuta, 2002; Booij et al., 2007; Segvic et al., 2007; Royer et al., 2005; Goedeme et al., 2005). A potential area of application is robotic couriers where human operators drive the robot from one destination to another with the robot subsequently able to repeat the routes autonomously. This paper presents a robust appearance based solution to the problem of route following in large scale outdoor environments. The navigation system targets urban environments during daylight hours. Passive vision has been chosen as the primary sensing modality due to it possessing a number of attractive properties. In comparison to laser range finders, which have finite range, vision is not constrained by distance to features. The appearance of the environment is also richer in information than the geometry of the environment. Vision sensors are small in size, low in power consumption, inexpensive as well as being passive in nature. The challenge associated with vision however, is mainly two fold: loss of information when a 3D world is projected onto a 2D image plane, and the variation of appearance of visual features depending on lighting conditions. Possible solutions to vision based route following are discussed next.

Building a geometric map of the environment is a possibility. If the robot can construct a geometric map of the environment during route teaching and record its trajectory through this map, then route following can be achieved by localising the robot in this map during the autonomous phase and compensating for the offset from the route. However, vision based SLAM (Simultaneous Localisation And Mapping) algorithms are still far from mature. A number of additional constraints make the route following problem a member of a superset of problems solved by SLAM. Construction of a globally consistent geometric map is the goal of SLAM. Whereas in route following, provided that a route could be retraced successfully, the way the environment is represented is irrelevant. Significant deviations from the route is not expected to occur during normal operation. Thus there is no need to predict the appearance of features from different viewing angles. A convergent feedback system only requires the correct sign of the lateral offset from the teaching route. Furthermore, the ground surface can be assumed to be locally flat, resulting in the vertical orientation of the camera being the same when revisiting the same part of the

route. Thus any differences in robot orientation while retracing the route introduce only a shift in the azimuth angle of the visual features. Under these constraints the authors propose that the approach presented in this paper, being algorithmically simpler than the map based approaches, would suffice for the route following problem. It operates extremely fast and at the same time achieves robustness against occlusion and lighting variations to an extent not seen in other literature. It is interesting to note that many apparently simple vision algorithms perform remarkably well in practice (Srinivasan et al., 2004, 1996, 2006).

## II. HARDWARE

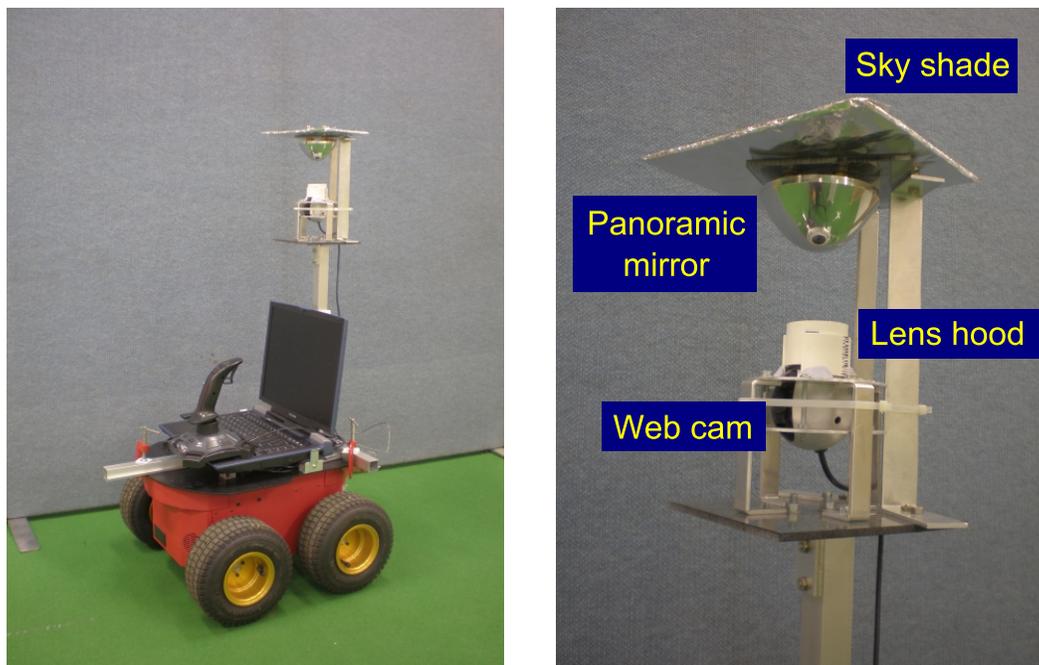


Fig. 1: Hardware setup consisting of a Pioneer P3-AT outdoor robot, camera/mirror assembly for panoramic vision, a laptop and a joystick for manual route teaching.

Figure 1 shows the Pioneer P3-AT outdoor robot that was used as the mobile platform. A web cam directed towards a panoramic mirror of the profile given in (Chahl and Srinivasan, 2000) provides panoramic vision with a vertical Field Of View (FOV) ranging from  $-90^\circ$  to  $48^\circ$  in elevation and  $360^\circ$  in azimuth (the mirror support beams cause some occlusion). A sky shade blocks out the sky to improve the camera automatic gain control. It is covered in aluminium foil such that its brightness changes with ambient lighting. A lens hood was used to mitigate the problem of lens flares. The camera/mirror assembly is 130cm above the ground.

## III. RELATED WORK

As stated earlier, SLAM solves a set of problems of which route following is a subset. A number of pure visual SLAM implementations in the literature could perform route following but most work only indoors. That

number reduces drastically when operation in outdoor environments are required. As this paper does not use SLAM techniques, only a few prominent vision only systems are discussed here to highlight their shortcomings as solutions to route following. SLAM using monocular vision alone has been achieved only recently and some emphasis has been placed on real-time map generation (Davison, 2003). Since route following does not require online map building Royer et al. (2005) used a time consuming but more accurate method to build maps offline. However, image feature tracking and association still needed to be done in real-time for localisation. The systems of Davison (2003); Royer et al. (2005); Herath et al. (2006) and most other geometric mapping based systems use points/corners as visual features. For example, the Harris corner detector was used in (Royer et al., 2005) to detect interest points. Feature association was done using normalised cross-correlation of a neighbourhood of pixels around the interest point. These features are sensitive to the choice of scale at which they are detected. Ideally, the features should be detected for all scales by progressive image subsampling. Point features offer the advantage of being easy to detect and localise. However, features such as the outlines of buildings and vegetation are significant visual features. But these features, especially the outlines of vegetation, are difficult to model geometrically because they change quickly with viewing angle. The work presented in (Goedeme et al., 2005) is capable of recovering the depth of vertical lines in panoramic images. The geometry of straight lines in 3D are recovered in (Smith et al., 2006) using a perspective camera. However, the use of free-form curves that are typical of the outlines of trees and bushes has yet to be demonstrated. The detection of more complex features is also more computationally expensive.

The approach presented in this paper is best classified as an appearance based approach to route following. A related area of research is visual homing where a robot is to return to a reference position when it is initialised at a nearby position. This can be achieved with a simple feedback mechanism without geometric reconstruction of the environment. One of the first to address this problem is Collett (1992). The basic idea is to take visual snapshots at the home/reference position, find the displacement of features in the current visual scene with the reference snapshot and then move in a direction that reduces the average feature displacement. This basic algorithm and its variants can be proven to be globally convergent provided that the feature associations are correct. Insects appear to employ this strategy for navigation (Srinivasan, 1998; Judd and Collett, 1998). The route following problem is then decomposed into a series of reference positions such that visual homing from one reference position to the next is effectively route following. This is a valid solution but it is less than ideal. The algorithm in (Argyros et al., 2001) is representative of most visual homing methods where image features need to be associated between current and reference images. However, feature association is difficult when occlusions and lighting variations are considered. A more fundamental issue is that the direction of movement at any one time cannot be proven to be in the direction of the reference position, even when the final position is convergent towards the reference position. It is therefore difficult to ensure a smooth trajectory while servoing. Recovering the essential matrix (Svoboda et al., 1998) between the current and reference image yields a homing vector in the direction of the reference position (Booij et al., 2007; Goedeme et al., 2005). But this approach implicitly recovers feature depth and should be classified as a SLAM method.

Our approach is similar to that of Matsumoto's 'View-Sequenced Route Representation' (Matsumoto et al., 1996,

2000b, 1999, 2000a) in the sense that the basic premises for convergence are the same and both use cross-correlation for image matching. However, the approach presented here is *provably* convergent and robust against occlusion and lighting changes in outdoor environments. In contrast to most existing works in the literature that are often not accompanied by adequate experimental results, this paper presents a comprehensive set of experimental results.

IV. APPEARANCE BASED ROUTE FOLLOWING

A. System Overview

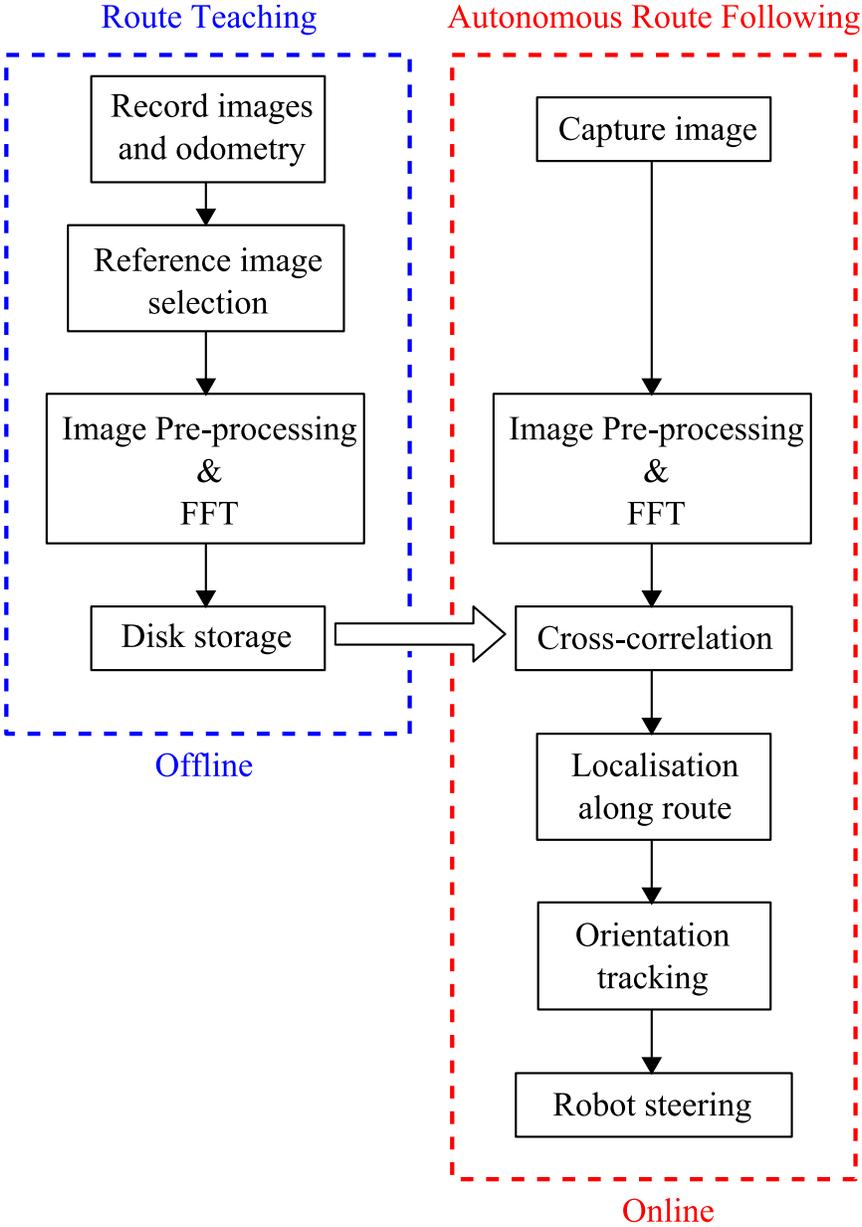


Fig. 2: System Overview

Figure 2 shows the overall system architecture. Route following consists of two phases: route teaching and autonomous route following. During the teaching phase, an operator manually drives the robot from the starting location to the end location. Manual teaching is an intuitive method for humans to communicate to the robot important information regarding the route. For example, the operator may choose to avoid parts of the footpath because it is frequently flooded, make pedestrians more comfortable by keeping to the left/right of the foot paths, or slowing down the robot when it is approaching a blind corner. These are otherwise difficult concepts to communicate, particularly in an outdoor environment. During this teaching phase the robot captures a dense video and records the odometry readings. Reference images are then selected, processed, transformed into Fourier domain and stored. The autonomous navigation phase starts by initialising the robot at the starting point. A new image is captured, pre-processed, then matched against a few reference images in front and behind the current estimated location of the robot along the route. Results of matching provide both a new estimate of the robot location along the route as well as a relative orientation estimate of the robot's current pose with respect to the reference images. A robot steering command that zeros this relative orientation achieves convergence towards the path. Theoretical basis for convergence is established in the next section. Each component in Figure 2 is covered in the subsequent sections, namely: image pre-processing; image cross-correlation; along route localisation; relative orientation tracking and robot steering. Experimental results are presented in Section V. Section VI discusses possible future work and Section VII draws some concluding remarks.

### B. Convergence Analysis

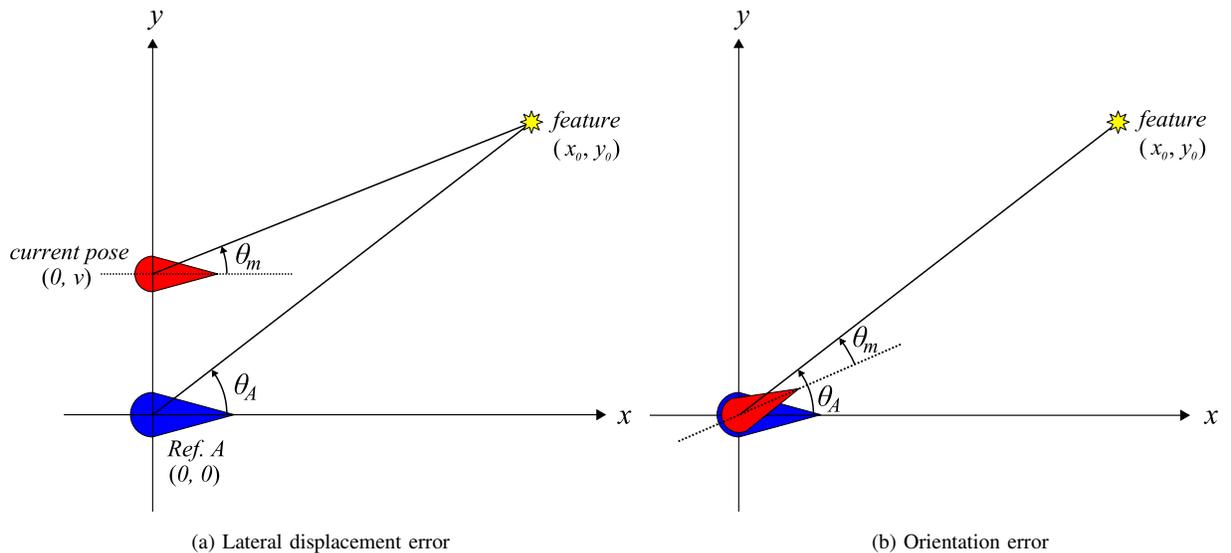


Fig. 3: By zeroing the difference between the angles  $\theta_m$  and  $\theta_A$ , the correct behaviour of rotating clockwise in both cases is achieved even though the two sources of errors cannot be distinguished.

The basis for convergence is illustrated in Figure 3, where the route is a straight line along the  $x$ -axis. Robot

pose where reference images are taken is in blue and the current robot pose in red. Angle to feature are designated  $\theta_m$  and  $\theta_A$  for the current and reference robot poses. The robot steers to zero the difference  $\theta_m - \theta_A$ . Figures 3a and 3b show that lateral displacement error and orientation error cannot be distinguished. But both cases result in a clockwise rotation that would increase  $\theta_m$  to match  $\theta_A$  which is the correct behaviour for convergence back onto the route. It is obvious that orientation errors can be corrected by correcting  $\theta_m$  to match  $\theta_A$ . A convergent behaviour under lateral displacement however, requires a more indepth analysis.

Assuming the presence of only lateral displacement as illustrated in Figure 3a. To ensure convergence towards the reference route the following condition must be met:  $\theta_A > \theta_m$  when  $v > 0$ . The proof is trivial. Since  $\tan$  is monotonic, the following inequality needs to be proven:

$$\begin{aligned} \tan(\theta_A) &> \tan(\theta_m) \\ \iff \frac{y_0}{x_0} &> \frac{y_0 - v}{x_0} \end{aligned} \quad (1)$$

Further restricting features to only the forward facing 180 degrees, i.e.  $x_0 > 0$ :

$$\begin{aligned} y_0 &> y_0 - v \\ \iff v &> 0 \quad \text{as by definition} \end{aligned} \quad (2)$$

This result can be generalised to any number of features by calculating their average:

$$\frac{1}{N} \sum_{i=1}^N \theta_A^i > \frac{1}{N} \sum_{i=1}^N \theta_m^i \quad (3)$$

where  $N$  is the number of features. Because each  $\theta_A^i > \theta_m^i$  is true, the inequality in Eq. 3 is also true. Convergence is therefore preserved. The algorithm's behaviour under displacement along the  $x$ -axis is analysed next.

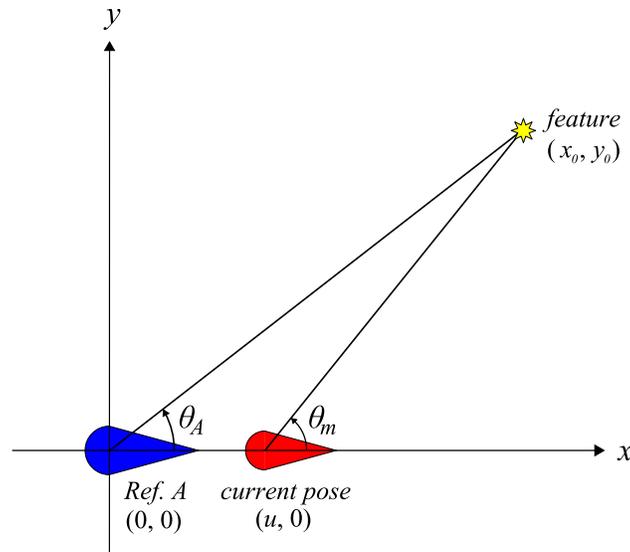


Fig. 4: Along route displacement

Figure 4 shows the robot with displacement along the  $x$ -axis only. By inspection, it is obvious that  $\theta_m > \theta_A$  if the feature is above the  $y$ -axis and  $\theta_m \leq \theta_A$  otherwise. The desired condition is that the average  $\theta_m$  is equal to the average  $\theta_A$  because the robot is already on the right path and it simply needs to travel straight ahead. This condition could be met if the distribution of the features are symmetric about the  $x$ -axis. i.e. the biases in  $\theta_m$  cancel out. However, this is not a valid assumption in reality. A novel, provably convergent algorithm that deals with this situation is presented next.

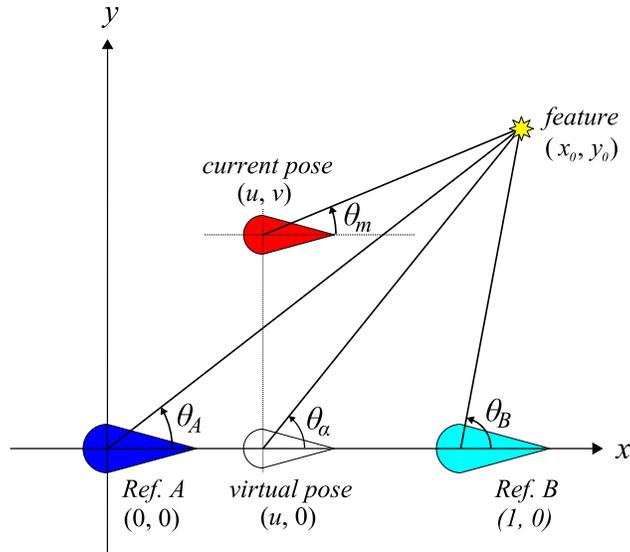


Fig. 5: Along route displacement

Consider Figure 5 where displacement in both axes are present. In order to obtain convergence in these cases, two reference poses are utilised. Referring to Figure 5, if a virtual reference point  $\alpha$  is added between reference points  $A$  and  $B$ , then the configuration is equivalent to that of Figure 3a, for which convergence has already been proven. The angle of the feature at the virtual reference point is then:

$$\theta_\alpha = \tan^{-1} \left( \frac{y_0}{x_0 - u} \right) \quad (4)$$

However, without access to feature location  $(x_0, y_0)$ , an approximation for  $\theta_\alpha$  is required:

$$\begin{aligned} \frac{d}{du}\theta_\alpha &= \frac{y_0}{(x_0 - u)^2 + y_0^2} \\ \text{let: } r &= \sqrt{x_0^2 + y_0^2}, \quad x_0 = r \cos \theta_A, \quad y_0 = r \sin \theta_A \\ \Rightarrow \frac{d}{du}\theta_\alpha &= \frac{r \sin \theta_A}{(r \cos \theta_A - u)^2 + (r \sin \theta_A)^2} \\ &= \frac{r \sin \theta_A}{r^2 + u^2 - 2ru \cos \theta_A} \\ &= \frac{r \sin \theta_A}{r^2 \left(1 + \left(\frac{u}{r}\right)^2 - 2\left(\frac{u}{r}\right) \cos \theta_A\right)} \\ \text{if } u \ll r \text{ then: } \frac{d}{du}\theta_\alpha &\approx \frac{\sin \theta_A}{r} \end{aligned} \quad (5)$$

Thus, the gradient  $\frac{d}{du}\theta_\alpha$  is approximately constant if the distance between the reference images are small compared to the distance to the feature. The angle  $\theta_\alpha$  can then be approximated by a linear interpolation between the boundary conditions  $\theta_A$  and  $\theta_B$ :

$$\theta'_\alpha = (1 - u)\theta_A + u\theta_B \quad (6)$$

where  $\theta'_\alpha$  is the approximation and  $u \in [0, 1]$ . Setting the virtual reference point as the origin reduces the robot offset to only in the direction of the  $y$ -axis. This situation is identical to that in Figure 3a for which convergence has already been proven. Thus, it has been shown that:

- (i) in the reference images, if only features in the forward facing  $180^\circ$  field of view are used;
- (ii) and that the distance between reference poses are small compared to the distance to the features;

then convergence towards the reference route is guaranteed. Condition (i) is numerically analysed in a simulation where reference  $A$  is placed at  $(0, 0)$ , reference  $B$  at  $(1, 0)$  and the virtual reference at  $(0.5, 0)$ . Features are placed uniformly along a semi-circle centred around reference  $A$ . The largest difference between  $\theta_\alpha$  and its approximation  $\theta'_\alpha$  is recorded for each radius of the semi-circle. Figure 6 is a plot of the maximum approximation errors against distance-to-feature from reference  $A$ . The graph clearly shows an asymptotic reduction in approximation errors as the distance to features increased. When features are at a distance of 4 times the separation between references the largest error falls to below 0.5 degrees.

Note that the robot control algorithm minimises the difference between  $\theta_m$  and  $\theta'_\alpha$ :

$$\begin{aligned} \theta_m - \theta'_\alpha &= \theta_m - ((1 - u)\theta_A + u\theta_B) \\ &= (1 - u)\theta_{\delta A} + u\theta_{\delta B} \end{aligned} \quad (7)$$

where  $\theta_{\delta A} = \theta_m - \theta_A$  and  $\theta_{\delta B} = \theta_m - \theta_B$ . This implies that the measurement image is to be compared against reference images  $A$  and  $B$  to obtain  $\theta_{\delta A}$  and  $\theta_{\delta B}$ . But  $A$  and  $B$  are not compared against each other to explicitly recover  $\theta'_\alpha$ .

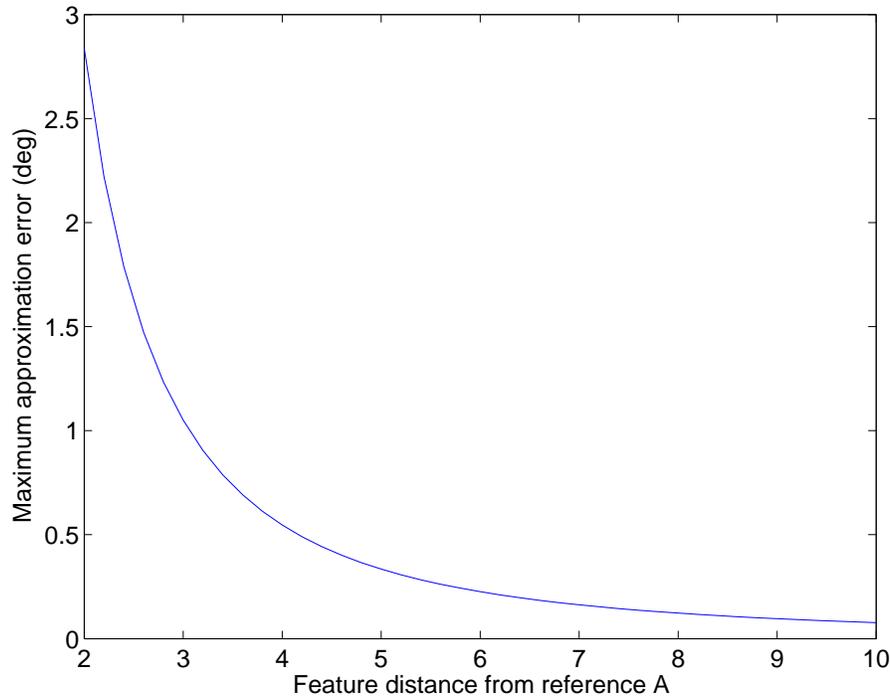


Fig. 6: Maximum approximation error at virtual reference point as a function of distance to feature from reference A.

Curved sections of a route can be divided into straight line segments where the convergence analysis presented above applies. Linear interpolation between references also serves to smooth out robot motion when the route contains sharp curves.

Simulations have also been conducted to directly verify convergence. Similar works in the literature ignore the effect of displacement from reference images along the direction of the route. Only a single closest matching reference image is used to find the relative orientation in (Matsumoto et al., 1999). Simulation results using the new algorithm is compared with the method used in (Matsumoto et al., 1999). The simulated route to follow is a straight line. An initial lateral offset was imposed to demonstrate convergence. For simplicity cumulative odometry was used to decide the location of the robot along the route. Robot dynamics are not simulated. Figure 7 shows simulation results from a symmetric feature distribution. Robot trajectory using the algorithm in (Matsumoto et al., 1999) is shown in black and the proposed algorithm in blue. As expected, there is very little difference in the behaviour of the algorithms under a symmetric feature distribution. Trajectories under *asymmetric* feature distribution is shown in Figure 8. In this case the new algorithm clearly provides smoother motion and at no instant is the robot heading away from the reference route.

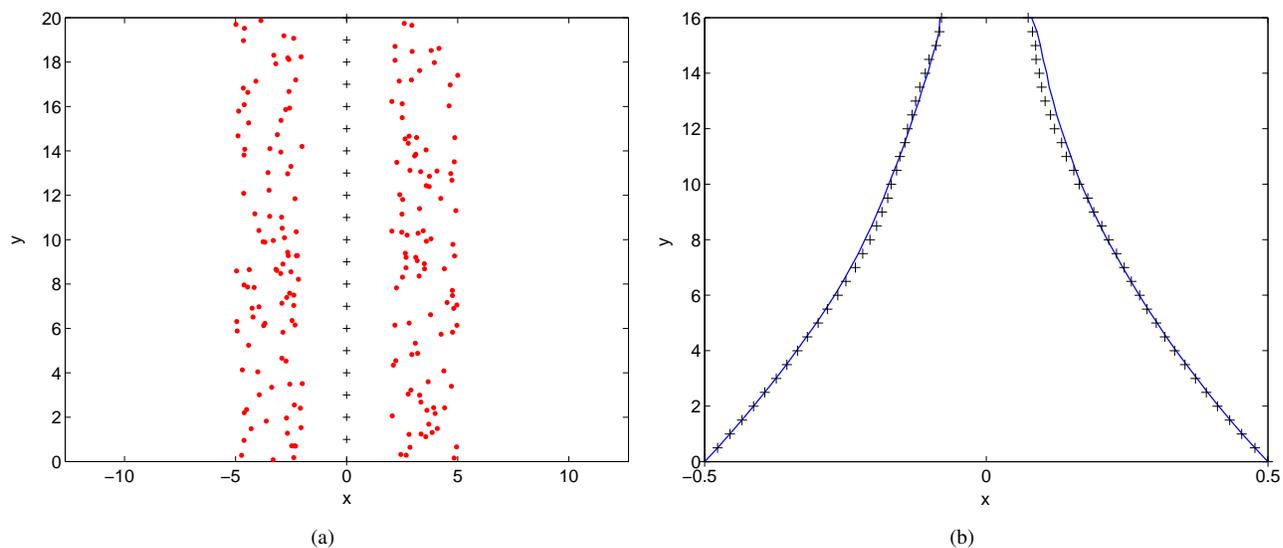


Fig. 7: Symmetric feature distribution. (a) Red points indicate feature locations. Black crosses are the reference poses. (b) Robot trajectory using the proposed algorithm is in blue and that of the algorithm in (Matsumoto et al., 1999) is shown with black crosses.

### C. Image Pre-processing

Identical image pre-processing steps are applied to both reference and measurement images. Input colour image is first converted into greyscale (colour information is unstable under changing lighting conditions). The panoramic image is then “unwarped” (i.e. remapped) onto the azimuth-elevation coordinates. Matching measurement and reference images at varying orientations involve only a shift along the azimuth axis in the unwarped images, whereas an image rotation is needed if matching was performed in the original image. An example of the original colour image and its unwarped greyscale image is shown in Figures 9a and 9b respectively, where horizontal axis is azimuth and vertical axis is elevation. Vertical field of view is restricted to  $[-50^\circ, 20^\circ]$ . The panoramic mirror used has the profile given in (Chahl and Srinivasan, 1999). While it has the advantage of offering constant resolution in elevation, there is no single effective view point. However, as an appearance based method, no geometric constraints are placed on the imaging system.

Robustness against lighting variations is a difficult problem in outdoor environments. It is made even more challenging by the fact that the teaching phase takes place only *once*, while subsequent autonomous runs can take place under a wide variety of lighting conditions. There is no opportunity to learn the appearance of the environment under differing lighting conditions. The burden is then placed on a good heuristic image pre-processing method to remove lighting variations as much as possible. It has been observed that lighting changes generally affect relatively large regions of the scene as a whole, such as shadows from buildings which generally cover large areas.

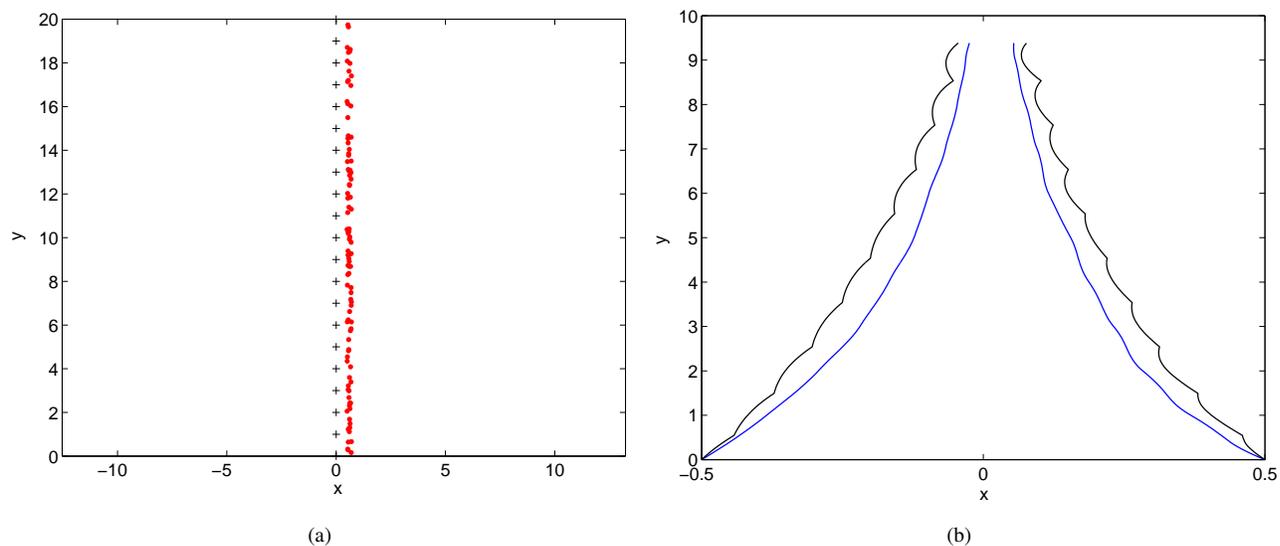


Fig. 8: Asymmetric feature distribution. (a) Red points indicate feature locations. Black crosses are the reference poses. (b) Robot trajectory using the proposed algorithm is in blue and that of the algorithm in (Matsumoto et al., 1999) is shown in black.

This observation motivates the use of ‘patch normalisation’ which transforms the pixel values as follows:

$$I'(x, y) = \frac{I(x, y) - \mu(x, y)}{\sigma(x, y)} \quad (8)$$

where  $I(x, y)$  and  $I'(x, y)$  are the original and normalised pixels respectively,  $\mu(x, y)$  and  $\sigma(x, y)$  are the mean and standard deviation of pixel values in a neighbourhood centred at  $(x, y)$ . A neighbourhood size of 17 by 17 pixels has worked well in the experiments. Figure 9c shows the result after patch normalisation is applied to Figure 9b. Edges can be seen to be preserved and details are brought out in areas with low contrast. However, in areas of near uniform intensity, pixel noise and video compression artifacts are greatly emphasised, as is evident in the lower central portion of Figure 9c. This noise is randomly distributed and does not affect the performance of the system. Patch normalisation exhibits behaviours similar to that of a high-pass filter. Low frequency components, such as shadows that affect large areas, are removed while the high frequency intensity transitions are preserved. The size of the neighbourhood determines the low frequency cut-off threshold. The values  $\mu(x, y)$  and  $\sigma(x, y)$  are calculated efficiently using integral images of pixel sums and squared pixel sums.

#### D. Image Cross-Correlation

This section addresses the problem of measuring an orientation difference between the current image and the reference images. Ground surface along the route is assumed to be locally flat such that the radial axis of the panoramic vision system is perpendicular to the ground plane. The elevation difference between matching features is small because the distance-to-features are large compared to the distance between reference images. Orientation



(a) Original image



(b) Unwarped greyscale image



(c) Patched normalised

Fig. 9: (a) Original colour image. (b) Converted to greyscale and mapped to azimuth-elevation coordinates, where the azimuth-axis is horizontal. (c) Patch normalised to remove lighting variations, using a neighbourhood of 17 by 17 pixels.

difference between reference and measurement images is therefore only a shift along the azimuth axis. This shift is recovered using Image Cross-Correlation (ICC) performed efficiently in the Fourier domain. Let  $\theta$  denote azimuth and  $\phi$  elevation. The frontal  $180^\circ$  field of view of the reference image serves as the template, i.e.  $\theta \in [-90^\circ, 90^\circ]$ . Let the search range be  $\pm\theta_{srch}$  such that the measurement image is limited to the angular range  $[-90^\circ - \theta_{srch}, 90^\circ + \theta_{srch}]$ . Because only a 1D cross-correlation along the azimuth axis is performed, each row in the image is transformed into Fourier domain separately. Reference image is padded with zeros to the same size as the measurement image.

If the measurement image is  $N_\phi$  by  $N_\theta$  pixels, then the Fourier domain image consists of  $N_\phi$  sets of 1D Fourier coefficients, each of a single row. The algorithmic complexity for a single image is  $O(N_\phi N_\theta \log N_\theta)$ . Convolution in spacial domain is equivalent to multiplication in Fourier domain:

$$\begin{aligned}
\lambda &= \sum_{i=1}^{N_\phi} R^i * M^i \\
\Rightarrow \mathcal{F}\{\lambda\} &= \mathcal{F}\left\{\sum_{i=1}^{N_\phi} R^i * M^i\right\} \\
&= \sum_{i=1}^{N_\phi} \mathcal{F}\{R^i\} \cdot \mathcal{F}\{M^i\} \\
\Rightarrow \lambda &= \mathcal{F}^{-1}\left\{\sum_{i=1}^{N_\phi} \mathcal{F}\{R^i\} \cdot \mathcal{F}\{M^i\}\right\} \tag{9}
\end{aligned}$$

where  $\lambda$  is the ICC coefficients,  $R^i$  and  $M^i$  are the  $i$ 'th row in the reference and measurement images respectively,  $*$  is the convolution operator and  $\mathcal{F}\{\bullet\}$  is the Fourier transform operator. Eq. 9 states that each corresponding row of the measurement and reference images are multiplied in Fourier domain. The results are then summed column-wise followed by an inverse Fourier transform to obtain the spacial domain cross-correlation coefficients. Complexity for the multiplication in Fourier domain is  $O(N_\phi N_\theta)$  and for inverse Fourier transform is  $O(N_\theta \log N_\theta)$ . Fourier transforms for the reference images are calculated offline after the teaching run and stored. The complexity of a complete ICC is thus  $O(N_\phi N_\theta \log N_\theta) + O(m N_\phi N_\theta) + O(m N_\theta \log N_\theta)$  where  $m$  is the number of reference images to compare against. This is significantly better than the complexity of ICC performed in spatial domain which is  $O(m N_\phi N_\theta \theta_{srch})$ .

Convergence analysis in Section IV-B used the average angle to features. But the algorithm is still valid if the average operator is replaced by the mode operator (i.e. the relative orientation that most features vote for). The ICC is actually a mixture of both the average and the mode operators. The exact effects are difficult to model but the overall effect is that it at least provides the correct sign for orientation correction. Alternatively, the cross-correlation can be viewed as treating the entire image as a single feature. There is a number of advantages of this approach compared with associating individual features as in (Tang and Yuta, 2002; Argyros et al., 2001; Segvic et al., 2007; Booij et al., 2007). The choice of scale at which to detect individual features like points or lines is particularly important. A multi-scale method is often employed, leading to increased computational burden. Incorrect feature associations also need to be rejected. Image cross-correlation avoids both of these problems. There is also no need to reinitialise tracking of temporarily occluded features. Additionally, the Fast Fourier Transform (FFT) is extremely well developed. Fast and off-the-shelf implementations are freely available. This paper uses the FFTW library version 3.1.2<sup>1</sup>. The total time of performing FFT on a measurement image and comparing against 11 reference images took a total of only 2.3ms on a 2.4GHz Mobile Pentium 4. More timing results are presented in Section V-E.

<sup>1</sup>Distributed under the GNU General Public License, <http://www.fftw.org>

Results from the convergence analysis in Section IV-B requires the distance-to-features to be large to guarantee convergence. It has been observed that large sections of a typical route are along footpaths where the features in front of the robot are much farther than features on the side. Based on this observation a weighting function is multiplied with the reference image during the image preprocessing stage to de-emphasises features on the side, i.e. a weight of unity at  $0^\circ$  azimuth and decreases to zero towards  $\pm 90^\circ$ . This heuristic has been empirically shown to improve performance. A Gaussian weighting function has performed well in the experiments.

#### *E. Along Route Localisation*

Existing works localise the robot to the nearest reference image (Matsumoto et al., 2000b; Jones et al., 1997), where a measurement image is compared to the current nearest reference image and the next one along the route. If the next reference is a better match then the robot is relocalised to this new reference. This method is far too sensitive to errors in image matching and does not provide position estimate in between two reference images. This paper uses a Markov localisation filter (Fox et al., 1999) for along route localisation. The state space is only 1D with the state variable being the distance from the starting point. Robot orientation is tracked separately with a Kalman filter covered in the next section. The state variable is finely discretised at a uniform resolution of 7cm per state. Separation between reference images are larger than the distance between localisation states and may not be uniform. Selecting reference images is itself non-trivial. They should be allocated densely at turns or when visual features are close to the robot. One possible selection method is presented by Matsumoto et al. (2000b), where during the teaching phase, new images are compared with the last reference image. When the matching error exceeds a certain threshold the new image is assigned as a new reference image. But the problem of selecting reference images is a separate offline process and left as an open problem for future work. For the experiments in this paper, a simple allocation method is used. Reference images are allocated with a maximum separation of 35cm and  $5^\circ$  rotation according to odometry. The localisation states being more densely discretised at 7cm per state allows for robot position estimate in between reference images.

Markov localisation filter assigns a probability for each state. Filter update consists of two steps: prediction update and observation update. Robot is assumed to follow the route closely such that odometry provides good estimate of distance traveled along the route. Prediction update involves shifting the probability distribution along the route according to odometry. Errors in distance traveled as measured by odometry is approximated by a Gaussian. This error is modeled in the filter by convolving the states with a Gaussian kernel. Prediction update is only performed when the distance traveled is larger than the distance between states. Since the "kidnapped robot problem" is not one of the operation scenarios and that the robot is always initialised at the starting point, there is no need for global localisation. Thus at any instant in time, only a local neighbourhood of localisation states centred at the most likely robot location is considered. The probabilities of the rest of the states are assigned zero. Computational complexity is therefore constant regardless of route length. During observation update, in order to obtain the observation likelihood, ICC is performed on the measurement image against 11 reference images centred around the estimated current robot location. Peaks are extracted from the ICC coefficients by detecting local maxima with a window size

of 21 degrees, i.e. a peak is the highest value within a window of 21 degrees centre around the peak's position. For each reference image, the height of a single peak with a relative orientation value closest to that of the current estimated robot orientation relative to the route determines the matching likelihood of that reference image. Because the localisation states are denser than the reference images, each state is then given a score via linear interpolation of scores of the reference images in front and behind that state. However, the ICC scores are only an indicator for the actual observation likelihood. The actual observation likelihood is obtained by first normalising the scores such that they sum to unity, followed by addition of a constant, and renormalisation. A larger additive constant has the effect of reducing the confidence placed in the observations. This constant was experimentally determined.

#### F. Relative Orientation Tracking

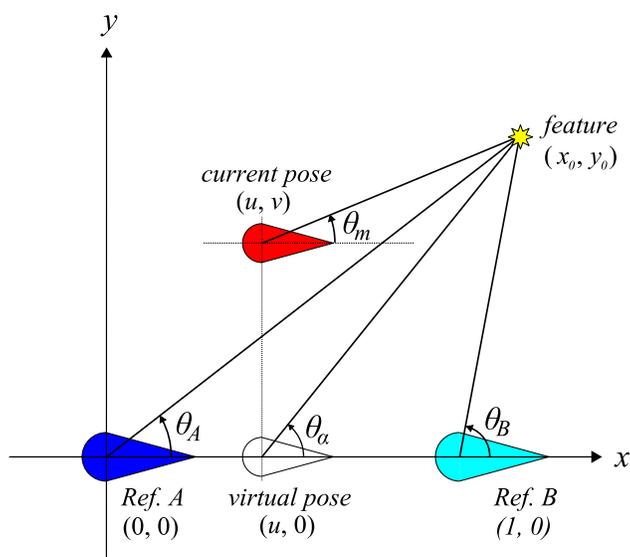


Fig. 10: Robot under both lateral and along route displacement.

Figure 5 is reproduced here in Figure 10 for clarity. With reference to Figure 10 the robot's 'relative orientation' refers to the difference between the robot's current orientation and that of the virtual reference image, i.e.  $\theta_m - \theta'_\alpha$ , where  $\theta'_\alpha$  is the linear approximation of  $\theta_\alpha$ . The relative orientation is designated  $\theta_\delta$ . The control algorithm steers the robot to zero this quantity. This section describes the tracking of  $\theta_\delta$ . A multi-modal tracker is not necessary since the robot must make steering decisions based on the current best estimate of  $\theta_\delta$ . Route following will diverge if the most probable mode does not track the true orientation. A Kalman filter is used to track  $\theta_\delta$ . Prediction and correction updates are presented next.

Prediction update uses readings from odometry as follows:

$$\theta_\delta(d + \Delta d) = \theta_\delta(d) + \Delta\theta_{msur}(d + \Delta d) - \Delta\theta_{ref}(d + \Delta d) \quad (10)$$

where  $d$  is the distance from the start of the route,  $\Delta d$  is the distance traveled since the last update,  $\Delta\theta_{msur}(d+\Delta d)$  and  $\Delta\theta_{ref}(d+\Delta d)$  are the changes in orientation measured by odometry in the distance interval  $[d, d+\Delta d]$  during the autonomous and teaching runs respectively. Change in robot orientation in a differentially steered robot can be modeled as:

$$\Delta\theta = \frac{d_R - d_L}{W} \quad (11)$$

where  $d_L$  and  $d_R$  are the distances traveled by the left and right wheels,  $W$  is the effective wheel separation and the variance is:

$$Var(\Delta\theta) = \frac{Var(d_R) + Var(d_L)}{W^2} \quad (12)$$

Variance of a distance measurement is directly proportional to the distance measured:

$$Var(d_R) = \alpha \cdot |d_R| \quad (13)$$

$$Var(d_L) = \alpha \cdot |d_L| \quad (14)$$

where  $\alpha$  controls the size of the odometry error.

Parameters such as effective wheel separation and distance per encoder tick change overtime due to loss of tire pressure. These systematic errors could be calibrated prior to the experiments. Another type of systematic errors are primarily caused by the changing ground surface characteristics along the route. eg. more wheel slippage occurs on gravel than on concrete surfaces leading to larger turns being registered by odometry. Because the ground surface characteristics cannot be predicted, these systematic errors cannot be compensated prior to the experiments. However, they are *repeatable* when the route is traversed again. Due to their repeatability, these systematic errors canceled out by the subtraction in Eq. 10. This is a distinct advantage compared to the geometric map building approaches where the map would be biased by systematic odometric errors. Other sources of noise include uncertainties in localisation along the route and the discretisation of localisation states. These are modeled collectively by choosing an appropriate  $\alpha$  in Eq. 13 and Eq. 14.

Observed value of  $\theta_\delta$  is obtained by evaluating Eq. 7, which is reproduced here for clarity:

$$\theta_m - \theta'_\alpha = (1 - u)\theta_{\delta A} + u\theta_{\delta B}$$

where reference images  $A$  and  $B$  are immediately behind and in front the current robot location respectively,  $u \in [0, 1]$  indicates the robot position in between  $A$  and  $B$ . The measurement image is compared against both reference images using ICC. Each local maxima in the ICC coefficients is potentially an observation. One local maximum from each of  $A$  and  $B$  that are nearest to the current predicted  $\theta_\delta$  are considered observations of the quantities  $\theta_{\delta A}$  and  $\theta_{\delta B}$ . Evaluation of Eq. 7 then provides an observation of  $\theta_\delta$ . The observation needs to pass a validation gate set at 95% confidence before being accepted for state update. A large number of variables effect the uncertainty of the ICC local maxima, such as occlusion, lighting variations, pixel noise, visual contents of the current scene and image distortion caused by wobbling of the mirror. These variables make accurate modelling of the observation errors extremely difficult. Nevertheless, it has been observed that the shape of the local maximum,

rather than its height, appears to be a reasonable indicator of the observation variance. A parabolar was thus fitted to model its shape. A heuristic then estimates the variance as follows:

$$Var(\Delta\theta_{msur}) = \beta \cdot a_2 \quad (15)$$

where  $a_2$  is the second order coefficient of the fitted parabola and  $\beta$  an empirically determined constant. Problem arises when the shape of the peak is very wide. The variance according to Eq 15 becomes large, resulting in an *increase* in measurement likelihood. This is not desirable because a wide peak is unlikely to be the correct measurement. Therefore an upper limit is imposed on  $Var(\Delta\theta_{msur})$ . Parabola fitting also improves the resolution of the peak position estimate.

Note that relative orientation tracking depends on good localisation along the route. It is particularly sensitive to localisation errors while executing sharp turns due to the large orientation differences between reference images. Thus, while making sharp turns, only the prediction updates are made and no observation updates take place. This ensures that the robot's location estimate moves smoothly along the route during the turn. Sections of the route are labeled as "sharp turns" during the reference image selection step if a curvature threshold is exceeded. This is a part of the reference image selection problem to be addressed in future work.

### G. Robot Steering

The steering algorithm aims to zero the robot's relative orientation using a proportional controller:

$$\omega = -\kappa \cdot \theta_\delta \quad (16)$$

where  $\omega$  is the robot's rate of rotation and  $\kappa$  is the experimentally determined system gain that depends on the system's processing speed and the robot dynamics. A value of 1.5 has worked well in the experiments at a processing rate of 4fps and the robot moving at 60cm/s in straight sections and 45cm/s at sharp turns. With a top speed of 60cm/s this simple proportional controller has worked well in practice.

## V. EXPERIMENTAL RESULTS

As a consequence of the proposed approach being behaviour based, any improvement to the algorithm needs to be tested online. The development process consists of conducting experiments which expose failure scenarios, improving the algorithm, experimental validation, and more experiments that expose further problems. Only the latest version of the algorithm has been described in Section IV. The effect of any improvement to the algorithm can only truly be verified with online experiments, under as many different lighting conditions as possible. It is clearly infeasible to repeat all the experiments after any improvement was made. So the procedure was to first apply the improved algorithm to the recorded experiments that exhibited a specific failure scenario to verify that a better behaviour is achieved. The improved algorithm is then applied to recordings of other successful experiments to ensure that the behaviour does not deviate drastically from the old algorithm. This offline verification process significantly increases the confidence in the improvement not introducing other failure modes. Although some

experiments presented in this section used earlier versions of the algorithm, they nevertheless validate various aspects of the final version. Differences in the algorithms will be explicitly stated for each set of experiments. The cumulative length of the experiments conducted is over 20km.

#### A. Robustness of Relative Orientation Measurement

This set of experiments demonstrates the robustness of the ICC method for relative orientation measurement against occlusion and lighting variations in outdoor environments. In order to isolate the causes of failures, prediction update based on odometry information in orientation tracking was ignored, making the system more sensitive to errors in relative orientation measurement between images. A modified Kalman filter was also used, the details of which is presented in Appendix A.



Fig. 11: Teaching route

Figure 11 shows the teaching route in the Monash University Clayton campus. Figure 12 shows the environment at corresponding spots along the route. The green circle and the red square are the start and end points. The route is 416 meters in length. A total of eight successful runs were made under differing lighting conditions. Cumulative odometry for these experiments are shown together in Figure 13. Odometry of the teaching run is shown in black. The routes' end points are indicated with asterisks. Errors of the end points are in orders of tens of meters.

The experiments were performed during the summer months with very strong sun light. All images in Figure 14 were captured from the same position. Figure 14a is the reference image, Figures 14b, 14c, 14d, 14e and 14f were at different times of sunny days, and 14g was on an overcast day. These images illustrate significant changes in appearance typically associated with changing lighting conditions in outdoor environments. The patch-normalised unwarped images of Figure 14 together with their ICC results against the reference image are shown in Figure 15. Units of  $x$ -axis are in degrees where positive values correspond to the reference image shifting to the right relative to the measurement image. Global maxima are clearly visible near  $0^\circ$  in the ICC coefficients. Note that the maxima in Figure 15e and 15f benefit from the parabola fitting mentioned in Section IV-F in obtaining a higher resolution estimate on the position of the maxima.



Fig. 12: Visual appearance of various points along the route.

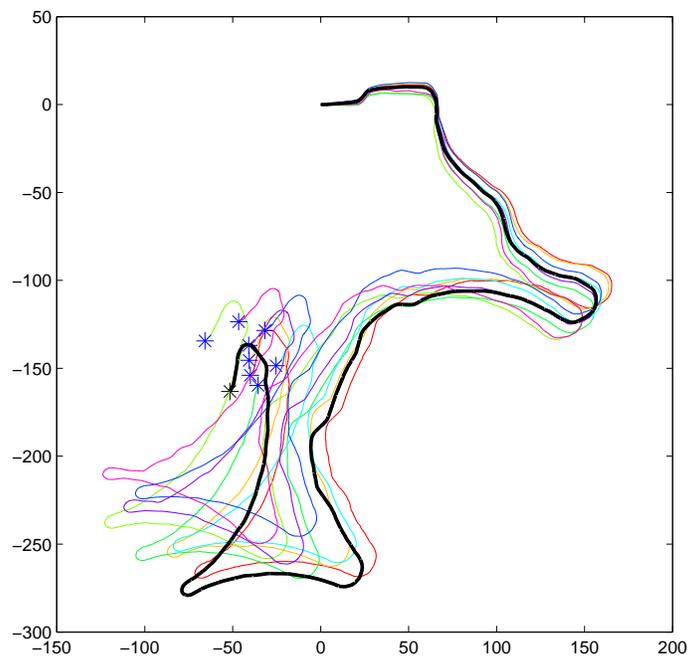


Fig. 13: Cumulative odometry. Units are in meters. Reference run is drawn in black. End point of routes are indicated by asterisks.

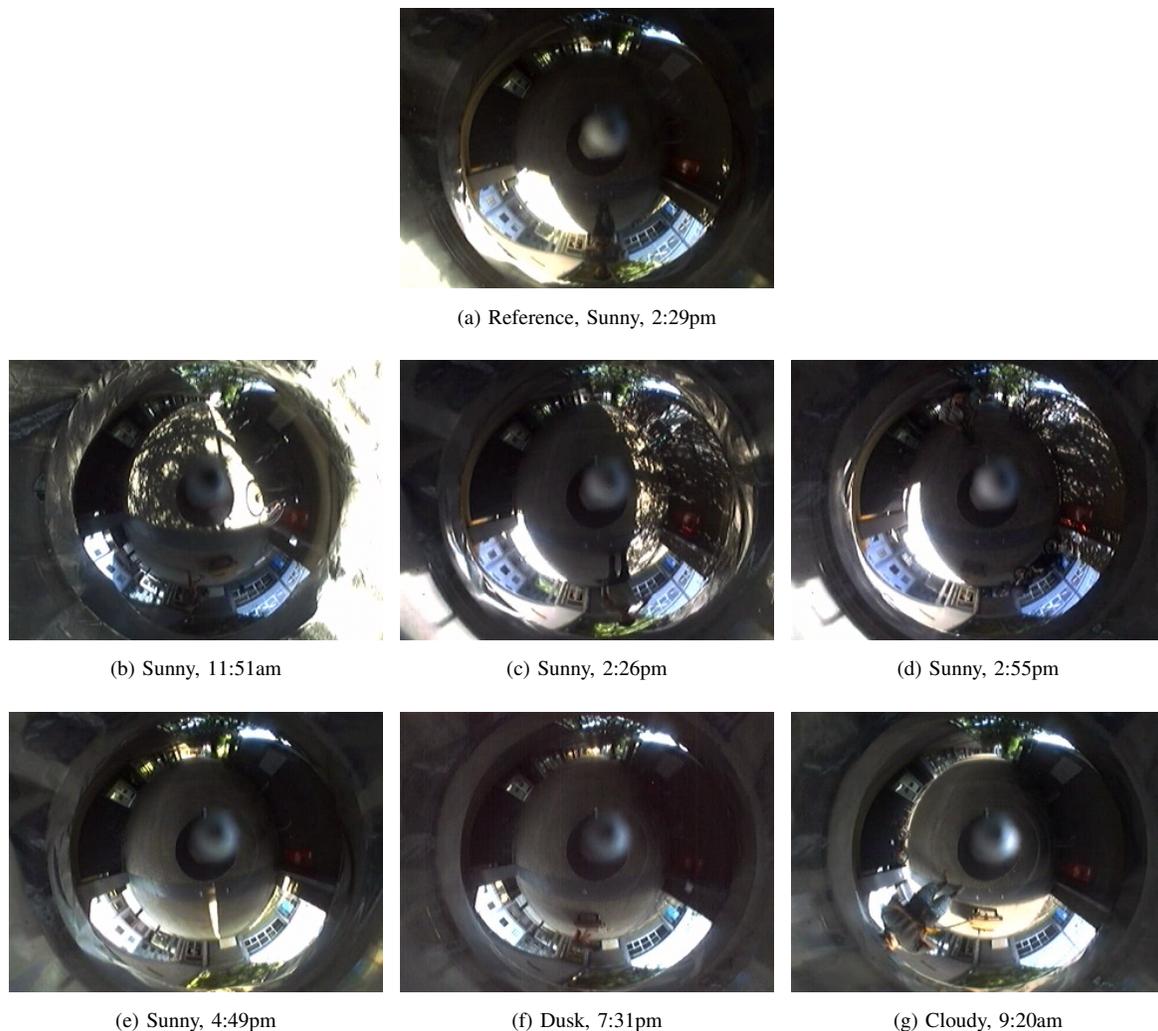


Fig. 14: Variation in lighting conditions. All images were captured at the same position.

Figure 16 shows sections of the autonomous runs containing large amounts of occlusion by pedestrians. The figures show the reference images (top, showing only the frontal  $180^\circ$ ), measurement images subjected to occlusion and the ICC coefficients. Lighting changes are also quite significant as can be seen in Figure 16e, where the outline of the building in the right half of the image is obscured by over saturation. Yet cross-correlation is robust enough to recover the orientation in each case.

### B. Experiments Under Various Lighting Conditions

Algorithm robustness against lighting changes is tested more exhaustively here than the previous set of experiments. The 322 meters long teaching route is shown in Figure 17. Figures 18a and 18b show changes in lighting conditions at two locations along the route. Table I summarises the 23 successful experiments conducted over a period of 9 days with each at a different time of day or weather condition. Sunny days were given greater sampling

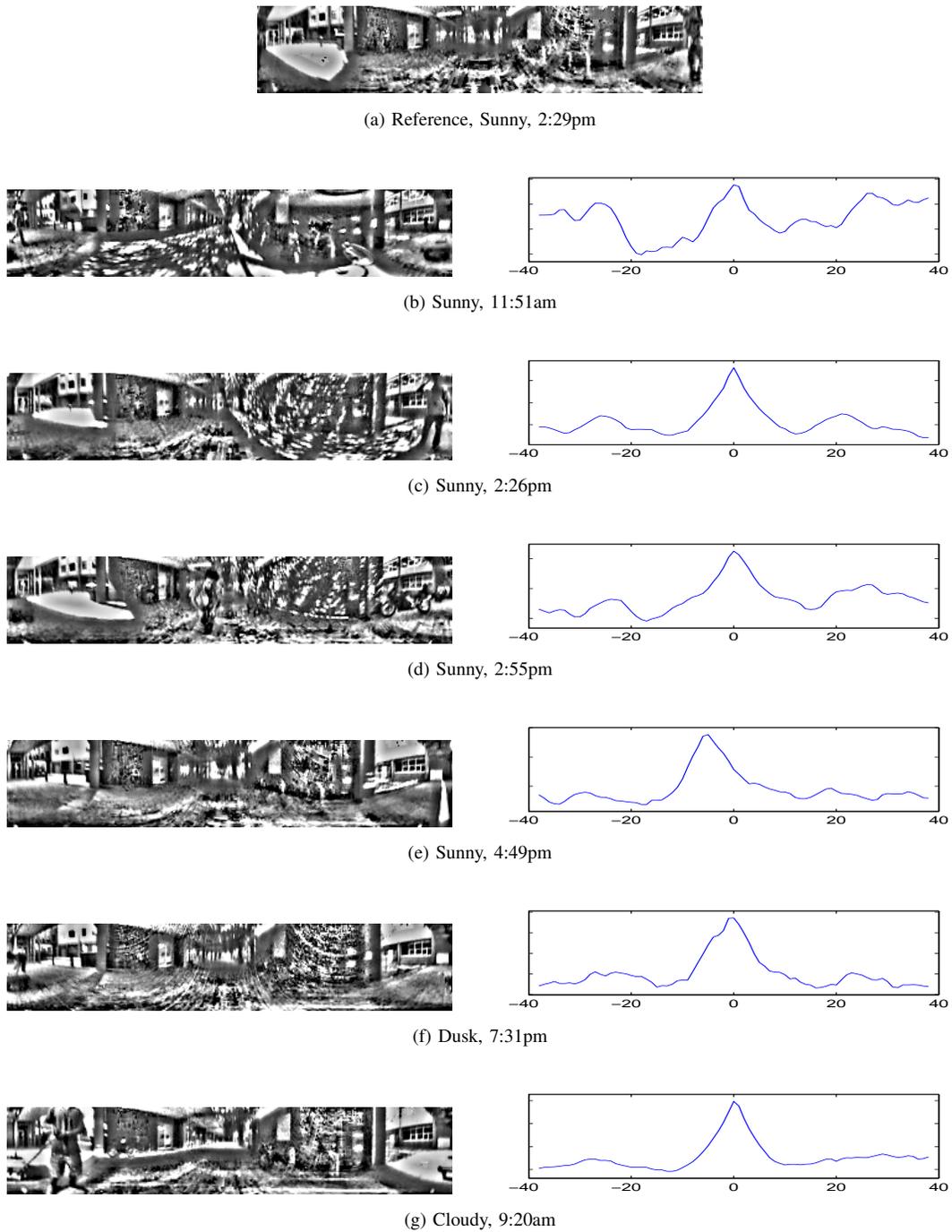


Fig. 15: Robustness of image cross-correlation (ICC) against lighting variations. Patch-normalised unwarped images in the left column. ICC coefficients against the reference image in the right column where units of  $x$ -axis are in degrees and positive values correspond to the reference image shifting to the right relative to the measurement image. Global maxima are clearly present in the ICC coefficients despite the dramatic changes in lighting conditions.

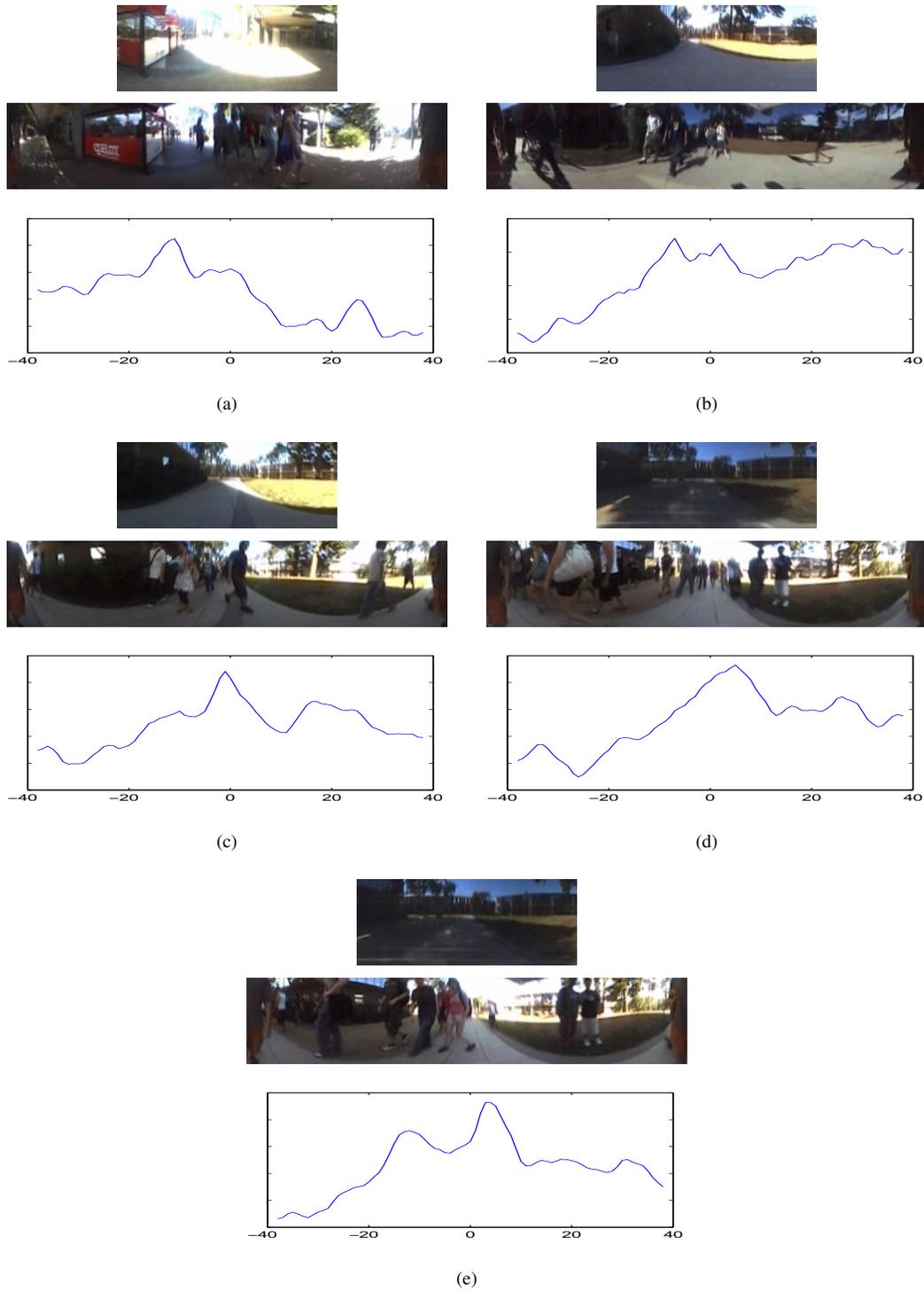


Fig. 16: Robustness of image cross-correlation (ICC) against occlusion. Images of each sub-figure: frontal  $180^\circ$  of reference image (top); measurement image; ICC coefficients where  $x$ -axis is in degrees and positive values correspond to the reference image shifting to the right relative to the measurement image. Global maxima in the ICC coefficients correctly recover the relative orientation between reference and measurement images.

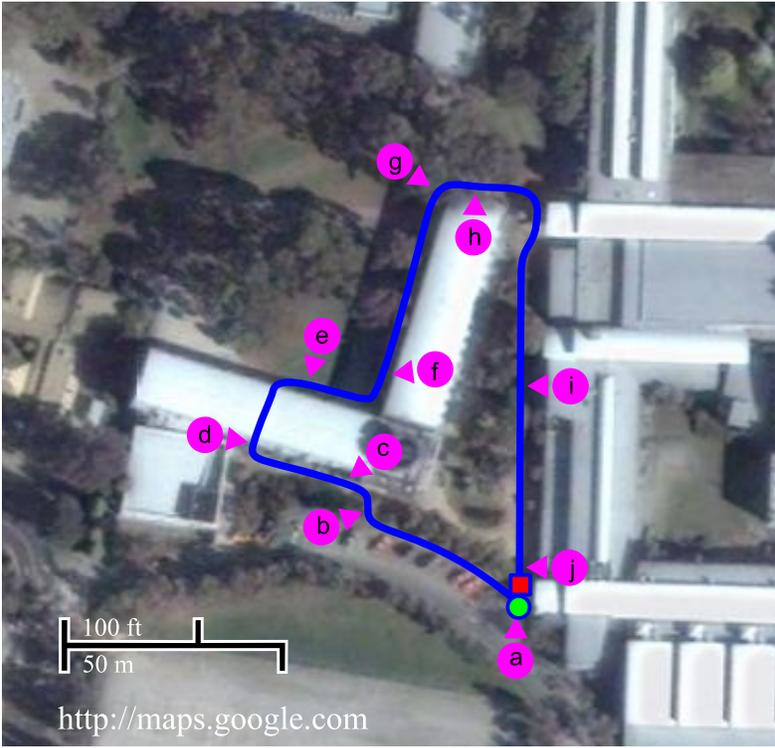


Fig. 17: Route for comprehensive testing under various lighting conditions.



Fig. 18: Top image of each column is the reference, taken at 14:29. Images were taken from experiments carried out at 9:14, 10:20, 11:17, 12:13, 13:13, 15:03, 16:47 and 17:34.

density because they generate a much greater variety of visual appearances as the position of the sun changes with time. These experiments used the modified Kalman filter in Appendix A for relative orientation tracking.

### C. Ground Truth

The ground truth experiments provide quantitative measurements on the accuracy of the algorithm. Ground truth is difficult to obtain in this case because the experiments should be conducted in realistic urban environments. Real-time kinematic GPS used by Royer et al. (2005) provides enough resolution but is inapplicable here due to buildings occluding GPS signals. Accurate laser range finders used in land surveying were also considered. But line-of-sight is difficult to maintain. The chosen method is to mark the robot position at waypoints along the route then manually measure the deviation at these waypoints during autonomous runs.

Exp.	Time	Weather	Exp.	Time	Weather
1	9:14	Sunny	13	13:13	Sunny
2	9:24	Sunny	14	13:21	Partly Sunny
3	10:11	Sunny	15	13:44	Sunny
4	10:20	Partly Sunny	16	13:57	Overcast
5	10:38	Overcast	17	14:54	Sunny
6	10:43	Sunny	18	14:55	Sunny
7	11:17	Sunny	19	15:03	Sunny
8	11:47	Sunny	20	15:21	Sunny
9	12:13	Sunny	21	16:47	Dusk
10	12:40	Sunny	22	17:34	Dusk
11	12:49	Overcast	23	17:57	Dusk
12	13:11	Sunny			

TABLE I: Weather conditions and time of day during which experiments were conducted for comprehensive testing of robustness against lighting variations.

The route followed is the same as in the previous set of experiments. A new teaching run was made to allow for ground truth measurements during which the robot was made to stop at regular intervals along the route. Robot positions at these waypoints were marked on the ground. During the autonomous phase, the robot stopped at these waypoints such that positional errors with respect to the ground truth could be measured. No waypoints were allocated in the middle of turns because stopping the robot during the autonomous phase changes the robot dynamics, hence interfering with the robot control algorithm and could potentially decrease the accuracy of route following. The algorithm used here is the final version as described in Section IV. A total of 9 experiments were conducted. Table II summarises the results. The direction of the  $x$ -axis is perpendicular to the route. Because the localisation states are discretised at 7cm between states, this limits the accuracy along the  $y$ -axis. Lateral deviation ( $x$ -axis) is more representative of the algorithm's accuracy. Reference sequence was captured at 12:37pm on a sunny day. The fact that Table II shows no clear correlation between the amount of deviation and the weather conditions or time of day demonstrates robustness against changing lighting conditions.

#### D. A Longer Experiment

The route is visualised in Figure 19. At 732 meters in length this is the longest route in the experiments. The robot travels through a variety of environments including big open spaces (Figure 19f), footpaths predominantly surrounded by vegetation (Figure 19c), and non-uniformly distributed features (Figure 19d where a building facade is close by on one side and the other side is open space). A total of 7 experiments were conducted, of which 5 were completely successful. The rest two experiments completed 730 meters but failed at 2 meters before the end of the route just before entering the robotics laboratory. One of these failures was caused by the slow response of the camera automatic gain control leading to the inside of the laboratory appearing completely saturated. The other

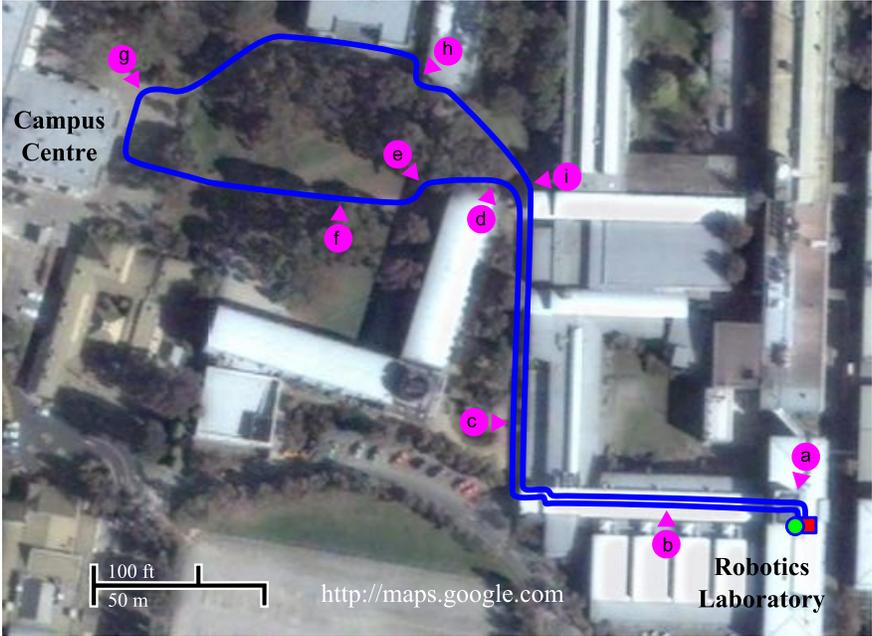


Fig. 19: A route 732 meters long.

Exp. No.	Time	Weather	x  (cm)			y  (cm)		
			Max.	Avg.	Std. Dev.	Max.	Avg.	Std. Dev.
1	16:20	Overcast	10.0	4.6	4.8	24.0	8.4	10.4
2	16:46	Overcast	11.0	5.1	5.5	24.0	9.3	11.2
3	11:20	Overcast	7.0	4.5	4.0	22.0	7.3	9.0
4	12:46	Sunny	6.0	3.4	3.0	20.5	8.5	10.1
5	14:28	Sunny	14.0	5.8	5.0	26.0	8.2	9.9
6	15:22	Sunny	13.0	4.9	4.8	25.5	7.9	9.4
7	15:37	Sunny	11.0	4.6	3.7	23.5	8.4	10.1
8	9:38	Sunny	21.5	8.4	7.4	25.5	9.3	10.9
9	10:27	Sunny	20.0	8.1	6.7	16.0	9.1	10.1

TABLE II: Error during autonomous route following compared with training route. Where  $|x|$  is the magnitude of the lateral offset from the route at the waypoints and  $|y|$  is localisation error along the route.

failure is suspected to be caused by clutter in the laboratory. Again the experiments were conducted under various lighting conditions (Figure 20). Tolerance to occlusion is demonstrated in Figure 21 where a market was setup in the square in front of the campus centre during the autonomous run.

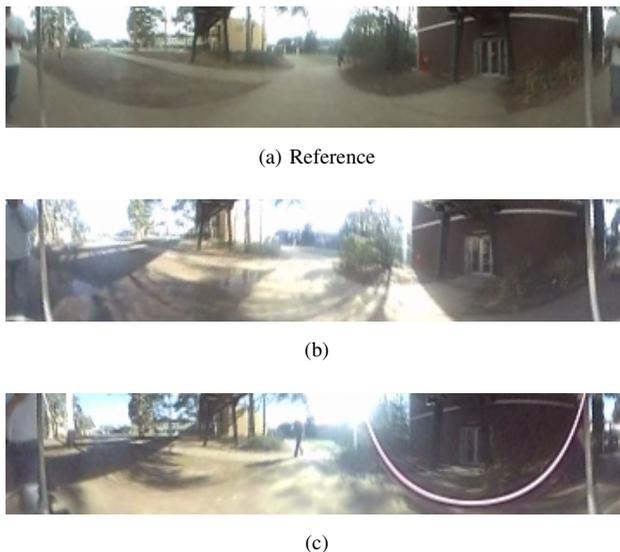


Fig. 20: Examples of changing lighting conditions.

*E. Timing and Storage Requirement*

Below is a summary of system parameters and timing results:

- **Reference image size:** 70 x 180 pixels ( = 70° elevation FOV, 180° azimuth FOV)
- **Measurement image size:** 70 x 256 pixels (i.e. ±38° azimuth search range)



Fig. 21: A market was set up in front of the campus centre that was not present in the teaching run.

- **No. of ref. images to compare against for along route localisation:** 11
- **Computer configuration:** 2.4GHz Mobile Pentium 4, 500MB memory
- **FFT library:** FFTW<sup>2</sup> version 3.1.2
- **Image preprocessing:** 7.2ms per image, including conversion to greyscale, unwarping and patch normalisation
- **Image cross-correlation per measurement image:** 2.3ms, including FFT of measurement image and comparing against 11 reference images

Because the reference images are pre-processed and stored as Fourier coefficients the overhead is only disk access. So image pre-processing plus cross-correlation takes only 9.5ms per measurement image. There are still much optimisation possible with image pre-processing. Offline processing runs at 60fps. During the autonomous phase, with every measurement image recorded in a video, online processing runs at 4fps. At this processing rate the robot is able to travel at 60cm/s in straight sections and 45cm/s at sharp turns. All experiments were conducted at 4fps with logging of every measurement image.

With reference images at a resolution of 70 x 180 pixels and allocated at 35cm intervals, a 1km route requires 36MB of storage space. This requirement is considered low by contemporary storage capacities.

## VI. FUTURE WORK

An obvious improvement to the route following system is a more adaptive method of reference image selection that allows for more closely spaced reference images when the environment becomes enclosed, eg. when moving in cluttered indoor environments. Following the route in reverse is also useful and can be easily achieved. Another useful ability is being able to merge routes and create a topological map that would allow the robot to travel from any starting point to any end point, greatly increasing its flexibility.

<sup>2</sup>Distributed under the GNU General Public License, <http://www.fftw.org>

## VII. CONCLUSION

A vision based navigation system capable of repeating a route previously taught under the control of a human operator in outdoor environments has been presented in this paper. It uses a much simpler algorithm compared to existing visual SLAM methods, made possible by exploiting constraints specific to the route following problem. The algorithm is appearance based and provably convergent. The system has shown a high level of robustness against occlusion and lighting variations in an extensive set of experiments. The longest route was 732 meters in length and was successfully repeated in 7 experiments.

## APPENDIX A

### A MODIFIED KALMAN FILTER

As stated in Section IV-B, the state observation for the relative orientation Kalman filter is the quantity:

$$\widehat{\theta}_\delta = (1 - u)\theta_{\delta A} + u\theta_{\delta B}$$

where  $0 \leq u \leq 1$ ,  $\widehat{\theta}_\delta$  is the observation on the tracked state  $\theta_\delta$ ,  $\theta_{\delta A}$  and  $\theta_{\delta B}$  are the relative orientations of the measurement image with respect to the reference image behind and in front of the current robot location respectively. It is possible for only one of  $\theta_{\delta A}$  and  $\theta_{\delta B}$  to be erroneous. In this case, the correct observation still provides useful information and should be used for filter update. Thus, a mechanism that is capable of independently rejecting observations from reference  $A$  or  $B$  is required. When both are accepted, the filter update should have the same outcome as using a single interpolated observation. Observation update using a single interpolated observation is as follows:

$$S = HPH^T + R_\alpha$$

$$W = HPS^{-1}$$

$$V = \widehat{\theta}_\delta - \theta_\delta$$

$$\theta'_\delta = \theta_\delta + WV$$

$$P' = P - WP^T H^T$$

where  $S$  is the observation likelihood,  $P$  the state variance,  $W$  the filter gain,  $V$  the measurement innovation and  $H$  the observation matrix which is unity in this case since the state is directly observed.  $R_\alpha$  is the observation variance and it is given a fixed value in this version of the algorithm. The innovation can be broken up into two parts:

$$V_A = \theta_{\delta A} - \theta_\delta$$

$$V_B = \theta_{\delta B} - \theta_\delta$$

$$V = (1 - u)V_A + uV_B$$

Substituting into state update:

$$\theta'_\delta = \theta_\delta + (1 - u)WV_A + uWV_B$$

Grouping the weightings  $(1 - u)$  and  $u$  together with  $W$  produces new *weighted* filter gains  $W_A = (1 - u)W$  and  $W_B = uW$ . State variance update can be broken up into two parts as well:

$$P' = P - W_A P^T H^T - W_B P^T H^T$$

The interpolation weights can be treated as a scaling on the observation likelihood:

$$W_A = HP \left( \frac{S}{1 - u} \right)^{-1}$$

$$W_B = HP \left( \frac{S}{u} \right)^{-1}$$

The validation gate then makes use of these scaled observation likelihoods:

$$e_A^2 = V_A^T \left( \frac{S_A}{1 - u} \right)^{-1} V_A$$

$$e_B^2 = V_B^T \left( \frac{S_B}{u} \right)^{-1} V_B$$

If both observations pass the validation gate, then update is made with a single interpolated observation. If only observation from reference  $A$  passes the gate then the following update is made:

$$S_A = \frac{1}{1 - u} (HPH^T + R_\alpha)$$

$$W_A = HPS_A^{-1}$$

$$V_A = \theta_{\delta A} - \theta_\delta$$

$$\theta'_\delta = \theta_\delta + W_A V_A$$

$$P' = P - W_A P^T H^T$$

Likewise if only observation from reference  $B$  passes the gate:

$$S_B = \frac{1}{u} (HPH^T + R_\alpha)$$

$$W_B = HPS_B^{-1}$$

$$V_B = \theta_{\delta B} - \theta_\delta$$

$$\theta'_\delta = \theta_\delta + W_B V_B$$

$$P' = P - W_B P^T H^T$$

## REFERENCES

- Antonis Argyros, Kostas Bekris, and Stelios Orphanoudakis. Robot homing based on corner tracking in a sequence of panoramic images. In *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001.
- O. Booij, B. Terwijn, Z. Zivkovic, and B. Krose. Navigation using an appearance based topological map. In *IEEE International Conference on Robotics and Automation*, pages 3927–3932, 10-14 April 2007.
- J. S. Chahl and M. V. Srinivasan. Panoramic vision system for imaging, ranging and navigation in three dimensions. In *Proceedings of the International Conference on Field and Service Robotics FSA'99*, Pennsylvania, USA, 1999.
- J. S. Chahl and M.V. Srinivasan. A complete panoramic vision system, incorporating imaging, ranging, and three dimensional navigation. In *IEEE Workshop on Omnidirectional Vision*, 2000.
- T. S. Collett. Landmark learning and guidance in insects. *Royal Society of London Philosophical Transactions Series B*, 337:295–303, September 1992.
- Andrew J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *International Conference on Computer Vision, Nice, October 13 - 16*, October 2003.
- Dieter Fox, Wolfram Burgard, and Sebastian Thrun. Markov localization for mobile robots in dynamic environments. *Journal of Artificial Intelligence Research*, 11:391–427, 1999.
- T. Goedeme, T. Tuytelaars, L. Van Gool, G. Vanacker, and M. Nuttin. Feature based omnidirectional sparse visual path following. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1806–1811, 2-6 Aug. 2005.
- D. C. Herath, Sarath Kodagoda, and Gamini Dissanayake. Simultaneous localisation and mapping: A stereo vision based approach. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 922–927, Oct. 2006.
- Stephen D. Jones, Claus Andresen, and James L. Crowley. Appearance based processes for visual navigation. In *1997 IEEE/RSJ International Conference on Intelligent Robots and Systems*, volume 2, pages 551–557, 1997.
- S. P. D. Judd and T. S. Collett. Multiple stored views and landmark guidance in ants. *NATURE*, 392:710–714, 1998.
- Yoshio Matsumoto, Masayuki Inaba, and Hirochika Inoue. Visual navigation using view-sequenced route representation. In *1996 IEEE International Conference on Robotics and Automation*, volume 1, pages 83–88, 22-28 Apr 1996.
- Yoshio Matsumoto, Kazunori Ikeda, Masayuki Inaba, and Hirochika Inoue. Visual navigation using omnidirectional view sequence. In *1999 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 1999.
- Yoshio Matsumoto, Katsuhiko Sakai, Masayuki Inaba, and Hirochika Inoue. Exploration and navigation in corridor environment based on omni-view sequence. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2000a.
- Yoshio Matsumoto, Katsuhiko Sakai, Masayuki Inaba, and Hirochika Inoue. View-based approach to robot

- navigation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2000b.
- E. Royer, J. Bom, M. Dhome, B. Thuilot, M. Lhuillier, and F F. Marmoiton. Outdoor autonomous navigation using monocular vision. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, (IROS 2005)*, pages 1253– 1258, 2005.
- Sinisa Segvic, Anthony Remazeilles, Albert Diosi, and Francois Chaumette. Large scale vision-based navigation without an accurate global reconstruction. In *Computer Vision and Pattern Recognition*, pages 1–8, 17-22 June 2007.
- Paul Smith, Ian Reid, and Andrew Davison. Real-time monocular slam with straight lines. In *British Machine Vision Conference*, volume 1, pages 17–26, September 2006. ISBN 1-904410-14-6.
- M. V. Srinivasan, S. W. Zhang, M. Lehrer, and T. S. Collett. Honeybee navigation enroute to the goal: visual flight control and odometry. *The Journal of Experimental Biology*, 199:237–244, 1996.
- M. V. Srinivasan, S. W. Zhang, J. S. Chahl, G. Stange, and M. Garratt. An overview of insect-inspired guidance for application in ground and airborne platforms. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 218(6):375–388, 2004.
- Mandyam V. Srinivasan. Animal navigation: Ants march as they march. *NATURE*, 392:660–661, apr 1998.
- M.V. Srinivasan, S. Thurrowgood, and D. Soccol. An optical system for guidance of terrain following in uavs. In *IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 51–56, Sydney, 2006.
- T. Svoboda, T. Pajdla, and V. Hlavac. Epipolar geometry for panoramic cameras. In *6th European Conference on Computer Vision*, pages 218–231, 1998.
- Lixin Tang and S. Yuta. Indoor navigation for mobile robots using memorized omni-directional images and robot's motion. In *IEEE/RSJ International Conference on Intelligent Robots and System*, pages 269–274, 2002.