2007 IEEE International Conference on
Robotics and Automation
Roma, Italy, 10-14 April 2007



# High Precision Relative Localization Using a Single Camera

Munir Zaman
School of Electronics and Physical Sciences,
University of Surrey, Guildford, GU2 7XH, UK.
email: m.zaman@surrey.ac.uk

*Abstract*— **In this paper a method for high precision relative localization using a single camera is proposed. The method provides pose estimates comparable in resolution to wheel odometry, but being independent of the kinematics and based on an exteroceptive sensor, is resistant to wheel slippage. The concept is based on extracting the planar transformations between frames from a sequence of ground images, which correspond to the change in robot pose. Results on a plain colored carpeted surface provide the proof of concept of the method as an alternative to wheel odometry. The contributions in this paper include a method to estimate the planar transformations between images to a high degree of precision (e.g., 0.01 degrees), and a method to calibrate the system using a 1D calibration object and known motions of the robot.**

## I. INTRODUCTION

Wheel odometry is a commonly used sensor input for mobile robot localization. Wheel odometry provides *relative* localization as it detects changes in pose relative to its previous pose. The advantage of wheel odometry is that it is high resolution and simple to use. It can typically detect movements of the order of tenths of millimetres. However, it has some known drawbacks. The localization errors increase without bound, and non-systematic errors, such as wheel slippage, are undetected. This is because wheel odometry data rely proprioceptive sensors (i.e., sensors which provide information on the internal kinematic position of the robot) and are therefore oblivious to the external environment. Exteroceptive sensing methods for relative localization, which sense the environment, may be used to overcome this drawback.

Exteroceptive sensing methods for relative localization, include optical mice [1]. These are mounted alongside a mobile robot either on its own, or in combination with wheel odometry [2] and a range sensor [3]. The optical mice are *forced* onto the ground to minimise variations in distance between the ground plane and the optical sensor. This increases friction affecting the traction of the wheels, increasing the likelihood of wheel slippage. Other ground-sensing methods include extracting the orientation of ground tiles to correct for orientation errors [4], but this is limited to surfaces known to contain such features.

The author is with QinetiQ Ltd, UK and is a Visiting Researcher at the University of Surrey. The work presented in this paper are the views of the author and no connection with QinetiQ Ltd is expressed or implied.

A method called 'visual odometry' has come into vogue [5] and [6]. Generally, visual odometry aims to estimate the motion of the robot by extracting and tracking *ground point* features, to compute a number of optical flow vectors. These are computed from images taken from forward facing camera inclined towards the ground. The image processing is complicated as a large number of the weak features need to be filtered out for matching to be robust. In addition, on planar surfaces, solutions may fail where the optical flow vectors lie on a single plane [7]. This can be addressed by defining a ground planar constraint in known environments [8]. Visual odometry has been used on Mars Exploration Rovers (MER), but fail on planar surfaces (e.g., Meridiana plains) because they are devoid of sufficient visual texture for reliable feature extraction [9]. On these surfaces the fallback to wheel odometry is used on the assumption that wheel slippage is minimal on planar regions.

Recently, a method coined 'visiodometry' has been published which is based on the sensor inputs from two *ground-pointing* cameras [10]. The method is based on extracting pairs of inter-frame translations *only*, called 'shift vectors', using phase correlation [11]. The change in pose is computed from each pair of shift vectors.

In this paper a comparable method for relative localization, but based on a *single* ground-pointing camera, is proposed. The method relies on estimating the translational *and* rotational component between frames. These estimates are then transformed to the robot frame to update the robot pose. The proposed method provides a sensor data input similar in modality and resolution to wheel odometry, but does not rely on the integrity of a kinematic model. It can be used on planar regions and has the potential to be used on those surfaces which do not contain sufficient visual texture for 'visual odometry'.

The proposed method relies on estimating inter-frame transformations to a high degree of precision. The translation component is estimated using phase correlation. For estimating the rotational component to a high precision, an extension to phase correlation is also presented. The technique is based on finding the maximum of a curve fitted to the peak value of the inverse Fourier of the normalized cross power spectrum (NCPS) at various image orientations. The precision of the rotational estimates are of the order of $0.01^\circ$.

The rest of the paper is structured as follows. The proposed

1-4244-0602-1/07/$20.00 ©2007 IEEE.

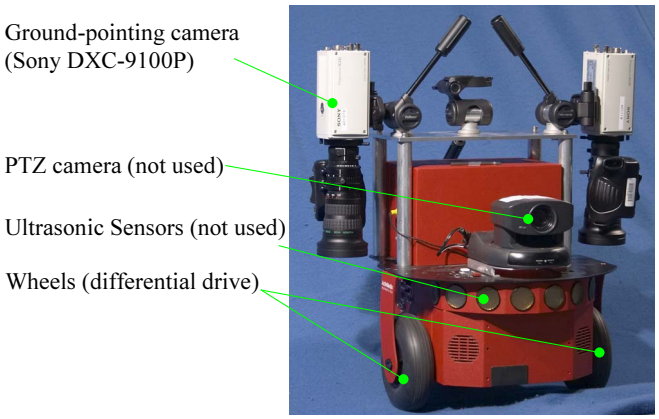



Fig. 1. Visiodometry implemented on a Pioneer DXe differential drive robot, augmented by a ground-pointing camera (the other camera is for load balance).

method for relative localization is described in section II, followed by a description of the method to extract planar roto-translations to high precision in section III. A calibration method for the vision system and robotic platform is described in section IV. In section V, experimental results are presented and discussed. Finally, conclusions and areas for future work are presented in VI.

## II. Relative Localization Using a Single Camera

### A. Experimental Setup

Fig. 1 shows a Pioneer DXe differential drive robot augmented by two ground-pointing cameras. However, only the images from one camera are required for input to the proposed method (the other camera is to balance the load on the robot). The image plane of the camera is ≈0.5m above the ground set at a focal length of ≈30mm. The camera is a Sony DXC-9100P, compact 1/2" 3CCD progressive scan RGB camera.

*a) The Coordinate Reference Frames:* Three 2D coordinate reference frames are defined. They are (i) the **world** frame, which is an arbitrary frame fixed in the world on the ground plane, (ii) the **robot** frame, which is the origin of the robot, typically the position corresponding to the centre of rotation during a spot-rotational motion, with the x-axis being the forward direction, and (iii) the **visiodometry** frame which corresponds to the image plane projected onto the ground, with the origin as the projected centre of the image.

*b) Notation:* The notation used is described as follows. The robot **pose** is defined as its position $(x, y)$ and orientation $(\theta)$ in the world frame. This is represented by the vector

$$\boldsymbol{p}(x, y, \theta) = \begin{pmatrix} x \\ y \\ \theta \\ 1 \end{pmatrix} \tag{1}$$

The matrix $\mathbf{T_{r2w}}$ transforms a pose in the robot frame to the world frame

$$\mathbf{T_{r2w}}(x, y, \theta) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 & x \\ \sin\theta & \cos\theta & 0 & y \\ 0 & 0 & 1 & \theta \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2}$$

Note the equivalence between the elements in the pose vector $\boldsymbol{p}$ and the transformation matrix, as the robot pose in the robot frame is simply the null vector

$$\boldsymbol{p_0} \triangleq \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \tag{3}$$

therefore $\boldsymbol{p} = \mathbf{T_{r2w}}\,\boldsymbol{p_0}$. The matrix $\mathbf{T_{v2r}}$ transforms a pose in the visiodometry frame to the robot frame

$$\mathbf{T_{v2r}}(x_v, y_v, \theta_v) = \begin{pmatrix} \cos\theta_v & -\sin\theta_v & 0 & x_v \\ \sin\theta_v & \cos\theta_v & 0 & y_v \\ 0 & 0 & 1 & \theta_v \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{4}$$

### B. Relative Localization

In this section the method to estimate the change in pose of the robot (i.e., 'relative localization') is presented. The principle of the method relies on the basis that the image plane is sufficiently parallel to the ground plane for changes in the pose of the robot to correspond to planar transformations of the scene (i.e., the visiodometry frame) viewed by the camera. The rotational component of the change in the robot pose corresponds to the rotational component of the transformation of the visiodometry frame, *regardless of the position of the origin of the visiodometry frame*. The relationship between a change in the visiodometry frame and the new robot pose in the world frame can be expressed by the following

$$\boldsymbol{p}' = \mathbf{T_{r2w}}\,\mathbf{T_{v2r}}\,\mathbf{V}\,\mathbf{T_{v2r}^{-1}}\,\boldsymbol{p_0} \tag{5}$$

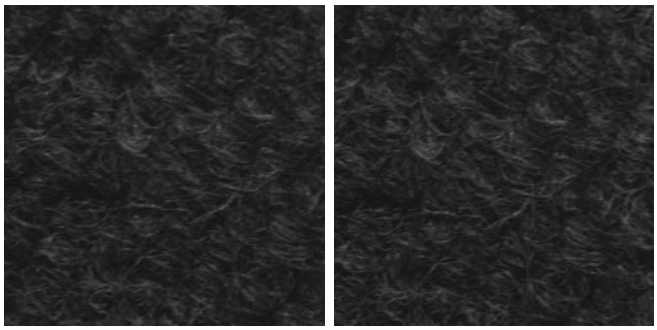
and

$$\mathbf{V}(\delta x_v, \delta y_v, \delta\theta_v) = \begin{pmatrix} \cos\delta\theta_v & -\sin\delta\theta_v & 0 & \delta x_v \\ \sin\delta\theta_v & \cos\delta\theta_v & 0 & \delta y_v \\ 0 & 0 & 1 & \delta\theta_v \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

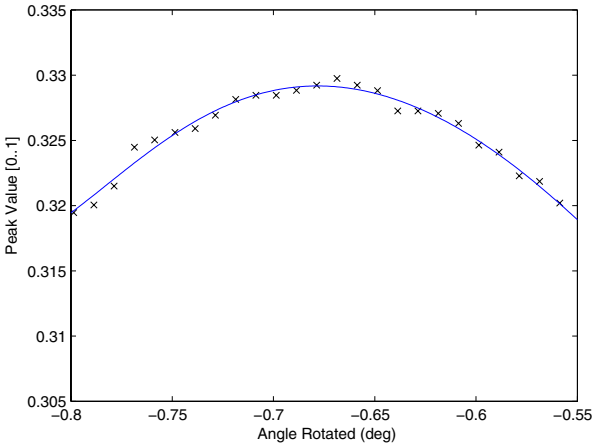
where $\mathbf{V}$ is the rotation $\delta\theta_v$, and translation $(\delta x_v, \delta y_v)$ of the visiodometry frame. The accuracy of the method relies on estimating the inter-frame planar transformations to a high level of precision and accuracy. This is described in section III. The parameters of $\mathbf{T_{v2r}}$ are estimated using a calibration method described in section IV-B.

## III. High Precision Planar Roto-Translation

In this section the method for estimating the planar roto-translations to a high degree of precision is described. It is also noted that the images taken are of the ground - where strong point features cannot be relied on to exist. Hence, methods to estimate the planar transformations from



(a) (b)




(c)

Fig. 2. **Estimating Rotation to High Precision.** Fig. 2(b) is a planar roto-translation of the image in Fig. 2(a). The images are 255x255 pixels representing an area of carpet 20mm x 20mm. Fig. 2(c) shows a curve fitted to the *peak value* of the inverse Fourier of the normalized cross power spectrum at relative rotational steps of $0.01^\circ$ in the range $[-0.8, -0.55]$ degrees. The x-coordinate of the maximum is the estimate of the rotational component between the images.

extracting and matching *point* features are not considered to be sufficiently robust. The alternative to point feature based methods are global correlation based methods.

Phase correlation based methods have been proven to be very robust with a claimed accuracy of $\approx 0.25^\circ$ [12]. However, this method is clearly unsuitable where the rotational component itself is at sub-degree level. In addition, there is no evaluation of the error distribution, nor is the accuracy validated using *naturally* rotated images (the accuracy is evaluated using the same image rotated by interpolation).

The method proposed here is an extension of the normalised phase correlation method published by Kuglin and Hines as far back as 1975 [11].

The basic concept of the method is to apply phase correlation to pairs of images $I_1, I_2$, where $I_2$ is a planar roto-translation of $I_1$. The effect of rotating the images in small steps on the peak value of the inverse Fourier of the NCPS is observed. This can be seen in Fig. 2(c) where Fig. 2(b) is a planar roto-translation of Fig. 2(a). It is considered that as the amount rotated approaches the true rotational component between the images, the peak value approaches a maximum. The peak value is denoted as $f_\cdot$ below.

1) **Inputs.** The method takes as input two images $I_1, I_2$, where $I_2$ is a planar transformation of $I_1$.
2) **First Pass: Bound**. The first stage is to bound the rotational component between the two images. Prior knowledge on the maximum angular velocity, limits the range of possible rotation values to $[-1, 1]$ degrees, which can be further narrowed if the previous rotation $\tau$ and the angular acceleration is known.
   a) **Plot function $f_1$.** Plot function $f_1(\theta; I_1, I_2)$ for values of $\tau - 0.5 < \theta < \tau + 0.5$ in steps of $0.05^\circ$. Each image is rotated by $\frac{\theta}{2}$ in opposite directions to reduce any possibility of bias in the errors during rotations (rotations are generally performed using interpolation).
   b) **Find max $f_1$.** A median filter is applied and a polynomial $\tilde{f}_1$ is fitted. The angle corresponding to the peak is denoted as $\alpha_1 \triangleq \arg\max_\theta\{\tilde{f}_1(\theta)\}$. Apply the phase correlation function and denote the translation estimate at $\alpha_1$ as $\Delta s_{x_1}, \Delta s_{y_1}$.
3) **Second Pass: Refine.** The second pass is refinement of the first pass. The rotations are performed using images $\tilde{I}_1, \tilde{I}_2$, which are the respective *cropped* images $I_1, I_2$ to exclude non-overlapping regions. As non-overlapping regions are considered as 'noise' by phase correlation, cropping not only increases the peak value but ensures that the peak value only represents 'noise' in the image. This provides a value representing a measure of confidence in the estimate which could be used as a form of input variance in a sensor fusion system. A narrow search window with finer step sizes is used.
   a) **Plot function $f_2$.** Plot function $f_2(\theta; \tilde{I}_1, \tilde{I}_2)$ for values of $\alpha_1 - 0.25 < \theta < \alpha_1 + 0.25$ in steps of $0.01^\circ$ (An example is in Fig. 2(c)).
   b) **Find max $f_2$.** A median filter is applied and a polynomial $\tilde{f}_2$ is then fitted. The position of the peak is $\alpha_2 = \arg\max_\theta\{\tilde{f}_2(\theta)\}$. Recalculate, and denote the translation estimate at $\alpha_2$ as $\Delta s_{x_2}, \Delta s_{y_2}$.
4) **Output.** The output is $s_x = \Delta s_{x_1} + \Delta s_{x_2}$, $s_y = \Delta s_{y_1} + \Delta s_{y_2}$ and $s_\theta = \alpha_2$, where $s_x, s_y$ is the pixel shift, and $s_\theta$ the rotation.

The values of $s_x, s_y$ are scaled by $\alpha_x, \alpha_y$, which provides the estimate of the translational component, and $s_\theta$ provides the rotational component of the visiodometry frame. The parameters $\alpha_x, \alpha_y$ are estimated from camera calibration (see section IV-A).

## IV. System Calibration

There are two separate elements for calibrating the system:

1) **Camera Calibration**. This provides an estimate of the scale factor parameters transforming pixels in the image to known units in the visiodometry frame.

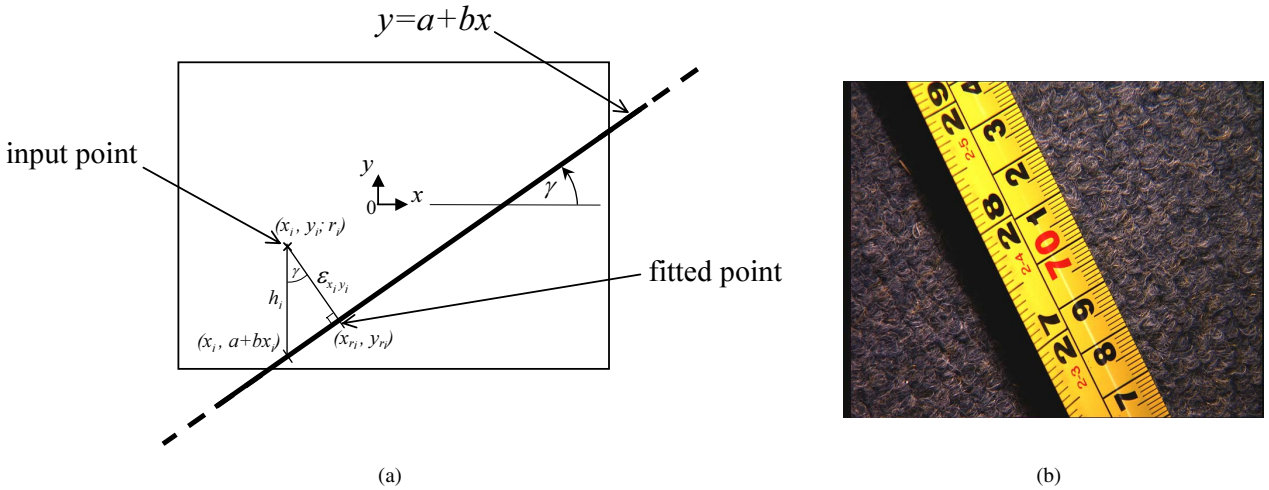



Fig. 3. **Camera Calibration.** The coordinates of points on the edge with the corresponding reading is input shown in Fig. 3(b). These are then fitted to a straight line representing the estimated position of the rule edge as shown in Fig. 3(a). A Least Squares method is used to estimate the equation of the line which provides the scale factor of pixel units to known units in the visiodometry frame.

2) **Kinematic Calibration**. This provides an estimate of the position and origin of the visiodometry frame with respect to the robot frame. The robot frame is defined as the kinematic origin of the robot (i.e., a point on the robot where there is no displacement during spot-rotation).

### A. Camera Calibration

The objective of camera calibration is to estimate the calibration matrix $\mathbf{K}$ containing the scale factors $(\alpha_x, \alpha_y)$ which transforms a pixel translation to a translation in known units in the scene (i.e., visiodometry frame). Lens distortion is ignored because, not only are the lenses of high quality (Fujinon lenses), but the images are cropped from a PAL image of 720(H)x576(V) pixels to a central 256x256 pixels (≈16%) where lens distortion is considered to be negligible.

A number of methods using 1D objects for camera calibration have been published and were considered. The earliest published method was in 2002 [13]. The method relies on six snapshots of a straight stick rotated around a fixed 3D point. The method requires several views of the calibration object, at various orientations. In 2003, a formal analysis of camera calibration using 1D objects was [14]. This was for the purposes of calibrating a multiple camera network. In 2004, a method was published which did not require metric information on the 1D object[15]. However, the method required taking two images of the object in two fixed positions and from two viewpoints (i.e. four images). Therefore the calibration object needs to be positioned at the same two positions after the camera is moved to its second viewpoint. All these methods were unsuitable because they either required the calibration object to be positioned in a non-planar environment or were unnecessarily complex. A much more simpler method using 1D object, which takes advantage of the known constraints of the system (i.e., planar environment), is proposed.

A 1D object is essentially a set of two or more collinear points. In the proposed method this is in the form of a graduated directional line cutting diagonally across the view of the camera. The 1D calibration object used is a metal tape measure. An example calibration image can be seen in Fig. 3. The pixel positions of points are manually extracted from the rule edge with the corresponding ruler reading noted. This provides an estimate of the scale factor of pixels to known units in the visiodometry frame.

As an affine camera model is assumed, with the origin of the scene corresponding to the image centre projected onto the ground. Projection is therefore a scale factor:

$$\mathbf{X_i} = \begin{pmatrix} X_i \\ Y_i \\ 1 \end{pmatrix} = \mathbf{K} \begin{pmatrix} x_i \\ y_i \\ 1 \end{pmatrix} \tag{6}$$

and

$$\mathbf{K} = \begin{pmatrix} \alpha_x & 0 & 0 \\ 0 & \alpha_y & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \alpha_x,\, \alpha_y > 0 \tag{7}$$

where $\alpha_x$ and $\alpha_y$ (e.g. units of mm/pixel) scale the pixel units to a known unit of measure, $\mathbf{X_i}$ the vector of corresponding image coordinates and $\mathbf{K}$ the calibration matrix.

1) **The Inputs $(x, y, r)$.** Take $n$ ruler readings $r_i$ $(i = 1, \ldots, n)$ and their corresponding image-centred coordinates at the rule edge $(x_i, y_i)$.
2) **The Equation of the Rule Edge $(y = a + bx)$.** It is known that the rule edge is a straight edge, and the points $(x_i, y_i)$ are fitted to a straight line, by minimising the perpendicular distance to the fitted line $y = a + bx$ from the points $(x_i, y_i)$ (e.g., [16]). From the estimate of the line representing the rule edge, the fitted positions of the observations on the line, denoted as $\boldsymbol{p_{r_i}} = (x_{r_i}, y_{r_i})$, are derived (see figure 3(a)).

$$x_{r_i} = x_i + \epsilon_{x_i y_i} \sin\gamma \tag{8}$$
$$y_{r_i} = y_i - \epsilon_{x_i y_i} \cos\gamma \tag{9}$$

where $\epsilon_{x_i y_i}$ is the perpendicular distance from the observed point to the fitted line.

3) **The Calibration Scale Factor** $(\alpha_x, \alpha_y)$**.** Taking the case for estimating $\alpha_x$, define $\tilde{x}_i \triangleq x_{r_i}$ then the relationship between the *true* points on the line as projected in the scene and the *true* ruler reading is

$$\alpha_x \left|\bar{x}_{i+1} - \bar{x}_i\right| = \left|r_{i+1} - r_i\right| \cos\gamma \tag{10}$$

and $\tilde{x}_i = \bar{x}_i + \varepsilon_{r_i} \cos\gamma$, where $(\bar{x}_i, \bar{y}_i)$ is the true value, and $\varepsilon_{r_i}$ the error. Assuming the error to be a Normal random variable with zero mean and variance $\sigma^2$, combining (8) and (10) and re-arranging

$$\Delta\varepsilon_{r_i} = \Delta\tilde{x}_i - \frac{\cos\gamma}{\alpha_x}\Delta r_i \tag{11}$$

where

$$\begin{aligned} \Delta\varepsilon_{r_i} &= \varepsilon_{r_{i+1}} - \varepsilon_{r_i}, \\ \Delta\tilde{x}_i &= \left|\tilde{x}_{i+1} - \tilde{x}_i\right|, \\ \Delta r_i &= \left|r_{i+1} - r_i\right|. \end{aligned}$$

The variance and covariance of $\Delta\varepsilon_i$ is derived, noting that by definition

$$\begin{aligned} \sigma^2 &\triangleq E[\varepsilon_{r_i}^2] \\ E[\varepsilon_{r_i}\varepsilon_{r_j}] &= E[\varepsilon_{r_i}]E[\varepsilon_{r_j}], \qquad i \neq j. \end{aligned}$$

The probability density function is therefore

$$p(\boldsymbol{\varepsilon}_\Delta) = \frac{1}{\sqrt{|2\pi\mathbf{P}|}} e^{-\frac{1}{2}\boldsymbol{\varepsilon}_\Delta^{\mathbf{T}}\mathbf{P}^{-1}\boldsymbol{\varepsilon}_\Delta} \tag{12}$$

and

$$\boldsymbol{\varepsilon}_\Delta = \begin{pmatrix} \Delta\varepsilon_{r_i} \\ \vdots \\ \Delta\varepsilon_{r_{n-1}} \end{pmatrix}, \mathbf{P} = \begin{pmatrix} 2\sigma^2 & -\sigma^2 & & \\ -\sigma^2 & \ddots & \ddots & \\ & \ddots & \ddots & -\sigma^2 \\ & & -\sigma^2 & 2\sigma^2 \end{pmatrix}$$

where $\boldsymbol{\varepsilon}_\Delta$ is the vector-valued Gaussian random variable and $\mathbf{P}$ the covariance matrix. Finding the Maximum Likelihood estimate of (12) leads to

$$\min_\alpha \left\{ \frac{1}{2}\boldsymbol{\varepsilon}_\Delta^{\mathbf{T}}\mathbf{P}^{-1}\boldsymbol{\varepsilon}_\Delta \right\} \tag{13}$$

The minimum of (13) must satisfy

$$\mathbf{J}^{\mathbf{T}}\mathbf{P}^{-1}\boldsymbol{\varepsilon}_\Delta = 0, \tag{14}$$

where $\mathbf{J}$ is the Jacobian of $\boldsymbol{\varepsilon}_\Delta$ with $\frac{1}{\alpha_\cdot}$ the variable.

**Estimating** $\alpha_x$. Vectorising and re-arranging (11) and combining with (14) we can solve for $\alpha_x$

$$\frac{1}{\alpha_x} = -(\mathbf{J}^{\mathbf{T}}\mathbf{P}^{-1}\mathbf{J})^{-1}\mathbf{J}^{\mathbf{T}}\mathbf{P}^{-1}\Delta\tilde{\boldsymbol{x}} \tag{15}$$

**Estimating** $\alpha_y$**.** The above principle is repeated for computing $\alpha_y$ by substituting (11) with

$$\Delta\varepsilon_{r_i} = \Delta\tilde{y}_i - \frac{\sin\gamma}{\alpha_y}\Delta r_i. \tag{16}$$

Hence, the required parameters for the calibration matrix $\mathbf{K}$ have been estimated.

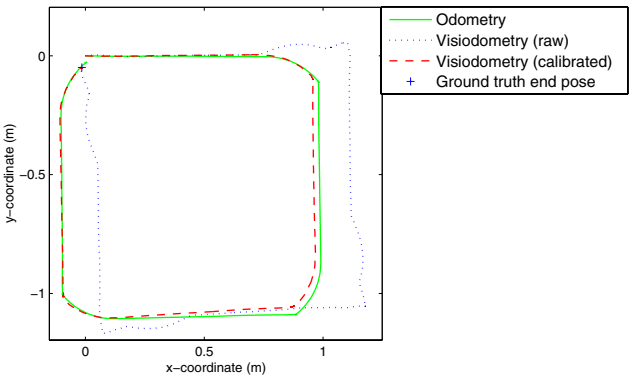



Fig. 4. **Kinematic Calibration.** As the origin of the visiodometry frame does not coincide with the robot origin, position estimates during rotational motion results in a 'swinging out' (as in the figure) or a trajectory which is rotated around the start position (not shown). This can be corrected by kinematic calibration where the trajectory matches that estimated from wheel odometry.

### *B. Kinematic Calibration*

The objective of kinematic calibration is to estimate the parameters of $\mathbf{T_{v2r}}$ (see (4)). Fig. 4 illustrates how the visodometry based trajectory estimates differ from the position of the robot. It can be observed that the position estimates from the raw visiodometry result in a deviation of the true trajectory of the robot where there is a rotational component in the robot motion. This makes it difficult to compare or combine the two position estimates. Kinematic calibration transforms the change in the visiodometry frame to a change in the robot frame. There are two parts for kinematic calibration (i) estimating $\theta_v$, and (ii) estimating $x_v, y_v$.

*1) Estimating* $\theta_v$*:* The parameter $\theta_v$ aligns the axes of the visiodometry frame with the robot frame. If forward motion is along the x-axis then the vector corresponding to the translation in the visiodometry frame $(x_i, y_i)$ during known *pure translational* motion needs to be rotated by $\theta_v$, thus

$$\theta_v = \arg\min_\phi \sum_{i=1}^{n} (\theta_i - \phi)^2, \tag{17}$$

where

$$\theta_i = \tan^{-1}\left(\frac{y_i}{x_i}\right). \tag{18}$$

*2) Estimating* $(x_v, y_v)$*:* The parameters $(x_v, y_v)$ corresponds to the position of the visiodometry origin in the robot frame. The concept of the method is to estimate $(x_v, y_v)$ by computing the displacement of the robot at each *spot rotation* by $\tau_i$. The key observation is that at the true value of $(x_v, y_v)$ spot rotations do not cause any displacement of the robot relative to the robot frame.

Let $\mathbf{V}_i(x_i, y_i, \tau_i)$ be the transformation of the visiodometry frame after a rotation by $\tau_i$. Now, if $\boldsymbol{p} = (p_x, p_y)$ is the unknown coordinates of the visiodometry origin in the

*pre-rotated* robot frame then

$$\boldsymbol{p} = \mathbf{T_{v2r}}\,\mathbf{V}_i\,\boldsymbol{p_0} \tag{19}$$

However, $\boldsymbol{p}$ and $(x_v, y_v)$ are unknown, but the rotation $\tau_i$ may be computed using the method described earlier in section III. Now the spot rotation of the robot by $\tau_i$ is the same as the rotational component of the visiodometry frame. The point $\boldsymbol{p}$ can be computed relative to the robot frame as

$$\boldsymbol{p} = \mathbf{R}(-\tau_i)\begin{pmatrix} x_v \\ y_v \end{pmatrix} \tag{20}$$

where

$$\mathbf{R}(-\tau_i) = \begin{pmatrix} \cos(-\tau_i) & \sin(-\tau_i) \\ -\sin(-\tau_i) & \cos(-\tau_i) \end{pmatrix}. \tag{21}$$

Combining (19) and (20) to eliminate $\boldsymbol{p}$, and re-arranging

$$\boldsymbol{x_i} = \begin{pmatrix} x_{v_i} \\ y_{v_i} \end{pmatrix}. \tag{22}$$

This can be solved as a linear Least Squares equation of the form $\mathbf{A}\boldsymbol{x}_i = \boldsymbol{b}$, for *each* step $i$.

Now for $i = 1, \ldots, n$ for $n$ rotation steps, the Least Squares estimate for the coordinates of the visiodometry origin, is simply the mean of $(x_{v_i}, y_{v_i})$ estimated at each step $i$, thus

$$\begin{pmatrix} x_v \\ y_v \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x_i}. \tag{23}$$

## V. Experimental Results and Discussion

In this section experimental results are presented followed by a discussion of the proposed method.

### A. Experimental Results

The proof of concept of the proposed method was demonstrated by conducting two closed loops each of 4 one-metre legs with a 90° rotation towards the end of each leg. Run-1 is a counter-clockwise run and consists of a sequence of around 1800 frames lasting 70s. Run-2 is a clockwise run and consists of around 3000 frames lasting 120s. Runs 1 and 2 were taken at different times under different camera configurations and speed. Examples of the type of images can be seen in Fig. 2.

Fig. 5 is a plot of the robot path estimated from visiodometry against the path estimated from wheel odometry data. It is considered that the runs were conducted under fairly ideal conditions of minimal wheel slippage, and that the plots from wheel odometry data are sufficiently reliable to provide a qualitative assessment of the validity of the proposed method. The ground truth of the end pose, however, is computed from measurements taken from the ground positions of two markers affixed on either side of the robot, taken at the start and end of each run using the method in [10].

**Position Accuracy.** Although strictly not ground truth, under the controlled conditions, the position estimates from wheel odometry are used as a reference against which a reasonable qualitative assessment of the accuracy of visiodometry can be made. The ground plots for the faster run-1 broadly follow the trajectory of the wheel odometry, however, there is clearly a much greater congruence for the slower run-2 (see Fig. 5).

**Heading Accuracy.** Ground truth was used to estimate the heading error. The heading error is **3.2%** and **1.2%** for run-1 and run-2 respectively, over a total change in heading of 351.3° and 362.0° respectively. The error for wheel odometry is **2.1%** and **0.5%** respectively.

### B. Discussion

A qualitative analysis of the results (see Fig. 5) demonstrates the feasibility of the concept of the method. Although the proposed method provides a simple to use data stream (i.e., a change in heading and position), estimating the rotational component is computationally expensive.

In the implementation the resolution is of the order of 0.1mm and 0.01° for translational and rotational components respectively, which is comparable to wheel odometry. The key distinction between the proposed method and wheel odometry is the use of an exteroceptive sensor (i.e., a camera) over a proprioceptive sensing mechanism (i.e., optical wheel encoders). This makes the method independent of the kinematics and has therefore the potential to be applied to non-wheeled robots and robots whose kinematic positions are not known or modelled.

In comparison to the **optical mice** method, the non-contact method of sensing is major advantage as there is no increase friction, and can be used on a wider range of surfaces, including friable and rough surfaces. The **shift vector** method and the optical mice method require two sensors with sufficient separation. This acts as a limit on the minimum size of the sensor system. The use of a single camera is therefore less complex and more compact. In addition, the shift vector method approximates the translation vector over the entire frame, even where, due to a rotational component, each pixel is translated by a slightly different amount. The proposed method does not suffer from this drawback. The rotational component is compensated for *before* estimating the translational component using phase correlation.

The advantage over **visual odometry** is its simplicity. There is no requirement to extract, store and match large numbers of point features over several frames, and perform complicated statistical processing. The precision of visual odometry implementations are generally an order of magnitude less precise (i.e., centimetres) than visiodometry (i.e., sub-millimetres, making it less robust to detecting displacements of the order of millimetres.

Visiodometry does not suffer from drift, is passive, non-contact, kinematic independent and uses a ubiquitous sensor (i.e., CCD array). A key pre-requisite is that the ground surface must contain sufficient texture for motion to be detectable from a sequence of images at the resolution of the camera. Images of most ground surface at resolutions of 0.1mm per pixel are likely to contain discernible texture and therefore this requirement is not considered to be major obstacle to a practical implementation.

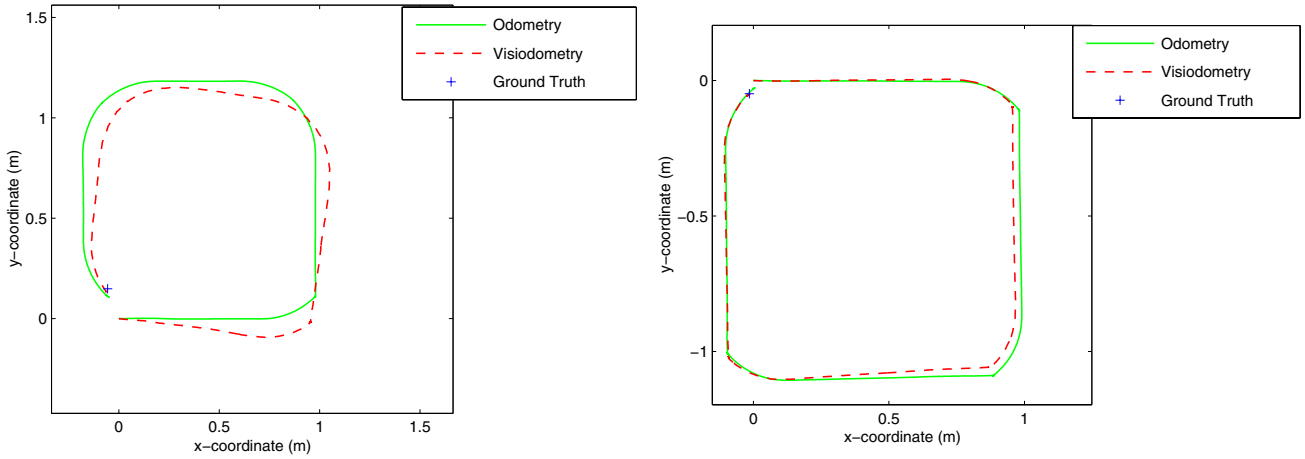



Fig. 5. **Ground Plots.** The position estimates from visiodometry are plotted against the position estimates computed from wheel odometry. Fig. 5(a) and Fig. 5(b) are plots from run-1 and run-2 respectively. The solid line represents the plot from wheel odometry and the dashed line the plot from visiodometry. The ground truth position at the end of the run is shown.

## VI. Conclusions and Future Works

In this paper a new exteroceptive sensor data input for relative localization was presented. The results demonstrate the feasibility of the proposed method for relative localization which avoids modelling the kinematics and is comparable to the resolution and accuracy of wheel odometry.

A basic calibration method using a 1D object and using known motions of the robot was described. An extension to phase correlation for estimating the planar transformations between images to a high precision and accuracy (which formed the basis of the proposed localization method), was also proposed.

**Future Work.** The use of the method to detect non-systematic errors and its applications to other types of mobile platforms (e.g. non-wheeled robots) to expand its application area is of interest. Improving the efficiency and accuracy of the method are also areas which require further investigation. Research on the accuracy of the method on different surfaces and degree of image blur will be useful in determining the envelope of use of the method.

## VII. Acknowledgment

The author is grateful to Prof. John Illingworth for the use of the Visual Media Lab (VML) of the Centre for Vision, Speech and Signal Processing (CVSSP) at the University of Surrey, UK. Thanks are due to Dr. Yi Liao for critiquing the draft manuscripts. The author is also grateful to Dr. Ilias Kolonias and Dr. Alexey Kostin for their assistance in setting up the equipment.

## References

[1] S. Lee, "Mobile robot localization using optical mice," in *Proceedings of IEEE International Conference on Robotics, Automation and Mechatronics*, December 2004, pp. 1192–1197.

[2] S. Lee and J.-B. Song, "Robust mobile robot localization using optical flow sensors and encoders," in *Procedings of IEEE Internationl Conference on Robotics and Automation*, April 2004, pp. 1039–1044.

[3] S. Baek, H. Park, and S. Lee, "Mobile robot localization based on consecutive range sensor scanning and optical flow measurements," in *Proceedings of IEEE International Conference on Advanced Robotics*, July 2005, pp. 17–22.

[4] C. Schroeter, H. J. Boehme, and H. Gross, "Extraction of orientation from floor structure for odometry correction in mobile robotics," in *German Association for Pattern Recognition (DAGM 2003)*, 2003, pp. 410–417.

[5] D. Nister, O. Naroditsky, and J. Bergen, "Visual odometry," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, vol. 1, July 2004, pp. 652–659.

[6] J. Campbell, R. Sukthankar, I. Nourbaksh, and A. Pahwa, "A robust visual odometry and precipice detection system using consumer-grade monocular vision," in *Proceedings of IEEE International Conference on Robotics and Automation*, April 2005, pp. 3421–3427.

[7] P. Torr, A. Fitzgibbon, and A. Zisserman, "The problem of degeneracy in structure from and motion recovery from uncalibrated image sequences," in *International Journal of Computer Vision*, vol. 32, 1999, pp. 27–45.

[8] H. Wang, K. Yuan, W. Zou, and Q. Zhou, "Visual odometry based on locally planar ground assumption," in *Proceedings of IEEE International Conference on Information Acquisition*, 2005, pp. 59–64.

[9] Y. Cheng, M. Maimone, and L. Matthies, "Visual odometry on the Mars Exploration Rovers," in *IEEE/RSJ Robotics and Automation Magazine*, vol. 13, June 2006, pp. 54–62.

[10] M. Zaman, A. Kadyrov, and J. Illingworth, "Odometric relative localisation using two cameras," in *Proceedings of Towards Autonomous Robotic Systems (TAROS)*. Imperial College, U.K., 2006, pp. 254–261.

[11] C. D. Kuglin and D. C. Hines, "The phase correlation image alignment method," in *Proceedings of IEEE International Conference on Cybernetics and Society*, 1975, pp. 163–165.

[12] Y. Keller, Y. Shkolnisky, and A. Averbuch, "The angular difference function and its application to image registration," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, June 2005, pp. 969–976.

[13] Z. Zhang, "Camera calibration with one-dimensional object," in *Proceedings of European Conference on Computer Vision*, 2002, pp. 161–174.

[14] P. Baker and Y. Aloimonos, "Calibration of a multi-camera network," in *Proceedings of Computer Vision and Pattern Recognition Workshop CVPR*, vol. 7, 2003, p. 72.

[15] X. Cao and H. Foroosh, "Camera calibration without metric information using 1D objects," in *Proceeding of International Conference on Image Processing ICIP*, vol. 2, 2004, pp. 1349–1352.

[16] D. Sardelis and T. Valahas, *Least Squares Fitting: Perpendicular Offsets*. http://library.wolfram.com/infocenter/MathSource/5292.