Robot Localization using Uncalibrated Camera Invariants *

Michael Werman[1] Subhashis Banerjee[2] Sumantra Dutta Roy[2] Maolin Qiu[1]

[1]Institute of Computer Science
Hebrew University, 91904 Jerusalem, Israel
Email: {werman,qiuml}@cs.huji.ac.il

[2]Department of Computer Science and Engineering
Indian Institute of Technology, New Delhi 110016, India
Email: {suban,sumantra}@cse.iitd.ernet.in

Abstract

We describe a set of image measurements which are invariant to the camera internals but are location variant. We show that using these measurements it is possible to calculate the self-localization of a robot using known landmarks and uncalibrated cameras. We also show that it is possible to compute, using uncalibrated cameras, the Euclidean structure of 3-D world points using multiple views from known positions. We are free to alter the internal parameters of the camera during these operations. Our initial experiments demonstrate the applicability of the method.

1 Introduction

Improving robot efficiency in advanced industrial automation requires the inclusion of sensory information at many levels, the most important of which is on-line task control. Range finders, proximity sensors or wrist force sensors can only address the sensing problem in a reliable way when the robot is already at its target. Visual sensors can be used on-line using the visual servoing approach. This technique, which consists of realizing an image-based feedback control loop is of increasing popularity. However, generally speaking, it is currently based on the use of a single camera and operates only on the premise that all the components are fully calibrated: camera, robot, and world. The related calibration procedures [14, 4] are tedious, non-flexible and expensive. Mobile robot navigation and docking require the robot to know where it is. This is usually achieved through some combination of odometry and landmark observations. The possibility of using visual landmarks viewed through uncalibrated cameras for self-localization will reduce the cost and enlarge the potential market of a robot system.

*Supported by Israel - India Scientific Cooperation 8970-1-97 and Espirit project 26247 - Vigor

Our main objective is to remove the calibration bottleneck in visually-guided robots. We address the problem of accurately positioning a robot, with up to 6 degrees of freedom, using only pictorial information from an uncalibrated camera. We are interested in localization even in the absence of an *a priori* pose estimate, obviating the need of using Kalman filtering or optical flow techniques to estimate odometry. We present a new method for self-localization based on image measurements which are invariant to the internal parameters of the camera. The invariant measures depend only on (i) the external parameters of the camera which relate the camera and world coordinate systems through rotations and translations, and (ii) the Euclidean coordinates (structure) of the landmarks. Given the landmarks, we can compute the external parameters of the camera from the image measurements. Conversely, given the robot position in a world coordinate system and the point correspondences in multiple views, we can compute the 3-D Euclidean structure of the points.

The problem of landmark-based position recognition was first formalized by Sugihara [11] as a computational geometry problem, assuming a completely calibrated camera. Recently, Quan and Lan [9] addressed the problem of finding linear algorithms for pose determination using calibrated cameras and known landmarks. Many robot localization methods rely on landmarks which are either artificially added to the environment or based on strong assumptions with respect to the environment [1, 7, 12]. For example, Krotkov [6] relies on the assumption that the environment is structured in such a way that vertical lines can be easily extracted as landmarks. This assumption is problematic in two ways. First, it places a restriction on the kinds of environments that can be explored, and second, it places a restriction on the pose of the camera. Basri and Rivlin [2] have also exploited the geometric behavior of landmarks in selected model images to provide navigation information. Exploiting an assumption of global invertibility of the imaging function and it being an affine projection, Nayar [8] has shown that subspace methods can provide accurate positional feedback in sufficiently constrained environments. A key assumption in his work is that each possible viewing position gives rise to a unique image. In a similar work, Thrun [13] derives a probabilistic approach to obtain a pose estimate using a neural net. In the works of both Nayar and Thrun, however, all significant variations in the set of possible images, including those due to lighting variations, must be explicitly sampled and encoded. In other work, it has been shown that localization can be achieved despite unanticipated illumination variations. That method can also deal with non-invertibility of the imaging transform, a problem that is typical in unconstrained environments.

Our approach is close to that of Dudek et al. [3, 10] which uses image landmarks to perform position estimation, but learns these landmarks from a preliminary traversal of the environment (i.e. an off-line mapping phase). They have to use the same camera on subsequent localization tasks as they do not compute an invariant but just use the feature image.

2 Image invariants with respect to camera internals

Consider the case when the image projection $\mathbf{m}^T = (u, v, 1)^t$ of a world point $\mathbf{P}^T = (X, Y, Z, 1)^t$ is described by the following model of imaging [4]

$$\lambda \mathbf{m} = \mathbf{A}[\mathbf{R} \mid \mathbf{t}]\mathbf{P} \tag{1}$$

where $\mathbf{A}$ is the matrix of camera internals given by

$$\mathbf{A} = \begin{bmatrix} f_x & 0 & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

$\mathbf{R}$ and $\mathbf{t}$ are the rotation and the translation between the world and the camera frames, and $[\mathbf{R} \mid \mathbf{t}]$ is the matrix of camera externals given by

$$[\mathbf{R} \mid \mathbf{t}] = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \end{bmatrix} \tag{3}$$

Rewriting Equation 1, we have:

$$\begin{array}{l} u = f_x \frac{\mathbf{r}_1\mathbf{P}}{\mathbf{r}_3\mathbf{P}} + u_0 \\ v = f_y \frac{\mathbf{r}_2\mathbf{P}}{\mathbf{r}_3\mathbf{P}} + v_0 \end{array} \tag{4}$$

Suppose we know three 3-D points, $\mathbf{P}_p^T = (X_p, Y_p, Z_p, 1)^T$, $p \in \{i, j, k\}$, and their images on the image plane, $\mathbf{m}_p^T = (u_p, v_p, 1)^T$, $p \in \{i, j, k\}$, then by eliminating the internals of the camera, we obtain

$$\begin{array}{l} K_{ijk} = \frac{u_i - u_j}{u_i - u_k} = \frac{\frac{\mathbf{r}_1\mathbf{P}_i}{\mathbf{r}_3\mathbf{P}_i} - \frac{\mathbf{r}_1\mathbf{P}_j}{\mathbf{r}_3\mathbf{P}_j}}{\frac{\mathbf{r}_1\mathbf{P}_i}{\mathbf{r}_3\mathbf{P}_i} - \frac{\mathbf{r}_1\mathbf{P}_k}{\mathbf{r}_3\mathbf{P}_k}} \\ J_{ijk} = \frac{v_i - v_j}{v_i - v_k} = \frac{\frac{\mathbf{r}_2\mathbf{P}_i}{\mathbf{r}_3\mathbf{P}_i} - \frac{\mathbf{r}_2\mathbf{P}_j}{\mathbf{r}_3\mathbf{P}_j}}{\frac{\mathbf{r}_2\mathbf{P}_i}{\mathbf{r}_3\mathbf{P}_i} - \frac{\mathbf{r}_2\mathbf{P}_k}{\mathbf{r}_3\mathbf{P}_k}} \end{array}, \tag{5}$$

in which K_{ijk} and J_{ijk} are image measurements that are functions of $\mathbf{R}$, $\mathbf{t}$ and $\mathbf{P}_p$, $p \in \{i, j, k\}$, and are independent of the internals of the camera. Thus the above equations can be re-written as the following constraints

$$\begin{array}{l} K_{ijk} = f_{ijk}(\mathbf{R}, \mathbf{t}, \mathbf{P}_i, \mathbf{P}_j, \mathbf{P}_k) \\ J_{ijk} = g_{ijk}(\mathbf{R}, \mathbf{t}, \mathbf{P}_i, \mathbf{P}_j, \mathbf{P}_k) \end{array} \tag{6}$$

The left hand sides of Equation 6 represent image measurement based on three points which are invariant to the camera internals. The right hand sides are nonlinear trigonometric expressions which are functions of only the camera externals and the structure (Euclidean coordinates) of the three points.

Given $n \geq 3$ control points in the 3-D world, we get $2(n-2)$ independent constraints from one view of the 3-D scene. Suppose the number of views is N, then the total number of the independent constraints is $2N(n-2)$.

2.1 Pose estimation using Euclidean landmarks

Now suppose that we know the Euclidean coordinates (X_i, Y_i, Z_i) of 5 points in the world coordinate system. Six independent invariant measurements give us six equations in terms of the six unknowns in $(\mathbf{R}, \mathbf{t})$. The six equations can be solved numerically for a complete pose estimation using an uncalibrated camera and known landmarks.

In the case of constrained planar motion, $\mathbf{R}$ has only one degree of freedom and $\mathbf{t}$ has two degrees of freedom. The total number of unknowns in such a case is three. Then four control points are sufficient for pose estimation.

It turns out that in some special cases it is possible to obtain closed-form or linear solutions to the pose estimation problem using the image invariants. We will discuss some of these special cases in Section 3.

2.2 3-D Euclidean reconstruction from known ego-motions

Conversely, suppose that we have a robot whose ego-motion and odometry can be known exactly from positional encoders and we know the $\mathbf{R}$ and $\mathbf{t}$ at N locations of the camera. With n control points we have $3n$ unknowns. In order to compute the Euclidean coordinates of these 3-D points, the relationship $2N(n-2) \geq 3n$ must be satisfied. Some possibilities of number of views N and number of control points n are $N=2$ and $n \geq 8$; $N=3$ and $n \geq 4$; $N=5$ and $n \geq 3$.

2.3 Interpolation of camera motion

In the case discussed in Section 2.2, suppose we do not know the $\mathbf{R}$ and $\mathbf{t}$ at one of the N locations of the camera, say the first. For the first camera station, we need at least 5 points to solve for the pose if the 3-D coordinates of the points were known. We can solve for the 3-D coordinates using the rest of the $N-1$ camera stations (as in Section 2.2). We substitute these values in the equations for the first camera station, and find out its pose (i.e., $2(N-1)(n-2) \geq 3n$ and $n \geq 5$).

2.4 Interpolation of the image measurements

Suppose we know the $\mathbf{R}$ and $\mathbf{t}$ for all stations but do not know the image measurements K's and J's corresponding to one of the stations, say the first. Had the 3-D coordinates been known, we could solve for the $2(n-2)$ K's and J's from the equations for the first camera station. Here again, we may solve for the 3-D coordinates from the rest of the $N-1$ camera stations (as in Section 2.2), and substitute these values in the equations for the first camera station. Thus, we have the inequality $2(N-1)(n-2) \geq 3n$.

In each of the above cases it is possible to use a larger number of control points to robustly estimate the parameters using a suitable nonlinear optimization technique.

3 Special cases

In what follows we discuss some of the special cases.

3.1 Rotation only about Z-axis

Let $\mathbf{X}_w\mathbf{Y}_w\mathbf{Z}_w$ be the world coordinate system. Consider restricted motion with rotation θ only about the $\mathbf{Z}_w$-axis. This would be the case of a robot moving on the ground with a special camera looking vertically up at the landmarks on the ceiling to facilitate localization.

Suppose there are three 3-D control points lying on the $\mathbf{X}_w\mathbf{Y}_w$ plane (ceiling) of the 3-D world coordinate system, denoted as $(X_i, Y_i, 0, 1)$, $(X_j, Y_j, 0, 1)$ and $(X_k, Y_k, 0, 1)$, with the corresponding image coordinates being $(u_i, v_i, 1)$, $(u_j, v_j, 1)$ and $(u_k, v_k, 1)$, respectively. The image projection is

$$\begin{array}{l} u_p = f_x \frac{X_p cos\theta + Y_p sin\theta + t_1}{t_3} + u_0 \\ v_p = f_y \frac{-X_p sin\theta + Y_p cos\theta + t_2}{t_3} + v_0 \end{array}, p \in \{i, j, k\}. \quad (7)$$

By eliminating the internals of the camera from Equation 7, we obtain

$$\begin{array}{l} \frac{u_i - u_j}{u_i - u_k} = \frac{(X_i - X_j)cos\theta + (Y_i - Y_j)sin\theta}{(X_i - X_k)cos\theta + (Y_i - Y_k)sin\theta} \\ \frac{v_i - v_j}{v_i - v_k} = \frac{(-X_i + X_j)sin\theta + (Y_i - Y_j)cos\theta}{(-X_i + X_k)sin\theta + (Y_i - Y_k)cos\theta} \end{array}. \quad (8)$$

Let $K = (u_i - u_j)/(u_i - u_k)$ and $J = (v_i - v_j)/(v_i - v_k)$. Let $\mathcal{P}_{pq}$ stand for $\mathcal{P}_p - \mathcal{P}_q$. We obtain

$$tan\theta = \frac{-KX_{ik} + X_{ij}}{KY_{ik} - Y_{ij}} = \frac{-JY_{ik} + Y_{ij}}{JX_{ki} - X_{ji}}. \quad (9)$$

We can easily compute θ (the rotation about $\mathbf{Z}$) from Equation 9. Once θ is obtained, the translation vector $\mathbf{t}$ can be computed as follows. Suppose we get another 3-D control point which is not on the plane $\mathbf{X}_w\mathbf{Y}_w$, say $(X_j, Y_j, Z_j, 1)$ (let us not stop at just painting the ceiling, but hang a few sticks as well). We compute K and J in the same way as in Equation 8. Then we have

$$\begin{array}{l} \frac{t_3}{Z_j} = \frac{X_i cos\theta + Y_i sin\theta + t_1 + K_{ik}}{K_{ik} + X_j cos\theta + Y_j sin\theta - (X_i cos\theta + Y_i sin\theta)} \\ \frac{t_3}{Z_j} = \frac{-X_i sin\theta + Y_i cos\theta + t_2 + J_{ik}}{J_{ik} - X_j sin\theta + Y_j cos\theta - (-X_i sin\theta + Y_i cos\theta)} \end{array} \quad (10)$$

where $K_{ik} = K(X_i cos\theta + Y_i sin\theta - (X_k cos\theta + Y_k sin\theta)) = KK_0$; and $J_{ik} = J(-X_i sin\theta + Y_i cos\theta - (-X_k sin\theta + Y_k cos\theta)) = JJ_0$. By equating the two equations in 10, and rearranging the terms, we arrive at

$$(b_j - d_j)t_1 + (c_j - a_j)t_2 = a_j d_j - c_j b_j \quad (11)$$

where $a_j = X_j cos\theta + Y_j sin\theta$; $b_j = -X_j sin\theta + Y_j cos\theta$; $c_j = X_i cos\theta + Y_i sin\theta - K_{ik}$; and $d_j = -X_i sin\theta + Y_i cos\theta - J_{ik}$. If we know more 3-D control points which are not on the $\mathbf{X}_w\mathbf{Y}_w$ plane, we can get more equations in the form of Equation 11. In such a case, a linear least squares technique can be used to solve for t_1 and t_2. Further, t_3 can be computed using the following relationship

$$t_3 = Z_j \frac{c_j + t_1}{a_j - c_j} = Z_j \frac{d_j + t2}{b_j - d_j} \quad (12)$$

Thus, in this special case, we can linearly solve the localization problem given three Euclidean landmarks on the $\mathbf{X}_w\mathbf{Y}_w$ plane and at least one landmark off the $\mathbf{X}_w\mathbf{Y}_w$ plane.

3.2 Motion and Structure from planar motion and rotation about Z

In the case described above, suppose we do not know the coordinates of the control points on the $\mathbf{X}_w\mathbf{Y}_w$ plane (the ceiling). Consider planar motion on a plane parallel to the $\mathbf{X}_w\mathbf{Y}_w$ plane and rotation only about $\mathbf{Z}_w$. By rearranging the terms in Equation 8 and letting $\mathcal{P}_{pq}$ stand for $\mathcal{P}_p - \mathcal{P}_q$, we obtain

$$\begin{array}{l} X_{ij} + Y_{ij} tan\theta - K_{ijk} X_{ik} - K_{ijk} tan\theta Y_{ik} = 0 \\ X_{ij} tan\theta - Y_{ij} - J_{ijk} tan\theta X_{ik} + J_{ijk} Y_{ik} = 0 \end{array}. \quad (13)$$

Now suppose we know the image coordinates of any 4 3-D points on the $\mathbf{X}_w\mathbf{Y}_w$ plane, which are in general position. We then have totally $2(4-2)=4$ equations as follows

$$\mathbf{M}(\theta)\mathbf{X} = \mathbf{0} \tag{14}$$

where $\mathbf{M}$ is a 4×8 matrix whose entries are dependent on $\tan(\theta)$ and the invariants K's and J's,

$$\mathbf{X} = \left[\begin{array}{cccccc} X_{12} & Y_{12} & X_{13} & Y_{13} & X_{14} & Y_{14} \end{array}\right]^t. \tag{15}$$

If the camera undergoes rotation with another angle θ', and we know the matching points of the first view, then we have a relationship similar to Equation 14

$$\mathbf{N}(\theta')\mathbf{X} = \mathbf{0} \tag{16}$$

Putting Equation 14 and Equation 16 together we obtain

$$\mathbf{A}(\theta,\theta')\mathbf{X} = \mathbf{0}, \mathbf{A}(\theta,\theta') = \left[\begin{array}{c} \mathbf{M}(\theta) \\ \mathbf{N}(\theta') \end{array}\right] \tag{17}$$

Equation 17 is a set of 8 nonlinear equations in $\mathbf{X}$, $tan(\theta)$ and $tan(\theta')$. Because $\mathbf{X}$ is nonzero vector, we have that $det(\mathbf{A}^T\mathbf{A}) = 0$, which is a polynomial equation in $tan(\theta)$ and $tan(\theta')$. Given a reference view with $\theta' = \theta_0$, e.g., $\theta' = 0$, which we can take as the orientation initialization of the 3-D world coordinate system, we can easily get any other rotation angle θ by solving the polynomial equation. If the initial orientation is not set, the rotation angle can only be determined up to a shift of any given angle.

Moreover, $\mathbf{X}$ is the eigenvector of the smallest eigenvalue of $\mathbf{A}^T\mathbf{A}$ and the Euclidean coordinates can be determined up to a scale assuming the first point to be the origin. The scale can be set by knowing any one length.

3.3 Planar motion and rotation about Y axis

Let $\mathbf{X}_w\mathbf{Y}_w\mathbf{Z}_w$ be the world coordinate system. Consider restricted motion with rotation θ only about the $\mathbf{Y}_w$-axis. This would be the case of a robot moving on the ground with a camera mounted horizontally (we look normally, at last). Consider three points in the world coordinate system $(X_i, Y_i, Z_i, 1)$, $(X_j, Y_j, Z_j, 1)$ and $(X_k, Y_k, Z_k, 1)$, with the coordinates of their images being $(u_i, v_i, 1)$, $(u_j, v_j, 1)$ and $(u_k, v_k, 1)$, respectively.

By projecting the 3-D control points on to the image plane, we have

$$\begin{array}{l} u_p = f_x \frac{X_p cos\theta + Z_p sin\theta + t_1}{-X_p sin\theta + Z_p cos\theta + t_3} + u_0 \\ v_p = f_y \frac{Y_p + t_2}{-X_p sin\theta + Z_p cos\theta + t_3} + v_0 \end{array}, p \in \{i,j,k\}. \tag{18}$$

Now, let the first two points i and j be on the $\mathbf{Y}_w$ axis and the third point k be on the $\mathbf{X}_w\mathbf{Y}_w$ plane. By eliminating the internal parameters of the camera from the above equation we obtain

$$\begin{array}{l} K_k = \frac{u_i-u_j}{u_i-u_k} = 0 \\ J_k = \frac{v_i-v_j}{v_i-v_k} = \frac{Y_i-Y_j}{(Y_i+t_2) - \frac{Y_k+t_2}{1-X_k\frac{sin\theta}{t_3}}} \end{array}. \tag{19}$$

If we consider another point l on the $\mathbf{X}_w\mathbf{Y}_w$ plane, we obtain, similarly,

$$\begin{array}{l} K_l = \frac{u_i-u_j}{u_i-u_l} = 0 \\ J_l = \frac{v_i-v_j}{v_i-v_l} = \frac{Y_i-Y_j}{(Y_i+t_2) - \frac{Y_l+t_2}{1-X_l\frac{sin\theta}{t_3}}} \end{array}. \tag{20}$$

Because K_k and K_l vanish, we can separate the terms of $sin\theta$ and t_3 to obtain

$$\begin{aligned} \frac{t_3}{sin\theta} &= \frac{X_k((1-J_k)Y_i - Y_j - J_k t_2)}{Y_i - Y_j + J_k(Y_k - Y_i)} \\ &= \frac{X_l((1-J_l)Y_i - Y_j - J_l t_2)}{Y_i - Y_j + J_l(Y_l - Y_i)} = T_s. \end{aligned} \tag{21}$$

t_2 can be easily found from Equation 21. In order to find θ, we have to know another control point lying on the plane $\mathbf{Y}_w\mathbf{Z}_w$, say m. Following exactly the derivation process of Equation 21, we can obtain a similar relationship as follows

$$\frac{t_3}{cos\theta} = -\frac{Z_m((1-J_m)Y_i - Y_j - J_m t_2)}{Y_i - Y_j + J_m(Y_m - Y_i)} = T_c. \tag{22}$$

Since we know t_2, T_s and T_c could be computed, then $tan\theta = T_c/T_s$ and $t_3 = T_s sin\theta = T_c cos\theta$.

In Equation 18, K_k vanishes because both 3-D points i and j are on the $\mathbf{Y}_w$ axis. K_k would not be zero if we choose 3 points in general position. Now that we know θ, t_2 and t_3, solving for t_1 is trivial. If three points $(X_i, Y_i, Z_i, 1)$, $(X_j, Y_j, Z_j, 1)$ and $(X_k, Y_k, Z_k, 1)$ are in general position, then the image measurement K_{ijk} does not vanish and we have

$$t_1 = \frac{K_{ijk}(\frac{b}{a} - \frac{f}{e}) - (\frac{b}{a} - \frac{d}{c})}{(\frac{1}{a} - \frac{1}{c}) - K_{ijk}(\frac{1}{a} - \frac{1}{e})}, \tag{23}$$

where $a = -X_i sin\theta + Z_i cos\theta + t_3$; $b = X_i cos\theta + Z_i sin\theta$; $c = -X_j sin\theta + Z_j cos\theta + t_3$; $d = X_j cos\theta + Z_j sin\theta$; $e = -X_k sin\theta + Z_k cos\theta + t_3$; and $f = X_k cos\theta + Z_k sin\theta$.

Thus, in this special case, we obtain a closed form solution to the localization problems with specially located landmarks - two on the $\mathbf{Y}_w$ axis, two on the $\mathbf{X}_w\mathbf{Y}_w$ plane and one on the $\mathbf{Z}_w\mathbf{Y}_w$ plane.

4 Results

In Figure 1, we show images of two positions of a calibration object obtained using a stationary Pulnix CCD camera. These two positions are kept a fixed distance apart ($200mm$), for calibration. Data from two such positions constitutes one data set. The world coordinates corresponding to both the positions of the calibration object are known and they serve as Euclidean landmarks. We refer to the world points (corners) on our calibration objects with numbers assigned according to a row major order. The calibration data was obtained using Tsai's method [14]. We assume that all correspondence information is known (hand picked).

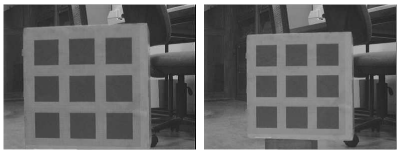

Figure 1: Calibration images

For the pose estimation problem (Section 2.1), Table 1 shows some sample results using 5 points. In Table 2, we show some results for recovering the 3-D structure of points from known egomotions(Section 2.2). We have considered N=4 and n=3. Figure 2 shows images taken from the 4 viewpoints around our calibration object. Table 3 shows some results for Section 2.3, for N=4 and n=5. Figure 2 shows the corresponding images taken from the camera stations around our calibration object. These were used for the experiments on 3-D and 2-D interpolation. In Table 4, we show results for 2-D interpolation(Section 2.4) for N=4 and n=5.

Calibration Data			Reconstructed Data		
X	Y	Z	X	Y	Z
500	0	200	495.884	3.501	197.995
700	70	0	704.152	69.4679	-1.348
500	100	200	507.501	98.997	203.334
670	0	200	669.324	0.190	199.888

Table 2: Structure estimation: Some sample results (error of objective function = -0.802)

For the first special case (Section 3.1), we have taken two camera stations, with two sets of 5 points each for each camera station. Table 5 summarizes the results. Figure 3 shows images from three such camera stations (Constrained rotation about the Z-axis only: We have kept our camera horizontal and rotated our calibration object). For the second special case (Section 3.2), we show some results in Table 6. Here, the first image of Figure 3 is taken was the reference camera station, while the other two are the stations numbered I and II, respectively. In Table 7, we show results for the third special case (Section 3.3). We have taken two sets of six points each, for three camera stations(shown in Figure 4).

Calibration Data			*Reconstructed Data*		
X	Y	Z	X	Y	Z
500	0	200	500.089	-0.012	200.039
700	70	0	701.062	70.362	-1.036
500	100	200	500.0873	99.849	200.039
670	0	200	670.978	0.0388	199.963
700	0	200	700.663	0.020	199.989

Calibration Data		*Reconstructed Data*	
R	**T**	**R**	**T**
-0.083	676.421	-0.083	606.801
0.046	14.203	0.0172	3.559
0.079	-1076.164	0.072	-1061.471

Table 3: 3-D Interpolation Results : error = 0.773, N=4, n=5

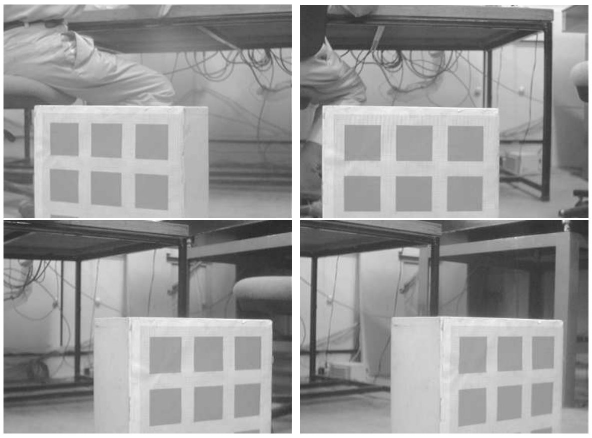

Figure 2: The 4 viewpoints used for interpolation experiments

5 Conclusion

We have described image measurements which are camera invariant but location variant. Using such uncalibrated image measurements we have presented a method of estimation of camera externals from known Euclidean landmarks. We have also shown that it is possible, using the uncalibrated image invariants, to compute the Euclidean structure of world points from multiple views with known camera positions. Our initial experiments show that the invariants are stable and

Item	*Rx*	*Ry*	*Rz*	*Tx*	*Ty*	*Tz*
Calib. Data	0.376	-4.795	0.649	-694.802	24.189	1041.215
Estimated Data	1.906	-6.518	0.509	-728.196	-17.798	1087.956
Calibration Data	0.003	-28.727	0.316	-434.520	18.000	822.988
Estimated Data	-0.252	-28.449	0.454	-434.542	17.971	822.989

Table 1: Pose estimation: Some sample results

Calibration Data			*Reconstructed Data*		
X	*Y*	*Z*	*X*	*Y*	*Z*
500	0	200	500.089	-0.012	200.039
700	70	0	701.062	70.362	-1.036
500	100	200	500.087	99.849	200.039
670	0	200	670.979	0.039	199.963
700	0	200	700.663	0.020	199.989

	Calibration Data	*Reconstructed Data*
K_{123}	29.001	28.984
J_{123}	0.664	0.676
K_{124}	1.315	1.309
J_{124}	5.187	4.988
K_{125}	1.114	1.109
J_{125}	4.536	4.363

Table 4: 2-D Interpolation Results : error = 0.774094, N=4, n=5

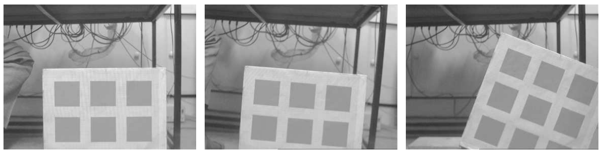

Figure 3: Special Cases 1 and 2: Images

it is indeed possible to compute the position information reliably.

The landmark based self-localization presented in this paper does not require camera calibration making it suitable for robot navigation applications where accurate calibration is not possible due to continually changing parameters and parts of the system. The method is also ideally suited for visual servoing using a feed-back loop.

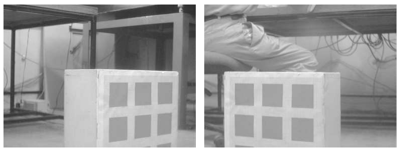

Figure 4: Special Case 3: Images

Camera Station I: Calibration Data: Tx = -634.112, Ty = 16.029, Tz = 1057.223, Rx = 1.080, Ry = 0.837, Rz = -0.317				
Points	*Rz*	*Tx*	*Ty*	*Tz*
5,10,15,40,48	-0.578	-654.555	15.526	934.162
8,12,19,39,44	-0.524	-648.851	39.923	1063.521
Camera Station II: Calibration Data: Tx = -633.105, Ty = -35.049, Tz = 1053.015, Rx = 0.831, Ry = 0.672, Rz = 4.701				
Points	*Rz*	*Tx*	*Ty*	*Tz*
6,13,18,25,37	4.975	-634.835	-29.289	990.971
4,15,20,27,48	2.047	-575.648	-37.7681	1062.035

Table 5: Pose estimation results, Special Case 1(Section 3.1) (All distances in mm, all angles in degrees)

The method can capitalize on the benefits of both image and landmark-based methods. Images can be encoded as a set of visual features called landmarks. Potential landmarks can be detected using an attention mechanism implemented as a measure of uniqueness. These can then be selected and represented by an appearance-based encoding. Localization can be performed using a landmark tracking and interpolation method which obtains an estimate accurate to a fraction of the environment sampling density. Localization can thus be a two-step process consisting of an off-line preprocessing stage and an on-line estimation stage. The off-line stage may consist of building a representation of the environment in the form of an invariants database, which is later used for positioning. The on-line stage can then use the database to match currently observed invariants to previously stored invariants using an interpolation method. These matches can then used to compute individual position estimates in a robust fashion.

We believe that that the camera invariant measurements described in this paper can be used in a variety of robotic applications.

Reference Camera Station : Calibration Data: Rx = 1.080, Ry = 0.837, Rz = -0.317		
Camera Station I: Calibration Data: Rx = 0.831, Ry = 0.672, Rz = 4.701		
Observation Set A: Points 9, 17, 19, 22		
Parameter	*Cal. Data*	*Observed*
Rz	4.701	5.005
X_{12} / Y_{14}	1.0	0.956
Y_{12} / Y_{14}	0.3	0.302
X_{13} / Y_{14}	-1.0	-0.942
Y_{13} / Y_{14}	1.0	0.998
X_{14} / Y_{14}	0.7	0.683
Camera Station II: Calibration Data: Rx = 1.146, Ry = 0.624, Rz = 21.009		
Observation Set B: Points 11, 14, 18, 20		
Parameter	*Cal. Data*	*Observed*
Rz	21.009	22.191
X_{12} / Y_{14}	-1.3	-1.239
Y_{12} / Y_{14}	0.3	0.309
X_{13} / Y_{14}	0.7	0.679
Y_{13} / Y_{14}	0.3	0.304
X_{14} / Y_{14}	-1.3	-1.234

Table 6: Pose estimation results, Special Case 2(Section 3.2) (All angles in degrees)

References

[1] D. Avis and H. Imai, "Locating a robot with angle measurements", *Journal of Symbolic Computation*, no. 10, pp. 311–326, 1990.

[2] R. Basri and E. Rivlin, "Localization and homing using combinations of model views", *Artificial Intelligence*, vol. 78, no. 1-2, pp. 327–354, October 1995.

[3] G. Dudek and C. Zhang, "Vision-based robot localization without explicit object models", in *Proc. IEEE Int. Conf. on Robotics and Automation*, 1996.

[4] O. Faugeras, "Three-Dimensional Computer Vision: A Geometric Viewpoint", The MIT Press, 1996.

[5] O. Faugeras, L. Quan and P. Strum, "Self-calibration of a 1D Projective Camera and its Application to the Self-Calibration of a 2-D Projective Camera", in *Proc. ECCV'98*.

[6] E. Krotkov, "Mobile robot localization using a single image", in *Proc. 1989 IEEE Int. Conf. on Robotics and Automation*, pp. 978–983. 1989.

[7] J. J. Leonard and H. F. Durrant-Whyte, "Mobile robot localization by tracking geometric beacons", *IEEE Transactions on Robotics and Automation*, vol. 7, no. 3, pp. 376–382, 1991.

Camera Station I: Calibration Data: Tx = 21.948, Ty=31.988, Tz=1034.709 Rx = -0.931, Ry = -39.155, Rz = 1.314				
Points	*Ry*	*Tx*	*Ty*	*Tz*
7,19,6,21,31,42	-30.843	138.861	14.758	883.234
7,13,4,22,37,47	-33.573	-633.137	20.163	957.999
Camera Station II: Calibration Data: Tx = -210.449, Ty = 18.132, Tz = 1195.758 Rx = 0.345, Ry = 23.262, Rz = 0.504				
Points	*Ry*	*Tx*	*Ty*	*Tz*
7,19,8,22,37,44	28.266	-99.024	47.190	1420.395
13,19,15,20,43,47	27.275	-153.946	45.356	1311.451

Table 7: Pose estimation results, Special Case 3(Section 3.3) (All distances in mm, all angles in degrees)

[8] S.K. Nayar, H. Murase, and S.A. Nene, "Learning, positioning, and tracking visual appearance", in *Proc. IEEE Conf on Robotics and Automation*, San Diego, CA, May 1994, pp. 3237–3246.

[9] L. Quan and Z. Lan, "Linear $N \geq 4$ Point Pose Determination", in *Proc. ICCV'98*, pp. 778–783.

[10] R. Sim and G. Dudek, "Mobile Robot Localization from Learned Landmarks", *Proc. IEEE/RSJ Conf. on Intelligent Robots and Systems (IROS)*, Victoria, BC, Oct. 1998.

[11] K. Sugihara, "Some location problems for robot navigation using a single camera", *Computer Vision, Graphics, and Image Processing*, vol. 42, pp. 112–129, 1988.

[12] K. T. Sutherland and W. B. Thompson, "Pursuing projections: Keeping a robot on path", in *Proc. IEEE Int. Conf. on Robotics and Automation*, San Diego, CA, May 1994, pp. 3355–3361, IEEE Computer Society Press.

[13] S. Thrun, "Finding landmarks for mobile robot navigation", in *Proc. IEEE Int. Conf. on Robotics and Automation*, Leuven, Belgium, May 1998, pp. 958–963.

[14] Roger Y. Tsai, "A versatile Camera Calibration Technique for High-Accuracy 3-D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", *IEEE Journal of Robotics and Automation*, Vol. RA-3, No. 4, August 1987, pp. 323–344.