Visual Navigation using Planar Homographies

Bojian Liang and Nick Pears

Department of Computer Science

University of York

York, YO10 5DD, UK

email: bojian@cs.york.a.uk, nep@cs.york.ac.uk

Abstract— **We introduce three new results, which allow homographies of the ground plane to support visual navigation functions for mobile robots using uncalibrated cameras. Firstly, we illustrate how, for pure translation, a homography can be computed from just two pairs of corresponding corner features. Secondly, we show how, for pure translation, we can determine the height of corner features above the ground plane using the recovered homography and a construct based on the cross ratio. This allows us to detect points which can be driven over, as their height is measured to be close to zero, and points which are sufficiently high to drive under. Finally, we show how, in the case of general planar motion, homographies can be used to determine the rotation of the camera and robot.**

Keywords— **Visual navigation, mobile robots, monocular vision, uncalibrated cameras.**

I. Introduction

The aim of our work is to develop new algorithms for visual navigation of mobile robots in indoor environments, with particular emphasis on achieving robustness of operation. Our focus on indoor environments means that planar regions in the scene will be common. In particular, floors which are planar to some approximation is a fundamental assumption. Apart from this ground planarity requirement, we impose no further restrictions and ultimately aim to be able to navigate in a broad range of indoor scenes.

This paper describes and experimentally validates three new results related to planar homographies. The first result, described in section III, relates to the computation of the homography itself using just two point pair correspondences. The next two results allow planar homographies to support mobile robot navigation functions using uncalibrated cameras. Section IV describes computation of height above the ground plane of corner features relative to the height of the camera optical centre, while section V shows how camera and robot rotation can be directly computed from planar homographies.

II. Corner based approaches

The structural information that can be extracted by tracking 2D features, such as corner points, is central to our mobile robot navigation system. Therefore, we briefly review two methods which use corner tracking (or correspondences) to elicit structural information.

A. Navigation using F

The most common approach used to track corner features through an image sequence is the so called "F-based tracker", where "**F**" represents the fundamental matrix. The fundamental matrix models the epipolar geometry between two views taken by uncalibrated cameras and the F-based tracker is an iterative process which simultaneously estimates **F** and and the correspondences consistent with that **F**. Once **F** is estimated, it may be used to reconstruct 3D position of the points in the scene up to an ambiguity of a projective transformation [1]. Furthermore, if camera parameters are approximately known, the projective skew can be "unwound" to give a "quasi-euclidean" structure which may be used for navigation purposes.

It is well known that estimation of F is degenerate when the scene points are coplanar. In indoor visual navigation problems many scenarios are encountered in which the ground plane is the dominant structural feature and there are only a minimal set of non ground plane features. For this reason we have aimed to use primarily planar homography (**H**) relations in our mobile robot navigation system.

B. Navigation using H

Early work on exploiting coplanar relations has been presented by Tsai and Huang [2], Longuet-Higgins [3] and Faugeras and Lustman [4]. We summarise the coplanar relation as follows: If a set of corner features in the scene lie in a plane, and they are imaged from two viewpoints, then the corresponding points in the two images (separated by k frames) are related by a plane-to-plane projectivity or planar homography, H, such that:

$$\lambda \mathbf{x_i} = \mathbf{H}\mathbf{x_{i-k}} \qquad (1)$$

where **x** represents a homogenous image coordinate $(x, y, 1)^T$, **H** is a 3 by 3 matrix representing the homography and λ is a scalar. Since this equation is valid

up to a scale factor, $\mathbf{H}$ has only eight degrees of freedom. Eight degrees of freedom requires that we have four corresponding coplanar features in general position (no three collinear), since each pair of corresponding points then provides two independent constraints, and $\mathbf{H}$ can be determined by standard linear methods.

Equation 1 suggests a method of grouping corner features into coplanar sets. Namely, if we can select a set of four coplanar corresponding point pairs which are in a sufficiently general configuration in both images (each point is unique and no three are collinear), then $\mathbf{H}$ can be computed and used to check whether other points in the scene lie in the same plane [5].

Recent work concerning planar homographies includes that of Gurdjos and Payrissat [6] which discusses the conditions for recovering the 3D structures of perpendicular planes in the scene, such as walls and the floor, using homographies. Ha and Kweon [7] proposed an approach for 3D structure recovery and camera calibration using planar homographies and known angles in the scene.

III. H FROM TWO CORNER MATCHES

We have used a four point method in a sample consensus approach to bootstrap our ground plane detection system. Once tracking of ground plane features is initiated, we always have a good estimate of which features are on the ground plane and which are not [5]. Here, we present a new result which allows us to compute $\mathbf{H}$, the ground plane homography, using just two ground plane corner matches, assuming that the robot undergoes pure translation.

This approach yields a number of benefits.
• We don't need to find four point correspondences in general position i.e. no three (near) collinear.
• As the vanishing line is fixed in planar motion, we don't need to compute the vanishing line at each iteration.
• We don't need to solve an overdetermined system to compute H at each iteration. If pure translation is maintained, we only need to compute a scalar to recover the H matrix as, in this case, the vanishing point and the vanishing line are both fixed.
• Once the vehicle changes direction and returns back to a pure translation motion, computation of the vanishing point can be applied both to check for pure translation and to recover the H matrix.

Consider two camera centered coordinate systems, frame 1 and frame 2, so that we can write

$$\mathbf{X}_2 = \mathbf{R}\,\mathbf{X}_1 + \mathbf{T}, \qquad (2)$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are the coordinates of the same 3D point, expressed in frames 1 and 2 respectively and where $\mathbf{R}$ and $\mathbf{T}$ are the rotation and the translation matrices encoding the relative position of the two coordinate systems. Now assume that $\mathbf{X}_1$ is a point on the plane defined by:

$$A\,X_1 + B\,Y_1 + C\,Z_1 + 1 = 0. \qquad (3)$$

This is a plane which does not pass through the origin (i.e. the optical center of the camera) and $\mathbf{N} = (A, B, C)^T$ is the plane normal. Thus we have $\mathbf{N}^T\mathbf{X}_1 = -1$ and denoting $\mathbf{T} = k\mathbf{t}$, where k is a scalar and $\mathbf{t}$ is a unit vector, we have:

$$
\begin{aligned}
\mathbf{X}_2 &= \mathbf{R}\,\mathbf{X}_{c1} - k\mathbf{t}\,N^T\,\mathbf{X}_1 \qquad (4)\\
&= (\mathbf{R} - k\mathbf{t}\,N^T)\,\mathbf{X}_1.
\end{aligned}
$$

The images of the scene point can be written as:

$$
\begin{aligned}
\mathbf{x}_2 &= \mathbf{P}(\mathbf{R} - k\mathbf{t}\,N^T)\mathbf{P}^{-1}\mathbf{x}_c \qquad (5)\\
&= \mathbf{H}\mathbf{x}_c.
\end{aligned}
$$

where $\mathbf{P}$ is the (unknown) camera model. For a pure translation, $\mathbf{R} = \mathbf{I}$, and so $\mathbf{H}$ has the form

$$
\begin{aligned}
\mathbf{H} &= \mathbf{P}(\mathbf{I} - k\mathbf{t}\,N^T)\mathbf{P}^{-1}\\
&= \mathbf{I} - k\mathbf{P}\mathbf{t}\,N^T\mathbf{P}^{-1}. \qquad (6)
\end{aligned}
$$

We note that $\mathbf{Pt}$ is the vanishing point, $\mathbf{v_p}$, and $N^T\mathbf{P}^{-1}$ is the vanishing line, $\mathbf{v_l}^T$, in the image. Thus, we have

$$\mathbf{H} = \mathbf{I} - k\mathbf{v_p}\mathbf{v_l^T} \qquad (7)$$

As shown in Fig-1, two corresponding point pairs fully define the vanishing line and the vanishing point. Given that we know the vanishing point and vanishing line, scalar k can be recovered by substituting any one known corresponding point pair and thus the H matrix can be recovered. From 7 we have

$$\mathbf{x}_2 = \mathbf{x}_1 - k\mathbf{v_p}\mathbf{v_l^T}\mathbf{x}_1 \qquad (8)$$

Since this equation is defined up to a scale factor we have

$$\lambda\mathbf{x}_2 = \mathbf{x}_1 - ks\mathbf{x}_t \qquad (9)$$

where $s\mathbf{x}_t = \mathbf{v_p}\mathbf{v_l^T}\mathbf{x}_1 = [sx_t, sy_t, s]^T$. Normalising homogenous vector $\mathbf{x}_2$ gives

$$x_2 = \frac{x_1 - ksx_t}{1 - ks}, \quad y_2 = \frac{y_1 - ksy_t}{1 - ks} \qquad (10)$$

Thus we have two estimates of the scalar k as

$$k_x = \frac{x_2 - x_1}{s(x_2 - x_t)}, \quad k_y = \frac{y_2 - y_1}{s(y_2 - y_t)} \qquad (11)$$
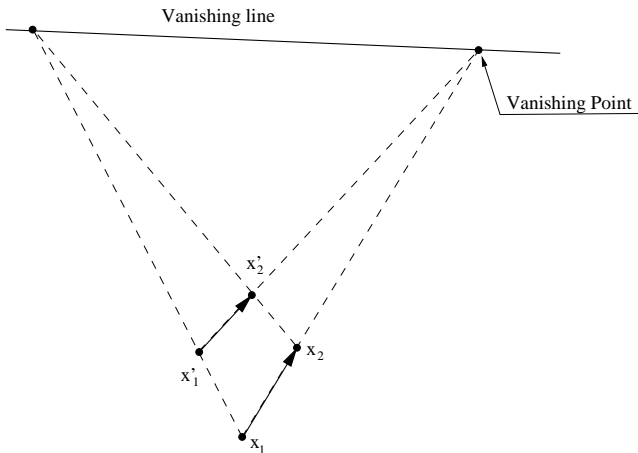
Fig. 1. Two corresponding point pairs fully define the vanishing point and the vanishing line.

Fig. 2. Computation of height of point A.

Now suppose there are n ($n \geq 2$) sets of corresponding point pairs, indexed as ($0 \leq i < n$), then a least squares fit can be applied to obtain the scalar k, as

$$k = \frac{1}{2n} \sum_{i=0}^{n-1} (k_{x_i} + k_{y_i}) \tag{12}$$

Once k has been computed, H can be recovered by Eq-7. Compared with using 4 point correspondences to compute H, this approach generates a "well formed" H matrix. By this we mean that it encodes a motion of pure translation and its eigenvectors are the points on the vanishing line. This is valuable in terms of 3D reconstruction relative to the ground plane.

In practice, the vanishing point can be computed by using all corner correspondences, not just those on the ground plane. Intersection of the two lines which join each pair of end points of the loci of the co-planar point pair is a point on the vanishing line (see fig 1). These intersection points can generate the horizon line using robust approaches such as RANSAC.

IV. Height above ground plane.

In this section, we present a method to compute the height of features above the ground plane in terms of units of the height of the camera optical center from the ground plane. This uncalibrated method assumes pure translation and uses the recovered H matrix and a construction based on the cross ratio.

Our aim is to recover the height of corner point A shown in figure 2, when the robot undergoes pure (forward) translation, t. Point A is the actual position of the corner point relative to the camera before the translation and point C is the position of the corner after the translation. Points A' and C' are the projections of these actual corner positions onto the ground
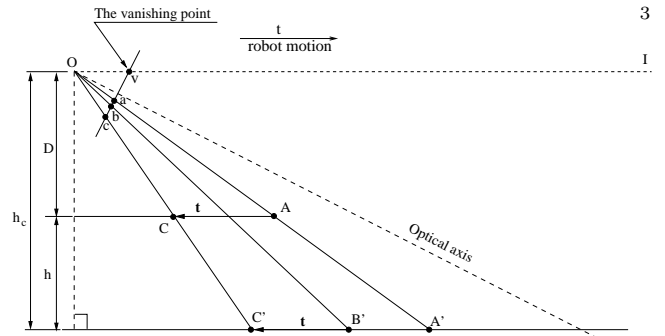
plane. Points a and c are the image positions of the corner at positions A and C respectively and b is the predicted image position of the corner point, if the corner point were to lie in the ground plane. Image point b is computed from the recovered H matrix as $\mathbf{b} = \mathbf{Ha}$.

Now the height of the corner point relative to the height of the camera optical centre is

$$h_r = \frac{h}{h_c} = 1 - \frac{D}{h_c} \tag{13}$$

Using similar triangles we note that:

$$\frac{D}{h_c} = \frac{OC}{OC'} = \frac{AC}{A'C'} \tag{14}$$

For pure translation, $d(A, C) = d(B', C')$, so that

$$h_r = 1 - \frac{B'C'}{A'C'} \tag{15}$$

Now, the four image points (a, b, c, v), where v is the vanishing point, and the corresponding four ground plane points (A', B', C', ∞) are collinear. The cross ratio for this set of points remains invariant under projection and so we can write:

$$\frac{d(B', C')}{d(A', C')} = \frac{d(b, c)\, d(a, v)}{d(a, c)\, d(b, v)}. \tag{16}$$

Where function $d(x, y)$ denotes the distance between points x and y. Hence we can compute relative height as:

$$h_r = 1 - \frac{d(b, c)\, d(a, v)}{d(a, c)\, d(b, v)} \tag{17}$$

This can be interpreted as the height of point A units of height h_c.

Note that this approach only needs the ground plane homography, H, and the tracked image correspondences a and c of the feature to determine the height above the ground plane. By thresholding the measured height above the plane, the method can be used to check for ground plane points, which can be driven
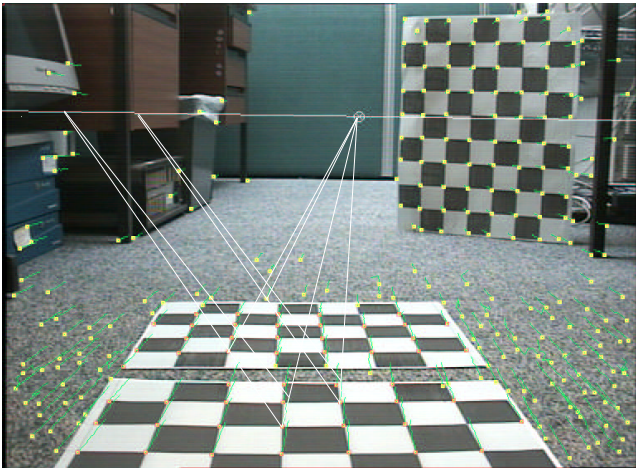
Fig. 3. Frame-11.

	Column							
	1	2	3	4	5	6	7	8
1	1.73	1.75	1.73	1.81	1.75	1.79	1.77	x
2	1.63	1.63	1.62	1.60	1.71	1.59	1.63	1.58
3	1.49	1.44	1.46	1.39	1.40	x	1.40	1.42
4	1.28	1.28	1.22	1.23	1.21	1.23	1.23	1.20
5	1.03	1.02	1.03	x	1.02	1.01	1.01	1.02
6	0.81	0.79	0.84	0.80	0.82	0.81	0.79	0.79
7	0.60	0.59	0.59	0.58	0.59	0.59	0.59	0.61
8	x	0.31	0.33	0.41	0.44	0.45	0.43	0.38
9	0.18	0.17	0.15	0.16	0.21	0.17	0.17	0.20
10	x	0.01	0.01	0.01	0.01	0.03	0.03	0.04

	Column							
	1	2	3	4	5	6	7	8
1	1.67	1.67	1.65	1.68	1.67	1.67	1.67	x
2	1.49	1.50	1.48	1.50	1.52	1.48	1.48	1.48
3	1.34	1.31	1.32	1.32	1.32	x	1.31	1.33
4	1.14	1.15	1.14	1.15	1.14	1.14	1.14	1.13
5	0.97	0.97	0.97	x	0.97	0.96	x	0.97
6	0.82	0.81	0.82	0.79	0.80	0.79	0.76	0.76
7	0.66	0.62	0.63	0.63	0.61	0.60	0.60	0.60
8	x	0.43	0.41	0.43	0.43	0.42	0.42	0.39
9	0.28	0.23	0.20	0.23	0.26	0.24	0.22	0.23
10	x	0.09	0.05	0.08	0.08	0.03	0.09	0.04

over, and for sufficiently high feature points which can be driven under. Note that this is achieved without camera calibration.

A. Height measurement validation

Fig-3 shows an example of height measurement from a pure translation image sequence. A chess board which has squares of dimension equal to 20% of the height of the camera optical center is placed in the front of the vehicle such that it is perpendicular to the ground plane. Two more boards are placed on the ground to obtain stable corner features for tracking. The recovered vanishing line and vanishing point are superimposed on the image and the H matrix is then computed from Eq-7. The heights of corners extracted on the chess board perpendicular to the ground are listed in table I for frame 11 and table II for frame 21. The columns are marked from left to right and the rows from top to the bottom. Ideally, the measurement in each row should increase by 0.2 units of the camera height. Note that, due to errors in the measured corner positions and the radial distortion near the image boundary, a number of measurements errors are observed. The mean, $\bar{x}_r$, and variance, s_r^2, for each row and the increment mean $\Delta\bar{x}_r$ between adjacent rows is shown in table III.

V. ROBOT ROTATION FROM H

A general planar motion includes both rotation and translation. In this case, the translation vector is always perpendicular to the plane normal and the axis of rotation (screw axis) is parallel to the plane normal. The H matrix of a planar motion generally has one real and two complex eigenvalues. We first prove that the real eigenvalue is unity and its corresponding eigenvector is the image of the intersection of the screw axis

and the plane of motion. We then show that the two complex conjugate eigenvalues have unity amplitude and their phase angles are equal to the actual angle of rotation that the robot (or camera) undergoes. Thus we can measure the robot's rotation, independent of any camera calibration.

First, a provisional coordinate system, C_{p1} is chosen such that its $x - y$ plane describes the *plane of motion* i.e the plane parallel to the ground plane, whose origin coincides with the optical centre i.e. the origin of a camera-centered coordinate system. Then, a general

	Frame 11			Frame 21		
Row	$\bar{x}_r$	s_r^2	$\Delta\bar{x}_r$	$\bar{x}_r$	s_r^2	$\Delta\bar{x}_r$
1	1.761	0.0008	0.137	1.669	0.0001	0.178
2	1.624	0.0014	0.195	1.491	0.0002	0.170
3	1.429	0.0012	0.194	1.321	0.0001	0.180
4	1.235	0.0008	0.215	1.141	0.0000	0.173
5	1.020	0.0001	0.214	0.968	0.0000	0.174
6	0.806	0.0003	0.214	0.794	0.0005	0.175
7	0.592	0.0001	0.199	0.619	0.0004	0.200
8	0.393	0.0026	0.217	0.419	0.0002	0.183
9	0.176	0.0003	0.156	0.236	0.0005	0.170
10	0.020	0.0001	0.000	0.066	0.0005	0.000

planar motion from C_{p1} to C_{p2} is given by,

$$\mathbf{X}_{p2} = \mathbf{R}_\theta \mathbf{X}_{p1} + \mathbf{t}, \qquad (18)$$

where $\mathbf{R}_\theta$ describes a rotation about the z axis over an angle θ and can be written as,

$$\mathbf{R}_\theta = \begin{bmatrix} cos\theta & sin\theta & 0 \\ -sin\theta & cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad (19)$$

and $\mathbf{t}$ is a translation matrix with the form

$$\mathbf{t} = (t_x, t_y, 0)^T. \qquad (20)$$

As the the ground plane does not pass through the origin and its unit normal is $\mathbf{a}_0 = (0, 0, 1)^T$, we can write equation 18 in the form of equation 5 to give:

$$\mathbf{X}_{p2} = (\mathbf{R}_\theta - k\mathbf{t}\mathbf{a}_0^T)\mathbf{X}_{p1}, \qquad (21)$$
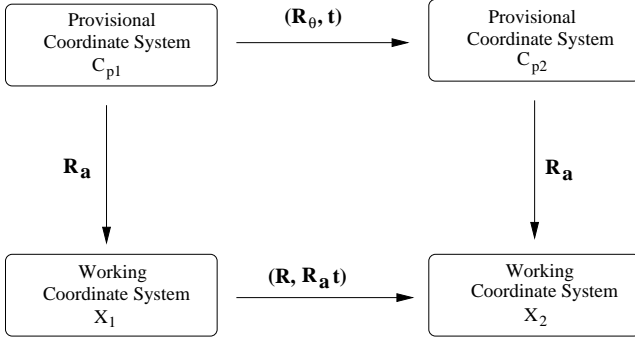

Fig. 4. Relations between provisional and working coordinate systems.

Fig-4 illustrates that the two camera-centered working coordinate systems, $\mathbf{X}_1$ and $\mathbf{X}_2$, which are the coordinate systems before and after the planar motion, are obtained by the same rotation $\mathbf{R}_a$ from their provisional coordinate systems, C_{p1} and C_{p2}. Thus, for all points on the ground plane, we have,

$$\begin{aligned} \mathbf{X}_2 &= \mathbf{R}_a(\mathbf{R}_\theta - k\mathbf{t}\mathbf{a}_0^T)\mathbf{R}_a^{-1}\mathbf{X}_1 \\ &= \mathbf{R}_a\mathbf{H}_1\mathbf{R}_a^{-1}\mathbf{X}_1, \end{aligned} \qquad (22)$$

where $\mathbf{H}_1 = \mathbf{R}_\theta - k\mathbf{t}\mathbf{a}_0^T$. Applying the (unknown) camera projection matrix $\mathbf{P}$, the image of points on the ground plane are given by.

$$\mathbf{x}_{c2} = \mathbf{P}\mathbf{R}_a\mathbf{H}_1\mathbf{R}_a^{-1}\mathbf{P}^{-1}\mathbf{x}_{c1}, \qquad (23)$$

Hence, the H matrix for a general planar motion is written as:

$$\mathbf{H} = \mathbf{P}\mathbf{R}_a\mathbf{H}_1\mathbf{R}_a^{-1}\mathbf{P}^{-1}, \qquad (24)$$

Clearly, $\mathbf{H}$ and $\mathbf{H}_1$ have the same eigenvalues and the eigenvectors, ξ, of $\mathbf{H}$ are related to the eigenvectors, ξ_1, of $\mathbf{H}_1$ by

$$\xi = \mathbf{P}\mathbf{R}_a\xi_1, \qquad (25)$$

It is straightforward to show that the eigenvalues of $\mathbf{H}_1$ are 1 and $e^{\pm i\theta}$. The eigenvector of $\mathbf{H}_1$ corresponding to the real (unity) eigenvalue is

$$\xi_{1a} = (\frac{1}{2}(t_x + t_y cot\frac{\theta}{2}), \frac{1}{2}(t_x cot\frac{\theta}{2} - t_y), \frac{1}{k}, \qquad (26)$$

for all $\theta \neq 0$. Geometrically, this can be interpreted as the image of the intersection of the screw axis and the plane of motion. The two conjugate eigenvectors corresponding to the complex eigenvalues are,

$$\xi_{1a,b} = (1, \pm i, 0)^T. \qquad (27)$$

These are the two circular points [8]. As the conjugate eigenvalues are invariant to projection, the angle of rotation can be obtained directly from the eigenvalues of the H matrix. There is no need to recover the camera parameters. Notice that the H matrix is defined up to a scale factor, and so the eigenvalues are also defined up to the the same scale factor. It is necessary to recover θ by the angle defined by both the real and imaginary parts of the eigenvalues. The image of the intersection point of the screw axis and the plane of motion combines with one corresponding point pair to define two equal length vectors which are the corresponding vector rotated around this intersection point. Hence, the sign of the rotation angle is defined by two corresponding vectors and this can be determined by evaluating the sign of their cross product.
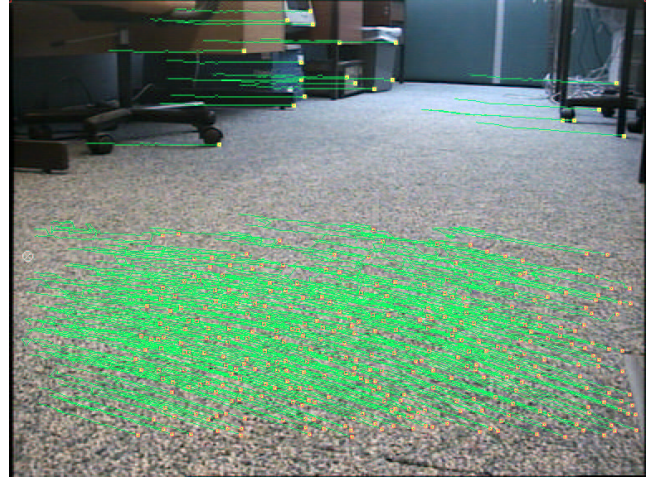

Fig. 5. Image of the 50th frame.

An example of the recovery of the rotation angle is now given. The movement of the tracked corners in
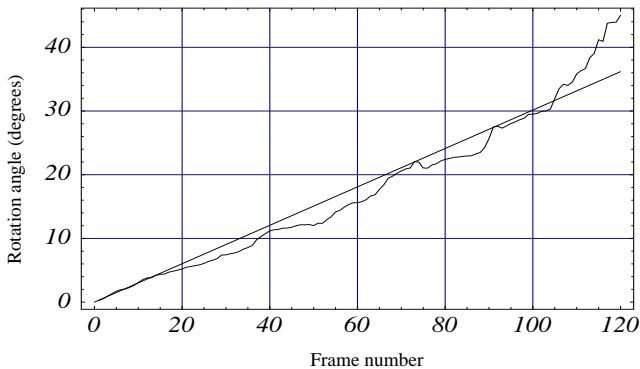
Fig. 6. Accumulated angle of the motion.

the experiment is shown in figure 5. Our mobile robot was programmed to move at a constant angular velocity and a set of corner features on the ground plane was extracted and tracked through the image sequence. (Note, however, that there are small variations in this angular velocity.) The H matrix of the each frame relative to the first frame was computed. The rotation angle from first frame to the current frame was then recovered and plotted in Fig-6. A least squares regression fit to the data is also plotted in the figure. We assess the method by checking the linear increment of the rotation angle and the final angle made by the vehicle positions before and after motion. The real angle was measured to be 41 degrees, and our measured angle was 45 degrees.

VI. Conclusions

For mobile robots operating in indoor environments, the ground plane is often the most dominant structural feature in the robot's field of view. Therefore, we have argued that using multi-view relations of the ground plane, in particular the planar homography, is a rational approach to take. We have shown how planar homographies can be extracted from two point pair correspondences. We have then shown how the height above the ground plane can be computed using homographies when the camera undergoes pure translation. For general planar motion, we have shown how to extract camera rotation from homographies. Our initial experimental results have validated that these techniques can support mobile robot visual navigation functions.

Acknowledgments

References

[1] Hartley R Gupta R and Chang T, "Stereo from uncalibrated cameras," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1992, pp. 761–764.

[2] Tsai R and Huang T, "Estimating three-dimensional motion parameters of a rigid planar patch," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 29, no. 6, pp. 1147–1152, 1981.

[3] Longuet-Higgins H C, "The reconstruction of a plane surface from two perspective projections," *Proc. Royal Society London*, vol. B227, pp. 399–410, 1986.

[4] Faugeras O and Lustman F, "Motion and structure from motion in a piecewise planar environment," *Int. Journ. Pattern Recognition and Artificial Intelligence*, vol. 2, no. 3, pp. 485–508, 1988.

[5] Pears N.E. and Liang B, "Ground plane detection in mobile robot visual navigation," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2001, pp. 1513–1518.

[6] P Gurdjos and R Payrissat, "About conditions for recovering the metric structure of perpendicular planes from the single ground plane to image homography," in *Int. Conf. Pattern Recognition*, 2000, pp. 358–361.

[7] J. E. Ha and I.S. Kweon, "3d structure recovery and calibration under varying intrinsic parameters using known angles," *Pattern Recognition*, vol. 34, no. 2, pp. 351–359, 2001.

[8] R Hartley and A Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge, 2001.