Available online at www.sciencedirect.com
SCIENCE DIRECT
Image and Vision Computing 24 (2006) 1301–1312
image AND vision COMPUTING
www.elsevier.com/locate/imavis

Monocular obstacle detection using reciprocal-polar rectification

Zezhi Chen *, Nick Pears, Bojian Liang

Department of Computer Science, University of York, York YO10 5DD, UK

Received 26 July 2004; received in revised form 10 February 2006; accepted 9 April 2006

Abstract

Our obstacle detection method is applicable to deliberative translation motion of a mobile robot and, in such motion, the epipole of each image of an image pair is coincident and termed the focus of expansion (FOE). We present an accurate method for computing the FOE and then we use this to apply a novel rectification to each image, called a reciprocal-polar (RP) rectification. When robot translation is parallel to the ground, as with a mobile robot, ground plane image motion in RP-space is a pure shift along an RP image scan line and hence can be recovered by a process of 1D correlation, even over large image displacements and without the need for corner matches. Furthermore, we show that the magnitude of these shifts follows a sinusoidal form along the second (orientation) dimension of the RP image. This gives the main result that ground plane motion over RP image space forms a 3D sinusoidal manifold. Simultaneous ground plane pixel grouping and recovery of the ground plane motion thus amounts to finding the FOE and then robustly fitting a 3D sinusoid to shifts of maximum correlation in RP space. The phase of the recovered sinusoid corresponds to the orientation of the vanishing line of the ground plane and the amplitude is related to the magnitude of the robot/camera translation. Recovered FOE, vanishing line and sinusoid amplitude fully define the ground plane motion (homography) across a pair of images and thus obstacles and ground plane can be segmented without any explicit knowledge of either camera parameters or camera motion.

Keywords: Homography; Fundamental matrix; Obstacle avoidance; Segmentation; Reciprocal-polar rectification; Image rectification

1. Introduction

Various visual cues have been employed to facilitate navigational functions with uncalibrated cameras. These include navigation down corridors, both by using the focus of expansion of non-vertical scene lines [1] and wide field peripheral flow [2]. Other approaches have been used such as time-to-contact from image divergence [3], combination of central flow divergence and peripheral flow [4], and quantitative planar region detection using point correspondences [5]. Most of these techniques work in some types of scene, but will fail when a particular type of feature is not well supported within the image data. Perhaps the most common approach used is to track corner features through an image sequence, based on recovery of the fundamental matrix (**F**-matrix). The **F**-matrix models the epipolar geometry between two views taken by uncalibrated cameras and an iterative process can be used to simultaneously estimate the **F**-matrix and the correspondences consistent with **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 [6–10]. Furthermore, we can use the **F**-matrix to detect and track a small number of independently moving objects [11] and to determine whether or not a collision will occur [12]. Early work based on coplanar relations has been presented by Tsai and Hung [13], Longuet-Higgins [14] and Faugeras and Lustman [15]. More recently, Lhuillier and Quan [16,17] have developed a quasi-dense approach to surface reconstruction from a sequence of uncalibrated images. The method gives a more robust and accurate geometry estimation, using fewer images than other approaches.

Previously, we have presented mobile navigation methods in indoor environments based on multiple visual cues, such as colour, texture and region boundaries [18,19]. Our research is particularly aimed at highly robust visual navigation, using a single standard CCD camera mounted on a mobile robot. We envisage that this robot can make deliberative (near) pure translation motions, which greatly simplifies the **F**-matrix and **H**-matrix (homography matrix) structure and suggests robust mechanisms for the estimation of these matrices.

The essence of this paper is the application of the reciprocal-polar (RP) image rectification process and the use of sinusoidal models for planar segmentation within that space. This

* Corresponding author.
E-mail addresses: chen@macs.hw.ac.uk (Z. Chen), nep@cs.york.ac.uk (N. Pears), bojian@cs.york.ac.uk (B. Liang).

0262-8856/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2006.04.007

approach is novel, although related ideas have been presented in the literature. For example, Pollefeys et al. [20] suggested a polar rectification to aid stereo matching. Wolberg and Zokai [21] have used the well-known log-polar transformation to aid affine motion recovery. Both of these techniques allow more general motion but do not give the main benefit of the RP transformation, which allows correlation based matching over large camera motions. If intensity correlation is used then acceptable results can be obtained even if there are few or no corner features on the plane of interest (ground plane). Furthermore, we never encounter degenerate configurations, when implementing our method. In the absence of ground plane texture or local intensity variations, we cannot segment on a pixel-by-pixel basis, as we cannot locate a strong maximum in the correlation process. In this case, we extract contours of smooth regions, which are then matched along epipolar lines and classified as ground or obstacle using the extracted sinusoid model of ground plane motion.). Finally, we note that, in our method, obstacles and the ground plane can be segmented, without any explicit prior knowledge of either camera parameters or camera motion. The obvious restriction to our algorithm is that we require (near) pure translation. In certain circumstances, such as hand-held camera applications, this may be over restrictive, but when camera motion can be carefully controlled, as in mobile robot navigation, such motions can be deliberate and used to probe the environment.

The paper is structured as follows. In Section 2, we describe the relationship between the homography and the fundamental matrix under pure translation. In Section 3, we make an incremental contribution to the accurate estimation of the focus of expansion (FOE) or epipole. Section 4 presents the main results of the paper, where the analysis suggests the use of a sinusoidal motion model in RP image space to simultaneously segment/group ground plane pixels and recover the vanishing line of the ground plane and hence **H**-matrix representing ground plane motion. Section 5 describes the obstacle detection technique for a monocular mobile robot under pure translation. Section 6 validates the method through experimentation, first using simple point correspondences and then using 1D intensity correlation, which is where the power of the method lies. Also, it is shown that in terms of reprojection errors of the recovered homography, the method performs as well as the state-of-the-art. In Section 7, conclusions are drawn.

2. The relation of H and F under pure translation

There are two relations between two views of a 3D plane: (1) through the epipolar geometry. This is applicable to a general 3D scene, in which a point in one view determines a line in the other, termed the epipolar line; (2) through the planar homography [22–24]. This is specific to 3D scene planes, in which a point in one view determines a point in the other which is the image of the intersection of the original ray with the 3D plane.

In pure translation of the camera, how are these related? One may consider an equivalent situation in which the camera is stationary, and the world undergoes a translation $-\mathbf{t}$. In this situation, points in 3D space move on straight lines parallel to **t**, and the imaged intersection of these parallel lines is the focus of expansion (FOE) **v** in the direction of **t**. It is evident that **v** is also the epipole **e** and $\mathbf{e}'$ for both views and the imaged parallel lines (now radial in the image with respect to the FOE) are epipolar lines. The fundamental matrix for pure translation has a special skew-symmetric form such that

$$\mathbf{F} = [\mathbf{e}']_\times = [\mathbf{e}]_\times = [\mathbf{v}]_\times \tag{1}$$

where for a general three-vector, $\mathbf{a}=[a_1, a_2, a_3]$, then its skew-symmetric matrix is:

$$[\mathbf{a}]_\times = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}$$

Let $\mathbf{X}_i$ be a set of points that are coplanar in 3D Euclidean space. The images of $\mathbf{X}_i$ from two view points are related by a plane to plane projectivity or homography **H**, such that:

$$\lambda \mathbf{x}_i' = \mathbf{H}\mathbf{x}_i \tag{2}$$

Here, λ is a scalar, $\mathbf{x}_i$ and $\mathbf{x}'_i$ are homogenous image coordinates of the images of point $\mathbf{X}_i$ and **H** is a 3×3 matrix representing the homography. As homogenous coordinates are defined up to a scale factor, the **H**-matrix has only eight degrees of freedom (*dof*), and it can be determined by standard linear methods of four corresponding point pairs in general position (no three collinear). When the number of point pairs is more than four, a standard least square method or SVD (singular value decomposition) method can be used.

Suppose the cameras are calibrated with the origin of the world coordinate system at the first camera and the intrinsic parameters (**K**) constant. If the cameras are separated by a rotation (**R**) and translation (**t**), the projection matrices for each camera (position) are then:

$$\mathbf{P} = \mathbf{K}[\mathbf{I}|0], \qquad \mathbf{P}' = \mathbf{K}[\mathbf{R}|\mathbf{t}] \tag{3}$$

If the world plane π_E has normal, n, and distance to origin, d, so that its coordinates are $\boldsymbol{\pi}_E = (\mathbf{n}^{\mathrm{T}}, d)^{\mathrm{T}}$, then

$$\mathbf{H} = \mathbf{K}(\mathbf{R} - \lambda \mathbf{t}\mathbf{n}^{\mathrm{T}})\mathbf{K}^{-1} \tag{4}$$

where $\lambda = 1/d$, For a pure translation, $\mathbf{R}=\mathbf{I}$, and so **H** has the form:

$$\mathbf{H} = \mathbf{I} - \lambda(\mathbf{K}\mathbf{t})(\mathbf{K}^{-\mathrm{T}}\mathbf{n})^{\mathrm{T}} \tag{5}$$

We note that $\mathbf{K}\mathbf{t}$ is the FOE $\mathbf{v} = \eta\begin{pmatrix} x_f & y_f & 1 \end{pmatrix}^{\mathrm{T}}$, and $\mathbf{K}^{-\mathrm{T}}\mathbf{n}$ is the vanishing line $\mathbf{l} = v\begin{pmatrix} a_v & b_v & 1 \end{pmatrix}^{\mathrm{T}}$ corresponding to the plane π_E. Thus, we have

$$\mathbf{H} = \mathbf{I} - k\mathbf{v}\mathbf{l}^{\mathrm{T}} \tag{6}$$

where k is a constant scalar. Since two corresponding point pairs fully define the FOE and vanishing line, the **H**-matrix can, in theory, also be fully determined by two corresponding matches (see Fig. 1).

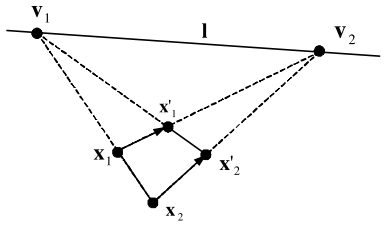

Fig. 1. Two corresponding point pairs fully define the points $\mathbf{v}_1$, $\mathbf{v}_2$ and vanishing line $\mathbf{l}$.

3. Accurate estimation of the FOE

Although robot and hence camera motion can be intentionally translational, we prefer to detect (near) pure translation by monitoring image motion, due to the potential for robot wheel slip. The simplest way to implement this is by intersecting all lines defined by all corner correspondences from the image pair. If a large proportion of intersections lie in a small area (for example, 85% of intersections should lie within a 50 pixels radius), then pure translation is assumed. However, this is not optimal and we implement a more sophisticated procedure. Under the assumption of Gaussian measurement noise, the maximum likelihood estimate (MLE) of the FOE and line segments is computed by determining a set of lines that do intersect in a single point and which minimize the sum of squared orthogonal distance from these lines. This minimization may be computed numerically using the Levenberg–Marquardt algorithm [25,26]. The question now is: how can we calculate the FOE with high accuracy and stability? A robust algorithm for estimating the FOE is summarized in Table 1.

4. Ground plane segmentation and ground plane motion recovery.

Once the FOE has been computed, we shift image coordinates so that each image is centred on the FOE $\mathbf{v} = \begin{pmatrix} x_f & y_f & 1 \end{pmatrix}^{\mathrm{T}}$. Let:

$$\mathbf{T}_c = \begin{bmatrix} 1 & 0 & -x_f \\ 0 & 1 & -y_f \\ 0 & 0 & 1 \end{bmatrix} \tag{7}$$

After translation $\mathbf{T}_c$ is applied, the FOE is at homogenous coordinates $\mathbf{v}_c = (0,0,1)^{\mathrm{T}}$ and vanishing line becomes $\mathbf{l}_c = \mathbf{T}_c^{-\mathrm{T}}\mathbf{l} = (a_v, b_v, \mathbf{v}^{\mathrm{T}}\mathbf{l})^{\mathrm{T}}$. The homography relating points in FOE centred coordinates is

$$\mathbf{H}_c = \mathbf{I} - k\mathbf{v}_c\mathbf{l}_c^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -ka_v & -kb_v & (1 - k\mathbf{v}^{\mathrm{T}}\mathbf{l}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ s & \mu & q \end{bmatrix} \tag{8}$$

Table 1
A robust method to estimate the FOE

(1) *Extract interest points*. Compute interest points in each image by using the Plessey–Harris corner detector [27] or KLT algorithm [28] or SUSAN [29] method
(2) *Putative correspondences*. Compute a set of interest point matches based on proximity and similarity of their intensity neighbourhood
(3) *RANSAC robust estimation*. Repeat for m samples, where m is determined adaptively by using a binning technique [6]
(a) Select a random sample of at least two correspondences and compute the FOE by using the simultaneous equations

$$\begin{cases} (\mathbf{x}_1 \times \mathbf{x}_1') \bullet \mathbf{v} = 0 \\ (\mathbf{x}_2 \times \mathbf{x}_2') \bullet \mathbf{v} = 0 \\ \quad \vdots \\ (\mathbf{x}_n \times \mathbf{x}_n') \bullet \mathbf{v} = 0 \end{cases}, \text{ where } \mathbf{x}_i \leftrightarrow \mathbf{x}_i' \quad (i = 1,2,\ldots,n) \text{ is any pair of matching}$$

points in two images
(b) Calculate the epipolar distance $f(\mathbf{v})$ for each putative correspondence $f(\mathbf{v}) = \left(\frac{1}{\sqrt{(\mathbf{F}\mathbf{x}_i)_1^2 + (\mathbf{F}\mathbf{x}_i)_2^2}} + \frac{1}{\sqrt{(\mathbf{F}^{\mathrm{T}}\mathbf{x}_i')_1^2 + (\mathbf{F}^{\mathrm{T}}\mathbf{x}_i')_2^2}} \right) \left| \mathbf{x}_i'^{\mathrm{T}} \mathbf{F} \mathbf{x}_i \right|$ where $\mathbf{F}$ is the fundamental matrix and the subscripts 1,2 denote vector components
(c) Compute the number of inliers consistent with $\mathbf{v}$, using the number of correspondences for which $f(\mathbf{v}) < threshold$. Choose the FOE with the largest number of inliers
(4) *Optimal estimation*. Re-estimate the FOE from all correspondences classified as inliers, by minimizing the object function $f(\mathbf{v})$
(5) *Repeat* steps (3)–(4) until the number of correspondences is stable or a maximum number of iterations is reached

where $q = 1 - k\mathbf{v}^{\mathrm{T}}\mathbf{l}$, $s = -ka_v$ and $\mu = -kb_v$. The homography $\mathbf{H}_\mathbf{c}$ has a very simple form and corresponding points $\mathbf{x}_\mathbf{c}$, $\mathbf{x}'_\mathbf{c}$, in FOE centered images are related as follows:

$$\lambda \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ s & \mu & q \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{9}$$

If the robot's translation direction is parallel to the ground plane, as is the case for normal mobile robot operation, $q = 1$, since the FOE lies on the vanishing line for these translation directions. In this case, the original $\mathbf{H}$-matrix has four *dof*, and is sometimes termed an elation [22]. Otherwise, the FOE is at a distance D from the vanishing line, where

$$D = \left| \frac{1 - q}{k\sqrt{a_v^2 + b_v^2}} \right| \tag{10}$$

and the five *dof* two-view planar relation is termed a homology. Under the new (FOE centered) homography, we have:

$$(sx_c' + \mu y_c' + q)^2 \left(x_c'^2 + y_c'^2 \right) = \left(x_c'^2 + y_c'^2 \right) \tag{11}$$

If we define $\rho = 1/r$, where r is the Euclidean distance between an image point and the FOE in a frame, then taking square roots of Eq. (11) yields the main result of this paper (showing that image motion in RP-space lies on a sinusoidal manifold)

$$f(\theta) = \rho - q\rho = s\frac{x_c'}{r} + \mu\frac{y_c'}{r} = k_{s\mu}\sin(\theta + \alpha) \tag{12}$$

where θ is the angular position of a pixel in a frame centred on the FOE and:

$$k_{s\mu} = \sqrt{s^2 + \mu^2}, \quad \sin\alpha = \frac{s}{k_{s\mu}}, \quad \cos\alpha = \frac{\mu}{k_{s\mu}}, \quad \tan\alpha = \frac{a_v}{b_v} \tag{13}$$

Eq. (12) indicates that we need to find three constants (q, $k_{s\mu}$, α) in order to recover the homography defining the ground plane image motion and that the computation should be implemented in (ρ, θ) image space (note that a planar homology has five *dof*, but two have been recovered in the FOE computation). We call this space reciprocal-polar (RP) image space. Thus, after computing the FOE, a linear interpolation procedure is used to generate an RP rectified image for each image in the original image pair. Fig. 5(a)–(d) show an image pair and their corresponding RP image pair. These images appear highly distorted, as the horizontal lines in the RP image represent radial (epipolar) lines centred on the FOE in the original image, but inverted, such that a point at r from the FOE in the original image is mapped to a horizontal position proportional to $1/r$ in the RP image. The second (vertical) dimension of the RP image corresponds to θ in the original image, which is the orientation of a FOE centred radial line (i.e. epipolar line).

In many mobile robot applications q may be taken as 1, but this parameter may be estimated if the translation is not parallel to the ground plane, for example, in detection of obstacles on a runway as an aircraft lands. In this case, along any radius from the FOE, $f(\theta)$ is constant, so for any two pairs of correspondences (i, j):

$$q = \frac{\rho_i - \rho_j}{\rho_i - \rho_j} \tag{14}$$

Values of ρ in above equation can be determined by 1D windowed correlation between the two images in RP image space, i.e. along lines of constant θ (horizontal lines) in the RP image space and q is obtained by using all the strong correlation results.

For the time being, lets assume that $q=1$. This allows correlations to be made along the ρ dimension of the RP image pair for each angle θ_i. For each pixel in image 1, its position in RP image space is computed, and a 1D window is created around this position along the (horizontal) ρ dimension. We then correlate this window along the ρ dimension in RP image 2, at the same value of θ. This correlation process is possible because of the 'pure-shift' relation, expressed in Eq. (12) and the position of the maximum value of the correlation is related as a value of $f_i(\theta)$.

Eq. (12) indicates that correlation maxima and feature correlations in RP space, which are associated with a planar surface, lie on a sinusoid and the constants ($k_{s\mu}$, α) can recovered by fitting a sinusoid to the data for $f(\theta)$.

Suppose that we have two values of $f(\theta)$, f_i, f_j measured at two angles, θ_i, θ_j, so that

$$f_i = k_{s\mu}\sin(\theta_i + \alpha), \qquad f_j = k_{s\mu}\sin(\theta_j + \alpha) \tag{15}$$

$$\frac{f_i}{f_j} = \frac{\sin\theta_i + \cos\theta_i \tan\alpha}{\sin\theta_j + \cos\theta_j \tan\alpha} \tag{16}$$

collecting terms in $\tan\alpha$ and rearranging gives:

$$\tan\alpha = \frac{f_j \sin\theta_i - f_i \sin\theta_j}{f_i \cos\theta_j - f_j \cos\theta_i} \tag{17}$$

Thus, in theory, a pair of f values, at different angular positions, for pixels belonging to the same plane, allows us to estimate the orientation of the vanishing line of that plane. Then, given the phase angle, α, corresponding to the orientation of the vanishing line, we can compute $k_{s\mu}$ from Eq. (15).

In order to robustly and accurately estimate the vanishing line orientation from many correlation maxima and feature correspondences in RP space, many of which will not be associated with the ground plane, a RANdom SAmple Consensus (RANSAC) method [30] and iterated least-squares process are used. We define an optimization object function as

$$\phi(\delta) = \sum_{i,j} \left| \delta - \frac{f_j \sin\theta_i - f_i \sin\theta_j}{f_i \cos\theta_j - f_j \cos\theta_i} \right| \tag{18}$$

where $\delta = \tan\alpha$ can be used to minimize $\phi(\delta)$ and determine the best set of inliers in the $f(\theta)$ data to a putative sinusoid. In this way, co-planar pixels may be grouped without explicit construction of a homography matrix, although this is easily recovered in the FOE centered frame from the FOE and the two parameters of the sinusoid.

Let:

$$f_i = k_{s\mu}\sin(\theta_i + \alpha) = s\cos\theta_i + \mu\sin\theta_i \tag{19}$$

Thus, for the all n inliers of the sinusoid model, we can write

$$\mathbf{AZ} = \mathbf{b} \tag{20}$$

where

$$\mathbf{A} = \begin{bmatrix} \cos\theta_1 & \sin\theta_1 \\ \vdots & \vdots \\ \cos\theta_n & \sin\theta_n \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} s \\ \mu \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} f_1 \\ \vdots \\ f_n \end{bmatrix}$$

$$\mathbf{Z} = (\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{b} \tag{21}$$

The original homography matrix (non-FOE centered) can then be explicitly expressed as

$$\mathbf{H} = \mathbf{T}_c^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ s & \mu & q \end{bmatrix} \mathbf{T}_c \tag{22}$$

It is useful to compute this homography, so that we can compare our method with other homography estimation techniques, in terms of noise sensitivity. A summary of the recovery of the H-matrix using RP rectification is given in Table 2. Note that the vanishing line of the ground plane also gives a good visual check of the quality of the homography, when overlaid on the original images. This can be determined from the FOE and the phase of the sinusoid, which represents

Table 2
Recovery of **H** matrix

Compute the FOE
Shift image coordinates so that each image is centred on the FOE
Convert both images from Cartesian space to RP space
Determine the best set of inliers to a putative sinusoid using RANSAC
Get a highly accurate vanishing line orientation of the ground plane by estimation of the phase of the sinusoid
Obtain the ground plane motion H matrix by formula (22)

the orientation of the vanishing line, or alternatively from Eq. (6), we have

$$k\mathbf{v}\mathbf{l}^{\mathrm{T}} = \mathbf{I} - \mathbf{H} \tag{23}$$

If the *j-th* row of the matrix $\mathbf{I}-\mathbf{H}$ is denoted by $h^{j\mathrm{T}}$, then we may write:

$$k\begin{pmatrix} x_f\mathbf{l}^{\mathrm{T}} \\ y_f\mathbf{l}^{\mathrm{T}} \\ \mathbf{l}^{\mathrm{T}} \end{pmatrix} = \begin{pmatrix} \mathbf{h}^{1\mathrm{T}} \\ \mathbf{h}^{2\mathrm{T}} \\ \mathbf{h}^{3\mathrm{T}} \end{pmatrix} \tag{24}$$

Then we can get the vanishing line

$$\mathbf{l}^{\mathrm{T}} = \mathbf{h}^{3\mathrm{T}} \tag{25}$$

5. Segmentation of obstacles from the ground plane

Although, we have discussed explicit reconstruction of a homography matrix to represent ground plane motion, this is not strictly necessary to detect obstacles, as we can leave the ground plane motion explicitly modelled as a sinusoid in RP space. In this case, we define residual error as

$$r = |(\rho' - q\rho) - k_{s\mu}\sin(\theta + \alpha)| \tag{26}$$

In order to classify a point (corner or pixel) we threshold the above metric and if $r <$ threshold (threshold $= 0.00025$ in our experiments), then that point is deemed to lie on the ground plane. Otherwise the point is classified as part of an obstacle.

6. Experimental results

A large amount of synthetic data and real images were selected and intensive experimental work was carried out in order to test the robustness and accuracy of the method proposed in this paper. Only selections of our experimental results are presented here: one simulated image and some real images.

6.1. Simulated experimental results

The simulated experiment was carried out on a 3D Euclidean model, consisting of 234 points (Fig. 2). The intrinsic simulated camera parameters were fixed as follows: aspect ratio of one with no skew, $f/\mathrm{d}x = f/\mathrm{d}y = 500$, principal point at (225, 225) pixels, and the image size is 512×512 pixels. Fig. 3 shows the images taken both when the robot's translation direction is parallel and not parallel to the ground. Note that the FOE positions and vanishing lines are also shown in this figure.

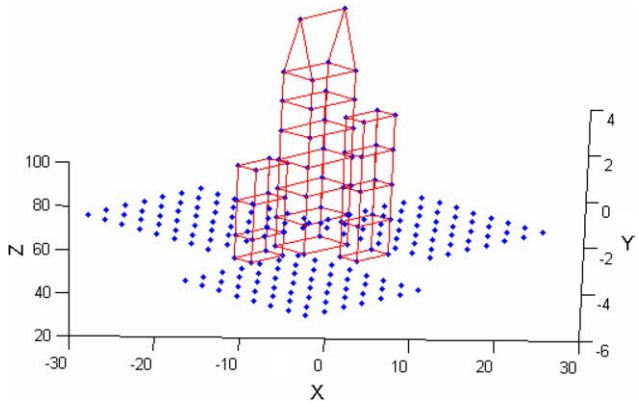

Fig. 2. A synthetic 3D scene.

In order to analyse the influence of noise on the RP algorithm, several methods were utilized to compute the **H**-matrix such as the pseudo-projective transform homography transformation algorithm (PPTH) [31]; the linear normalized transformation (LNT) method and the minimization reprojection error (MRPE) method [22]. For each method, the variation of the reprojective error was computed using the symmetric transfer error (STE) formula

$$d = \sum_{i=1}^{N} d(\mathbf{x}_i, \hat{\mathbf{x}}_i)^2 + d(\mathbf{x}'_i, \hat{\mathbf{x}}'_i)^2 \tag{27}$$

where $\hat{\mathbf{x}}_i = \mathbf{H}^{-1}\mathbf{x}'_i$ and $\hat{\mathbf{x}}'_i = \mathbf{H}\mathbf{x}_i$, $d(\mathbf{x},\mathbf{x}')^2 = (x-x')^2 + (y-y')^2$. The results under different Gaussian noise conditions are shown in Fig. 4(a). For all the tests that were performed, we added Gaussian noise with zero mean, and variance from zero (meaning no noise) to 10 pixels-squared. Note that all 180 of the feature points on the ground plane are used. Fig. 4(b) shows the variation of reprojective error with the same Gaussian noise condition, zero mean and the variance equal to 5.0 pixels-squared. The number of matching points used is varied from 2 to 180 and all of the computed STE values are the average values computed from 10,000 executions of the main simulation loop. Fig. 4(a) shows that if we have enough matching points, the results of the RP method, the LNT method and the MRPE method are almost the same. The MRPE method is more accurate than the LNT method, but it cannot always converge to a solution. Fig. 4(b) shows that if there are only a limited number of matching points on the ground plane (in particular, less than 40 points), the RP method is the best method to use.

6.2. Real images experimental results and applications

In the first experiment, $q = 1$ was assumed, since the camera was moved parallel to the ground plane. Fig. 5(a) and (b) shows the original images. After the FOE was obtained, the images were then converted to RP (ρ, θ) form, as shown in (c) and (d). Fig. 5(e) shows us one of the sinusoidal forms of $f(\theta)$ for a fixed

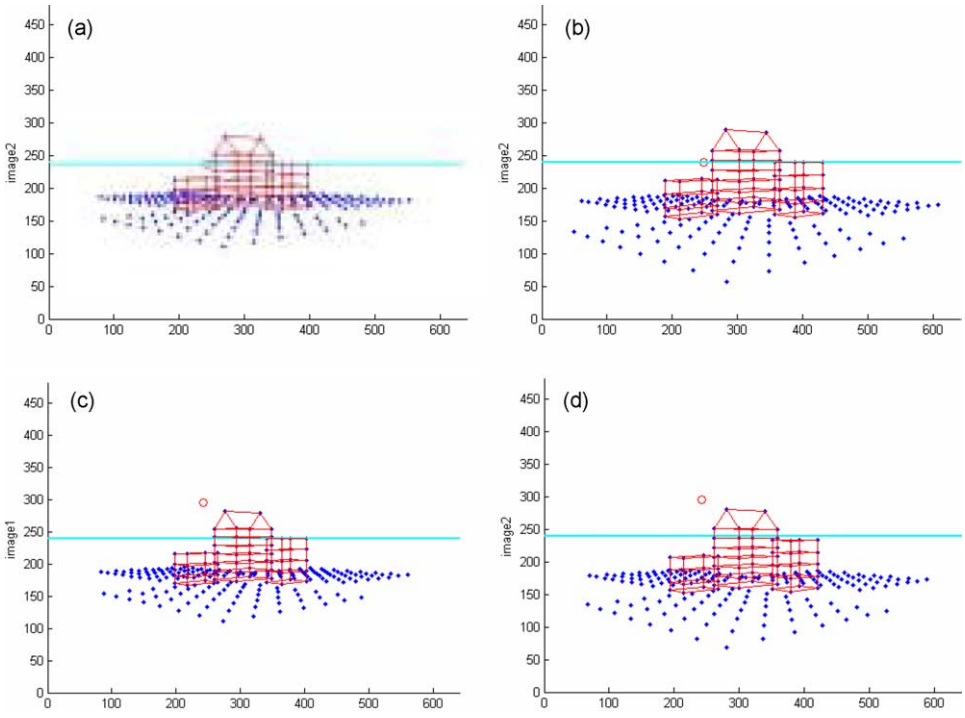

Fig. 3. (a) and (b) Robot's translation direction is parallel to the ground. (c) and (d) Robot's translation direction is not parallel to the ground.

ρ and hence fixed r. The partial sinusoidal curve ($\theta \in [193.25, 315.00]$) is clearly shown and represents the motion of the ground plane in RP space. The phase is shown close to 180° rather than 0° because the direction of y in the image is directed upwards from the FOE rather than downwards. The main result of the paper, showing that ground plane motion, plotted over RP image space, lies on a 3D sinusoidal manifold, is shown in Fig. 5(f). Here, the value of the function relates to image motion in RP space, i.e. the pure shift described by Eq. (12). Note that when we recover the amplitude and phase of this sinusoid, we use the data over the full range of (reciprocal) radii, as shown in this 3D plot. We can compute the **H**-matrix, if required, using Eqs. (21) and (22). Furthermore, the vanishing line can also be obtained from the Eq. (25), which is useful for overlaying on the original images to check the quality of the recovered ground plane homography. Fig. 6 shows one of the images with matching points (·), the FOE (o) and the vanishing line. In this set of experiments, the number of coplanar matching points varied from 54 to 112. The mean of residual errors of coplanar points is 1.558×10^{-5}, the standard deviation is 1.393×10^{-5} and the maximum value is 6.3×10^{-5}. But the mean of the residual errors of non-coplanar points is 0.00101, the standard deviation is 0.00183 and the maximum value is 0.011. The variation of residual errors of matching points is shown in Fig. 7.

The reprojection errors of several methods are shown in Fig. 8, which shows us that the accuracy of the RP method is very similar to the LNT method and much better than the other two methods. Note that the vanishing lines of the other three methods are not correct, since the FOE should lie on the vanishing line for this particular situation. Quantitative results of our experiments are given in Table 3. Finally, ground plane points and points that lie on obstacles can be segmented by using Eq. (26), as shown in Fig. 9. Ground plane points and

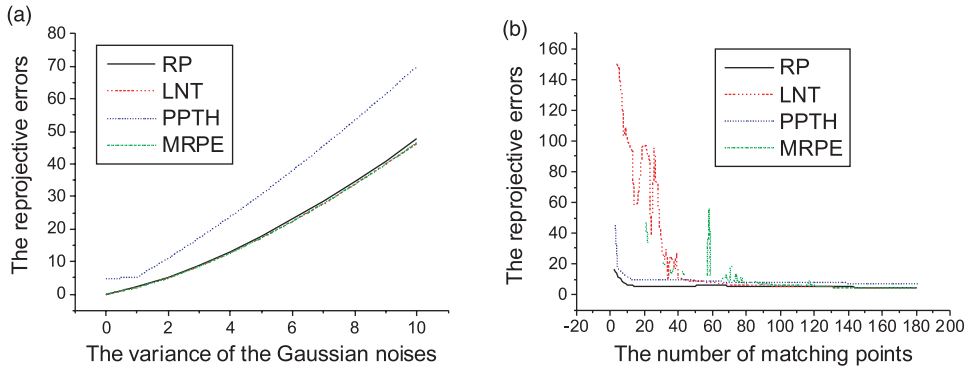

Fig. 4. The variation of reprojective errors computed under different Gaussian noise conditions and different number of matching points. (a) Gaussian noise with zero mean and variance from 0 to 10 pixels-squared. (b) All image points include the same Gaussian noise with zero mean and variance 5.0 pixels-squared.

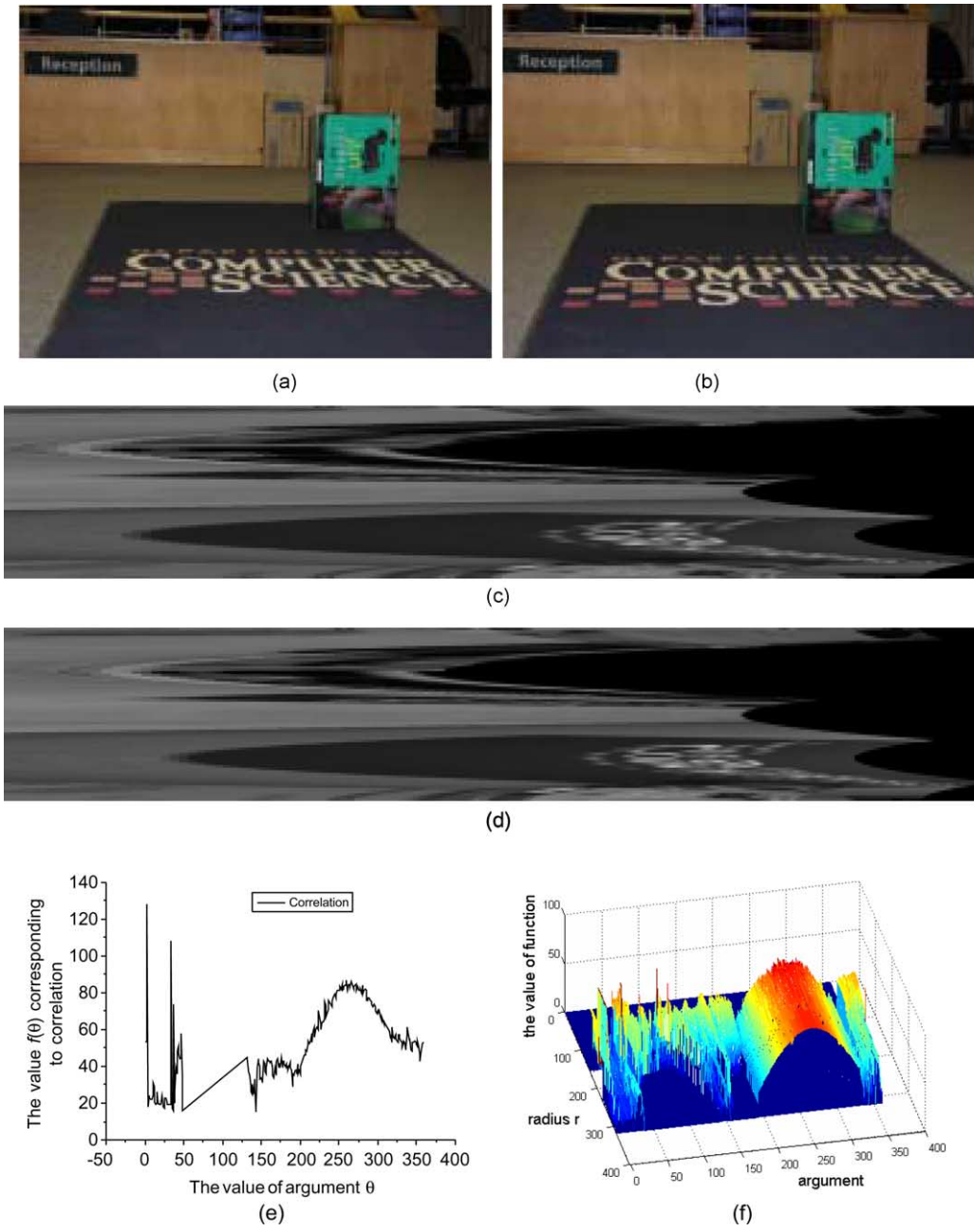

Fig. 5. (a) and (b) are the original images; (c) and (d) are the images in the RP space corresponding to (a) and (b), respectively; (e) shows the value of $f(\theta)$ at fixed radius. (f) shows $f(\theta)$ over a range of radii and angles and clearly shows the ground plane motion stand out as a 3D sinusoid.

points, which lie on potential obstacles are marked as $(\cdot)$ and $(+)$, respectively.

In the second experiment, the camera was moved in a forward translation mode, but in a direction not parallel to the ground. The camera motion is inclined downwards towards the ground plane. In this situation, at least two matching points are needed to calculate q and μ. The H-matrix can then be obtained as:

$$\mathbf{H} = \begin{bmatrix} 1.0000 & -0.1522 & 35.9500 \\ 0 & 0.8343 & 39.1454 \\ 0 & -0.0005 & 1.1294 \end{bmatrix}$$

Also, we computed the FOE as $\mathbf{v}=(277.7252, 302.4104)$ and the vanishing line as $\mathbf{l}=(0, 0.0005, -0.1294)$. Some correspondences, feature tracks, the vanishing line and the FOE (o) are shown in Fig. 10. The segmented coplanar points $(\cdot)$ and non-coplanar points $(+)$ are shown in Fig. 11.

The final experiment presented in this paper uses correlations within locally textured regions and contour matching to determine whether smooth (textureless) regions should be grouped with the ground plane. Note that an additional process is required, not described in this paper, which is our own region segmentation algorithm, which extracts homogenous regions of colour–texture and their

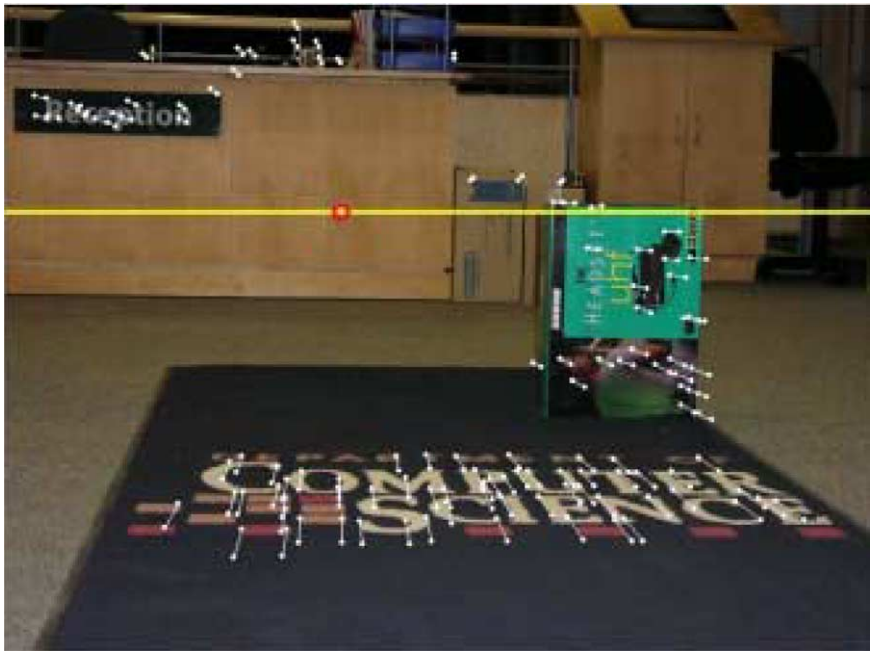

Fig. 6. The matching points, the moving track of feature points and the FOE.

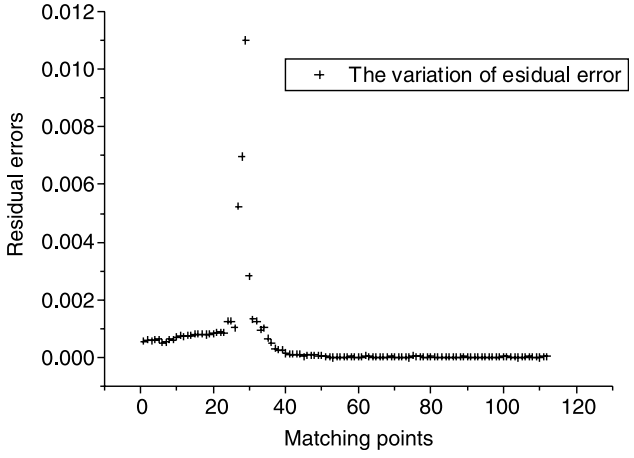

Fig. 7. The variation of residual errors (the number of coplanar points is between 54 and 112).

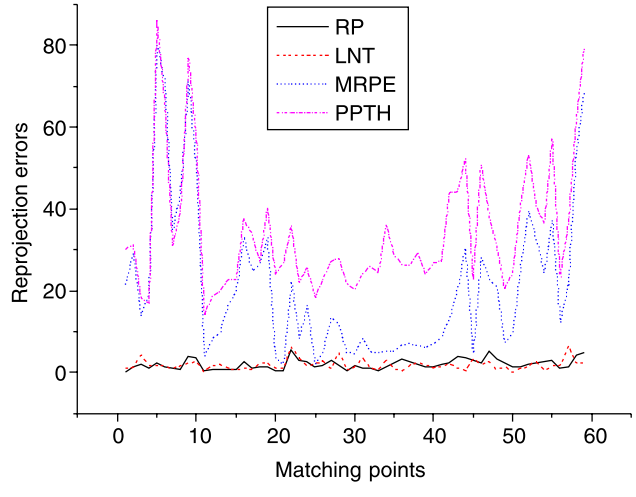

Fig. 8. Reprojection errors of several methods.

boundaries. Smooth (textureless and featureless) regions cannot be classified as ground plane or non-ground plane as they cannot be matched across an image pair. Their boundaries, however, can be and, in the case of pure translation, this matching is easily done by 'casting' rays (epipolar lines) from the FOE (epipole) recovered from all corner matches (note that there may be few or even no corner correspondences on the ground!). Fig. 12(a) shows an image with two regions on the floor, which have little texture. The first is a circular piece of white paper, which can be driven over, and the second is a small cardboard box, which cannot. The boundaries of these regions are extracted and the FOE is used to cast a ray (epipolar line) in order to match points along corresponding boundaries using an adaptive windowing technique [32]. If the motion of all matched boundary points falls within the threshold of the sinusoid model in RP space, it is classified as belonging to the ground plane, otherwise the region is classified as an obstacle. Fig. 12(b) shows the extracted ground region, where the textured carpet has been classified on a pixel by pixel basis, and the smooth white paper region has been included by virtue of its boundary motion being consistent with ground plane motion in RP space. A second example of pixel-based segmentation is shown in Fig. 12(c) and (d). Note that there are some 'drop outs' in the foreground of the image, but the shape of the segmentation is excellent to the extent that even the small black doorstop to the centre right of the original image, Fig. 12(c), has been correctly classified as an obstacle and removed in the image, Fig. 12 (d).

Table 3
The distance between the FOE and the vanishing line (pixels)

Methods	RP	LNT	MRPE	PPTH
Distance	-9.1054×10^{-14}	3.2690×10^{4}	1.9652×10^{5}	2.4590×10^{4}

Fig. 9. The coplanar points (·) and the points lie on potential obstacles (+).

Fig. 13 shows some more obstacle detection results in which a rectangular bounding box for 'obstacle region contours' is determined and highlighted with a coloured translucent overlay on one of the original images in the image pair.

7. Conclusions

We have presented a practical method for monocular mobile robot obstacle detection, which requires, as input, a pair of uncalibrated images, with viewpoints separated by a (near) pure translation. The main contributions in this paper are: (1) a robust method for estimating the FOE was developed yielding an accurate result. (2) We have presented a novel reciprocal-polar (RP) image rectification, which transforms planar image motion under pure translation into a pure shift, irrespective of the degree of perspective distortion of the planar surface. Hence correlation can be done over large translations, when correlation in the original image space would fail. Since the method is correlation based, corner matches on the ground plane are not a necessity. (3) We have shown that across the RP

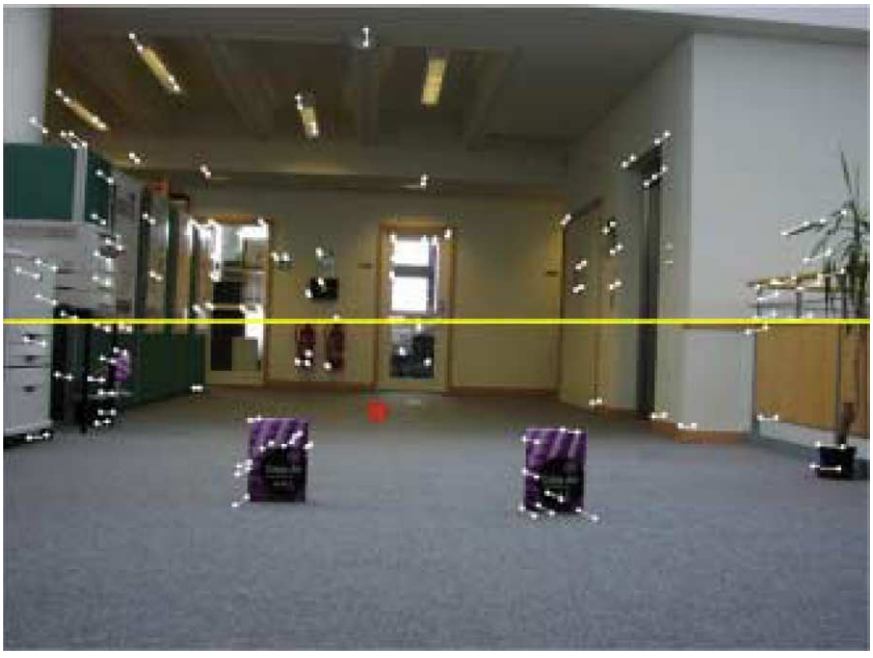

Fig. 10. Some correspondences, moving tracks (small line segment), the vanishing line and the FOE (o).

Fig. 11. The segmented coplanar points (·) and non-coplanar points (+).

image pair, the magnitude of the image motion follows a sinusoidal form along the θ direction over a maximum of π radians. Simultaneous planar pixel grouping and recovery of the planar homography thus amounts to accurately finding the FOE, and then robustly fitting a sinusoid to shifts of maximum correlation in RP space. The phase of the recovered sinusoid corresponds to the orientation of the vanishing line of the plane and the amplitude is related to the magnitude of the camera translation. (4) Intensive experimental work was carried out in order to test the accuracy of the method proposed in this paper. The results show that our algorithm performs very well to outliers and noise and the stability, accuracy and robustness performs favourably to other methods in terms of the reprojection errors of the recovered homography. (5) We

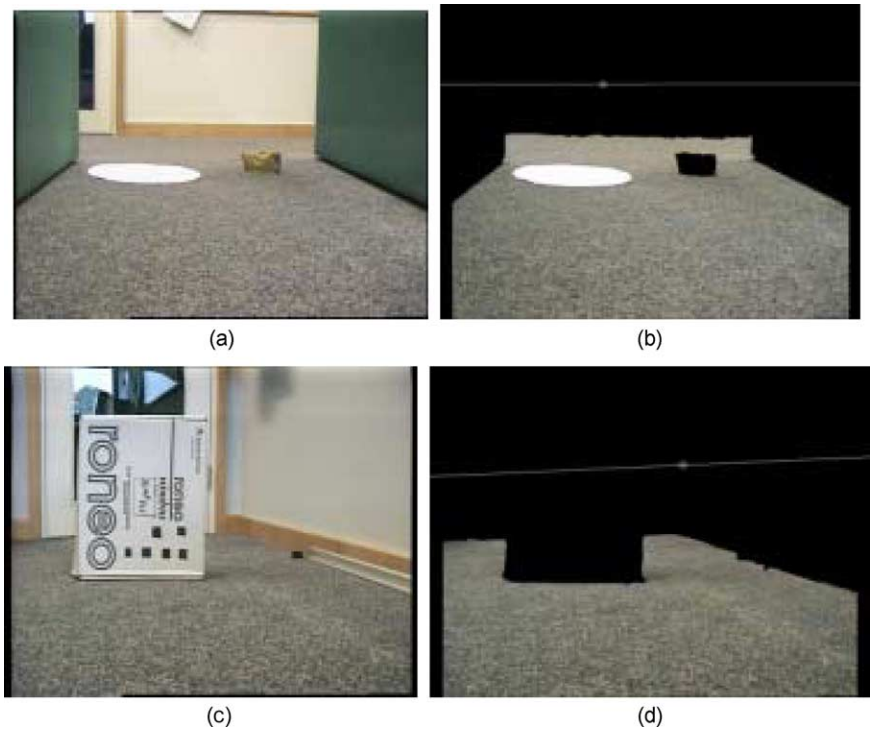

Fig. 12. (a) Original image 1. (b) The segmented ground plane region corresponding to image 1. (c) Original image 2. (d) The segmented ground plane region corresponding to image 2.

Fig. 13. Some obstacle detection results.

have demonstrated that the technique is effective in detecting obstacles on planar surfaces in real images.

Acknowledgements

The authors acknowlede the support of the UK DTI Aeronautics Research Programme.

References

[1] J. Guerrero, C. Sagues, Navigation from uncalibrated monocular vision, in: Proceedings of the Third IFAC Symposium on Intelligent Autonomous Vehicles, 1998, pp. 210–215.
[2] J. Santos-Victor, G. Sandini, F. Curotto, S. Garibaldi, Divergent stereo in autonomous navigation: from bees to robots, International Journal of Computer Vision 14 (1995) 159–177.
[3] R. Cipolla, A. Blake, Surface orientation and time to contact from image divergence and deformation, in: Proceedings of the Second European Conference on Computer Vision, 1992, pp. 187–202.
[4] D. Coombs, M. Herman, T. Hong, M. Nashman, Real-time obstacle avoidance using central flow divergence and peripheral flow, in: Fifth International Conference on Computer Vision, Los Alamitos, CA, 1995, pp. 276–283.
[5] D. Sinclair, A. Blake, Quantitative planar region detection, International Journal of Computer Vision 18 (1) (1996) 77–91.
[6] Zhengyou Zhang, Determining the epipolar geometry and its uncertainty: a review, International Journal of Computer Vision 27 (2) (1998) 161–195.
[7] Zezhi Chen, Chengke Wu, Yong Liu, From an uncalibrated image sequence of a building to virtual reality modeling language (VRML), The Journal of Imaging Science and Technology 46 (4) (2002) 365–374.
[8] From an uncalibrated image sequence of a building to virtual reality modeling language (VRML), The Journal of Imaging Science and Technology 46 (4) (2002) 365–374.
[9] A. Criminisi, I. Reid, A. Zisserman, Single view metrology, International Journal of Computer Vision 40 (2) (2000) 123–148.
[10] M. Wilczkowiak, E. Boyer, F. Sturm, Camera calibration and 3d reconstruction from single images using parallelepipeds, International Conference on Computer Vision, Vancouver (2001) 142–148.
[11] R. Vidal, S. Soatto, Y. Ma, S. Sastry, Segmentation of dynamic scenes from the multibody fundamental matrix, in: Proceedings of the ECCV Workshop on Vision and Modeling of Dynamic Scenes, 2002.
[12] J.C. Clarke, Applications of Sequence Geometry to Visual Motion, PhD thesis, University of Oxford, 1997.
[13] R. Tsai, T. Huang, Estimating three-dimensional motion parameters of a rigid planar patch, IEEE Transactions on Acoustics, Speech and Signal Processing 29 (6) (1981) 1147–1152.
[14] H.C. Longuet-Higgins, The reconstruction of a plane surface from two projections, in: Proceedings of the Royal Society London, vol. B227, 1996, pp. 399–410.
[15] O. Faugeras, F. Lustman, Motion and structure from motion in a piecewise planar environment, International Journal of Pattern Recognition and Artificial Intelligence 2 (3) (1998) 485–508.
[16] M. Lhuillier, L. Quan, A quasi-dense approach to surface reconstruction from uncalibrated images, IEEE Transaction on Pattern Analysis and Machine Intelligence 27 (3) (2005) 1–16.
[17] M. Lhuillier, L. Quan, Mach propagation for image-based modeling and rending, IEEE Transaction on Pattern Analysis and Machine Intelligence 24 (8) (2002) 1140–1146.
[18] B. Liang, N.E. Pears, Visual navigation using planar homographies, Proceedings of IEEE International Conference on Robotics and Automation, Washington DC, USA, vol. 1, 2002, pp. 205–210.
[19] B. Liang, N.E. Pears, Ground plane segmentation from multiple visual cues, Second International Conference on Image and Graphics, Hefei, China, Proceedings of SPIE, vol. 4875, 2002, pp. 822–829.
[20] M. Pollefeys, R. Koch, L. Van Gool, A simple and efficient rectification method for general motion, in: Proceedings of the ICCV 1999, IEEE Computer Society Press, 1999, pp. 496–501.
[21] G. Wolberg, S. Zokai, Robust image registration using log-polar transform, Proceedings of the IEEE International Conference on Image Processing, Vancouver, Canada, vol. I, 2000, pp. 493–496.
[22] Richard Hartley, Andrew Zissermain, Multiple View Geometry in Computer Vision, Cambridge University Press, reprinted 2001.
[23] O. Faugeras, Q.-T. Luong, T. Papadopoulo, The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications, The MIT Press, Cambridge, MA, 2001.
[24] A. David, Forsyth and Jean Ponce, Computer Vision: A Modern Approach, Prentice–Hall, Upper Saddle River, NJ, 2003.
[25] K. Levenberg, A method for the solution of certain non-linear problems in least squares, Quarterly of Applied Mathematics 2 (2) (1944) 164–168.
[26] D.W. Marquardt, An algorithm for the least-squares estimation of nonlinear parameters, SIAM Journal of Applied Mathematics 11 (2) (1963) 431–441.
[27] C.J. Harris, M. Stephens, A combined corner and edge detector, in: Fourth Alvey Vision Conference Manchester, 1988, pp. 147–151.

[28] Jianbo Shi, Carlo Tomasi, Good features to track, IEEE Conference on Computer and Pattern Recognition, Seattle, USA, 1994, pp. 593–600.
[29] S. Smith, J. Brady, SUSAN-a new approach to low level image processing, International Journal of Computer Vision 23 (1) (1997) 45–78.
[30] M.A. Fischer, C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Communications of the Association for Computing Machinery 24 (6) (1981) 381–395.
[31] H. Nakai, N. Takeda, H. Hattori, Y. Okamoto, K. Onoguchi, A practical stereo scheme for obstacle detection in automotive use, in: Proceedings of 17th International Conference of Pattern Recognition (ICPR'2004), Cambridge, United Kingdom, vol. 3, August 2004, pp. 346–350.
[32] Li Tang, Chengke Wu, Zezhi Chen, Image dense matching based on region growth with adaptive window, Pattern Recognition Letters 23 (2002) 1169–1178.