

Ground plane estimation, error analysis and applications

Stephen Se^{a,*}, Michael Brady^b

^a MD Robotics, 9445 Airport Road, Brampton, Ont., Canada L6S 4J3

^b Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, UK

Received 19 January 2001; received in revised form 7 January 2002

Communicated by T.C. Henderson

Abstract

Ground plane perception is of vital importance to human mobility. In order to develop a stereo-based mobility aid for the partially sighted, we model the ground plane based on disparity and analyze its uncertainty. Because the mobility aid is to be mounted on a person, the cameras will be moving around while the person is walking. By calibrating the ground plane at each frame, we show that a partial pose estimate can be recovered. Moreover, by keeping track of how the ground plane changes and analyzing the ground plane, we show that obstacles and curbs are detected. Detailed error analysis has been carried out as reliability is of utmost importance for human applications. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Ground plane; Stereo; Mobility aids; Error analysis; Pose estimation; Obstacle detection; Curb detection

1. Introduction

The motivation for this work is to build a portable mobility aid for the partially sighted using vision as the primary sensor. There are over 40 million partially sighted people worldwide who could benefit from some form of aid. The most obvious problems they face are moving around in their environment without bumping into obstacles and reaching their destinations.

There has been some progress in using sonar [16,25,26,30] for mobility aids since the 1960s, but the development of vision is comparatively recent due to lowered camera cost and increased processing power. Vision-based mobility aids projects [5,32] in the 1980s had considerable limitations due to the lack of technology and powerful processors then.

Gibson [14] noted that ground plane perception is of vital importance to human and aviation mobility. Quoting from Gibson, “there is literally no such thing as a perception of space without the perception of a continuous background surface”. His “ground theory” hypothesis suggested that the spatial character of the visual world is given not by the objects in it but by the ground and the horizon.

Obstacle detection using the ground plane for autonomous guided vehicles (AGVs) has been investigated by various researchers [4,6,13,11,33]. However, as we will discuss in Section 6.2, obstacle detection algorithms for AGVs cannot be used readily in mobility aids.

Moreover, the ground plane constraint has been used for traffic and pedestrian tracking on road scenes [29,31], and also in planetary rover perception [18]. They have only used the ground plane as a constraint for the extraction of objects, but did not model the ground plane itself.

* Corresponding author. Tel.: +1-905-790-2800x4270; fax: +1-905-790-4400.

E-mail address: sse@mdrobotics.ca (S. Se).

In recent years, there has been quite a number of ongoing research projects [3,8,9,17,19] on mobility aids. Some of them use vision while some use sonar. Some of them provide merely navigational assistance whereas some also provide obstacle detection function. However, ground plane modeling has not been investigated in any of these projects.

Among the sensing devices to obtain 3D representation of an environment, vision has distinct advantages over sonar, as it is passive with higher angular resolution and nearly all surfaces have diffuse reflectances. We use a grayscale stereo vision system in our backpack system [22,23].

A mobility aid needs to synchronize with the world frame to frame and it should track the ground plane fairly accurately. It should also provide accurate information about the flatness of the ground plane ahead to ensure user safety [7].

The objective of this paper is to present the theoretical foundation and analysis used to develop our mobility aid. The prototype specifications and performance characteristics are discussed in [22,23]. We will demonstrate that the modeling and tracking of the ground plane are essential for a mobility aid, because it allows obstacle detection and curb detection, which are the most basic requirements of any mobility aid to enable independent travel.

In the next section, we provide an analytical formulation of the ground plane disparity. Random sample consensus (RANSAC) is proposed for fitting ground planes in arbitrary scenes. We then carry out error analysis on ground plane fitting and analyze how the ground plane changes when the user moves. Moreover, three applications using this ground plane model are outlined: pose estimation, obstacle detection and curb detection. Finally, we conclude and discuss some future work.

2. Ground plane model

Li et al. [20] have shown that the ground plane disparity can be expressed as a linear relationship in terms of the image pixel coordinates. In Fig. 1, the camera tilt is θ and h is the camera height. For a scan line subtended at α from the optic axis, its depth with respect to the camera centered coordinates system is $Z(\alpha) = h \cos \alpha / \cos(\phi + \alpha)$ where $\phi = 90^\circ - \theta$. Therefore its disparity is

$$\delta(\alpha) = \frac{fI}{Z(\alpha)} = \frac{fI}{h} (\cos \phi - \sin \phi \tan \alpha),$$

where f is the focal length and I the interocular distance. For a typical pinhole camera model, $\tan \alpha = y_c/f$ where y_c is the vertical camera coordinate.

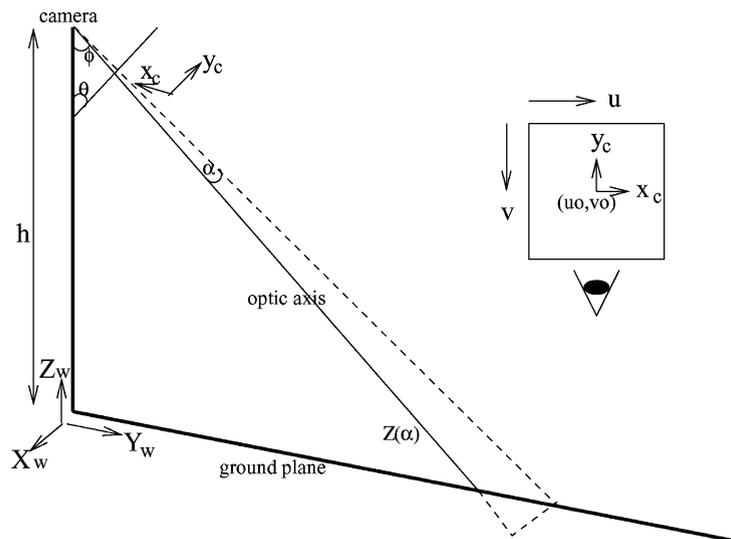


Fig. 1. The ground plane geometry and transformations between world coordinates and image coordinates.

The inset in Fig. 1 shows the relationship between the camera coordinates (x_c, y_c) and the image pixel coordinates (u, v) , we have $v_0 - v = k_v y_c$ and $d = k_u \delta$ where (u_0, v_0) are the image center coordinates, k_u and k_v are the camera intrinsic parameters, d is the disparity in pixels. We have

$$d = \left(\frac{fIk_u}{h} \cos \phi - \frac{Iv_0 k_u}{hk_v} \sin \phi \right) + \left(\frac{Ik_u}{hk_v} \sin \phi \right) v = k_1 + k_2 v.$$

Because the camera tilt, height and the intrinsic parameters are fixed, k_1 and k_2 denote some constant values. Allowing noise and errors in camera geometry, we use u as well for a better fit. The ground plane disparity map is

$$d = au + bv + c, \quad (1)$$

in which (a, b, c) are the ground plane parameters and a should be very small. Therefore, a ground plane can be fitted to image points that have been stereo matched without knowing the intrinsic camera parameters.

Three points are enough to define a plane, but the more points we have, the more accurate the ground plane is. Given more than three points, we do least-squares fitting for the parameters by orthogonal regression [1], which minimizes the shortest distances from the points to the plane. Re-arranging Eq. (1), we have $au + bv - d + c = 0$, and so we minimize

$$C = \sum_{i=1}^n \frac{(\mathbf{p}_i^\top \mathbf{l})^2}{l_a^2 + l_b^2 + l_c^2},$$

where $\mathbf{p}_i = (u_i, v_i, d_i, 1)$ and $\mathbf{l} = (l_a, l_b, l_c, l_d)$.

This problem is equivalent to

$$\min_{\mathbf{l}} C = \mathbf{l}^\top \mathbf{Q} \mathbf{l},$$

subject to $l_a^2 + l_b^2 + l_c^2 = 1$, in which $\mathbf{Q} = \sum_{i=1}^n \mathbf{p}_i \mathbf{p}_i^\top$. The Lagrangian for this is given by

$$C' = \mathbf{l}^\top \mathbf{Q} \mathbf{l} + \mu(l_a^2 + l_b^2 + l_c^2 - 1),$$

and setting $\partial C' / \partial \mathbf{l} = 0$, we have

$$\mathbf{Q} \mathbf{l} + \mu[l_a, l_b, l_c, 0]^\top = \mathbf{0}.$$

The solution for (l_a, l_b, l_c) is the eigenvector corresponding to the smallest eigenvalue of

$$\begin{bmatrix} q_{11} - \frac{q_{14}^2}{q_{44}} & q_{12} - \frac{q_{14}q_{42}}{q_{44}} & q_{13} - \frac{q_{14}q_{43}}{q_{44}} \\ q_{12} - \frac{q_{14}q_{42}}{q_{44}} & q_{22} - \frac{q_{24}^2}{q_{44}} & q_{23} - \frac{q_{24}q_{43}}{q_{44}} \\ q_{13} - \frac{q_{14}q_{43}}{q_{44}} & q_{23} - \frac{q_{24}q_{43}}{q_{44}} & q_{33} - \frac{q_{34}^2}{q_{44}} \end{bmatrix},$$

and $l_d = -(l_a q_{41} + l_b q_{42} + l_c q_{43}) / q_{44}$ where q_{ij} denotes the (i, j) th element of matrix \mathbf{Q} . Then, the ground plane parameters (a, b, c) are obtained as $(-l_a/l_c, -l_b/l_c, -l_d/l_c)$.

3. RANSAC ground plane fitting

We employ the Sobel edge detector and PMF [24] feature-based matching algorithm to the stereo images to obtain feature disparity. These algorithms are chosen because of their performance and computational efficiency.

In order to obtain the ground plane parameters for a certain configuration, we need to calibrate using a ground scene without any obstacles. However, it may occasionally be difficult to find an obstacle-free scene, we propose using RANSAC [12] to calibrate arbitrary scenes and the procedure is as follows [28]:

1. Randomly select three feature points to fit a plane, check each of all the feature points whether or not it satisfies this plane and count the number of supporting points.
2. Repeat step 1 for m times, select the triple with maximum support and do least-squares ground plane fitting to this triple with all its supporting points.

Provided that there are sufficiently many ground plane features, a plane will be fitted to the ground plane features while obstacle features are regarded as outliers, because they are unlikely to be all lying on the same plane, but the ground plane is the dominant plane.

The probability of a good sample (all inliers) is given by $1 - (1 - (1 - e)^p)^m$ where e is the contamination fraction, p the sample size and m the number of samples. The sample size in this case is 3 because we

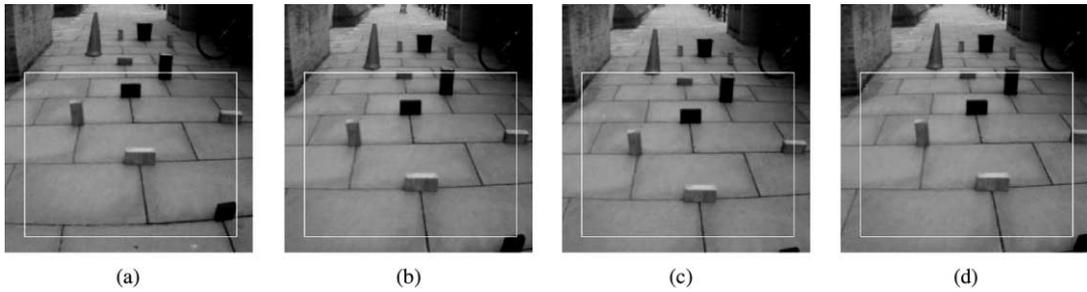


Fig. 2. Left images from an image sequence with slight camera motion between frames.

need three points to define a plane. Assuming the percentage of contamination for a typical scene is 75%, to achieve 99% probability of a good sample, we need to repeat the sampling around 300 times.

Fig. 2(a) shows a ground scene with some obstacles and the region of interest is indicated. The ground plane disparity map found is $d = -0.0127u + 0.2917v + 3.2582$.

4. Error analysis

There are many sources of error in the process of capturing images, finding features, stereo matching and then disparity fitting. Some investigation of image quantization and error models has been carried out in [15,21]. Here, we will look into feature detection inaccuracies and ground plane fitting errors.

In Fig. 3(a), Y represents the ground plane distance and y represents the image coordinate. Expressing $Y = f(y)$, because y is sampled uniformly in the image, $y_{i+1} = y_i + \delta y$ where δy is a constant, then $Y_{i+1} - Y_i = f(y_i + \delta y) - f(y_i) = \delta Y_i$. It is intuitive that δY_i is not constant, but it becomes larger as the pixel moves higher in the image.

Assuming the features are detected and matched to pixel accuracy, Fig. 3(b) shows that one pixel represents a different area on the ground from another [21]. The camera sampling process averages the intensity and so the same feature further away will have a smaller effect on the image as a result of averaging over a larger region. Moreover, we cannot determine accurately the position of the feature as it can be anywhere in the region. Furthermore, there are also image distortions on the sides due to vignetting of the camera lens. These inaccuracies affect the ground plane disparity fitting.

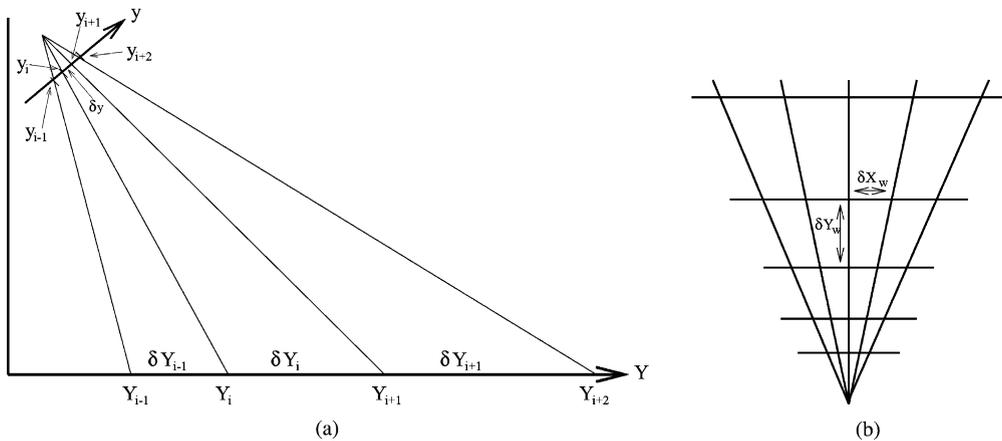


Fig. 3. Image pixel backprojections. (a) Image coordinate y corresponding to varying ground plane distance Y . (b) Pixels corresponding to different sizes of region on the ground.

Referring to Fig. 1, to transform from the world coordinates (X_w, Y_w, Z_w) to the image coordinates (u, v) :

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \equiv M_{\text{intrinsic}} M_{\text{projection}} M_{\text{rotation}} M_{\text{translation}} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \quad (2)$$

where \equiv denotes projective equality,

$$M_{\text{translation}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -h \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$M_{\text{rotation}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(90 - \theta) & \sin(90 - \theta) & 0 \\ 0 & \sin(90 - \theta) & -\cos(90 - \theta) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

$$M_{\text{projection}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix},$$

$$M_{\text{intrinsic}} = \begin{bmatrix} fk_u & 0 & u_0 \\ 0 & -fk_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Considering points on the ground plane only for which $Z_w = 0$, we have

$$u = \frac{fk_u X_w}{Y_w \cos \theta + h \sin \theta} + u_0, \quad (3)$$

$$v = \frac{-fk_v (Y_w \sin \theta - h \cos \theta)}{Y_w \cos \theta + h \sin \theta} + v_0, \quad (4)$$

\Rightarrow

$$X_w = \frac{(u - u_0)(Y_w \cos \theta + h \sin \theta)}{fk_u},$$

$$Y_w = \frac{h(fk_v \cos \theta + (v_0 - v) \sin \theta)}{fk_v \sin \theta + (v - v_0) \cos \theta}. \quad (5)$$

Using these equations, we can compute the distance on the ground between two adjacent pixels on two consecutive scan lines in the image and hence the area of the ground region corresponding to a particular pixel.

For the experiments carried out, $h = 1.25$ m, $\theta = 15^\circ$, $fk_u = 560$, $fk_v = 470$, $u_0 = 120$ and $v_0 = 120$. We compare among three positions: 2.5 m, 4 m and 5.5 m. Because the pixel intensity is proportional to the photon energy arriving from the corresponding region, it is obtained by averaging the energy from that region. Considering its analogue in the 1D case, we take the square root of the region area and look at this averaging effect. It can be regarded as a standard step function for the convolution operator centered at the origin and stretching between $\pm b$.

Fig. 4(a) shows the smoothing operators and Fig. 4(b) shows their Fourier transforms. An edge detector acts like a high pass filter. Arbitrarily assuming that it allows frequencies higher than 10 to pass through, as shown in Fig. 4(b), the amplitude of the 2.5 m case above $w = 10$ is much higher than that of the 4 m case which is in turn higher than the 5.5 m case. As a result, the probability of an edge being detected in the 2.5 m region is higher than in the 4 m region which in turn is higher than in the 5.5 m region. This implies that features in the closer region are fitted better in the ground plane calibration.

In Section 2, we obtained a least-squares fit for the ground plane parameters. However, as the image coordinates and the disparities are not noise-free, we study the covariance of (a, b, c) because this tells us how much confidence we can have in our ground plane estimate.

To analyze the covariance of \mathbf{l} , we follow the technique from Faugeras [10, pp. 151–158]) for the constrained minimization case. Assuming that \mathbf{l}_0 has been obtained by minimizing the criterion function $C(\mathbf{p}_0, \mathbf{l})$ subject to the constraint in Section 2, this defines implicitly a function \mathbf{f} such that $\mathbf{l} = \mathbf{f}(\mathbf{p})$ in a neighborhood of $(\mathbf{p}_0, \mathbf{l}_0)$. We define the vector $\Phi(\mathbf{p}, \mathbf{l})$ by

$$\Phi(\mathbf{p}, \mathbf{l}) = \begin{bmatrix} \mathbf{Q}^{[1]}\mathbf{l} - \left(\frac{l_a}{l_c}\right) \mathbf{Q}^{[3]}\mathbf{l} \\ \mathbf{Q}^{[2]}\mathbf{l} - \left(\frac{l_b}{l_c}\right) \mathbf{Q}^{[3]}\mathbf{l} \\ \mathbf{Q}^{[4]}\mathbf{l} \\ l_a^2 + l_b^2 + l_c^2 - 1 \end{bmatrix},$$

where $\mathbf{Q}^{[i]}$ denotes the i th row of matrix \mathbf{Q} . The Jacobian of \mathbf{f} is given by $\nabla \mathbf{f} = -(\partial \Phi / \partial \mathbf{l})^{-1} (\partial \Phi / \partial \mathbf{p})$.

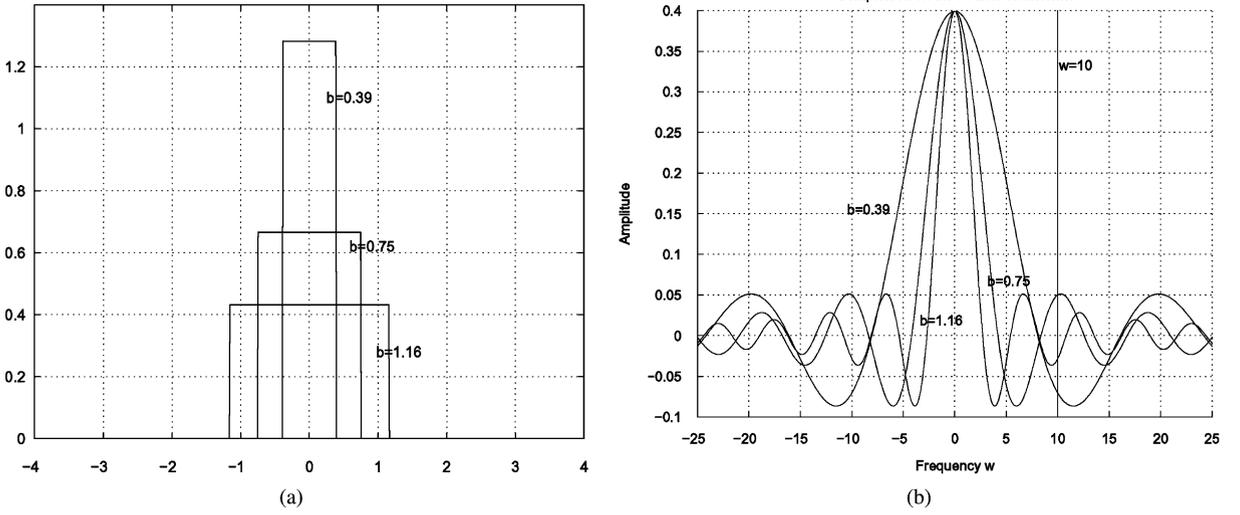


Fig. 4. Comparison of feature detection in different regions: 2.5 m ($b = 0.39$), 4 m ($b = 0.75$) and 5.5 m ($b = 1.16$). (a) The averaging operators. (b) Fourier transforms of (a).

Table 1

$V = \sigma^2$	$\begin{bmatrix} \left(1 + \frac{l_a^2}{l_c^2}\right)C + q_{11} + \frac{l_a^2}{l_c^2}q_{33} - \frac{2l_a}{l_c}q_{13} & q_{12} + \frac{l_a l_b}{l_c^2}q_{33} - \frac{l_a}{l_c}q_{23} - \frac{l_b}{l_c}q_{13} & q_{14} - \frac{l_a}{l_c}q_{34} & 0 \\ q_{12} + \frac{l_a l_b}{l_c^2}q_{33} - \frac{l_a}{l_c}q_{23} - \frac{l_b}{l_c}q_{13} & \left(1 + \frac{l_b^2}{l_c^2}\right)C + q_{22} + \frac{l_b^2}{l_c^2}q_{33} - \frac{2l_b}{l_c}q_{23} & q_{24} - \frac{l_b}{l_c}q_{34} & 0 \\ q_{14} - \frac{l_a}{l_c}q_{34} & q_{24} - \frac{l_b}{l_c}q_{34} & q_{44} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$			

Assuming that the error σ^2 at each point is independent and that the errors are isotropic, the covariance matrix Λ_p for the original data is block diagonal in form with p_i 's covariance matrix as the i th block, assuming all points have the same diagonal covariance matrix. Currently, σ is set to 1 pixel, but verification by further experiments is required. The covariance matrix for l is given by

$$\Lambda_l = \nabla f \Lambda_p \nabla f^T = \left(\frac{\partial \Phi}{\partial l} \right)^{-1} V \left(\frac{\partial \Phi}{\partial l} \right)^{-T},$$

where V is given in Table 1.

Because l_c is always close to unity and its variance is smaller than the others by a magnitude of at least 2, we take the variances for a , b and c to be σ_a^2 , σ_b^2 and σ_c^2 respectively [2].

This error analysis of the ground plane allows us to estimate the uncertainty in the various applications, so that the user can be informed about the reliability of the system.

5. Ground plane tracking

As the camera undergoes six degrees of freedom motion, the perceived ground plane changes. It is intuitive that translations in the horizontal directions and yaw (rotation about the vertical axis) have no effect on the ground plane.

Let (l, m, n) be the translational motion expressed in the world coordinates system (X_w, Y_w, Z_w) , we consider the effect in the Z_w direction. In Fig. 5(a) for the same (u, v) in the new configuration, the

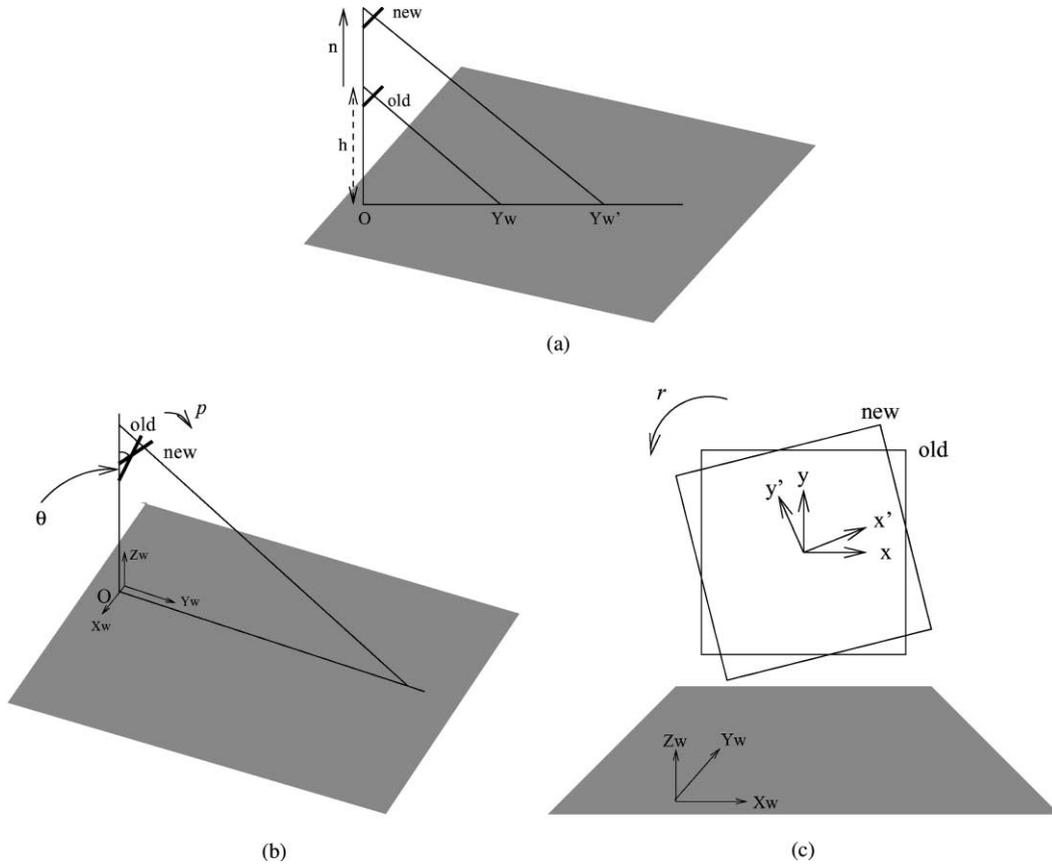


Fig. 5. New and old configurations for the ground plane. (a) Due to Z_w movement. (b) Due to pitch. (c) Due to roll.

backprojected point is further away, and so the disparity decreases. From Eq. (5), $Y_w = hF(v)$ where $F(v) = (fk_v \cos \theta + (v_0 - v) \sin \theta) / (fk_v \sin \theta + (v - v_0) \cos \theta)$ in which θ , fk_v and v_0 are fixed.

Similarly, for that same image point in the new configuration, $Y_w' = (h + n)F(v)$ and so the disparity at Y_w : $d = k/hF(v)$ where k is some constant. The disparity at Y_w' :

$$d' = \frac{k}{(h+n)F(v)} = \frac{k}{hF(v)} \left(\frac{1}{1+n/h} \right) = d \left(\frac{h}{h+n} \right). \tag{6}$$

Let (p, r, y) be the rotational motion corresponding to pitch, roll and yaw, we look at the effects of pitch and roll. The pitch is the rotation about the axis through

the camera optical centered parallel to the X_w axis, as shown in Fig. 5(b), where θ becomes $\theta + p$.

Taking partial differentials of Eqs. (3) and (4) with respect to θ and simplifying, we can find how the image coordinates u and v change when θ changes for the same 3D point:

$$\Delta u = \frac{fk_u X_w (-v - v_0) / fk_v}{fk_u X_w / (u - u_0)} p = \frac{(v - v_0)(u - u_0)}{-fk_v} p, \tag{7}$$

$$\Delta v = -fk_v p - \frac{(v - v_0)^2}{fk_v} p. \tag{8}$$

From Eq. (1), for the same 3D point now and hence the same disparity, we have $d = a(u' - \Delta u) + b(v' - \Delta v) + c$. But for some typical values such as $u_0 = 120$,

$v_0 = 120$, $p = 0.1$, $fk_v = 470$, the factor $a\Delta u$ is so small that it is negligible. Moreover, for Δv , the magnitude of the second part $((v - v_0)^2/fk_v)p$ is much smaller than the first part $-fk_v p$, and so it is relatively negligible. We take Δv as just $-fk_v p$.

Roll r is the rotation about the axis through the camera optical center parallel to the Y_w axis, as shown in Fig. 5(c). For small angle r , the transformation from the old axes to the new axes:

$$\begin{aligned} \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} &= \begin{bmatrix} \cos r & \sin r & 0 \\ -\sin r & \cos r & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \\ &\approx \begin{bmatrix} 1 & r & 0 \\ -r & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}. \end{aligned}$$

Using $M_{\text{intrinsic}}$ from Eq. (2), we have $u = u_0 + fk_u x$; $v = v_0 - fk_v y$

$$u' = u_0 + fk_u(x + yr) = u + \frac{fk_u}{fk_v}(v_0 - v)r, \quad (9)$$

$$v' = v_0 - fk_v(y - xr) = v + \frac{fk_v}{fk_u}(u - u_0)r. \quad (10)$$

From Eq. (1), for the same 3D point and hence the same disparity, we have

$$\begin{aligned} d &= a \left(u' - \frac{fk_u}{fk_v}(v_0 - v)r \right) \\ &\quad + b \left(v' - \frac{fk_v}{fk_u}(u - u_0)r \right) + c. \end{aligned}$$

Although $(fk_u/fk_v)(v_0 - v)r$ is of the magnitude 10, a is of magnitude 10^{-3} usually, so it can be ignored. From Eqs. (9) and (10)

$$\begin{aligned} d &= u' \left(a - \frac{fk_v}{fk_u} \frac{br}{1+r^2} \right) + v' \left(\frac{b}{1+r^2} \right) \\ &\quad + \left(c + \frac{br}{1+r^2} \left[rv_0 + \frac{fk_v}{fk_u} u_0 \right] \right), \end{aligned}$$

but the magnitude of rv_0 is a tenth of that of $(fk_v/fk_u)u_0$ and so is negligible. Using only the first order terms, the overall effects of translational and rota-

tional motion:

$$\begin{aligned} \begin{bmatrix} a' \\ b' \\ c' \end{bmatrix} &= f \left(\begin{bmatrix} a \\ b \\ c \end{bmatrix}, \begin{bmatrix} p \\ r \\ y \end{bmatrix}, \begin{bmatrix} l \\ m \\ n \end{bmatrix} \right) \\ &= \begin{bmatrix} \frac{h}{h+n} \left(a - \frac{fk_v}{fk_u} br \right) \\ \frac{h}{h+n} b \\ \frac{h}{h+n} \left(c + bfk_v p + \frac{fk_v}{fk_u} bru_0 \right) \end{bmatrix}. \quad (11) \end{aligned}$$

We have carried out numerous simulations to validate the appropriateness of these approximations and to verify this formula, which tells us how the ground plane changes as the camera moves around.

6. Applications

In this section, we will describe the various uses of the ground plane model and its error analysis in our mobility aid.

6.1. Pose estimation

Because only three of the six degrees of freedom affect the ground plane, tracking the ground plane gives us estimates for these three parameters. Using Eq. (11), from the ground plane parameters before and after motion, we can estimate its vertical translation n , roll r and pitch p :

$$\begin{bmatrix} n \\ r \\ p \end{bmatrix} = \begin{bmatrix} h \left(\frac{b}{b'} - 1 \right) \\ \left(\frac{fk_u}{fk_v} \right) \left(\frac{a}{b} - \frac{a'}{b'} \right) \\ \frac{1}{fk_v} \left(\frac{c'}{b'} - \frac{c}{b} \right) - \frac{u_0}{fk_v} \left(\frac{a}{b} - \frac{a'}{b'} \right) \end{bmatrix}. \quad (12)$$

Using the first order error propagation formulae [2], we can also compute the variances of these parameters.

Figs. 2(b)–(d) show the same scene with some camera motion between frames carried out. The relative movement between frames measured manually is shown in Table 2. Ground plane parameters (a, b, c)

Table 2

The camera movement between frames measured manually, where (l, m, n) are the translational motion and (p, r, y) are the rotational motion

Measured motion	l (cm)	m (cm)	n (cm)	p (deg)	r (deg)	y (deg)
Figs. 2(a) and (b)	0	20	0	-3	0	0
Figs. 2(b) and (c)	0	0	10	0	0	0
Figs. 2(c) and (d)	-10	0	-10	0	0	0

Table 3

The ground plane parameters and their standard deviations for the various frames

Image	a (σ_a)	b (σ_b)	c (σ_c)
Fig. 2(a)	-0.0127 (0.0023)	0.2917 (0.0015)	3.2582 (0.4453)
Fig. 2(b)	-0.0161 (0.0023)	0.2871 (0.0015)	-2.0204 (0.4489)
Fig. 2(c)	-0.0148 (0.0022)	0.2667 (0.0015)	-3.3461 (0.4140)
Fig. 2(d)	-0.0392 (0.0023)	0.2896 (0.0015)	2.3639 (0.4629)

fitted to each image together with their standard deviations are tabulated in Table 3.

Assuming a flat ground plane, we can compute the estimated camera motion between frames using Eq. (12). Therefore, by fitting ground plane at each frame, we can track the ground plane parameters and recover a partial pose estimate of the camera motion. From the ground plane error analysis, we can also estimate the standard deviation of the camera pose which provides an indication of the confidence of the system. The results shown in Table 4 show that fairly good estimates are obtained.

The more accurate the ground plane parameters are, the better the pose estimates will be. The accuracy of the ground plane parameters depends on the number of features found, how accurate these features are and also their distribution, for example a uniform distribution of the ground plane features is better than a clustered one. Because a more or less same scene is used in these images, we obtain ground plane parameters with similar variances in Table 3.

6.2. Ground plane obstacle detection

Since only ground plane features satisfy our disparity map, any feature that does not is considered as an obstacle, hence the ground plane can be used for obstacle detection [28]. From the error analysis in Section 4, we can estimate the variances for the ground plane parameters and know how reliable our ground plane parameters are. Therefore, we can set a threshold (e.g. a 95% confidence level range) for the obstacle detection process, i.e. only features whose measured disparities are outside of the expected ground plane disparity range will be considered as obstacles. As a result, the lower part of an obstacle may not be detected as its disparity difference may not be significant enough.

Horizontal translations and yaw rotation do not affect the ground plane, and a particular feature is still an obstacle irrespective of the camera horizontal translations or yaw rotation. On the other hand, vertical translation, pitch and roll do affect the obstacle detection

Table 4

Partial camera pose estimated and their standard deviation, where n is the vertical movement, p is the pitch and r is the roll

Estimated motion	n (σ_n) (cm)	p (σ_p) (deg)	r (σ_r) (deg)
Figs. 2(a) and (b)	1.9227 (0.8039)	-3.7376 (0.4913)	0.7338 (0.647)
Figs. 2(b) and (c)	9.3243 (0.9627)	-1.0289 (0.4970)	-0.0305 (0.6590)
Figs. 2(c) and (d)	-9.6133 (0.6726)	1.9815 (0.5014)	4.665 (0.6610)

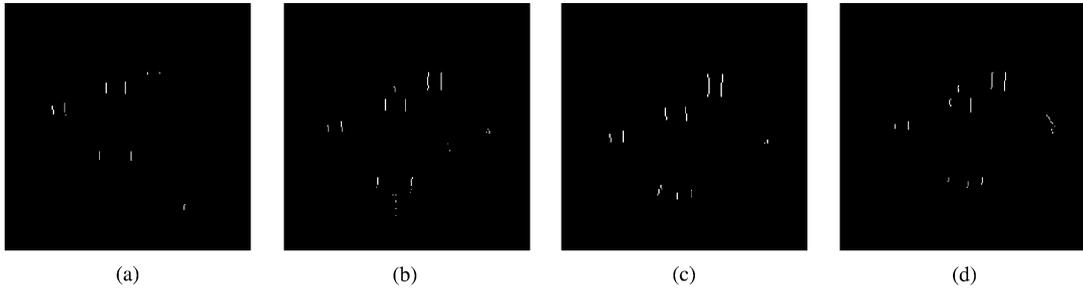


Fig. 6. Obstacles detected by dynamic ground plane recalibration for: (a) Fig. 2(a); (b) Fig. 2(b); (c) Fig. 2(c); (d) Fig. 2(d).

process. For example, if the camera moves downward, ground plane features are now treated as obstacles if the previous disparity map is used.

Therefore, for wheeled mobile robots whose cameras only undergo horizontal translations and yaw rotation, a fixed one-time ground plane calibration is sufficient. However, for cameras mounted on human or legged robots, dynamic ground plane recalibration is necessary, unless there are some other extrinsic sensors such as digital compass and inclinometer to measure how the cameras have moved, in which case Eq. (11) can be used to predict the new ground plane parameters.

Without ground plane recalibration at each frame, obstacles are detected only in the first frame but not

in the subsequent frames. Figs. 6(a)–(d) show the dynamic recalibration obstacle detection results for the scenes in Fig. 2, with most obstacle features correctly identified. From the ground plane error analysis, we can also estimate the obstacle detection probability to provide an indication of the reliability, as shown in Fig. 7.

6.3. Curb detection

A curb can be characterized as a step change in the height of the ground plane, i.e. there exists a discontinuity in the disparity map. Therefore, during the ground plane fitting stage, if the curb step is sufficiently large, identifying the discontinuity can

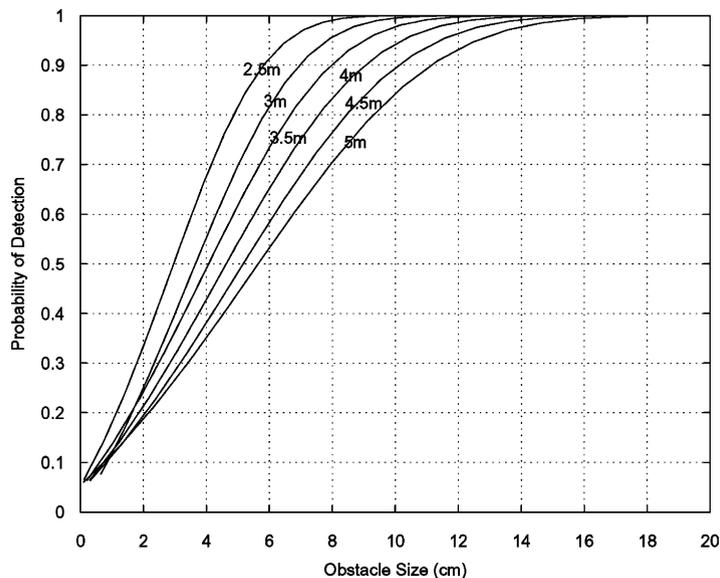


Fig. 7. Obstacle detection probability at various distances.

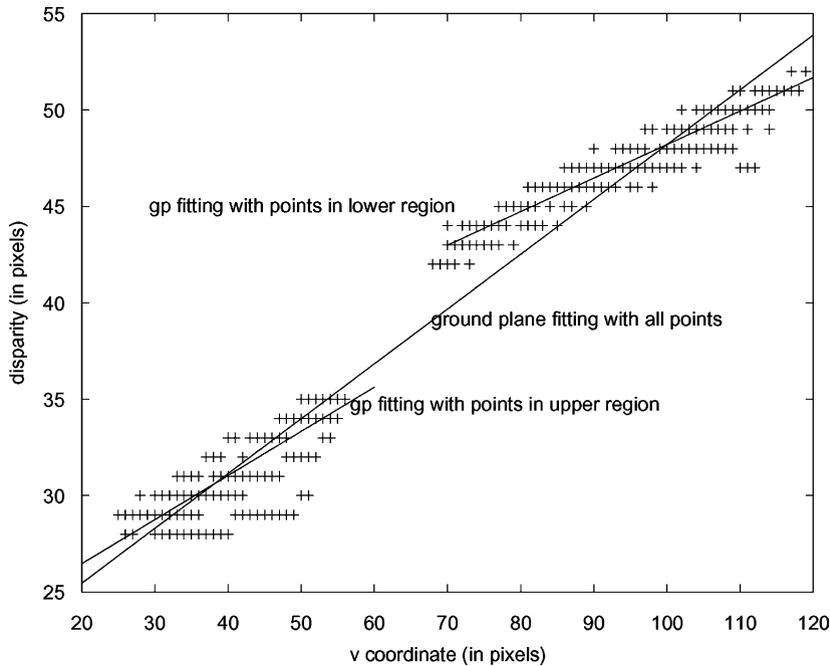


Fig. 8. Ground plane disparity fitting for a synthetic curb scene: using just points in the lower region, upper region and whole region.

localize the curb. Two planes of different heights can be fitted to features either sides of the hypothesized curb. Fig. 8 shows the disparity fit of a synthetic curb scene, a discontinuity and hence two regions can be observed.

However, this relies on having sufficiently many ground plane features on both sides of the curb. In real curb images, we may need to exploit the presence of curb edges and road markings, as image evidence for a local separation between the two planes [27].

Let d_{inside} and d_{outside} be the disparities for the inside and outside regions respectively:

- $H_{\text{step-down}}$ if d_{inside} is significantly greater than d_{outside} ;
- $H_{\text{step-up}}$ if d_{inside} is significantly less than d_{outside} ;
- $H_{\text{no-step}}$ otherwise.

Using the error propagation formulae [2] and the ground plane variance, we can compute the uncertainty for the disparities and test against these three hypotheses whether it is likely to be a step-down or step-up or no step at all.

7. Conclusion

In this paper, we firstly motivated the need for a vision-based mobility aid for the partially sighted and the importance of ground plane perception for human mobility. We described our ground plane disparity model as a planar map linear to the image coordinates. Error analysis has been carried out regarding the feature extraction and the ground plane fitting processes, which is important so that we know how reliable the estimation is.

Moreover, we analyzed the effect of camera movement on the ground plane. Then, three applications using the ground planes are described with some experimental results are shown. We obtained a partial pose estimate by tracking how the ground plane changes. Obstacles and curbs are hazards in the environment that the partially sighted need to be made aware of, which are detected using the ground plane.

Our contribution also includes the detailed error analysis carried out. Reliability and quantitative measures are of utmost importance for real applications to be used by human in particular. Although the

algorithms have been developed for a mobility aid, they are also useful for mobile robots and legged robots in general. However, real-time processing for these algorithms on our prototype has not yet been realized.

For our ground plane fitting, we have only investigated the feature-based stereo approach, but we should also consider the correlation-based approach which is useful for environments such as textured surfaces or carpeted floor. Using a combination of feature-based and correlation-based stereo should allow the system to maximize its usability in various scenes.

Acknowledgements

We thank Fuxing Li, David Lee, Nick Molton, and Penny Probert Smith for many useful discussions.

References

- [1] L. Ammann, J. Van Ness, A routine for converting regression algorithms into corresponding orthogonal regression algorithms, *ACM Transactions on Mathematical Software* 14 (1) (1988) 76–87.
- [2] P.R. Bevington, D.K. Robinson, *Data Reduction and Error Analysis for the Physical Sciences*, 2nd Edition, McGraw-Hill, New York, 1992.
- [3] C. Bonivento, L. Di Stefano, M. Ferri, C. Melchiorri, G. Vassura, The VIDET Project: Basic ideas and early results, in: *Workshop on Service Robotics, Proceedings of the Eighth International Conference on Advanced Robotics (ICAR'97)*, Monterey, CA, July 1997.
- [4] Y.C. Cho, H.S. Cho, A stereo vision-based obstacle detecting method for mobile robot navigation, *Robotica* 12 (3) (1994) 203–216.
- [5] C.C. Collins, On mobility aids for the blind, in: D.H. Warren, E.R. Strelow (Eds.), *Electronic Spatial Sensing for the Blind*, Martinus Nijhoff, Dordrecht, 1985, pp. 35–64.
- [6] S. De Paoli, R. Zigmann, T. Skordas, H. Huot Soudain, Evaluation of an obstacle detection technique based on stereo-vision, in: *Computer Vision for Industry*, Munich, Germany, SPIE, Vol. 1989, 1993, pp. 128–136.
- [7] M.F. Deering, Computer vision requirements in blind mobility aids, in: D.H. Warren, E.R. Strelow (Eds.), *Electronic Spatial Sensing for the Blind*, Martinus Nijhoff, Dordrecht, 1985, pp. 65–82.
- [8] L. Di Stefano, A. Eusebi, M. Ferri, C. Melchiorri, M. Montanari, G. Vassura, A robotic system for visually impaired people based on stereo-vision and force-feedback, in: *Proceedings of the IARP International Workshop on Medical Robots*, Vienna, October 1996.
- [9] M.R. Everingham, B.T. Thomas, Head-mounted mobility aid for low vision using scene classification techniques, *International Journal of Virtual Reality* 3 (4) (1999) 3–12.
- [10] O. Faugeras, *3-Dimensional Computer Vision—A Geometric Viewpoint*, MIT Press, Cambridge, MA, 1993.
- [11] F. Ferrari, E. Grosso, G. Sandini, M. Magrassi, A stereo vision system for real time obstacle avoidance in unknown environment, in: *Proceedings of the IEEE International Workshop on Intelligent Robots and Systems (IROS'90)*, Ibaraki, Japan, 1990, pp. 703–708.
- [12] M.A. Fischler, R.C. Bolles, Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography, *Communication of the ACM* 24 (1981) 381–395.
- [13] J.A. Gaspar, J. Santos-Victor, J. Sentieiro, Ground plane obstacle detection with a stereo vision system, in: *Proceedings of the International Workshop on Intelligent Robotic Systems (IRS'94)*, Grenoble, France, July 1994.
- [14] J.J. Gibson, *The Perception of the Visual World*, Houghton Mifflin, Boston, MA, 1950.
- [15] G.E. Healey, R. Kondepudy, Radiometric CCD camera calibration and noise estimation, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16 (3) (1994) 267–276.
- [16] L. Kay, An ultrasonic sensing probe as a mobility aid for the blind, *Ultrasonics* 2 (1964) 53–59.
- [17] S. Kotani, H. Mori, N. Kiyohiro, Development of the robotic travel aid 'HITOMI', *Robotics and Autonomous Systems* 17 (1996) 119–128.
- [18] E. Krotkov, M. Hebert, M. Buffa, F. Cozman, L. Robert, Stereo driving and position estimation for autonomous planetary rovers, in: *Proceedings of the IARP Workshop on Robotics in Space*, Montreal, Quebec, July 1994.
- [19] G. Lacey, K.M. Dawson-Howe, The application of robotics to a mobility aid for the elderly blind, *Robotics and Autonomous Systems* 23 (4) (1998) 245–252.
- [20] F. Li, J.M. Brady, I. Reid, H. Hu, Parallel image processing for object tracking using disparity information, in: *Proceedings of the Second Asian Conference on Computer Vision (ACCV'95)*, Singapore, December 1995, pp. 762–766.
- [21] L. Matthies, S.A. Shafer, Error modeling in stereo navigation, *IEEE Journal of Robotics and Automation* 3 (3) (1987) 239–248.
- [22] N. Molton, S. Se, J.M. Brady, D. Lee, P. Probert, A stereo vision-based aid for the visually impaired, *Image and Vision Computing* 16 (4) (1998) 251–263.
- [23] N. Molton, S. Se, M. Brady, D. Lee, P. Probert, Robotic sensing for the partially sighted, *Robotics and Autonomous Systems* 26 (1999) 185–201.
- [24] S.B. Pollard, J.E.W. Mayhew, J.P. Frisby, Implementation details of the PMF stereo algorithm, in: J.E.W. Mayhew, J.P. Frisby (Eds.), *3D Model Recognition from Stereoscopic Cues*, MIT Press, Cambridge, MA, 1991, pp. 33–39.
- [25] N. Pressey, Mowat sensor, *Focus* 3 (1977) 35–39.
- [26] L. Russell, Travel path sounder, in: *Proceedings of the Rotterdam Mobility Research Conference*, American Foundation for the Blind, New York, 1965.
- [27] S. Se, M. Brady, Vision-based detection of kerbs and steps, in: A.F. Clark (Ed.), *Proceedings of the British Machine Vision*

Conference (BMVC'97), Essex, England, September 1997, pp. 410–419.

- [28] S. Se, M. Brady, Stereo vision-based obstacle detection for partially sighted people, in: Proceedings of the Third Asian Conference on Computer Vision (ACCV'98), Vol. I, Hong Kong, January 1998, pp. 152–159.
- [29] G.D. Sullivan, Model-based vision for traffic scenes using the ground plane constraint, in: D. Terzopoulos, C. Brown (Eds.), Real-time Computer Vision, Cambridge University Press, Cambridge, 1994.
- [30] S. Tachi, K. Komoriya, Guide dog robot, in: H. Hanafusa, H. Inoue (Eds.), Proceedings of the Second International Symposium on Robotics Research, Kyoto, Japan, MIT Press, Cambridge, MA, 1984, pp. 333–340.
- [31] T.N. Tan, G.D. Sullivan, K.D. Baker, Recognising objects on the ground-plane, Image and Vision Computing 12 (1994) 695–704.
- [32] J.T. Tou, M. Adjouadi, Computer vision for the blind, in: D.H. Warren, E.R. Strelow (Eds.), Electronic Spatial Sensing for the Blind, Martinus Nijhoff, Dordrecht, 1985, pp. 83–124.
- [33] Y. Zheng, D.G. Jones, S.A. Billings, J.E.W. Mayhew, J.P. Frisby, Switcher: A stereo algorithm for ground plane obstacle detection, Image and Vision Computing 8 (1) (1990) 57–62.



Stephen Se is currently with the research and development department at MD Robotics in Canada, developing computer vision systems for space applications. He completed B.Eng. degree with first class honours in Computing at Imperial College of Science, Technology and Medicine, London in 1995 and D.Phil. degree in the Robotics Research Group at the University of Oxford in 1999. His D.Phil. thesis is on computer vision aids for the partially sighted. He then worked at the Department of Computer Science, University of British Columbia, as a Post-doctoral Researcher, carrying out research on vision-based localization and map building for mobile robots. His research interests include computer vision, robotics, image processing and artificial intelligence. He has published in international journals and conference proceedings. He is a member of the IEEE, a member of the ACM and an associate member of the Institute of Electrical Engineers, UK.



Michael Brady, FRS, FEng BP Professor of Information Engineering at the University of Oxford. Professor Brady's degrees are in Mathematics (B.Sc. and M.Sc. from Manchester University, and Ph.D. from the Australian National University). At Manchester University, he was awarded the Renold Prize as the outstanding undergraduate of the year. Professor Brady combines his work at Oxford University, where he founded the Robotics Laboratory and the Medical Vision Laboratory (MVL), with a range of entrepreneurial activities. He is the Director of the recently announced EPSRC/MRC Inter-disciplinary Research Consortium on "From Medical Images and Signals to Clinical Information". He was appointed Senior Research Scientist of the MIT Artificial Intelligence Laboratory in 1980, and founded its world famous Robotics Laboratory. In 1985, he left MIT to take up a newly created Professorship in Information Engineering. Professor Brady serves as a non-executive director and Deputy Chairman of Oxford Instruments plc, as a non-executive director of AEA Technology, and, until recently, Isis Innovation (Oxford University's intellectual property company). Professor Brady is a Founding Director of the start-up companies Guidance and Control Systems, Oxford Medical Image Analysis (OMIA) and Oxford Intelligent Visualisation and Analysis (OXIVA). Professor Brady is the author of over 275 articles in computer vision, robotics, medical image analysis, and artificial intelligence, and the author/editor of nine books, including: Robot Motion (MIT Press, 1984), Robotics Science (MIT Press, 1989), Robotics Research (MIT Press, 1984) and Mammographic Image Analysis (Kluwer, January 1999). He is Editor of the journal Artificial Intelligence, and Founding Editor of the International Journal of Robotics Research. He is a member of the Editorial Board of 14 journals, most recently Medical Image Analysis. Professor Brady was elected as a Fellow of the Royal Academy of Engineering (UK) in 1991 and a Fellow of the Royal Society (UK) in 1997. He is a Fellow of the Institution of Electrical Engineers and a Founding Fellow of the Association of Artificial Intelligence, and a Fellow of the Institute of Physics. He is a member of the Conseil Scientifique de l'INRIA, France. He has been awarded honorary doctorates by the universities of Essex, Manchester, Liverpool, Southampton and Paul Sabatier (Toulouse). He was awarded the IEE Faraday Medal for 2000 and the IEEE Third Millennium Medal for UK.

Professor Brady is a Founding Director of the start-up companies Guidance and Control Systems, Oxford Medical Image Analysis (OMIA) and Oxford Intelligent Visualisation and Analysis (OXIVA). Professor Brady is the author of over 275 articles in computer vision, robotics, medical image analysis, and artificial intelligence, and the author/editor of nine books, including: Robot Motion (MIT Press, 1984), Robotics Science (MIT Press, 1989), Robotics Research (MIT Press, 1984) and Mammographic Image Analysis (Kluwer, January 1999). He is Editor of the journal Artificial Intelligence, and Founding Editor of the International Journal of Robotics Research. He is a member of the Editorial Board of 14 journals, most recently Medical Image Analysis. Professor Brady was elected as a Fellow of the Royal Academy of Engineering (UK) in 1991 and a Fellow of the Royal Society (UK) in 1997. He is a Fellow of the Institution of Electrical Engineers and a Founding Fellow of the Association of Artificial Intelligence, and a Fellow of the Institute of Physics. He is a member of the Conseil Scientifique de l'INRIA, France. He has been awarded honorary doctorates by the universities of Essex, Manchester, Liverpool, Southampton and Paul Sabatier (Toulouse). He was awarded the IEE Faraday Medal for 2000 and the IEEE Third Millennium Medal for UK.