Advanced Robotics, Vol. 6, No. 2, pp. 197–213 (1992)

Paper

Road shape reconstruction by local flatness approximation

KENICHI KANATANI* and KAZUNARI WATANABE†

**Department of Computer Science, Gunma University, Kiryu, Gunma 376, Japan*
†NTT Communications and Information Processing Laboratories, 1-2356 Take, Yokosuka, Kanagawa 238-03, Japan

Received for *AR* 13 February 1991

Abstract—A new algorithm is presented for reconstructing the three-dimensional (3-D) road shape from camera images for the purposes of navigating autonomous land vehicles. The approximation that *the road surface is locally flat* enables us to determine a one-to-one correspondence between the two road boundaries, which together with our knowledge about roads (the 'model' of roads) determines the 3-D road shape. In order to cope with the inaccuracy of image data, a least-squares curve-fitting technique is proposed, and the behaviours of image noise are analysed. Some examples based on real images are given.

1. INTRODUCTION

Considerable attention has recently been paid to the research and development of autonomous land vehicles [1–4]. The ultimate aim of the research is to construct vehicles that navigate autonomously by taking video images of the scene ahead, identifying the road, computing the 3-D geometry, and determining the course of navigation. The use of some guidance system, e.g. guiding lines painted on the road surface, makes vehicle control easier [5], but if the vehicle is to move along an arbitrary road in an uncontrolled environment, we need several sophisticated modules—an image analysis module, a geometric reasoning module, a path-planning module, and a navigation control module, for example.

In the past, effort has been concentrated on the image analysis module—road boundary detection, in particular [6, 7]. The techniques developed so far include colour analysis [3], Hough transforms [8], model fitting [2], and high-level reasoning [9, 10]. In this paper we focus on the geometric reasoning module. Although the necessary 3-D data could be obtained by direct measurement such as stereo and range sensing, we propose an algorithm to reconstruct the 3-D road shape from a *single* image by combining the imaging geometry of perspective projection with an appropriate *model* that idealizes real roads. This approach has the advantage of not requiring any sophisticated sensing devices. Moreover, direct measurement such as stereo and range sensing is effective only in the vicinity of the sensing devices, while our method of model-based image analysis, as we will show, can reconstruct the 3-D road shape over a very long distance, for which direct sensing is almost impossible.

Many of previously proposed methods are based on finding pairs of road boundary segments that are supposedly parallel in the scene and then computing

their vanishing points [4, 11]. The computation is very easy if the road is assumed to be either horizontal (but not necessarily straight) or straight (but not necessarily horizontal). However, it is not easy to find such locally parallel pairs. In order to avoid this difficulty, DeMenthon [12] proposed a discrete numerical method based on the road model proposed by Ozawa and Rosenfeld [13]. His algorithm iteratively computes the 3-D road shape from a pair of starting points whose 3-D positions are known. Sakurai *et al.* [14] proposed a parametric fitting approach by preparing several prototypes of the 3-D road shape. Kanatani *et al.* [15, 16] proposed a differential approach, describing the constraints that ideal roads should satisfy in terms of differential equations, and reconstructing the 3-D road shape by numerically integrating them. Their method is very robust to noise even in the distant part of the road.

The discrete approach of DeMenthon and Davis [17] and the differential approach of Kanatani *et al.* [15, 16] both suffer the same problem: computational error grows rapidly in the course of reconstruction due to the inaccuracy of the original image data and approximations involved in the scheme. Later, DeMenthon and Davis [17] proposed a new scheme based on the assumption that *the road is locally flat* and showed that the solution can be determined pointwise. As a result, one part of the solution is not affected by the error involved in other parts of the solution. However, this very locality destroys the global consistency of the solution; locally constructed solutions can be inconsistent with each other. DeMenthon and Davis [17] proposed the use of dynamic programming to search for a globally consistent solution, but there is no guarantee that such a solution exists.

In this paper, we reformulate the local flatness approximation of DeMenthon and Davis [17] in terms of *N-vectors* representing points and lines in the image [18, 19]. The use of N-vectors ensures that computation is always done in a finite domain without the danger of computational overflow. At the same time, the 3-D implications of vanishing points and vanishing lines are straightforwardly understood in terms of N-vectors [18, 20].

First, we determine a one-to-one correspondence between the two road boundaries in the image by solving the equation that expresses the local flatness approximation. We also present an approximation technique to facilitate numerical computation. The resulting one-to-one correspondence point-wise determine the 3-D road shape independently of other points. The reconstructed 3-D shape may be wildly distorted, since a slight error of the detected road boundary can cause a very large deviation in the distant part. In this paper, we propose a technique to fit a globally defined smooth curve to the computed 3-D road shape by analysing the behaviours of image noise. In this curve fitting, estimation of the road vanishing point plays a crucial role, and its implications to human perception will be discussed. Examples based on real images are also given.

2. PERSPECTIVE PROJECTION AND BACK-PROJECTION OF ROADS

We use an XYZ-coordinate system fixed to the camera in such a way that the coordinate origin O coincides with the centre of the lens, which we call the *viewpoint*, and the Z-axis coincides with the optical axis of the camera. The camera imaging can be modelled as perspective projection of the scene onto an image plane

placed parallel to the XY-plane apart from the viewpoint O by a distance f, which we call the *focal length* (Fig. 1). The value of f (measured in image pixels) is assumed to be calibrated beforehand [21]. A point (X, Y, Z) in the scene is projected onto a point (x, y) on the image plane according to the following relationship:

$$x = f\frac{X}{Z}, \quad y = f\frac{Y}{Z}. \tag{1}$$

Let $P_l:(x(l), y(l))$ and $P_r:(x(r), y(r))$ be the projections of the road boundaries, where l and r are arbitrary parameters defined along the image curves (e.g. the arc lengths measured on the image plane). We assume that the road boundaries are obtained as parameterized smooth curves.

In general, 3-D geometry cannot be reconstructed directly from a single image because depth information is completely lost. For a unique 3-D reconstruction, we need some *a priori* information about the 3-D shape to be reconstructed. Following DeMenthon and co-workers [12, 17, 22] and Kanatani *et al.* [15, 16, 19], we assume that a road is generated by a line segment of known length w sweeping in the scene in such a way that the segment is always kept horizontal and the trajectories of its end-points meet the segment perpendicularly (Fig. 1). This means that:

- there exists a one-to-one correspondence between the two road boundaries;
- line segments joining the corresponding points, which we call the *cross-segments*, are horizontal;
- all cross-segments have the same length w; and
- all cross-segments meet the road boundaries perpendicularly at their end-points.

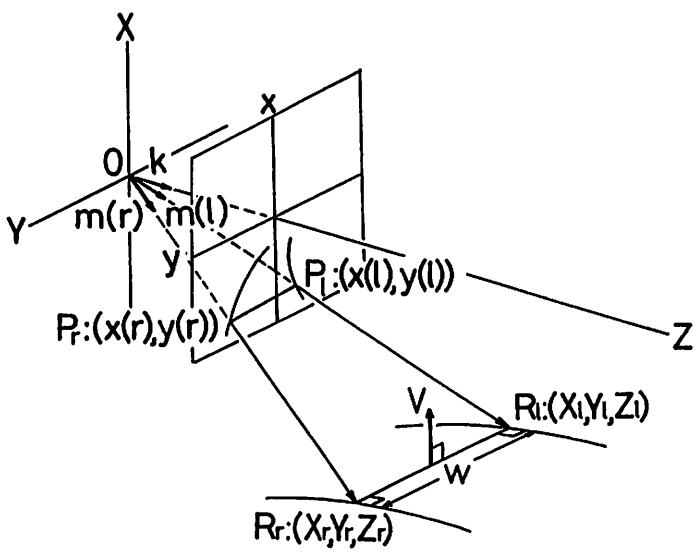


Figure 1. The scene is perspectively projected onto the image plane. A road is assumed to be generated by a horizontal cross-segment of fixed length w sweeping in the scene in such a way that the trajectories of its end-points meet it perpendicularly.

Let $P_l:(x(l), y(l))$ and $P_r:(x(r), y(r))$ be points on the left and right road boundaries, respectively, on the image plane. Define unit vectors

$$\mathbf{m}(l) = \mathrm{N}[\begin{pmatrix} x(l) \\ y(l) \\ f \end{pmatrix}], \quad \mathbf{m}(r) = \mathrm{N}[\begin{pmatrix} x(r) \\ y(r) \\ f \end{pmatrix}], \tag{2}$$

where $\mathrm{N}[\cdot]$ designates normalization into a unit vector (i.e. $\mathrm{N}[\mathbf{a}] = \mathbf{a}/\|\mathbf{a}\|$, where $\|\mathbf{a}\|$ is the norm of vector $\mathbf{a}$). These are the unit vectors starting from the viewpoint O and respectively pointing toward points P_l and P_r on the image plane $Z = f$ (Fig. 1). We call these vectors the *N-vectors* of points P_l and P_r [18, 21]. As DeMenthon and Davis [17] pointed out, the 3-D road shape is reconstructed straightforwardly once a one-to-one correspondence between the road boundaries is established. Indeed, there exists the following explicit relationship (in this paper, $(\mathbf{a}, \mathbf{b})$ denotes the inner product of vectors $\mathbf{a}$ and $\mathbf{b}$):

Proposition 1. If P_l and P_r are the projections of two corresponding points, and if $\mathbf{m}(r)$ and $\mathbf{m}(l)$ are their respective N-vectors, their 3-D positions are given by

$$\mathbf{R}_l = \frac{w\,|\,\mathbf{V}, \mathbf{m}(r))\,|}{D(\mathbf{V}, \mathbf{m}(l), \mathbf{m}(r))}\,\mathbf{m}(l), \quad \mathbf{R}_r = \frac{w\,|\,(\mathbf{V}, \mathbf{m}(\mathrm{l}))\,|}{D(V, \mathbf{m}(l), \mathbf{m}(r))}\,\mathbf{m}(r), \tag{3}$$

where

$$D(\mathbf{V}, \mathbf{m}(l), \mathbf{m}(r)) = \sqrt{(\mathbf{V}, \mathbf{m}(l))^2 + (\mathbf{V}, \mathbf{m}(r))^2 - 2(\mathbf{V}, \mathbf{m}(l))(\mathbf{V}, \mathbf{m}(r))(\mathbf{m}(l), \mathbf{m}(r))}. \tag{4}$$

Here, vector $\mathbf{V}$ is the unit vector indicating the vertical orientation and w is the width of the road. □

Proof. If we put $\mathbf{R}_l = c_l\mathbf{m}(l)$ and $\mathbf{R}_r = c_r\mathbf{m}(r)$, and determine the constants $c_l\,(>0)$ and $c_r(>0)$ in such a way that $\|\mathbf{R}_l - \mathbf{R}_r\| = w$ and $(\mathbf{V}, \mathbf{R}_l - \mathbf{R}_r) = 0$ are satisfied (Fig. 1), we obtain equations (3). □

Hence, the remaining issue is how to find the correct correspondence. Note that equations (3) become singular when the line of sight is horizontal. Namely, if $(\mathbf{V}, \mathbf{m}(l)) = 0$, we must necessarily have $(\mathbf{V}, \mathbf{m}(r)) = 0$ for the solution to exist, but then the right-hand sides of equations (3) become 0/0. Hence, if the road is uphill and its projection crosses over the horizon in the image, the intersections of the horizon with the two road boundaries must correspond to each other. Equations (3)

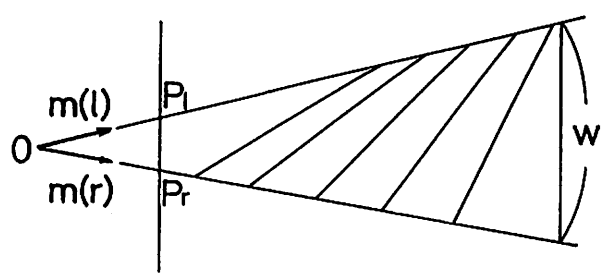


Figure 2. Infinitely many solutions exist if the road cross-segment lies entirely in the horizontal plane passing through the viewpoint O.

are indeterminate there, yielding infinitely many solutions (Fig. 2). This implies that numerical instability occurs near the horizon. In Section 5, we propose a curve-fitting approach to overcome this difficulty.

3. LOCAL FLATNESS APPROXIMATION

Let P_l and P_r be a pair of corresponding points on the road boundaries in the image. Let $\mathbf{m}(l)$ and $\mathbf{m}(r)$ be their respective N-vectors. Consider the tangents to the road boundaries at these points (Fig. 3). Their orientations are specified by 2-D vectors $(\dot{x}(l), \dot{y}(l))$ and $(\dot{x}(r), \dot{y}(r))$, where the dot indicates differentiation with respect to the parameter (e.g. $\dot{x}(l) = \mathrm{d}x(l)/\mathrm{d}l$ and $\dot{y}(r) = \mathrm{d}y(r)/\mathrm{d}r$). Since we are assuming that the road boundaries are obtained as parameterized smooth curves, their derivatives are readily available.

Consider the planes passing through the viewpoint O and intersecting the image plane $Z = f$ along these tangents. Let $\mathbf{n}(l)$ and $\mathbf{n}(r)$ be their respective unit surface normals (Fig. 3). We call these vectors the *N-vectors* of the tangents [18, 21].

Proposition 2. The N-vectors of the tangents at points $(x(l), y(l))$ and $(x(r), y(r))$ are respectively given by

$$\mathbf{n}(l) = \mathrm{N}\left[\begin{pmatrix} -f\dot{y}(l) \\ f\dot{x}(l) \\ x(l)\dot{y}(l) - y(l)\dot{x}(l) \end{pmatrix}\right], \quad \mathbf{n}(r) = \mathrm{N}\left[\begin{pmatrix} -f\dot{y}(r) \\ f\dot{x}(r) \\ x(r)\dot{y}(r) - y(r)\dot{x}(r) \end{pmatrix}\right]. \quad \square \quad (5)$$

Proof. Vectors $(x(r), y(r), f)^{\mathrm{T}}$ and $(\dot{x}(r), \dot{y}(r), 0)^{\mathrm{T}}$ both lie in the plane defined by the viewpoint O and the tangent at P_r. Hence, its unit surface normal $\mathbf{n}(r)$ is obtained by normalizing their vector product (Fig. 3). The unit surface normal $\mathbf{n}(l)$ is similarly obtained. $\square$

The correspondence between the road boundaries is easily established if we employ the *local flatness approximation* of DeMenthon and Davis [17], which states that every cross-segment and the tangents at its end-points define a tangent

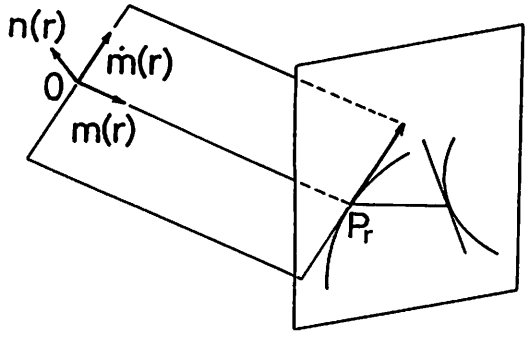


Figure 3. Vector $\mathbf{m}(r)$ starts from the viewpoint O and points toward P_r on the right road boundary, while vector $\mathbf{n}(r)$ is normal to the plane defined by the viewpoint O and the tangent to the right road boundary at P_r. Vectors $\mathbf{m}(l)$ and $\mathbf{n}(l)$ are similarly defined for the left road boundary.

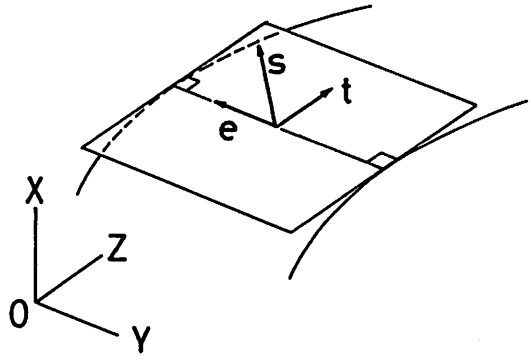


Figure 4. The local-flatness approximation: a cross-segment and the tangents at its end-points define a tangent plane. The moving frame $\{\mathbf{t}, \mathbf{e}, \mathbf{s}\}$ is defined for each cross-segment.

plane in the 3-D scene (Fig. 4). This is only an approximation because if this were strictly true, no twist would be allowed; hence a spiral-shaped road could not be described. However, this is a good approximation if the twist is not very wild. The following Proposition 3 essentially states the condition originally proposed by DeMenthon and Davis [17] but has a slightly different form (in this paper, $|\mathbf{abc}| = (\mathbf{a}, \mathbf{b} \times \mathbf{c}) = (\mathbf{b}, \mathbf{c} \times \mathbf{a}) = (\mathbf{c}, \mathbf{a} \times \mathbf{b})$ designates the scalar triple product of three vectors $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$):

Proposition 3. Under the local flatness approximation, corresponding points P_l and P_r are related by

$$(\mathbf{V}, \mathbf{m}(r))\,|\,\mathbf{m}(l)\mathbf{n}(l)\mathbf{n}(r)\,| + (\mathbf{V}, \mathbf{m}(l))\,|\,\mathbf{m}(r)\mathbf{n}(r)\mathbf{n}(l)\,| = 0. \quad \square \tag{6}$$

Proof. See Fig. 5. Under the local flatness approximation, the two tangents in the scene are parallel to each other. Hence, their vanishing point is at the intersection P_∞ of the projections of the tangents on the image plane. Its N-vector is given by

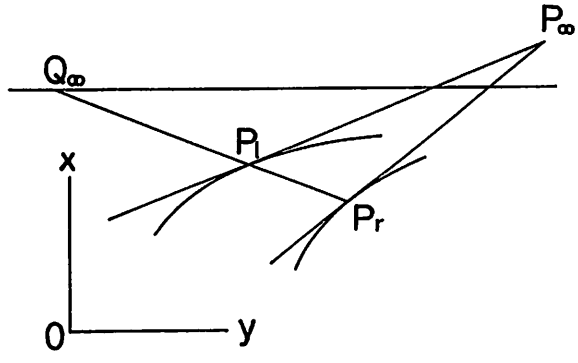


Figure 5. The vanishing point P_∞ of the tangents at the corresponding points P_l and P_r indicates the local 3-D orientation of the road surface. The vanishing point Q_∞ of the corresponding cross-segment lies on the horizon and indicates the 3-D orientation of P_lP_r. The tangents at P_l and P_r must be orthogonal to the cross-segment P_lP_r in the scene.

$N[\mathbf{n}(l) \times \mathbf{n}(r)]$. Since the N-vector of the vanishing point of a line in the scene designates its 3-D orientation, the 3-D orientation of the two tangents is given by $N[\mathbf{n}(l) \times \mathbf{n}(r)]$.

The horizon is a line on the image plane whose N-vector is $\mathbf{V}$. Since all cross-segments are horizontal in the scene, their vanishing points are at the intersections of the horizon with their projections on the image plane. Since the N-vector of line $P_l P_r$ is $N[\mathbf{m}(l) \times \mathbf{m}(r)]$, and since it is a projection of a cross-segment, the N-vector of its vanishing point Q_∞ is given by $N[\mathbf{V} \times N[\mathbf{m}(l) \times \mathbf{m}(r)]]$, which indicates the 3-D orientation of the cross-segment in the scene.

Since the cross-segment is orthogonal to the two tangents in the scene, we have

$$(N[\mathbf{V} \times N[\mathbf{m}(l) \times \mathbf{m}(r)]],\ N[\mathbf{n}(l) \times \mathbf{n}(r)]) = 0. \tag{7}$$

The normalization operation $N[\cdot]$ can be removed because it is an operation of

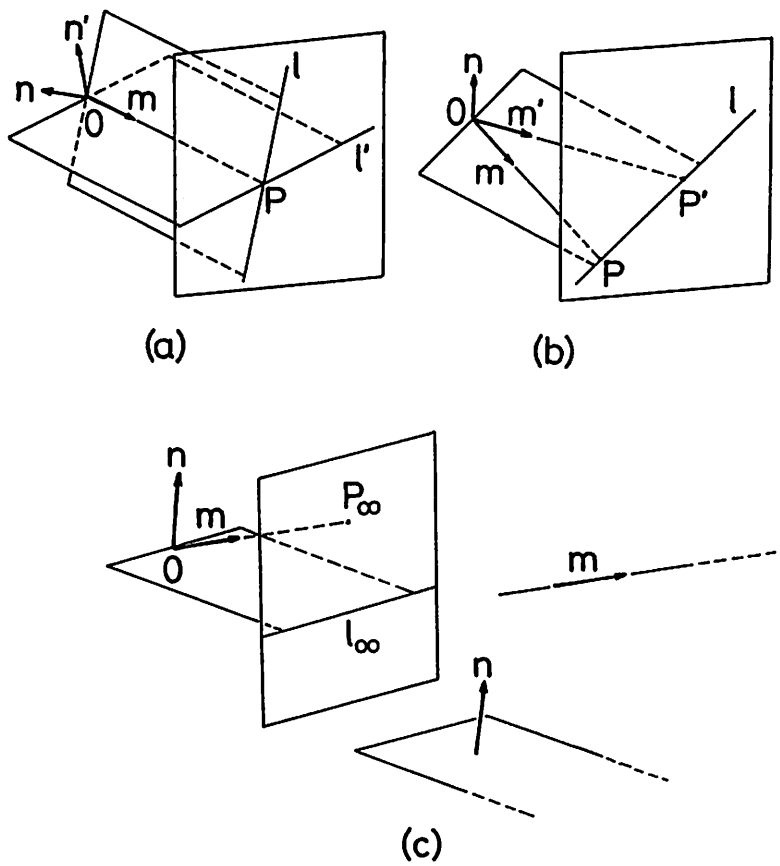


Figure 6. (a) The N-vector $\mathbf{m}$ of the intersection P of two lines l and l' is orthogonal to their N-vectors $\mathbf{n}$ and $\mathbf{n}'$. (b) The N-vector $\mathbf{n}$ of the line l passing through two points P and P' is orthogonal to their N-vectors $\mathbf{m}$ and $\mathbf{m}'$. (c) The N-vector $\mathbf{m}$ of the vanishing point P_∞ of a line in the scene indicates its 3-D orientation, while the N-vector $\mathbf{n}$ of the vanishing line l_∞ of a planar surface in the scene indicates its unit surface normal.

multiplying a constant. Using the identity $\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a}, \mathbf{c})\mathbf{b} - (\mathbf{a}, \mathbf{b})\mathbf{c}$, we obtain equation (6). □

In the above proof, we used the following fundamental facts [18, 20]:

- The N-vector $\mathbf{m}$ of the intersection P of two lines l and l' is obtained by normalizing the vector product $\mathbf{n} \times \mathbf{n}'$ of the N-vectors $\mathbf{n}$ and $\mathbf{n}'$ of the two lines l and l' (Fig. 6a).
- The N-vector $\mathbf{n}$ of the line l passing through two points P and P' is obtained by normalizing the vector product $\mathbf{m} \times \mathbf{m}'$ of the N-vectors $\mathbf{m}$ and $\mathbf{m}'$ of the two points P and P' (Fig. 6b).

We have also used the following fundamental relationships [18, 20] (Fig. 6c):

- The N-vector of the vanishing point of a line in the scene indicates its 3-D orientation.
- The N-vector of the vanishing line of a planar surface in the scene indicates its unit surface normal.

4. DETERMINATION OF THE CORRESPONDENCE

Since we assume that N-vectors $\mathbf{m}(l)$, $\mathbf{m}(r)$, $\mathbf{n}(l)$, and $\mathbf{n}(r)$ have been obtained as smooth functions of l and r, equation (6) establishes a correspondence between the two parameters l and r. Hence, we can use Newton iterations to determine numerically, say, the value of r for a given value of l. In order to do Newton iterations, the derivative of the equation must be evaluated. Let $F(l, r)$ be the left-hand side of equation (6). If the road boundaries are straight, we have $\dot{\mathbf{n}}(r) = 0$. Even if the road boundaries are curved, we can expect $\dot{\mathbf{n}}(r) \approx 0$ as long as the curvature is not very large. With this approximation, we have

Proposition 4.

$$\frac{\partial F}{\partial r}(l, r) \approx \nu(r)[|\mathbf{V}\mathbf{n}(r)\mathbf{m}(r)||\mathbf{m}(l)\mathbf{n}(l)\mathbf{n}(r)| + (\mathbf{V}, \mathbf{m}(l))(\mathbf{m}(r), \mathbf{n}(l))], \tag{8}$$

where $\nu(r) = \|\dot{\mathbf{m}}(r)\|$. □

Proof. If $\dot{\mathbf{n}}(r) = 0$, the derivative $\partial F(l, r)/\partial r$ is

$$(\mathbf{V}, \dot{\mathbf{m}}(r))|\mathbf{m}(l)\mathbf{n}(l)\mathbf{n}(r)| + (\mathbf{V}, \mathbf{m}(l))|\dot{\mathbf{m}}(r)\mathbf{n}(r)\mathbf{n}(l)|. \tag{9}$$

Since $\mathbf{m}(r)$ and $\mathbf{n}(r)$ are mutually orthogonal unit vectors, vector $\mathbf{n}(r) \times \mathbf{m}(r)$ is orthogonal to both of $\mathbf{n}(r)$ and $\mathbf{m}(r)$ and has unit length. Since vector $\dot{\mathbf{m}}(r)$ is orthogonal to both of $\mathbf{n}(r)$ and $\mathbf{m}(r)$ (Fig. 3), we have

$$\dot{\mathbf{m}}(r) = \nu(r)\mathbf{n}(r) \times \mathbf{m}(r). \tag{10}$$

From equations (5), it is easy to confirm that this gives the correct sign. Thus, we have

$$(\mathbf{V}, \dot{\mathbf{m}}(r)) + \nu(r)(\mathbf{V}, \mathbf{n}(r) \times \mathbf{m}(r)) = \nu(r)|\mathbf{V}\mathbf{n}(r)\mathbf{m}(r)|, \tag{11}$$

$$|\dot{\mathbf{m}}(r)\mathbf{n}(r)\mathbf{n}(l)| = (\dot{\mathbf{m}}(r) \times \mathbf{n}(r), \mathbf{n}(l)) = \nu(r)(\mathbf{m}(r), \mathbf{n}(l)), \tag{12}$$

where we have used the relationship $(\mathbf{n}(r) \times \mathbf{m}(r)) \times \mathbf{n}(r) = \mathbf{m}(r)$. If these are substituted into equation (9), we obtain equation (8). □

Proposition 5. If the parameter r is taken to be the arc length along the right road boundary on the image plane, we have

$$\nu(r) = \frac{1}{f} \frac{(\mathbf{m}(r), \mathbf{k})^2}{\sqrt{(\mathbf{m}(r), \mathbf{k})^2 + |\mathbf{n}(r)\mathbf{m}(r)\mathbf{k}|^2}}, \tag{13}$$

where $\mathbf{k} = (0, 0, 1)^T$. □

Proof. From the definition of the N-vector $\mathbf{m}(r)$, we can see that $\vec{OP}_r = f\mathbf{m}(r)/(\mathbf{m}(r), \mathbf{k})$ (Fig. 1). Hence,

$$\frac{d\vec{OP}_r}{dr} = \frac{f\dot{\mathbf{m}}(r)}{(\mathbf{m}(r), \mathbf{k})} - \frac{f(\dot{\mathbf{m}}(r), \mathbf{k})}{(\mathbf{m}(r), \mathbf{k})^2}\mathbf{m}(r) = \frac{f\nu(r)\mathbf{n}(r) \times \mathbf{m}(r)}{(\mathbf{m}(r), \mathbf{k})} - \frac{f|\mathbf{n}(r)\mathbf{m}(r)\mathbf{k}|\nu(r)}{(\mathbf{m}(r), \mathbf{k})^2}\mathbf{m}(r), \tag{14}$$

where equation (10) has been substituted. Since $\mathbf{n}(r) \times \mathbf{m}(r)$ and $\mathbf{m}(r)$ are mutually orthogonal unit vectors, we have

$$\left\| \frac{d\vec{OP}_r}{dr} \right\|^2 = \left(\frac{f\nu(r)}{(\mathbf{m}(r), \mathbf{k})} \right)^2 + \left(\frac{f|\mathbf{n}(r)\mathbf{m}(r)\mathbf{k}|\nu(r)}{(\mathbf{m}(r), \mathbf{k})^2} \right)^2. \tag{15}$$

If r is the arc length on the image plane, we have $\| d\vec{OP}_r/dr \| = 1$, from which we obtain equation (13). □

Using Propositions 4 and 5, we can compute $\partial F(l, r)/\partial r$ from $\mathbf{m}(l)$, $\mathbf{m}(r)$, $\mathbf{n}(l)$, and $\mathbf{n}(r)$ *without taking derivatives*. For a given l, the corresponding r is given by the Newton iteration formula $r \leftarrow r - F(l, r)/(\partial F(l, r)/\partial r)$. If we want to determine l as a function of r, we repeat the same procedure after exchanging the roles of l and r.

5. CURVE FITTING

The point-to-point correspondence between the road boundaries is determined point-wise as shown in the preceding section. The 3-D location of the resulting cross-segment is determined by Proposition 1. Hence, the 3-D road shape is point-wise reconstructed independently of other parts of the road. However, the reconstructed shape may not be smooth. Also, numerical instability occurs, as pointed out in Section 2, near the horizon if the road image crosses over it. In order to force smoothness, two stages are necessary:

(1) erroneous cross-segments are removed, and
(2) a smooth curve is fitted to the centres of the remaining cross-segments.

First, we must establish a criterion which tells us to what extent a cross-segment is erroneous. To this end, we introduce the *moving frame* of the cross-segment (Fig. 4). It is a triad of mutually orthogonal unit vectors $\{\mathbf{t}, \mathbf{s}, \mathbf{e}\}$:

- vector $\mathbf{t}$ indicates the 3-D orientation of the road direction;
- vector $\mathbf{e}$ indicates the 3-D orientation of the cross-segment; and
- vector $\mathbf{s}$ indicates the road surface normal.

Proposition 6.

$$\mathbf{t} = \mathrm{N}[\mathbf{n}(l) \times \mathbf{n}(r)], \quad \mathbf{e} = \mathrm{N}[\mathbf{V} \times \mathbf{t}], \quad \mathbf{s} = \mathrm{N}[\mathbf{t} \times \mathbf{e}]. \qquad \square \quad (16)$$

Proof. Since $\mathrm{N}[\mathbf{n}(l) \times \mathbf{n}(r)]$ is the N-vector of the intersection of the tangents to the road boundaries at the two end-points of the cross-segment, it indicates the 3-D orientation of the tangents. Since the cross-segment is horizontal and perpendicular to the tangents at its end-points in the scene, its direction **e** is orthogonal to both the vertical orientation **V** and the road direction **t**. Also, the road surface normal **s** is orthogonal to both the road direction **t** and the cross-segment orientation **e**. From these, we obtain equations (6). □

Since the three vectors $\{\mathbf{t}, \mathbf{e}, \mathbf{s}\}$ are determined for each cross-segment independently of others, we reject those cross-segments whose moving frames deviate too much from a standard one. As the standard, we consider the frame *at infinity*. We first estimate the vanishing point $P_\infty: (x_\infty, y_\infty)$ by extending the detected road boundaries by straight lines (Fig. 7). Let $\mathbf{m}_\infty$ be its N-vector. The frame at infinity is given by

$$\mathbf{t}_\infty = \mathbf{m}_\infty, \quad \mathbf{e}_\infty = \mathrm{N}[\mathbf{V} \times \mathbf{m}_\infty], \quad \mathbf{s}_\infty = \mathrm{N}[\mathbf{m}_\infty \times \mathbf{e}_\infty]. \qquad (17)$$

We ignore those cross-segments for which

$$|(\mathbf{t}, \mathbf{t}_\infty)| < \cos\theta_t, \quad |(\mathbf{e}, \mathbf{e}_\infty)| < \cos\theta_e, \quad |(\mathbf{s}, \mathbf{s}_\infty)| < \cos\theta_s, \qquad (18)$$

where θ_t, θ_e, and θ_s are preassigned maximum tolerable angles of deviation.

The 3-D positions of the surviving cross-segments are computed by the procedure of Proposition 1. Let

$$\mathbf{R}_c(i) = (X_c(i), Y_c(i), Z_c(i)) \qquad (19)$$

be the centre of the ith cross-segment. Its image point $(x_c(i), y_c(i))$ is given by

$$x_c(i) = f\frac{X_c(i)}{Z_c(i)}, \quad y_c(i) = f\frac{Y_c(i)}{Z_c(i)}. \qquad (20)$$

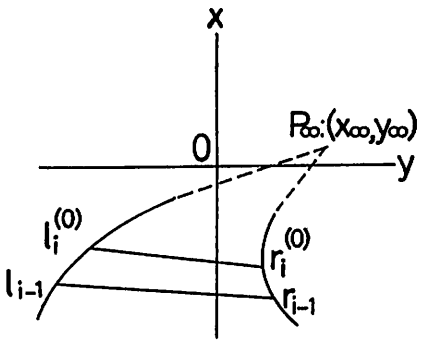


Figure 7. The vanishing point P_∞ of the road is estimated by extending the road boundaries by straight lines.

We fit a function of the form

$$\frac{1}{Z_c(x_c)} = \sum_{k=1}^{N} a_k (x_c - x_\infty)^k \tag{21}$$

to the computed pairs $\{(1/Z_c(i), x_c(i))\}$, where x_∞ is the x-coordinate of the estimated vanishing point P_∞. We use the following least-squares criterion:

$$\sum_i W_i \left(\frac{1}{Z_c(i)} - \sum_{k=1}^{N} a_k (x_c(i) - x_\infty)^k \right)^2 \to \min \tag{22}$$

(Fig. 8). It makes sense to set the weight W_i as a constant, as we will explain shortly. The coefficients $a_k, k = 1, \ldots, N$, are easily determined by solving the simultaneous linear equations obtained by differentiating the above expression with respect to a_k and putting the results equal to zero. The corrected depth $\hat{Z}_c(i)$ of the ith cross-segment is given by

$$\hat{Z}_c(i) = \frac{1}{\sum_{k=1}^{N} a_k (x_c(i) - x_\infty)^k}. \tag{23}$$

The end-points of the ith cross-segment are given by

$$\frac{\hat{Z}_c(i)}{Z_c(i)} \mathbf{R}_c(i) \pm \frac{w}{2} \mathbf{e}_i, \tag{24}$$

where w is the known road width.

Note that many types of function forms are conceivable for the purpose of fitting a curve. The most natural one may be fitting $Z_c = Z_c(X_c)$ to data $\{(Z_c(i), X_c(i))\}$. However, this choice makes it difficult to back-project the road image onto the computed curve, since non-linear equations must be solved in order to determine the 3-D position of $R_c(i)$ that satisfies equations (20) for given $(x_c(i), y_c(i))$. In view of this, relating x_c to Z_c is the best choice.

The reason why the reciprocal $1/Z_c$ is used instead of Z_c itself is that the right-hand side of equation (21) becomes linear if the road is flat. Indeed, if the road is such that $X_c = pZ_c - h$, where p is the slope of the road and h is the vertical height of the viewpoint O from the road, we have $1/Z_c = -(x_c - x_\infty)/fh$, where $x_\infty = fp$. In particular, we have $1/Z_c = -x_c/fh$ if the road is horizontal ($X_c = -h$).

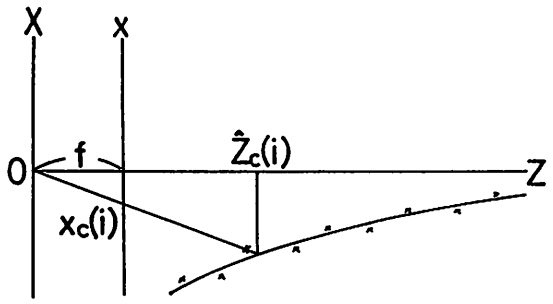


Figure 8. A smooth curve is fitted to the computed 3-D positions.

It is crucial to adopt a polynomial in $x_c - x_\infty$ instead of simply x_c on the right-hand side of equation (21) because the constraint $Z_c(x_\infty) = \infty$ must be strictly imposed. Otherwise, inconsistencies may arise near the vanishing point P_∞: the road may not exist where it should, or it may exist where it should not.

The weight W_i of the least-squares criterion (22) is chosen to be a constant by the following reasoning about the behaviours of image noise. The discrepancy of fitting for the ith cross-segment is

$$\varepsilon_i = \frac{1}{Z_c(i)} \sum_{k=1}^{N} a_k (x_c(i) - x_\infty)^k. \tag{25}$$

Using a constant weight means suppressing the discrepancy uniformly over all i: $\Sigma_{i=1}^{N} (\varepsilon_i)^2 \to \min$. Since $\Delta(1/Z_c) \approx -\Delta Z_c / Z_c^2$, the discrepancy $\varepsilon_i \approx \Delta(1/Z_c(i))$ gives rise to the discrepancy of $Z_c(i)$ by

$$\Delta Z_c(i) \approx -Z_c(i)^2 \varepsilon_i. \tag{26}$$

If the ith cross-segment is placed in the depth $Z_c(i)$ from the viewpoint O, its length measured on the image plane (i.e. the width of the road image) is approximately $\delta_i \approx fw/Z_c(i)$, where w is the true length of the cross-segment. Since δ_i is measured on the image plane, it may contain an error of a few pixels. The error $\Delta\delta_i$ of the width δ_i and the error $\Delta Z_c(i)$ it causes to the computed depth $Z_c(i)$ are related by

$$\Delta\delta_i \approx -fw\, \Delta Z_c(i)/Z_c(i)^2. \tag{27}$$

From equations (26) and (27), we have $\Delta\delta_i \approx fw\varepsilon_i$. Since the error $\Delta\delta_i$ results from image processing, it is reasonable to assume that its magnitude is approximately uniform over the entire road image. Hence, it makes sense to minimize $\Sigma_{i=1}^{N} (\Delta\delta_i)^2 \approx (fw)^2 \Sigma_{i=1}^{N} (\varepsilon_i)^2$.

6. EXAMPLES

We took several video images of real roads near Gunma University in Kiryu, Gunma, Japan. The road boundaries were detected by tracing the white lines painted on the road, and spline curves were fitted to them. The subsequent procedures are as follows:

(1) The spline curves are extended and the road vanishing point P_∞ is estimated (Fig. 7). The two curves are parameterized by their arc lengths; $0 \leqslant l \leqslant L$, $0 \leqslant r \leqslant R$ (the lengths of the left and right boundaries to the vanishing point P_∞ are set to be L and R, respectively).

(2) The correspondence is sought by Newton iterations. The starting values $l_i^{(0)}$ and $r_i^{(0)}$ for the ith cross-segment are chosen in such a way that

$$\frac{L - l_i^{(0)}}{L - l_{i-1}} = \frac{R - r_i^{(0)}}{R - r_{i-1}} \tag{28}$$

(Fig. 7), where l_{i-1} and r_{i-1} are the parameters of the $(i-1)$th cross-segment already established in the previous step (initially we set $l_{-1} = 0$ and $r_{-1} = 0$). The search is done so as to determine r_i for fixed $l_i = l_i^{(0)}$ if

$$F(l_i^{(0)}, r_i^{(0)}) \left. \frac{\partial F(l_i^{(0)}, r)}{\partial r} \right|_{r = r_i^{(0)}} > 0, \tag{29}$$

and l_i for fixed $r_i = r_i^{(0)}$ otherwise, where the derivative $\partial F(l, r)/\partial r$ is approximated by equation (8). In other words, a new cross-segment is always sought *ahead* of the preceding one so that the new cross-segment never crosses over the preceding one. If the iterations do not converge, we proceed forward to seek a new cross-segment.

(3) The moving frame $\{\mathbf{t}, \mathbf{e}, \mathbf{s}\}$ is computed for each cross-segment by equations (16), and those cross-segments that satisfy the criterion (18) are rejected as erroneous. Then the 3-D positions are computed for the remaining cross-segments by the procedure of Proposition 1, and the pair $(Z_c(i), x_c(i))$ is computed for each cross-segment by equations (19) and (20). We require that $Z_c(i)$ be monotonically increasing with respect to $x_c(i)$. If it happens that

$$\cdots < Z_c(i^* - 1) < Z_c(i^*) > \cdots > Z_c(j^*) < Z_c(j^* + 1) < \cdots, \tag{30}$$

those cross-segments for which $Z_c(j^*) - cw < Z_c(i^*) + cw$ are removed, where w is the road width ($c = 1$~2).

(4) A smooth curve is fitted to the surviving cross-segments in the form of equation (21) by the least-squares criterion (22). The 3-D positions of the end-points of each cross-segment are computed by equation (24).

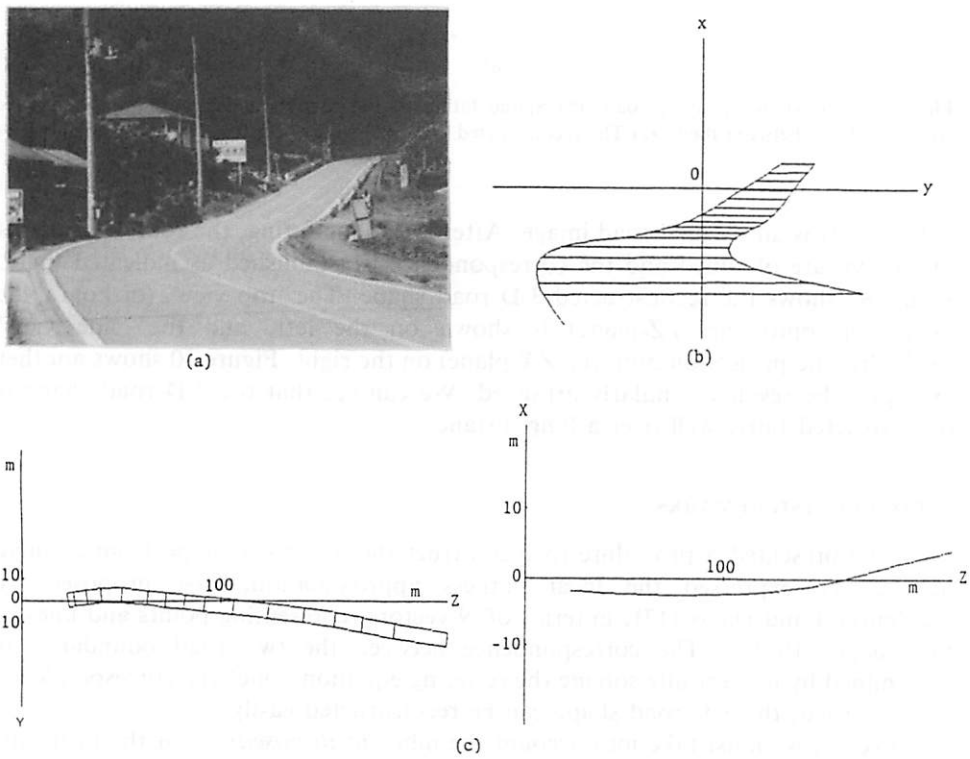


Figure 9. (a) An image of a road. (b) Spline fitting to the road boundaries and the computed correspondence between them. (c) The reconstructed 3-D road shape (a fifth-degree polynomial curve is fitted).

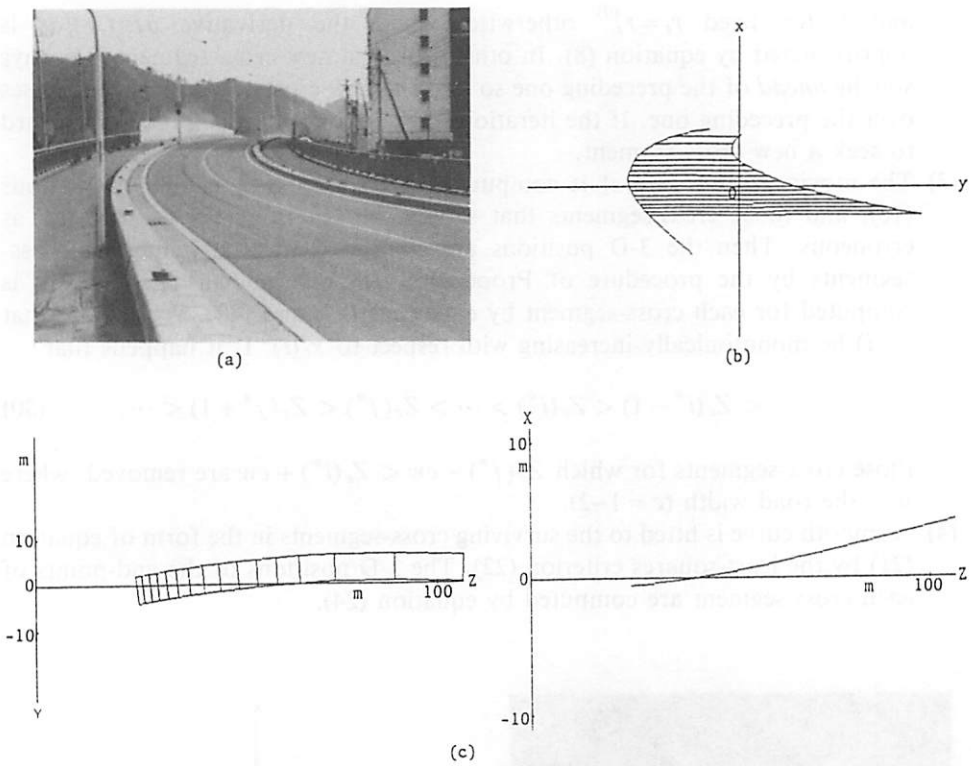


Figure 10. (a) An image of a road. (b) Spline fitting to the road boundaries and the computed correspondence between them. (c) The reconstructed 3-D road shape (a fifth-degree polynomial curve is fitted).

Figure 9a is an original road image. After the spline fitting, the road boundaries of Fig. 9b are obtained and the correspondence is established as indicated there. Figure 9c shows the reconstructed 3-D road shape. The 'top view' (orthographic projection onto the YZ-plane) is shown on the left, and the 'side view' (orthographic projection onto the ZX-plane) on the right. Figure 10 shows another example; the result is similarly arranged. We can see that the 3-D road shape is reconstructed fairly well over a long distance.

7. CONCLUDING REMARKS

We have presented a procedure to reconstruct the 3-D road shape from a single image. We expressed the local flatness approximation, first proposed by DeMenthon and Davis [17], in terms of N-vectors representing points and lines in the image [18, 19]. The correspondence between the two road boundaries is determined by numerically solving the resulting equation. Once the correspondence is established, the 3-D road shape can be reconstructed easily.

However, we must take into account the inherent *ill-posedness* of the problem. If the reconstruction is faithfully based on observed image data, the reconstructed shape is more and more sensitively affected by the image noise as the distance from

the viewer increases. Hence, we must give priority to our knowledge about the object. In this paper, we adopted a curve-fitting approach. The estimate of the road vanishing point P_∞ plays a crucial role for ensuring computational consistency. Note that human drivers can reconstruct the 3-D road shape in a very distant part while driving. Mathematically speaking, the road image has very little information in the distant part because a slight perturbation can cause a very large deviation to the reconstructed shape. It seems that humans compensate for this loss of information with some reasoning based on the estimated location of the road vanishing point. Our algorithm is, in a sense, simulating this human reasoning (if this is true).

This observation implies that the accuracy of reconstruction is determined primarily by the agreement of the 'model' with the road in question. If the road does not agree with the model, say if the road does not have a constant width, the reconstructed shape is bound to be false. On the other hand, if the model matches the road exactly, the reconstruction has 'super-accuracy', since the model is given priority over image data. In this sense, our algorithm is not precisely a 'sensing' or 'measurement' but rather a *model-based prediction* or *hypothesis generation*. The predicted hypothesis must be tested by other means (e.g. by actually driving the vehicle).

In our formulation, we have taken the behaviours of image noise into consideration and devised reasonable approximation techniques to facilitate the computation. Examples based on real images were also given. The solution is very robust to noise and fairly accurate over a long distance. The differential approach of Kanatani *et al.* [15, 16] is also very robust over a long distance and computationally very efficient, but accuracy decreases rapidly in the distant part. The present method attempts to increase accuracy in the distant part by constraining the solution by the estimated road vanishing point and to avoid error accumulation by using local information at the cost of increased computation time. For actual navigation, 3-D reconstruction of the part immediately in front of the vehicle may be sufficient [1–3, 6, 7]. For this purpose, direct measurement such as stereo and range sensing may be effective. However, prediction of the road geometry over a long distance, for which stereo and range sensing are not effective, is also necessary for many types of planning. Our scheme of model-based image analysis can serve this purpose very well.

Acknowledgements

This work was motivated by the ALV project of the University of Maryland, where K. K. stayed in 1985–1986. He thanks Azriel Rosenfeld, Larry Davis, and Daniel DeMenthon of the University of Maryland for helpful discussions. We also thank Chizuko Koyama for conducting real image processing.

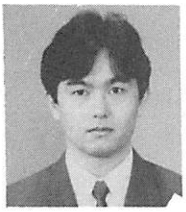
REFERENCES

1. D. Kuan, G. Phipps and A.-C. Hsueh, "Autonomous robotic vehicle road following," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, no. 5, pp. 648–654, 1988.
2. C. Thorpe, M. H. Hebert, T. Kanade and S. A. Shafter, "Vision and navigation of the Carnegie-Mellon Navlab," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, no. 3, pp. 362–373, 1988.

3. M. A. Turk, D. G. Morgenthaler, K. D. Gremban and M. Marra, "VITS—a vision system for autonomous land vehicle navigation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 3, pp. 342–361, 1988.
4. A. M. Waxman, J. LeMoigne, L. S. Davis, B. Srinivasan, T. Kushner, E. Liang and T. Siddalingaiah, "A visual navigation system for autonomous land vehicles," *IEEE J. Robotics Automation*, vol. 3, no. 2, pp. 124–141, 1987.
5. S. Ishikawa, H. Kuwamoto and S. Ozawa, "Visual navigation of an autonomous vehicle using white line recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, no. 5, pp. 743–749, 1988.
6. V. Graefe, "Dynamic vision systems for autonomous mobile robots," *Proc. Int. Workshop Intelligent Robots Systems*, September 1989, Tsukuba, Japan, pp. 12–23.
7. M. Ohzora, T. Ozaki, S. Sasaki, M. Yoshida and Y. Hiratsuka, "Video-rate image processing system for an autonomous personal vehicle system," *Proc. IAPR Workshop on Machine Vision Applications*, November 1990, Tokyo, Japan, pp. 389–392.
8. L. S. Davis, T. R. Kushner, J. LeMoigne and A. M. Waxman, "Road boundary detection for autonomous vehicle navigation," *Opt. Eng.*, vol. 25, pp. 409–414, 1986.
9. K. M. Andress and A. C. Kak, "Evidence accumulation and flow of control in a hierarchical spatial reasoning system," *AI Mag.*, vol. 9, no. 2, pp. 75–94, 1988.
10. J. LeMoigne, "Domain-dependent reasoning for visual navigation of roadways," *IEEE J. Robotics Automation*, vol. 4, no. 4, pp. 419–427, 1988.
11. S.-P. Liou and R. C. Jain, "Road following using vanishing points," *Comput. Vision Graphics Image Process*, vol. 39, pp. 116–130, 1987.
12. D. DeMenthon, "A zero-bank algorithm for inverse perspective of a road from a single image," *Proc. IEEE Int. Conf. Robotics Automation*, March–April 1987, Raleigh, NC, pp. 1444–1449.
13. S. Ozawa and A. Rosenfeld, "Synthesis of a road image as seen from a vehicle," *Pattern Recognition*, vol. 19, pp. 123–145, 1986.
14. K. Sakurai, H. Zen, H. Ohta, Y. Ushioda and S. Ozawa, "Analysis of a road image as seen from a vehicle," *Proc. 1st Int. Conf. Comput. Vision*, June 1987, London, U.K., pp. 651–656.
15. K. Kanatani, K. Watanabe and C. Koyama, "3D reconstruction of road shape from images for autonomous land vehicles (ALV)," *Proc. Int. Workshop Intelligent Robots Systems*, October–November 1988, Tokyo, Japan, pp. 639–644.
16. K. Kanatani and K. Watanabe, "Reconstruction of 3-D road geometry from images for autonomous land vehicles," *IEEE Trans. Robotics Automation*, vol. 6, no. 1, pp. 127–132, 1990.
17. D. DeMenthon and L. S. Davis, "Reconstruction of a road by local image matches and global 3D optimization," *Proc. IEEE Int. Conf. Robotics Automation*, May 1990, Cincinnati, OH, pp. 1337–1342.
18. K. Kanatani, "Computational projective geometry," *CVGIP: Image Understanding*, vol. 54, no. 3, pp. 333–348, 1991.
19. K. Kanatani and K. Watanabe, "Computing 3D road shape from images for ALV: a challenge to an ill-posed problem," *Proc. Int. Workshop Intelligent Robots Systems*, September 1989, Tsukuba, Japan, pp. 371–378.
20. K. Kanatani, *Group-Theoretical Methods in Image Understanding*. Berlin: Springer, 1990.
21. K. Kanatani and Y. Onodera, "Noise robust camera calibration by using vanishing points," *IEICE Trans. Inform. Syst.*, vol. 74, no. 10, pp. 3369–3378, 1991.
22. D. G. Morgenthaler, S. Hennessy and D. DeMenthon, "Range-video fusion and comparison of inverse perspective algorithms in static images," *IEEE Trans. Syst. Man Cybern.*, vol. 20, no. 6, pp. 1301–1312, 1990.

ABOUT THE AUTHORS

Kenichi Kanatani was born in Okayama, Japan and received B.Eng., M.Eng., and Ph.D. degrees in applied mathematics from the University of Tokyo, Japan in 1972, 1974, and 1979, respectively. In 1979, he joined Gunma University, Japan, where he is now Professor of Computer Science. He studied physics at Case Western Reserve University, U.S.A. from 1969 to 1970. He was a visiting researcher at the University of Maryland, U.S.A. from 1985 to 1986; at the University of Copenhagen, Denmark in 1988; and at the University of Oxford, U.K. in 1991. He is the author of *Group-Theoretical Methods in Image Understanding* (Springer, 1990), and *Geometric Computation for Machine Vision* (Oxford Univ. Press, 1992).



Kazunari Watanabe was born in Toyohashi, Aichi, Japan in 1963 and received B.Eng. and M.Eng. degrees in computer science from Gunma University, Japan in 1988 and 1990, respectively. In 1990 he joined NTT Corporation, Japan. He is currently engaged in the development of message processing systems at the NTT Communications and Information Processing Laboratory.