



ArticleReach Paging Slip / Cover Sheet

09/30/2008

Citation Information

Journal: SIAM journal on control and optimization

Article: A Perspective Theory for Motion and Shape Estimation in Machine Vision

Author: Ghosh, Bijoy

ISSN: 03630129

EISSN:

Volume: 33

Issue: 5

Quarter:

Season:

Number:

Month: 09

Day:

Year: 1995

Pages: 1530 -

Patron

Note:

Staff Note:

Paged Locations: >>6wash<<, 9geor, 9uliv, 9wsu0, 9glas, ... (1 of 8 sites paged)

Holdings Information

Location: Olin Level B Stacks

Call Number: Q350 .S6

HOLDINGS: v.14 (1976)-

Request ID: 97573

Central Ariel IP 205.227.91.137



Central Ariel arlir, delivery Patron:



Ariel IP Address

>> <<



Odyssey IP Address

Patron's Library Information

Address: Washington University Libraries, Interlibrary Loan, One Brookings Drive, St. Louis, MO 63130

Email: ill@wumail.wustl.edu

Fax: 314-935-4719

Patron Information

Patron Number: .p1137483@6wash

Patron Type: ARLIR grad

Patron Email: tre1@cec.wustl.edu

Preference: Not Specified

Courtesy of Library Services

For further assistance, please contact your library. NOTICE: This material may be protected by copyright law (Title 17 U.S. Code).

A PERSPECTIVE THEORY FOR MOTION AND SHAPE ESTIMATION IN MACHINE VISION*

B. K. GHOSH[†] AND E. P. LOUCKS[†]

Abstract. In this paper, we consider the problem of motion and shape estimation of a moving body with the aid of a monocular camera. We show that the estimation problem reduces to a specific parameter estimation of a perspective dynamical system. Surprisingly, the above reduction is independent of whether the data measured is the brightness pattern which the object produces on the image plane or whether the data observed are points, lines, or curves on the image plane produced as a result of discontinuities in the brightness pattern. Many cases of the perspective parameter estimation problem have been analyzed in this paper. These cases include a fairly complete analysis of a planar textured surface undergoing a rigid flow and an affine flow. These two cases have been analyzed for orthographic, pseudo-orthographic, and image-centered projections. The basic procedure introduced for parameter estimation is to subdivide the problem into two modules, one for "spatial averaging" and the other for "time averaging." The estimation procedure is carried out with the aid of a new "realization theory for perspective systems" introduced for systems described in discrete time and in continuous time. Finally, much of our analysis has been substantiated by computer simulation of specific algorithms developed in order to explicitly compute the parameters. Detailed simulation that would answer the perspective realizability question is a subject of future research.

Key words. perspective, vision, parameter identification

AMS subject classifications. 93B30, 93C10, 93C15, 93C60

1. Introduction. The problem that we consider in this paper is described as follows.

PROBLEM 1. *We have a textured surface which is moving in continuous time following a certain vector field where we assume that both the shape of the surface and the vector field are unknown. Assume that a camera produces a perfect image of the textured surface in continuous time. The problem of interest is to estimate the shape and motion parameters of the surface from the observed time-varying image produced by the camera.*

Two important assumptions regarding the surface being observed, the camera, and its imaging mechanism need to be emphasized. First, we assume that the surface is constantly under focus, i.e., there is no blurring of the image as a result of imperfect focusing. Second, we assume that the photometric effects on the image due to the light source and the physical properties of the surface are negligible and can be ignored. Thus, the process of image formation is such that the intensity corresponding to each pixel on the surface is transferred to the image plane unattenuated via the projection process.

The existing approaches to the estimation problem in the literature can be divided broadly into two categories depending upon what is assumed to be measured from the scene. If the data observed is assumed to be the brightness pattern which the object produces on the image plane, a well-known approach in the literature is based on analyzing the optical flow field (see [1], [32], [33]). For a system theoretic treatment [2] of the subject we refer the reader to [47]. On the other hand, if the data observed are assumed to be the discontinuity curves in the brightness pattern on the image

* Received by the editors April 2, 1993; accepted for publication (in revised form) May 6, 1994. This research was supported in part by the Department of Energy under grant DE-FG02-90ER14140.

[†] Department of Systems Science and Mathematics, Washington University, Campus Box 1040, One Brookings Drive, St. Louis, Missouri 63130.

plane, a well-known feature-based approach is to identify the correspondence of various features such as points, lines, and curves between one frame and the next (see [3]-[6], [8], [10], [12], [13], [40]). The former approach assumes that the image intensity is a smooth function and restricts attention to the smooth part of the image plane only. The latter approach assumes that the image intensity is a piecewise smooth function and restricts attention to the region of the image plane wherein the image intensity is separated by a discontinuity curve. Of course for each of the two approaches, there are various projection models that one might want to consider. The two projection models well known in the literature are called "orthographic" and "perspective."

There are also other projection models (see [11]) that generalize orthographic and perspective projections. They are described as "image centered projection" and "viewer-centered projection." There are still other projection models in the literature (see [48] not considered in this paper. In this paper, we consider a model of projection (see equation (3.1)) that generalizes the various projection models considered in the literature. The generalized projection degenerates to orthographic, pseudo-orthographic, and perspective projection under various limiting cases. The corresponding estimates of the parameters also degenerate and these have been studied in detail in this paper. Before we describe the main contribution of this paper, we survey some of the important contributions in the field of motion parameter estimation.

The problem of estimating the motion parameters in computer vision has a long history, initiated by the early works of Ullman [9]. The problem was tested subsequently with real images by Roach and Aggarwal [16]. Finally Nagel [17] reduced the problem to solving a single nonlinear equation. A fairly complete analytical solution for eight feature points was given independently by Longuet-Higgins [18] and Tsai and Huang [21]. Zhang [23], [24] proposed a simplified eight-point algorithm and discussed the uniqueness issue. On the question of uniqueness, Netravali et al. [25] introduced a numerical technique called the homotopy method and showed the existence of 10 solutions. Using projective geometry, Faugeras and Maybank [7] showed that at most 10 solutions can be obtained from 5 feature points. Using the quaternion representation of three-dimensional (3-D) rotation, Jerian and Jain [26] reduced the problem to solving the resultant of degree 16 of a pair of polynomials of degree 4 in 2 variables. Jerian and Jain [27] also compared known algorithms exhaustively and compared their performances with noisy data.

Many algorithms in the literature are known to perform poorly under noisy data. A robust algorithm was introduced by Weng, Huang, and Ahuja [28] and by Spetsakis and Aloimonos [14], [15]. They used optimization-based methods to compute "epipolar equations." Grzywacz and Hildreth [29] have also indicated that the effects of image noise on reconstruction from image velocities are severe in some cases. Jerian and Jain [26] and Murray and Buxton [30] proposed various schemes toward a stable reconstruction algorithm. The particular estimation problem has been summarized in two books by Maybank [31] and by Kanatani [11]. In fact, one of the reconstruction algorithms described in this paper has been initiated by Kanatani [11]. For some other related books and references we refer the reader to [39], [41], [45], [42], [43].

In this paper, we consider in detail the problem of estimating motion and shape parameters of a planar surface undergoing an affine motion. The proposed affine motion generalizes the rigid motion already considered in the literature (see [3], [17], [19]-[22]). While preserving the shape of the surface being observed, an affine motion adequately models many other nonrigid deformations. We also consider a generalized projection which includes as a special case both "image-centered projection"

oving
tion
uces
plane
ctive
plete
have
basic
one
d out
d by
eters.
ture

d as

time
and
the
shape
uced

era,
face
fect
ght
red.
ach
tion

ded
the
ject
on
ent
ved
age

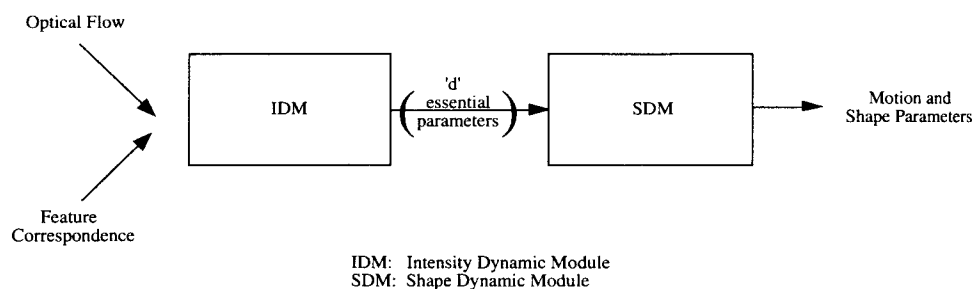


FIG. 1.1. A two-module approach to parameter identification.

and “viewer-centered projection,” together with orthographic and perspective projections. Finally, we consider both the “optical flow analysis” (see [6], [32], [33]) and the “feature-based analysis” (see [35], [34], [44], [46], [40]) and show as the main contribution of this paper that irrespective of what is assumed to be the nature of the data observed (within the class of data considered), and regardless of what is assumed to be the projection model (within the chosen class of models), the problem of motion and shape estimation for a moving textured surface can always be analyzed as a specific parameter estimation problem of a perspective system. The specific form of the perspective system depends on how the surface and the motion field have been parameterized. It may be recalled that perspective systems have already been introduced in [36] in order to study feature-based estimation of motion parameters. Roughly speaking, a perspective system is a linear system with a homogeneous observation function (see [36]).

The details about the estimation scheme proposed in this paper are explained as follows. As shown in Fig. 1.1, the estimation problem is broken up into two modules, known as the Intensity Dynamic Module (IDM) and the Shape Dynamic Module (SDM). Data from the observed surface are first processed in the IDM in order to estimate a set of “essential parameters.” Effectively, IDM performs a “spatial averaging” throughout the entire image plane from either the observed sequence of features or the optical flow data.

The essential parameters are functions of motion and shape parameters. The shape-dynamic module views them as an observation function corresponding to the “shape dynamics” introduced in this paper. The shape-dynamical system together with the essential parameters (viewed as an output) can be regarded as an example of a perspective system introduced in [36]. By observing the essential parameters over time, the SDM obtains an estimate of the motion and shape parameters.

Thus, via a dynamical systems approach, we characterize a complete set of identifiable parameters or functions of parameters for a planar surface undergoing an affine motion. Such a characterization is done both for a generalized projection (3.1) and for an orthographic projection (3.2). As a special case we consider the case when the motion is restricted to a rigid flow and recover many known results in the literature.

In summary, this paper introduces a new unified treatment of the estimation problem.

2. Shape dynamics of a surface patch. We assume throughout this paper that we have a textured surface patch which faces a camera without any occlusion. Furthermore, we assume that every point on the surface moves according to a certain differential equation. As a result of the motion of the individual points, the shape of

the surface undergoes deformation while the surface moves in time. In this section, we write down a differential equation that describes the motion of the surface. We also specialize the equation to a planar surface patch undergoing affine motion and subsequently to a planar surface patch undergoing rigid motion.

Let us assume that (X, Y, Z) is the world coordinate frame wherein we have a surface defined by the equation

$$(2.1) \quad Z = S(X, Y, t).$$

We assume that S is smooth enough so that its derivatives with respect to each of the variables are defined everywhere. We now assume that the motion field is given by the equation

$$(2.2) \quad \dot{X} = f(X, Y, Z), \quad \dot{Y} = g(X, Y, Z), \quad \dot{Z} = h(X, Y, Z).$$

We now describe how the surface (2.1) moves as points on the surface move following the motion field (2.2). This is given by

$$(2.3) \quad \frac{\partial S}{\partial t} + f(X, Y, S) \frac{\partial S}{\partial X} + g(X, Y, S) \frac{\partial S}{\partial Y} = h(X, Y, S).$$

The equation (2.3) is a quasilinear partial differential equation and is called the "shape dynamics." We consider the initial condition

$$(2.4) \quad S(X, Y, 0) = \phi(X, Y).$$

The pair (2.3), (2.4) constitutes an example of a Riccati partial differential equation introduced in [38]. In this paper, we shall assume that the surface (2.1) is a plane described as

$$(2.5) \quad Z = pX + qY + r,$$

where p, q, r are shape parameters that are changing in time as a result of the motion field (2.2). Furthermore we shall also assume that the motion field (2.2) is affine and is given by

$$(2.6) \quad \dot{\mathcal{X}} = A\mathcal{X} + b,$$

where

$$(2.7) \quad A = [a_{ij}], \quad b = \text{col}[b_1, b_2, b_3]$$

are respectively a 3×3 matrix and a 3×1 vector and where $\mathcal{X} = \text{col}[X, Y, Z]$. Thus in this paper, we do not assume that the shape undergoes any deformation as a result of the motion field. We now construct a differential equation that describes the motion of the shape parameters p, q, r . This is done as follows. Let us homogenize the vector

$$(2.8) \quad p = \bar{p}/\bar{s}, \quad q = \bar{q}/\bar{s}, \quad r = \bar{r}/\bar{s}$$

We rewrite (2.5) as $(\bar{p}, \bar{q}, \bar{r})^T \mathcal{X} = 0$ and (2.6) as $\dot{\mathcal{X}} = -\mathcal{A}^T \mathcal{X}$ where $\mathcal{X} = (\bar{X}, \bar{Y}, \bar{Z}, \bar{W})^T$ and

$$(2.9) \quad -\mathcal{A}^T = \begin{pmatrix} A & b \\ 0 & 0 \end{pmatrix}.$$

It follows that

$$(2.10) \quad \frac{d}{dt} \begin{pmatrix} \bar{p} \\ \bar{q} \\ -\bar{s} \\ \bar{r} \end{pmatrix}^T = \mathcal{A} \begin{pmatrix} \bar{p} \\ \bar{q} \\ -\bar{s} \\ \bar{r} \end{pmatrix}^T,$$

where \mathcal{A} is the 4×4 matrix in (2.9) and is defined up to addition by a scalar multiple of the identity matrix. If we assume initial condition to be $\bar{s}(0) = 1$, $\bar{p}(0) = p(0)$, $\bar{q}(0) = q(0)$, $\bar{r}(0) = r(0)$, it may be concluded that the dynamical system (2.10) describes the motion of the shape parameters p, q, r . In fact, from (2.8) and (2.10) the dynamics of p, q, r can be written as the following Riccati equation:

$$(2.11) \quad \begin{aligned} \dot{p} &= (a_{33} - a_{11})p - a_{21}q + a_{31} - a_{13}p^2 - a_{23}pq, \\ \dot{q} &= (a_{33} - a_{22})q - a_{12}p + a_{32} - a_{13}pq - a_{23}q^2, \\ \dot{r} &= -(a_{33} + a_{23}q + a_{13}p)r + (b_3 - b_2q - b_1p). \end{aligned}$$

In general, Riccati equation (2.3) or (2.11) propagates in time the relationship between coordinates X, Y , and Z expressed via the surface (2.1) or the plane (2.5). Note that the equation (2.11) is parameterized by 12 motion parameters and 3 initial conditions on shape parameters. Thus there is a total of 15 parameters describing the shape dynamics (2.10) for the affine motion.

An important special case of the affine motion (2.6) is the case when A is a skew symmetric matrix given by

$$(2.12) \quad \begin{pmatrix} 0 & \omega_1 & \omega_2 \\ -\omega_1 & 0 & \omega_3 \\ -\omega_2 & -\omega_3 & 0 \end{pmatrix} \triangleq \Omega.$$

Under this assumption, the motion field (2.6) describes a rigid motion. The shape dynamics (2.10) can be written as

$$(2.13) \quad \frac{d}{dt} \begin{pmatrix} \bar{p} \\ \bar{q} \\ -\bar{s} \\ \bar{r} \end{pmatrix} = \begin{pmatrix} \Omega & 0 \\ -b^T & 0 \end{pmatrix} \begin{pmatrix} \bar{p} \\ \bar{q} \\ -\bar{s} \\ \bar{r} \end{pmatrix}.$$

Note that the shape dynamics (2.11) reduces to $\dot{p} = -\omega_2(1 + p^2) + \omega_1q - \omega_3pq$, $\dot{q} = -\omega_3(1 + q^2) - \omega_1p - \omega_2pq$, and $\dot{r} = b_3 - b_1p - b_2q - r(\omega_3q + \omega_2p)$ which is parameterized by a total of six motion parameters and three initial conditions on shape parameters. Thus there is a total of nine parameters describing the shape dynamics (2.13) for the rigid motion.

3. Intensity dynamics of a moving textured surface. Assume that the surface described by (2.1) is textured, i.e., the intensity $E(X, Y, Z, t)$ of a point (X, Y, Z) on the surface at time t does not change along the integral curves of (2.2). We also assume that the camera is perfectly focused on the object surface, i.e., intensity from a surface on the object to the image plane is transferred unattenuated under the camera correspondence. The above two assumptions together imply that the intensity on the image plane does not change along the projection of the integral curves of (2.2). In this paper we consider the projection to be described as follows.

Let (x, y) be the coordinates of the image plane obtained under the projection of a point (X, Y, Z) on the surface of the object. We define

$$(3.1) \quad x = \frac{fX}{Z + \delta}, \quad y = \frac{fY}{Z + \delta},$$

where $\delta \in [0, f]$ and f is the focal length of the camera. Note that if $\delta = 0$ we obtain a viewer-centered projection. If $\delta = f$ we obtain an image-centered projection. These two projections have been described in [11]. Finally note that if $\delta = f$ and $f \rightarrow \infty$ we obtain

$$(3.2) \quad x = X, \quad y = Y$$

which is known in the literature [11] as the "orthographic projection."

In an orthographic projection, a point (X, Y, Z) is projected by dropping the Z coordinate information. In order to motivate the image-centered and viewer-centered projections, assume that the image plane is perpendicular to the Z axis and passes through the point $Z = a$. Assume that the optical axis is the Z axis and a point (X, Y, Z) is projected onto the image plane via the center of the camera located at $Z = -Z_0$. In order to derive the projected point, one computes the line l passing through the points (X, Y, Z) and $(0, 0, -Z_0)$ and computes the intersection of l with the image plane. The projection of the point (X, Y, Z) is this intersection. If the center of the camera is the origin of the coordinate axis, i.e., if $Z_0 = 0$, we obtain a viewer-centered projection. On the other hand, if we assume that the image plane passes through the origin of the coordinate axis, i.e., if $a = 0$, we obtain an image-centered projection.

For a given fixed value of f, δ we have a new set of coordinates (x, y, Z) . We now rewrite the shape equation (2.1) and the restriction of the motion field (2.2) on the image plane in the new set of coordinates as

$$(3.3) \quad Z = S(x, y, t)$$

and

$$(3.4) \quad \dot{x} = f(x, y, S(x, y, t)), \quad \dot{y} = g(x, y, S(x, y, t))$$

for some suitable functions S, f, g .

The integral curves of (3.4) are exactly the projection of the integral curves of the motion field under the generalized projection (3.1). The vector field described by (3.4) has been described in the literature (see Horn [1]) as "optical flow." Note in particular that the optical flow is in general a time-varying dynamical system described via the coordinates of the image plane. The time variation of the optical flow is a result of the motion of the surface (2.1), or equivalently (3.3).
 Let $e(x, y, t)$ be the intensity of a point (x, y) on the image plane at time instant t . Since $e(x, y, t)$ does not change along the integral curves of (3.4), it follows that $e(x, y, t)$ satisfies the partial differential equation given by

$$(3.5) \quad \frac{\partial e}{\partial t} + f(x, y, S(x, y, t)) \frac{\partial e}{\partial x} + g(x, y, S(x, y, t)) \frac{\partial e}{\partial y} = 0.$$

We shall call the dynamical system (3.5) as "intensity dynamics." Let us now assume that the initial condition is given by

$$(3.6) \quad e(x, y, 0) = \Psi(x, y).$$

We shall call the function $\Psi(x, y)$ the "texture function." The above pair (3.5), (3.6) is a linear partial differential equation, which describes the dynamics of the intensity function on the image plane.

Let us now restrict our attention to a planar surface (2.5) with affine motion (2.6) and assume a generalized projection (3.1). The "optical flow" equation for this special case can be written as follows:

$$(3.7) \quad \begin{aligned} \dot{x} &= d_1 + d_3x + d_4y + \frac{1}{f}(d_7x^2 + d_8xy), \\ \dot{y} &= d_2 + d_6y + d_5x + \frac{1}{f}(d_8y^2 + d_7xy), \end{aligned}$$

where

$$(3.8) \quad \begin{aligned} d_1 &= f(a_{13} + c_1), d_2 = f(a_{23} + c_2), d_3 = (a_{11} - a_{33}) - (c_3 + pc_1), \\ d_4 &= a_{12} - qc_1, d_5 = a_{21} - pc_2, \\ d_6 &= (a_{22} - a_{33}) - (c_3 + qc_2), d_7 = pc_3 - a_{31}, d_8 = qc_3 - a_{32} \end{aligned}$$

and where

$$(3.9) \quad c_i = (b_i - a_{i3}\delta)/(r + \delta), \quad i = 1, 2, 3.$$

Various limits of the optical flow equation have been considered in the literature. They all pertain to analyzing what happens when f tends to ∞ , assuming $f = \delta$. In the process of taking the limit, one would approximate the coefficients of the optical flow equation (3.7) up to order $\frac{1}{f}$, while neglecting the higher-order terms. If we define

$$(3.10) \quad h_j = \lim_{f \rightarrow \infty} d_j; \quad j = 1, 2, \dots, 8$$

we obtain the following:

$$(3.11) \quad \begin{aligned} h_1 &= a_{13}r + b_1, h_2 = a_{23}r + b_2, \\ h_3 &= a_{11} + a_{13}p, h_4 = a_{12} + a_{13}q, \\ h_5 &= a_{21} + a_{23}p, h_6 = a_{22} + a_{23}q, \\ h_7 &= -a_{31} - a_{33}p, h_8 = -a_{32} - a_{33}q. \end{aligned}$$

Thus when $f \rightarrow \infty$ and $f = \delta$, the optical flow equation can be approximated up to order $\frac{1}{f}$ by

$$(3.12) \quad \begin{aligned} \dot{x} &= h_1 + h_3x + h_4y + \frac{1}{f}(h_7x^2 + h_8xy), \\ \dot{y} &= h_2 + h_5x + h_6y + \frac{1}{f}(h_8y^2 + h_7xy). \end{aligned}$$

Of course if the focal length of the camera is fixed at ∞ , one observes the optical flow equation as

$$(3.13) \quad \dot{x} = h_1 + h_3x + h_4y, \dot{y} = h_2 + h_5x + h_6y.$$

The projection which produces the optical flow given by (3.13) is known as "orthographic projection." Such a projection described by (3.2) does not give any information about the quadratic component d_7 and d_8 of the optical flow (3.7) in general. The optical flow equation (3.12), on the other hand, is an approximation of (3.7) up to order $\frac{1}{f}$ assuming f is approaching ∞ . Thus if the focal length of a camera can

be varied, one can obtain the asymptotic values of d_7 and d_8 for large f and use this information to compute h_7 and h_8 . We shall call (3.12) the optical flow under "orthographic approximation," as opposed to (3.13), which is the optical flow under "orthographic projection."

We also introduce a "pseudo-orthographic approximation" of (3.7) originally introduced by Kanatani [11]. This is described as follows:

$$\begin{aligned} x &= d_1 + d_3x + d_4y + \frac{1}{f}(h_7x^2 + h_8xy), \\ y &= d_2 + d_6y + d_5x + \frac{1}{f}(h_8y^2 + h_7xy). \end{aligned} \tag{3.14}$$

"Orthographic approximation" and "pseudo-orthographic approximation" to the optical flow equation (3.7) is useful in the process of reconstructing the motion and shape parameters from the coefficients of the optical flow equation. The reconstruction algorithm has been described in §5 using an approach described by Kanatani [11].

4. Estimation of essential parameters based on intensity and feature measurements.

Assume as in §3 that we have a moving textured plane which produces a time-varying intensity profile on the image plane. In this section we consider the intensity dynamic module problem described as follows.

PROBLEM 2 (intensity dynamic module problem). Assume that the intensity function $e(x, y, t)$ is measured in a given region of the image plane over a given interval of time. The problem is to estimate the vector (d_1, \dots, d_8) from this data.

In subsequent sections, we shall see that the vector (d_1, \dots, d_8) is of paramount importance in estimating the motion and shape parameters. For this reason we shall call the vector (d_1, \dots, d_8) the "vector of essential parameters."

4.1. Estimation based on intensity measurements. Assume that the intensity function is smooth so that all its partial derivatives exist and can be computed. If the motion field is affine given by (2.6), it follows from (3.5), (3.7) that the intensity dynamics is given by

$$\frac{\partial e}{\partial t} + F(x, y) \frac{\partial e}{\partial x} + G(x, y) \frac{\partial e}{\partial y} = 0, \tag{4.1}$$

where $e(x, y, t)$ is the observed intensity function on the image plane and

$$\begin{aligned} F(x, y) &= d_1 + d_3x + d_4y + \frac{1}{f}(d_7x^2 + d_8xy), \\ G(x, y) &= d_2 + d_6y + d_5x + \frac{1}{f}(d_8y^2 + d_7xy). \end{aligned} \tag{4.2}$$

The parameters d_1, \dots, d_8 can be defined from (3.8). Combining (4.1) and (4.2), we now write

$$v^T p = - \frac{\partial e}{\partial t}, \tag{4.3}$$

where

$$v^T = \begin{pmatrix} e_{x^2} e_{y^2} e_{xy} e_x e_y e_{xx} e_{yy} e_{xy} \\ \frac{1}{f} (x^2 e_{xx} + xy e_{xy} + y^2 e_{yy}) \end{pmatrix}, \tag{4.4}$$

and

$$(4.5) \quad d = (d_1, \dots, d_8)^T.$$

In order to compute an estimate of the coefficient vector d , we proceed as follows. Choose $n \geq 8$ points on the image plane denoted by $(x_i, y_i), i = 1, \dots, n$. From the observed data $e(x, y, t)$ we now form the matrices

$$(4.6) \quad V = \begin{pmatrix} v(x_1, y_1) & v(x_2, y_2) & \dots & v(x_n, y_n) \end{pmatrix}$$

and

$$(4.7) \quad u = \begin{pmatrix} -e_t(x_1, y_1) & -e_t(x_2, y_2) & \dots & -e_t(x_n, y_n) \end{pmatrix}^T$$

From (4.3) it follows that $V^T d = u$. If the points (x_i, y_i) are chosen in such a way that $\text{rank } V = 8$, we compute

$$(4.8) \quad \hat{d} = (VV^T)^{-1}Vu$$

as an estimate of d . We therefore have the following theorem.

THEOREM 4.1. *Assume that the function $e(x, y, t)$ is such that all its partial derivatives are available and can be measured. Assume furthermore that the points $(x_i, y_i), i = 1, \dots, n$ are such that $\text{rank } V = 8$, where V is given by (4.6). It is possible to obtain a unique estimate of d .*

4.2. Estimation based on feature measurements: Curve correspondence. By the word "feature" we shall mean points or curves of discontinuity for the intensity function $e(x, y, t)$. We shall assume that, via edge detection, these features can be observed in real time. We shall assume that the moving textured surface produces a time-varying intensity function on the screen. The moving intensity function in turn would make the features move on the screen. The dynamical system which describes such a motion is called "feature dynamics." The main result of this section is to see that the coefficients of the feature dynamics are exactly the essential parameters introduced in (3.8). Thus under an appropriate technical condition, the essential parameters can be estimated from the feature dynamics as well, as was the case for intensity dynamics. In order to describe the feature dynamics we proceed as follows.

Let

$$(4.9) \quad y = \mathcal{I}(x, t)$$

be the curve along which the function $e(x, y, t)$ is discontinuous. We want to study how the feature curve (4.9) changes in time. Differentiating (4.9) with respect to time, we obtain

$$(4.10) \quad \dot{y} = \frac{\partial \mathcal{I}}{\partial x} \dot{x} + \frac{\partial \mathcal{I}}{\partial t}.$$

Recall that

$$(4.11) \quad \begin{aligned} \dot{x} &= F(x, y), \\ \dot{y} &= G(x, y), \end{aligned}$$

where $F(x, y), G(x, y)$ are given in (4.2). It follows that

$$(4.12) \quad \frac{\partial I}{\partial t} + \frac{\partial x}{\partial t} \left[d_1 + d_3x + d_4I(x, t) + \frac{f}{1} (d_7x^2 + d_8xI(x, t)) \right] = d_2 + d_6I(x, t) + d_5x + \frac{f}{1} (d_8I(x, t)^2 + d_7xI(x, t)).$$

The above equation (4.12) is referred to as the feature dynamics, which can be rewritten as

$$(4.13) \quad v^T d = - \frac{\partial I}{\partial t}$$

where d is defined as in (4.5) to be the vector of essential parameters. The vector v^T is given by

$$(4.14) \quad v^T = \left(I_x, -1, xI_x, I_x, -x, -I, \frac{f}{1} (x^2I_x - x - xI), \frac{f}{1} (-I^2 + xII_x) \right).$$

We now choose $n \geq 8$ points on the curve (4.9) denoted by $(x_i, y_i), i = 1, \dots, 8$. As in (4.6), (4.7) we construct the matrix V and vector u and obtain an estimate d of d given by (4.8), provided of course rank $V = 8$.

In order for the matrix V to have rank 8, the curve (4.9) has to be of sufficiently high order. In fact, if (4.9) is a polynomial, it cannot be of degree > 4 . On the other hand, if

$$(4.15) \quad \mathcal{I}(x, t) = a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4, a_4 \neq 0$$

in order for rank $V = 8$, one must have

$$(4.16) \quad a_3^2 \neq \frac{3}{8} a_2 a_4.$$

Thus we have essentially proved the following theorem.

THEOREM 4.2. Assume that the observed feature is a polynomial discontinuity curve (4.9) of degree 4 given by (4.15). It is possible to estimate d given by (4.8) iff (4.16) is satisfied.

If the observed discontinuity curve is of degree < 4 , we shall see that one needs to observe a larger number of features in order for rank $V = 8$. Two cases of interest are when the observed feature is a line and when it is a point. These two subcases are now considered.

4.3. Estimation based on line correspondence. Let

$$(4.17) \quad y = ax + b$$

be the line along which the function $e(x, y, t)$ is discontinuous. Assume furthermore that the line (4.17) is generated as a result of a discontinuity in the texture of the surface (2.5). We also assume that changes in x, y are given by (3.7). Thus, the feature dynamics is given by (4.12) or (4.13) where

$$(4.18) \quad \frac{\partial I}{\partial t} = ax + b$$

and

$$(4.19) \quad v^T = \left(a, -1, ax, a(ax+b), -x, -(ax+b), -\frac{b}{f}x, -\frac{b}{f}(ax+b) \right).$$

The vector d of essential parameters (see (4.5)) satisfies the ordinary differential equation

$$(4.20) \quad \begin{pmatrix} -a & 1 & 0 & -ab & 0 & b & 0 & \frac{b^2}{f} \\ 0 & 0 & -a & -a^2 & 1 & a & \frac{b}{f} & \frac{ab}{f} \end{pmatrix} d = \begin{pmatrix} \dot{a} \\ \dot{b} \end{pmatrix}.$$

If we assume that the motion of the line (4.17) is observed, we might infer that in (4.20), a, b, \dot{a}, \dot{b} is observed. Thus (4.20) represents a pair of equations in eight variables, the variables being the eight-parameter d vector. Choosing a set of four lines on the surface being observed and assuming that these four lines define a set of eight independent conditions on the d vector, one can obtain a unique estimate of the d vector. The procedure is similar to that outlined in §4.1 and described by (4.8). We now state the following theorem.

THEOREM 4.3. *Assume that the observed feature is a set of four lines on the image plane given by the equation*

$$(4.21) \quad y = a_i x + b_i, \quad i = 1, \dots, 4,$$

where the lines (4.21) are generated as a result of discontinuity in the texture of the surface (2.5). Define

$$(4.22) \quad \phi_i = \begin{pmatrix} -a_i & 1 & 0 & -a_i b_i & 0 & b_i & 0 & \frac{b_i^2}{f} \\ 0 & 0 & -a_i & -a_i^2 & 1 & a_i & \frac{b_i}{f} & \frac{a_i b_i}{f} \end{pmatrix}$$

$i = 1, \dots, 4$ and the 8×8 matrix $\phi = (\phi_1^T \phi_2^T \phi_3^T \phi_4^T)^T$. It is possible to estimate the vector d uniquely given by

$$(4.23) \quad \hat{d} = (\phi^T \phi)^{-1} \phi^T (\dot{a}_1 \quad \dot{b}_1 \quad \dot{a}_2 \quad \dot{b}_2 \quad \dot{a}_3 \quad \dot{b}_3 \quad \dot{a}_4 \quad \dot{b}_4)^T$$

iff rank $\phi = 8$.

4.4. Estimation based on point correspondence. If we assume that the texture function is discontinuous at a single point, one would observe this point as a discontinuity in the function $\epsilon(x, y, t)$. Tracking the discontinuity in real time would amount to tracking the projection of the feature point on the image plane. Thus we rewrite the optical flow (3.7) as

$$(4.24) \quad \begin{pmatrix} 1 & 0 & x & y & 0 & 0 & \frac{x^2}{f} & \frac{xy}{f} \\ 0 & 1 & 0 & 0 & x & y & \frac{xy}{f} & \frac{y^2}{f} \end{pmatrix} d = \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix},$$

where d is once again the vector of essential parameters given by (4.5). The point (x, y) is the projection of the feature point on the image plane. Assuming that we are able to observe x, y, \dot{x}, \dot{y} in real time, it follows that equation (4.24) represents a pair of equations in eight variables, the variables being the eight-parameter d vector of essential parameters. As in §4.3, if we choose a set of four feature points on the image plane that are projections of points of discontinuity in the texture of the surface (2.5),

and assume that they define a set of eight independent conditions on the vector d , it follows that one can uniquely obtain an estimate of the vector d . Thus we have the following theorem.

THEOREM 4.4. Assume that the observed feature is a set of four points on the image plane given by $(x_i, y_i), i = 1, \dots, 4$, where we assume that the points are generated as a result of discontinuity in the texture of the surface (2.5). Assume furthermore that the 8×8 matrix

$$\psi = (\psi_1^T \psi_2^T \psi_3^T \psi_4^T)^T \tag{4.25}$$

is nonsingular, where

$$\psi_i = \begin{pmatrix} 1 & 0 & x_i & y_i & 0 & 0 & \frac{f}{x_i y_i} & \frac{f}{x_i y_i} \\ 0 & 1 & 0 & x_i & y_i & \frac{f}{x_i y_i} & \frac{f}{x_i y_i} & \frac{f}{x_i y_i} \end{pmatrix} \tag{4.26}$$

$i = 1, 2, 3, 4$. It is possible to estimate the vector d uniquely given by

$$d = (\psi_1^T \psi_2^T \psi_3^T \psi_4^T)^{-1} \psi^T (x_1 \ y_1 \ x_2 \ y_2 \ x_3 \ y_3 \ x_4 \ y_4)^T \tag{4.27}$$

To summarize the main results of this section, we show that the vector d of essential parameters can be estimated from intensity and feature measurements. The task of the IDM is to estimate the vector d . It may be noted that the IDM requires information only at a given instant of time and performs "spatial averaging."

5. Estimating motion and shape parameters from the recovery equation.

In this section we shall assume that the essential parameter vector d has already been estimated by the intensity dynamic module. The problem that we would like to consider is to solve (3.8) for the motion and shape parameters. We would also like to study how the solution degenerates for $f = \delta$ as $f \rightarrow \infty$, i.e., when the projection model degenerates to that produced by orthographic projection. Some portion of our analysis in this section is an adaptation of earlier work due to Kanatani [11].

5.1. Estimation under general projection. We assume that we have a planar surface (2.5) undergoing a rigid motion (2.13). The essential parameter vector d given

by (3.8) for this case is given as follows:

$$\begin{aligned} d_1 &= f(\omega_2 + c_1), & d_2 &= f(\omega_3 + c_2), & d_3 &= -(c_3 + pc_1), & d_4 &= \omega_1 - qc_1, \\ d_5 &= -\omega_1 - pc_2, & d_6 &= -(c_3 + qc_2), & d_7 &= (\omega_2 + pc_3), & d_8 &= (\omega_3 + qc_3), \end{aligned} \tag{5.1}$$

where

$$c_1 = (b_1 - \omega_2 \delta) / (r + \delta), \quad c_2 = (b_2 - \omega_3 \delta) / (r + \delta), \quad c_3 = b_3 / (r + \delta) \tag{5.2}$$

The problem that we consider is described as follows.

PROBLEM 3. Assume that we are given (d_1, \dots, d_8) . Using the algebraic equation

$$(5.1), (5.2), \text{ solve for the parameters } c_1, c_2, c_3, \omega_1, \omega_2, \omega_3, p, q. \tag{5.1}$$

It may be noted that (5.1) describes exactly a set of eight nonlinear equations in eight parameters. This particular set of equations is known as the "recovery equation."

The following result is quite surprising, however.

THEOREM 5.1. Assume $c_3 \neq 0$; then (5.1) can be solved for exactly two real solutions. If

$$(c_1, c_2, c_3, \omega_1, \omega_2, \omega_3, p, q) \tag{5.3}$$

is one solution, then the other solution is given by

$$(-c_3p, -c_3q, c_3, \omega_1 - c_1q + c_2p, \omega_2 + c_1 + c_3p, \omega_3 + c_2 + c_3q, -c_1/c_3, -c_2/c_3). \quad (5.4)$$

It may be remarked that the existence of two solutions to the recovery equation (5.1) and described by Theorem 5.1 has been reported earlier in the literature by Waxman and Ullman [8] and by Kanatani [11]. In [8] the analytical steps leading up to the two solutions have not been documented. In [11] the analytical formula (5.4) of the two solutions has not been presented. The purpose of stating and proving Theorem 5.1 is therefore tutorial.

Before we prove Theorem 5.1, we proceed to solve the set of equations (5.1). Let us define

$$(5.5) \quad \begin{aligned} T &= d_3 + d_6, & R &= d_5 - d_4, & U_0 &= d_1 + id_2, \\ K &= \frac{1}{f}(d_7 + id_8), & S &= d_3 - d_6 + i(d_4 + d_5), \end{aligned}$$

and

$$(5.6) \quad P = p + iq, \quad V = c_1 + ic_2, \quad W = \omega_3 - i\omega_2, \quad L = fK - \frac{1}{f}U_0.$$

The equations (5.1) can be written as

$$(5.7) \quad \begin{aligned} U_0 &= f(V + iW), \\ S &= -PV, L = c_3P - V, \\ -iPV^* &= R + 2\omega_1 + i(T + 2c_3). \end{aligned}$$

Note that (5.7) is a set of four equations in complex variables that needs to be solved. From (5.7) we have

$$(5.8) \quad V^2 + LV + c_3S = 0.$$

Solving (5.8) for V and then using (5.7) for P we have

$$(5.9) \quad V = \frac{-L \pm \sqrt{L^2 - 4c_3S}}{2},$$

$$(5.10) \quad P = \frac{L \pm \sqrt{L^2 - 4c_3S}}{2c_3}.$$

From (5.7) we have

$$(5.11) \quad \omega_1 = [Im(PV^*) - R]/2,$$

$$(5.12) \quad T + 2c_3 = -Re(PV^*).$$

From (5.9) and (5.10) we have

$$(5.13) \quad Re(PV^*) = \frac{-|L|^2 + \sqrt{(L^2 - 4c_3S)(L^2 - 4c_3S)^*}}{4c_3}.$$

Combining (5.12) and (5.13) we have

$$(5.14) \quad |L|^2 - 4Tc_3 - 8c_3^2 = \sqrt{|L|^4 + 16c_3^2|S|^2 - 8c_3Re(L^2S^*)}.$$

Note that (5.14) as an equation in c_3 has two solutions. One solution is at $c_3 = 0$ and the other solution is at $c_3 = c_3^*$. Squaring (5.14) on both sides, we conclude that c_3^* is the middle root of the cubic equation

$$(5.15) \quad c_3^3 + Tc_3^2 + \frac{1}{8}(T^2 - |L|^2 - |S|^2)c_3 + \frac{1}{8}(Re(L^2S^*) - T|L|^2) = 0$$

Using c_3 , one can solve for a pair of solutions for P and V from (5.9) and (5.10). Finally, from (5.7) we have

$$(5.16) \quad W = -\frac{1}{2}V - \frac{1}{2}U_0$$

and from (5.11) one can solve for ω_1 . Thus the set of equations (5.7) can be solved for exactly two distinct solutions if $c_3 \neq 0$. If (5.1) is solved, these are exactly the two solutions that one would obtain.

Proof of Theorem 5.1. It can be easily checked that if (5.3) is one solution of (5.1), then the other solution is given by (5.4). However, since (5.1) has exactly two solutions, these are the only solutions. Moreover the solutions are obtained by solving the cubic polynomial equation (5.15) outlined as above.

From the two solutions to the recovery equation (5.7), it is easy to see what happens when $f \rightarrow \infty$. Note that

$$(5.17) \quad \lim_{f \rightarrow \infty} c_1 = -\omega_2, \quad \lim_{f \rightarrow \infty} c_2 = -\omega_3, \quad \lim_{f \rightarrow \infty} c_3 = 0.$$

It follows that one of the two solutions $(c_1, c_2, c_3, \omega_1, \omega_2, \omega_3, p, q)$ approaches the vector

$$(5.18) \quad (-\omega_2, -\omega_3, 0, \omega_1, \omega_2, \omega_3, p, q),$$

the first six components of the other solution approach the vector

$$(5.19) \quad (0, 0, 0, \omega_1 + \omega_2q - \omega_3p, 0, 0)$$

and the last two components of the other solution approach ∞ asymptotically along the line

$$(5.20) \quad d/q = \omega_2/\omega_3.$$

The parameters b_1, b_2, b_3, r are never recovered exactly. In fact, from the definition of d_1, d_2, c_3 we have, for a given f , the straight line

$$(5.21) \quad \omega_2r + b_1 = \left(1 + \frac{f}{r}\right)d_1, \quad \omega_3r + b_2 = \left(1 + \frac{f}{r}\right)d_2, \quad b_3 = c_3(r + f)$$

described in the (b_1, b_2, b_3, r) space corresponding to the solution $(c_1, c_2, c_3, \omega_1, \omega_2, \omega_3, p, q)$. On the other hand, corresponding to the other solution we have the straight line

$$(5.22) \quad \begin{aligned} \omega_2 + c_1 + c_3d &= r + b_1 \\ \omega_3 + c_2 + c_3d &= r + b_2 \end{aligned} \quad \begin{aligned} \left(1 + \frac{f}{r}\right)d_1 &= \\ \left(1 + \frac{f}{r}\right)d_2 &= \end{aligned} \quad \begin{aligned} c_3(r + f) &= \\ c_3(r + f) &= \end{aligned}$$

As $f \rightarrow \infty$, the straight line (5.21) tends to the straight line

$$(5.23) \quad \omega_2 r + b_1 = h_1, \omega_3 r + b_2 = h_2, b_3 = b_3^*,$$

where b_3^* is an arbitrary constant. To see (5.23) we need the following lemma.

LEMMA 5.2. *In the (b_3, r) space the straight line $b_3 = c_3(r + f)$ converges to the line $b_3 = b_3^*$ as $f \rightarrow \infty$, where b_3^* is an arbitrary constant.*

Proof. Recall that $d_3 = -c_3 - pc_1$, i.e.,

$$(5.24) \quad (d_3 + pc_1)r + b_3 = -(d_3 + pc_1)f.$$

As $f \rightarrow \infty$, we have $(d_3 + pc_1) \rightarrow 0$ and $(b_3 + (d_3 + pc_1)f) \rightarrow 0$. At a given f , the line (5.24) passes through the point $(0, -f)$ and $(-(d_3 + pc_1)f, 0)$. For large f , the line passes closely through the points $(0, -f)$ and $(b_3^*, 0)$ where b_3^* is a fixed constant, which is also the true value of b_3 . Thus as $f \rightarrow \infty$ the line (5.24) approaches the line $b_3 = b_3^*$. \square

The above calculation can be summarized via the following theorem.

THEOREM 5.3. *Consider the solution vector $(\omega_1, \omega_2, \omega_3, p, q)$ for the recovery equation (5.7). For a given fixed f there are exactly two solutions, one of which remains unchanged as $f \rightarrow \infty$ and the other of which goes off to infinity as described by (5.19), (5.20). For the parameter vector (b_1, b_2, b_3, r) , the recovery equation specifies these parameters up to a choice of two straight lines (5.21) and (5.22). The line (5.21) corresponds to the parameter vector $(\omega_1, \omega_2, \omega_3, p, q)$, which does not change with f . Moreover as $f \rightarrow \infty$, the line (5.21) changes with f and approaches the limit (5.23).*

Remark. It follows from Theorem 5.3 that for large f one recovers (b_1, b_2, r) up to a line given by (5.23) and b_3 exactly.

5.2. Estimation under pseudo-orthographic approximation. Under the pseudo-orthographic approximation, the equation we need to solve for instead of (5.1) is given by

$$(5.25) \quad \begin{aligned} d_1 &= f(\omega_2 + c_1), & d_2 &= f(\omega_3 + c_2), & d_3 &= -(c_3 + pc_1), & d_4 &= \omega_1 - qc_1, \\ d_5 &= -\omega_1 - pc_2, & d_6 &= -(c_3 + qc_2), & h_7 &= \omega_2, & h_8 &= \omega_3. \end{aligned}$$

Let us define T, R, U_0, S as in (5.5) and replace K by K_1 given by $K_1 = \frac{1}{f}(h_7 + ih_8)$. Furthermore let us define P, V, W as in (5.6) and replace L by L_1 given by $L_1 = fK_1 - \frac{1}{f}U_0$. The recovery equation (5.25) can be written as $U_0 = f(V + iW)$, $S = -PV$, $L_1 = -V$, and $-iPV^* = (R + 2\omega_1) + i(T + 2c_3)$, which can be easily solved (see [11]) and the solution is given by

$$(5.26) \quad \begin{aligned} V &= -L_1, \\ P &= S/L_1, \\ \omega_1 &= -[Im(SL_1^*/L_1) + R]/2, \\ W &= i \left(V - \frac{1}{f}U_0 \right). \end{aligned}$$

The following theorem describes an important property of the pseudo-orthographic approximation.

THEOREM 5.4. *The solution (5.26) of the pseudo-orthographic approximation, converges as $f \rightarrow \infty$ to one of the solution of the recovery equation (5.7), described by (5.9), (5.10), (5.11), and (5.16). The solution to which (5.26) converges to is exactly the one which does not change with f .*

Proof of Theorem 5.4. It is easy to see from (5.9), (5.10) that

$$\lim_{c_3 \rightarrow 0} \frac{-L - \sqrt{L^2 - 4c_3 S}}{2} = -L,$$

$$\lim_{c_3 \rightarrow 0} \frac{L - \sqrt{L^2 - 4c_3 S}}{2c_3} = S/L.$$

If $f \rightarrow \infty$ it follows that $c_3 \rightarrow 0$. Thus it may be concluded that if $f \rightarrow \infty$, the solution (5.26) approaches one of the two solutions of the recovery equation (5.7). Finally note that as $f \rightarrow \infty$, (5.26) remains finite. To see this we compute

$$V = -\lim_{f \rightarrow \infty} L_1 = -(h_7 + ih_8),$$

$$P = \lim_{f \rightarrow \infty} S/L_1 = \frac{h_7 + ih_8}{(h_3 - h_6) + i(h_4 + h_5)} \quad (5.27)$$

$$\omega_1 = [Tm(PV^*) - (h_5 - h_4)]/2,$$

$$W = iV.$$

Thus the solution (5.26) to the pseudo-orthographic approximation remains finite and approaches one of the two solutions to the recovery equation (5.7). It follows that it must approach the one which does not change with f because the other solution does not remain finite. \square

Remark. The limiting solution (5.27) is exactly the solution to the recovery equation under orthographic approximation. Such an equation will be given by

$$h_1 = \omega_2 r + b_1, h_2 = \omega_3 r + b_2, h_3 = \omega_2 p, h_4 = \omega_1 + \omega_2 q, h_5 = -\omega_1 + \omega_3 p, h_6 = \omega_3 q, h_7 = \omega_2, \text{ and } h_8 = \omega_3.$$

Remark. The advantage of using pseudo-orthographic approximation as opposed to solving the recovery equation (5.7) is that one needs to solve only linear equations in the former whereas one needs to solve a cubic equation in the latter.

6. Identifiability condition of a planar surface undergoing affine motion

We consider a planar surface undergoing an affine motion and note that the motion of the shape parameters is given by (2.10). In this section we shall consider identifying parameters of (2.10) by considering an output equation given by (3.8). However, since (3.8) is nonlinear in the parameters, we would like to homogenize the vector $(d_1, \dots, d_8)^T$ as follows. Let us define

$$d_j = \frac{y_j}{y_9}, \quad j = 1, \dots, 8 \quad (6.1)$$

so that the vector

$$(y_1, \dots, y_9) \quad (6.2)$$

is a homogenization of the essential parameters. Equation (3.8) can be written as

$$(6.3) \quad \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -b'_1 & 0 \\ 0 & -b'_2 & 0 \\ 0 & -b'_3 & 0 \\ -fb_1 & -fb_2 & -fb_3 \\ a_{11} - a_{33} & -\delta a_{12} & b_3 - \delta a_{22} \\ a_{21} & -\delta a_{21} & b_3 - \delta a_{22} \\ a_{31} & -\delta a_{31} & -\delta a_{32} \\ 1 & -\delta & 0 \end{pmatrix} \begin{pmatrix} p \\ q \\ s \end{pmatrix}$$

