MOTION AND STRUCTURE FROM MOTION IN A PIECEWISE PLANAR ENVIRONMENT

O. D. FAUGERAS and F. LUSTMAN
INRIA
Domaine de Voluceau, Rocquencourt BP 105, 78153 Le Chesnay Cedex, France

Received 16 June 1988

We show in this article that when the environment is piecewise linear, it provides a powerful constraint on the kind of matches that exist between two images of the scene when the camera motion is unknown. For points and lines located in the same plane, the correspondence between the two cameras is a collineation. We show that the unknowns (the camera motion and the plane equation) can be recovered, in general, from an estimate of the matrix of this collineation. The two-fold ambiguity that remains can be removed by looking at a second plane, by taking a third view of the same plane, or by using *a priori* knowledge about the geometry of the plane being looked at. We then show how to combine the estimation of the matrix of collineation and the obtaining of point and line matches between the two images, by a strategy of Hypothesis Prediction and Testing guided by a Kalman filter. We finally show how our approach can be used to calibrate a system of cameras.

Keywords: Motion; Structure from motion; Correspondence problem; Kalman filtering; Projective geometry.

1. INTRODUCTION

In solving the motion problem, one has to answer the following questions:

- which tokens can be computed from the images,
- which constraints can be used to match them between frames,
- how the motions of objects in the scene and the structure of the scene can be robustly recovered from the token matches.

A potentially useful constraint is exemplified by the idea of limiting the geometric complexity of the scenes and assuming that the kinds of surfaces which are present are *a priori* limited. For example, we could assume that they are well approximated by second degree surfaces, i.e. quadrics. Even though this might be an interesting line of approach for a number of industrial scenes, it is appealing to consider first an even simpler case in which the assumption is that surfaces present in the scene are planes. This is quite acceptable for many practical applications of a mobile vehicle in an urban or building environment where streets, walls, ceilings, and floors are fairly common.

This idea has already been exploited in a number of contexts. Assuming that the stereo problem has been solved, Saint-Vincent[1] searches for vertical planes in 3-D data obtained from stereo. Thonnat[2] searches for planes from 3-D data, first using a blind generation of hypotheses and then a phase of classification by an "expert system". For these two authors, the planarity constraint is not used to help solve the stereo matching problem but is used after, to perform piecewise linear approximation on the resulting 3-D data. Lorette and Gaudaire[3], on the other hand, explicitly proposed using the planarity constraint as a

International Journal of Pattern Recognition and Artificial Intelligence Vol. 2 No. 3 (1988) 485–508

guide to determine whether two shapes can be the perspective images of the same planar shape but did not relate that to motion.

The case of the motion problem, and more precisely the optical flow based approach to that problem, has been studied in great detail by Waxman and Ullman[4], Subbarao and Waxman[5], Tziritas[6], and Maybank[7]. The main result is that the motion is ambiguous, i.e. the measurements yield several interpretations of the motion of a planar patch.

For the token matching approach of the same problem, we refer to the very interesting work of the German school, among which is that of Josef Krames[8]. More recently, Tsai and Huang[9] have studied in detail the motion of a planar patch when snapshots are taken at discrete time instants and tokens have been matched between these snapshots. Again, the main result is that the solution is not unique. Longuet-Higgins[10] shows nevertheless that the solution is unique if not all the observed points are closer to one camera than to the other.

Our work extends that of Tsai and Huang in a number of directions. Like all the previous authors, we assume that, in the motion case, only one motion is present; in our case, even though it is not essential in the mathematical proofs, we assume for simplicity that the camera is moving in an otherwise static environment. We make full use of the fundamental property that if two cameras are forming the image of a plane, then the correspondence between the two retinas has a very simple analytical form: it is a collineation, i.e. it is linear in projective coordinates. This implies in particular that, because of the fundamental duality property of projective geometry, points and lines play exactly the same role as tokens. We characterize (like Tsai and Huang) the relationship between this collineation and the geometry of the problem. We then prove (in a somewhat simpler way than Tsai and Huang) that, given the matrix of the collineation, the positions and orientations of the second camera and the plane with respect to the first one can be recovered. We analyze in detail the number of solutions and provide a geometric interpretation of the degenerate cases.

We then study under which conditions the ambiguity of the solutions can be removed: we show that this can be done by looking at another plane or by using a third snapshot of the same plane. We also discuss the problem of estimating the matrix of the collineation from point and line matches and study the stability of the motion and structure parameters computed from this matrix.

We then present an algorithm for obtaining those matches by a technique of Hypothesis Prediction and Testing which makes full use of the analytical correspondence between the two retinas and recursively estimates the matrix of the collineation by Kalman filtering (this is the automation of the matching process). Results are presented on real images.

A slightly more detailed version can be found in Lustman's thesis[11].

2. FUNDAMENTAL THEOREM

Figure 1 illustrates the geometry of the problem. Two cameras form the images of a plane. The position and orientation of the second camera with respect to the first one is defined by the rotation matrix $\boldsymbol{R}$ and the translation vector $\boldsymbol{t}$. The plane is defined by its normal $\boldsymbol{n}$ and its distance d to the origin.

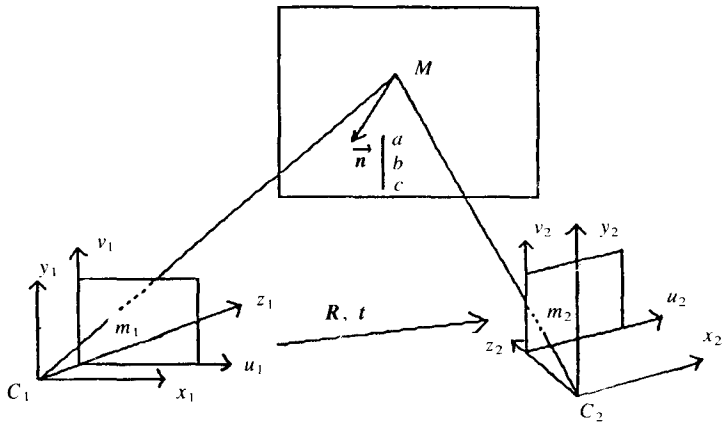


Fig. 1. Geometry of the problem.

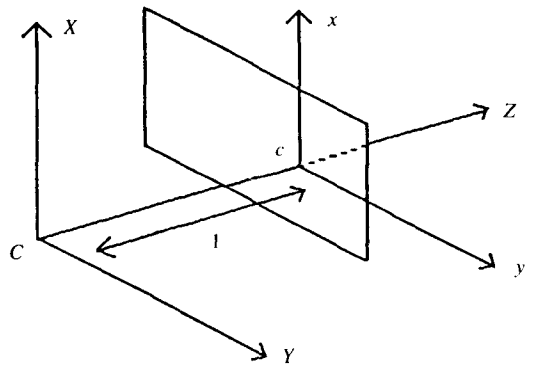


Fig. 2. Perspective transformation.

The coordinate system that we use is the one which can be canonically attached to a camera. It is defined, as shown in Fig. 2, as a coordinate system in which the projection of a point $M(X, Y, Z)$ is the point $m(x, y)$ determined by the relations:

$$X/x = Y/y = Z. \tag{1}$$

It can be shown that such an image coordinate system can be derived from the real image coordinate system (in pixels) by an affine transformation. The basic assumption is that the camera performs a perfect perspective transformation with respect to point C (the camera optic center) at a unit distance from the camera plane. A camera calibration process,

described in Ref. 12, enables us to compute this perspective transformation from 3-D coordinates to 2-D image coordinates in pixels. We are thus able to derive an affine transformation of the 2-D image coordinates such that the perspective transformation is expressed by Eq. (1).

As we show next, the correspondence between the images is a homography. Moreover, determining this homography allows us to explicitly compute the physical parameters:

- the normal $\boldsymbol{n}$ to the plane,
- the rotation $\boldsymbol{R}$,
- the ratio $\frac{\boldsymbol{t}}{d}$ of the translation to the distance of the plane.

We now show the following:

Proposition 1. Let M be a point of a plane $\mathcal{P}$, the equation of $\mathcal{P}$ being $\boldsymbol{n}^t\boldsymbol{X}_1 = d$ in the canonical coordinate system of the first camera. The projections m_1 and m_2 of M on the two retinas are related by a collineation, or homography, depending only on the relative positions of the two retinas and of the plane.

$$m_2 = \boldsymbol{A}m_1 = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} m_1. \tag{2}$$

Moreover, the matrix $\boldsymbol{A}$ is related to the geometric parameters of the problem by the relation:

$$\boldsymbol{A} = d\boldsymbol{R} + \boldsymbol{t}\boldsymbol{n}^t. \tag{3}$$

Proof. The coordinates $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ of M in the two canonical coordinate systems of the cameras are, by definition, related by:

$$\boldsymbol{X}_2 = \boldsymbol{R}\boldsymbol{X}_1 + \boldsymbol{t}. \tag{4}$$

If M belongs to $\mathcal{P}$, defined by its normal $\boldsymbol{n}$ and its distance d in the canonical coordinate system of the first camera, then:

$$\boldsymbol{n}^t\boldsymbol{X}_1 = aX_1 + bY_1 + cZ_1 = d. \tag{5}$$

$\boldsymbol{n}$ is oriented so that it points towards the first camera: $c \leq 0$.

Equations (4) and (5) yield:

$$\boldsymbol{X}_2 = \left(\boldsymbol{R} + \frac{\boldsymbol{t}\boldsymbol{n}^t}{d}\right)\boldsymbol{X}_1.$$

As $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are the projective coordinates of m_1 and m_2 through Eq. (1), the result is proved. □

Notice that (2) is a projective equality, true up to a multiplicative factor. $\boldsymbol{A}$ depends only upon eight independent coefficients, corresponding to the eight geometric unknowns of the problem, i.e. the three rotation parameters, the three coordinates of $\boldsymbol{t}/d$, and the two parameters describing the orientation of the plane.

3. USING LINES

3.1. Affine and Projective Geometries

In planar projective geometry, a point m is described by three coordinates X, Y, Z, not all equal to zero and defined up to a scale factor (i.e. λX, λY, λZ, $\lambda \neq 0$, represents the same point). When we try to relate projective and affine geometries, one of the projective coordinates has to play a special role. If, as it is usually assumed, it is the third coordinate, Z, points such that $Z = 0$ are affine points at infinity. Projectively, they are just the same as any other points. The affine coordinates of a projective point not at infinity can be computed by Eqs. (1).

The equation of a projective line is $\boldsymbol{n}'m = 0$ where $\boldsymbol{n}$ is a 3×1 nonzero vector (a, b, c), therefore:

$$\boldsymbol{n}'m = aX + bY + cZ.$$

The affine equation is obtained by dividing by Z (again points at infinity play a special role in the affine case which they do not project):

$$ax + by + c = 0.$$

Just like a projective point is described by three coordinates not all equal to zero, a projective line is also described by three coordinates not all equal to zero. This is the principle of duality of projective geometry, an extremely powerful tool for solving many geometric problems[13].

3.2. Matching Lines

Matching points or lines between the two images is in fact the same, which is not surprising since, as we just showed, points and lines are dual.

Let, indeed, D_1 be a line on the first retina $\boldsymbol{n}_1'm_1 = 0$ and let D_2 be the corresponding line on the second retina $\boldsymbol{n}_2'm_2 = 0$.

Proposition 2. If we form, in two retinas, the images of lines belonging to the same 3-D plane, the lines being defined by $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ in the images, then:

$$\boldsymbol{n}_1 = \boldsymbol{A}'\boldsymbol{n}_2. \tag{6}$$

Proof. The equation of the first line is $\boldsymbol{n}_1^t m_1 = 0$ and the equation of the second line $\boldsymbol{n}_2^t m_2 = 0$. Since $m_2 = \boldsymbol{A} m_1$, we have $\boldsymbol{n}_2^t \boldsymbol{A} m_1 = 0$, or $(\boldsymbol{A}'\boldsymbol{n}_2)' m_1 = 0$, which proves that $\boldsymbol{n}_1 = \boldsymbol{A}'\boldsymbol{n}_2$. □

Equation 6 is similar to Eq. (2), and can be exploited using the same methods.

4. SOLVING THE DECOMPOSITION PROBLEM

We show how to solve Eq. (3).

Proposition 3. Equation (3) has, in general, 8 different solutions. It has only 4 iff $\boldsymbol{A}$ has a singular value of multiplicity 2. The problem is partially undermined iff $\boldsymbol{A}$ has a singular value of multiplicity 3.

Proof. Using the singular value decomposition (SVD), $\boldsymbol{A}$ can always be decomposed as: $\boldsymbol{A} = \boldsymbol{U}\,\Lambda\boldsymbol{V}'$, Λ being a diagonal matrix, and $\boldsymbol{U}$ and $\boldsymbol{V}$ orthogonal matrices (satisfying $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{V}'\boldsymbol{V} = \boldsymbol{I}$). The diagonal elements of Λ are the square roots of the eigenvalues of $\boldsymbol{A}\boldsymbol{A}'$. These singular values d_i are positive and can be sorted in decreasing order: $d_1 \geq d_2 \geq d_3$.

Using this decomposition, we obtain the new equation:

$$\Lambda = d'\boldsymbol{R}' + \boldsymbol{t}'\boldsymbol{n}'' \tag{7}$$

$\boldsymbol{R}$, $\boldsymbol{t}$ and $\boldsymbol{n}$ being related to $\boldsymbol{R}'$, $\boldsymbol{t}'$ and $\boldsymbol{n}'$ by:

$$\begin{cases} \boldsymbol{R} = s\boldsymbol{U}\,\boldsymbol{R}'\boldsymbol{V}' \\ \boldsymbol{t} = \boldsymbol{U}\boldsymbol{t}' \\ \boldsymbol{n} = \boldsymbol{V}\boldsymbol{n}' \\ d = sd' \\ s = \det \boldsymbol{U} \det \boldsymbol{V}. \end{cases} \tag{8}$$

Notice that $\boldsymbol{R}'$ is a rotation (i.e. $\boldsymbol{R}'\boldsymbol{R}'' = 1$ and $\det \boldsymbol{R}' = 1$).
Using the canonical basis $(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$ and writing $\boldsymbol{n}' = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + x_3\boldsymbol{e}_3$, Eq. (7) gives us three vector equations:

$$d_i\boldsymbol{e}_i = d'\boldsymbol{R}'\boldsymbol{e}_i + \boldsymbol{t}'x_i \text{ for } i = 1, 2, 3. \tag{9}$$

Notice that, since $\boldsymbol{n}$ has a unit norm and $\boldsymbol{V}$ is orthogonal, $\boldsymbol{n}'$ has also a unit norm: $\Sigma_{i=1}^3 x_i^2 = 1$. Eliminating $\boldsymbol{t}'$ finally yields:

$$d'\boldsymbol{R}'(x_j\boldsymbol{e}_i - x_i\boldsymbol{e}_j) = d_i x_j \boldsymbol{e}_i - d_j x_i \boldsymbol{e}_j \text{ for } i \neq j. \tag{10}$$

As $\boldsymbol{R}'$ preserves the vector norm, we obtain the following set of equations:

$$\begin{cases} (d'^2 - d_2^2)x_1^2 + (d'^2 - d_1^2)x_2^2 = 0 \\ (d'^2 - d_3^2)x_2^2 + (d'^2 - d_2^2)x_3^2 = 0 \\ (d'^2 - d_1^2)x_3^2 + (d'^2 - d_3^2)x_1^2 = 0. \end{cases} \tag{11}$$

This can be considered as a linear system in the unknowns x_1^2, x_2^2 and x_3^2. As it must have a nonzero solution, its determinant must be zero:

$$(d'^2 - d_1^2)(d'^2 - d_2^2)(d'^2 - d_3^2) = 0.$$

We therefore obtain different cases, according to the order of multiplicity of the singular values of $\boldsymbol{A}$:

1. $d_1 \neq d_2 \neq d_3$ and then $d' = \pm d_2$.
2. $d_1 = d_2 \neq d_3$ or $d_1 \neq d_2 = d_3$, and $d' = \pm d_2$.
   For this case, we consider only the case $d_1 = d_2$, the case $d_2 = d_3$ being symmetrical.
3. $d_1 = d_2 = d_3$ and $d' = \pm d_2$.

The solutions $d' = \pm d_1$ or $d' = \pm d_3$ are indeed impossible; let us prove it in Case 1, for example. Assuming $d' = d_1$, the Eqs. (11) yield:

$$x_1 = 0 \text{ and } (d_1^2 - d_3^2)x_2^2 + (d_1^2 - d_2^2)x_3^2 = 0.$$

As $d_1 > d_2 > d_3$, this implies $x_2 = x_3 = 0$, which is impossible because $\boldsymbol{n}'$ has a unit norm.

If $d_1 \neq d_3$, we can express x_1, x_2 and x_3 using Eqs. (11) and remembering that $\boldsymbol{n}'$ has a unit norm:

$$\begin{cases} x_1 = \varepsilon_1 \sqrt{\dfrac{d_1^2 - d_2^2}{d_1^2 - d_3^2}} \\ x_2 = 0 \qquad\qquad \varepsilon_1,\ \varepsilon_3 = \pm 1. \\ x_3 = \varepsilon_3 \sqrt{\dfrac{d_2^2 - d_3^2}{d_1^2 - d_3^2}}. \end{cases} \tag{12}$$

Let us now study in detail the three cases. The study is divided into two parts, depending on the sign of d'.

- Case $d' > 0$.

1. Referring to Eqs. (9), we obtain: $\boldsymbol{R}'\boldsymbol{e}_2 = \boldsymbol{e}_2$. $\boldsymbol{R}'$ is, then, a rotation of axis $\boldsymbol{e}_2$.
   We can compute the matrix $\boldsymbol{R}'$ as:

$$\begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix}$$

Using Eqs. (10) and (12), we find:

$$\begin{cases} \sin\ \theta = (d_1 - d_3)\dfrac{x_1 x_3}{d_2} = \varepsilon_1 \varepsilon_3 \dfrac{\sqrt{(d_1^2 - d_2^2)(d_2^2 - d_3^2)}}{(d_1 + d_3)d_2} \\ \cos\ \theta = \dfrac{d_1 x_3^2 + d_3 x_1^2}{d_2} = \dfrac{d_2^2 + d_1 d_3}{(d_1 + d_3)d_2}. \end{cases} \tag{13}$$

Substituting these values into Eqs. (9) yields:

$$\boldsymbol{t}' = (d_1 - d_3)\begin{pmatrix} x_1 \\ 0 \\ -x_3 \end{pmatrix} \tag{14}$$

2. In this case, we obtain $x_1 = x_2 = 0$ and $x_3 = \pm 1$.
We then get:

$$\begin{cases} \boldsymbol{R}' = \boldsymbol{I} \\ \boldsymbol{t}' = (d_3 - d_1)\boldsymbol{n}'. \end{cases}$$

3. x_1, x_2 and x_3 are undefined and the Eqs. (10) and (9) provide us:

$$\begin{cases} \boldsymbol{R}' = \boldsymbol{I} \\ \boldsymbol{t}' = 0. \end{cases}$$

In this case, the motion is a pure rotation and the normal remains undefined.

- The same methods allow us to deal with the case $d' < 0$.

1. The Eqs. (9) yield $\boldsymbol{R}'\boldsymbol{e}_2 = -\boldsymbol{e}_2$, which implies that $\boldsymbol{R}'$ is a symmetry (i.e. a rotation of angle π), with respect to an axis v perpendicular to $\boldsymbol{e}_2$, i.e. in the plane $(\boldsymbol{e}_1, \boldsymbol{e}_3)$. Let $\varphi/2$ be the angle between $\boldsymbol{e}_1$ and v.
The rotation $\boldsymbol{R}'$ can be written as:

$$\begin{pmatrix} \cos\ \varphi & 0 & \sin\ \varphi \\ 0 & -1 & 0 \\ \sin\ \varphi & 0 & -\cos\ \varphi \end{pmatrix}$$

with

$$\begin{cases} \sin\ \varphi = \dfrac{d_1 + d_3}{d_2} x_1 x_3 = \varepsilon_1 \varepsilon_3 \dfrac{\sqrt{(d_1^2 - d_2^2)(d_2^2 - d_3^2)}}{(d_1 - d_3)d_2} \\ \cos\ \varphi = \dfrac{d_3 x_1^2 - d_1 x_3^2}{d_2} = \dfrac{d_1 d_3 - d_2^2}{(d_1 - d_3)d_2}. \end{cases} \tag{15}$$

Substituting these values into Eqs. (9) yields:

$$\boldsymbol{t}' = (d_1 + d_3)\begin{pmatrix} x_1 \\ 0 \\ x_3 \end{pmatrix}. \tag{16}$$

2. We have, once more, $x_1 = x_2 = 0$ and $x_3 = \pm 1$.
$\boldsymbol{R}'$ is therefore a symmetry with respect to $\boldsymbol{e}_3$:

$$\begin{cases} \boldsymbol{R}' = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ \mathrm{t}' = (d_3 + d_1)\boldsymbol{n}'. \end{cases}$$

3. The Eq. (7) imply that $\boldsymbol{R}'\boldsymbol{x} = -\boldsymbol{x}$ for all vectors $\boldsymbol{x}$ in the plane orthogonal to $\boldsymbol{n}'$. $\boldsymbol{R}'$ is therefore a symmetry with respect to the axis $\boldsymbol{n}'$. Therefore, $\boldsymbol{R}' = -\boldsymbol{I} + 2\boldsymbol{n}'\boldsymbol{n}'^t$ and, according to Eq. (7), $\boldsymbol{t}' = -2d'\boldsymbol{n}'$.

□

Since we have *a priori* no way of knowing the sign of d', we have a total of eight solutions when the singular values are distinct, four solutions when two singular values are equal, and an indetermination when the three are equal. But among these eight solutions to the problem, only two are possible if we go back to the physical interpretation of the results.

Proposition 4. The observed points being seen by the two cameras, the decomposition problem has in fact only 2 physical solutions in the general case, and 1 solution when there is a double singular value.

Proof. Let M be a point in the plane, visible by the two cameras, and Z_1 and Z_2 its z-coordinates in the canonical coordinate systems of Fig. 1. The constraint of visibility by the two cameras enforces $Z_1 > 0$ and $Z_2 > 0$. Combining Eqs. (1) and (2) we obtain $\alpha\boldsymbol{X}_2 = \boldsymbol{A}\boldsymbol{X}_1$ for some real number α. Using Eqs. (4) and (3), we then obtain $\alpha = d$, so that

$$\frac{Z_2}{Z_1} = \frac{a_{31}x_1 + a_{32}y_1 + a_{33}}{d} > 0.$$

Z_1 and Z_2 keep a constant sign when M varies in the plane, and d is constant, so $a_{31}x_1 + a_{32}y_1 + a_{33}$ also keeps a constant sign when M varies in the plane. The knowledge of one point therefore determines the sign of d. As $d = sd'$ and since s is determined by $\boldsymbol{A}$, we have only four solutions, depending on the sign of d'.

The corresponding solutions for the normal $\boldsymbol{n}$ are $\boldsymbol{n}_1$, $-\boldsymbol{n}_1$, $\boldsymbol{n}_2$ and $-\boldsymbol{n}_2$ (see Eqs. 12). $\boldsymbol{X}_1$ belongs to the plane so that $\boldsymbol{n}^t\boldsymbol{X}_1 = d$. Using Eqs. (1), writing $Z_1 > 0$ therefore yields:

$$\frac{\boldsymbol{n}^t\boldsymbol{m}_1}{d} > 0$$

which leaves only two solutions among the four. □

The previous result comes from writing the visibility constraint for one point. What happens if we write it for all the observed points? Longuet-Higgins[10] shows the following very interesting property:

Proposition 5. If not all the observed points are closer to one of the optical centers than to the other, then only one solution satisfies the visibility conditions for all points.

To summarize the process, the knowledge of $\boldsymbol{A}$ gives eight solutions. Writing the visibility constraint for one point leaves two solutions. If the points are not all closer to one of the cameras than to the other, writing the visibility constraint leaves only one solution.

Notice, finally, the duality found by Waxman and Ullmann[4] or Subbarao and Waxman[5] in the case of optic flow, between the translation and the normal in the SVD coordinate system: in this coordinate system, the normal and the translation play symmetrical roles and share the same directions. If d' is positive, for instance, the two solutions $(\boldsymbol{n}'_1, \boldsymbol{t}'_1)$ and $(\boldsymbol{n}'_2, \boldsymbol{t}'_2)$ are related by:

$$\begin{cases} \boldsymbol{t}'_2 = (d_1 - d_3)\boldsymbol{n}'_1 \\ \boldsymbol{n}'_2 = \dfrac{\boldsymbol{t}'_1}{d_1 - d_3} \end{cases}$$

5. THE DEGENERATE CASES

Let us now study the degenerate cases, i.e. the cases when there are singular values with an order of multiplicity greater than 1.

Proposition 6. $\boldsymbol{A}$ has 2 equal singular values iff $\boldsymbol{t} \wedge \boldsymbol{Rn} = 0$, which means that the translation is normal to the plane, as shown in Fig. 3.

$\boldsymbol{A}$ has 3 equal singular values iff either the translation is zero, i.e. in the case of pure rotation, or $\boldsymbol{t} = -2\boldsymbol{Rn}$, i.e. a transparent plane is observed from the two opposite sides and at the same distance, as in Fig. 4.

Proof.

- Double singular value: the singular values are the square roots of the eigenvalues of $\boldsymbol{AA}^t$. We have $\boldsymbol{AA}^t\boldsymbol{x} = d^2\boldsymbol{x} + (\boldsymbol{t} \cdot \boldsymbol{x})\boldsymbol{t} + d(\boldsymbol{Rn} \cdot \boldsymbol{x})\boldsymbol{t} + d(\boldsymbol{t} \cdot \boldsymbol{x})\boldsymbol{Rn}$. If $\boldsymbol{t} = \lambda\boldsymbol{Rn}$, it is clear that the plane orthogonal to $\boldsymbol{t}$ is a plane of eigenvectors of $\boldsymbol{AA}^t$ for the eigenvalue d^2.

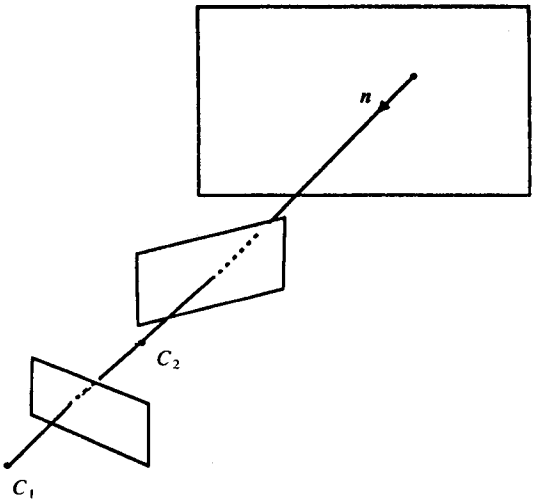


Fig. 3. An example of double degeneracy.

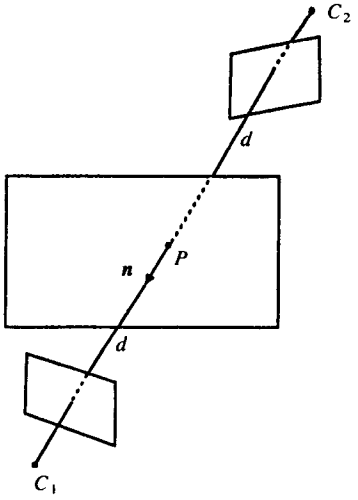


Fig. 4. An example of triple degeneracy.

Reciprocally, if there is a double singular value, we have already proved that:
— if $d' > 0$, $\boldsymbol{R}' = \boldsymbol{I}$ and $\boldsymbol{t}' = \pm(d_3 - d_1)\boldsymbol{n}'$, so that the Eqs. (8) yield:

$$\boldsymbol{R}\boldsymbol{n} = s\boldsymbol{U}\ \boldsymbol{R}'\boldsymbol{V}'\boldsymbol{V}\ \boldsymbol{n}' = \frac{s}{d_3 - d_1}\boldsymbol{U}\boldsymbol{t}' = \frac{s}{d_3 - d_1}\boldsymbol{t}.$$

— *if* $d' < 0$, $\boldsymbol{R}'\boldsymbol{e}_3 = \boldsymbol{e}_3$ and $\boldsymbol{t}' = (d_1 + d_3)\boldsymbol{n}'$, so that the same equations as above show the result.

- Triple singular value:
  If d' is positive, we have shown that $\boldsymbol{t}' = 0$, and therefore $\boldsymbol{t} = 0$.
  Reciprocally, if $\boldsymbol{t} = 0$ then $\boldsymbol{A} = d\boldsymbol{R}$ and $\boldsymbol{A}\boldsymbol{A}^t = d^2\boldsymbol{I}$ and all the singular values are equal to d.
  If d' is negative, we have shown that $\boldsymbol{t}' = -2d'\boldsymbol{n}'$ and that $\boldsymbol{R}'\boldsymbol{n}' = \boldsymbol{n}'$. It follows that $\boldsymbol{t} = \boldsymbol{U}\boldsymbol{t}' = -2d'\boldsymbol{U}\,\boldsymbol{R}'\boldsymbol{n}' = -2sd\boldsymbol{U}\,\boldsymbol{R}'\boldsymbol{V}^t\boldsymbol{n} = -2d\boldsymbol{R}\boldsymbol{n}$.
  Reciprocally, if $\boldsymbol{t} = -2d\boldsymbol{R}\boldsymbol{n}$, then $\boldsymbol{A} = d\boldsymbol{R} - 2d\boldsymbol{R}\boldsymbol{n}\boldsymbol{n}^t = d\boldsymbol{R}(\boldsymbol{I} - 2\boldsymbol{n}\boldsymbol{n}^t)$ and $\boldsymbol{A}^t\boldsymbol{A} = d^2\boldsymbol{I}$.

□

6. SOLVING THE AMBIGUITY PROBLEM

In the case where two solutions remain, what can be done to choose the right solution.

In the absence of *a priori* information to guide our choice, we can think of three ways to proceed:

- look at a second plane.
- use a third image.
- use known geometric relationships between tokens in the plane, for example, that two lines are orthogonal.

In the first two cases, we obtain two pairs of solutions (S_1, S_2) and (S_1', S_2'), and we must then find a compatible pair (S_i, S_j'), i.e. find a common plane equation if we look at a single plane from three positions, or find a common motion if we look at two planes from two positions. *In general,* there is only one compatible pair, and the problem therefore has a unique solution.

In the third case, let $(\boldsymbol{n}_1, \boldsymbol{n}_1')$ and $(\boldsymbol{n}_2, \boldsymbol{n}_2')$ be two pairs of corresponding lines between two images, given by their equations. Let $\boldsymbol{R}_1$ and $\boldsymbol{R}_2$ be the two solutions for the rotation. The direction of the reconstructed lines for a rotation $\boldsymbol{R}$ is $\boldsymbol{n}_1 \wedge \boldsymbol{R}^t\boldsymbol{n}_1'$ and $\boldsymbol{n}_2 \wedge \boldsymbol{R}^t\boldsymbol{n}_2'$. If we know the angle between these two vectors, checking this angle for $\boldsymbol{R} = \boldsymbol{R}_1$ and $\boldsymbol{R} = \boldsymbol{R}_2$ will, in general, give the right solution.

7. ESTIMATING THE TRANSFORMATION MATRIX

Matrix $\boldsymbol{A}$ is estimated by matching points or lines between the two images. In order to fully determine a planar projective transformation, we need at least four projectively independent points, i.e. such that any three of them are not aligned. In fact, in the process of finding planes we match a large number of points and lines and because of measurement errors and matching errors they do not always correspond to actual coplanar points or lines. Therefore we use mean square techniques.

Using Eqs. (2) to match points $m_1(x_1, y_1)$ and $m_2(x_2, y_2)$ yields the two equations:

$$\begin{cases} a_{11}x_1 + a_{12}y_1 + a_{13} - a_{31}x_2x_1 - a_{32}x_2y_1 - a_{33}x_2 = 0 \\ a_{21}x_1 + a_{22}y_1 + a_{23} - a_{31}y_2x_1 - a_{32}y_2y_1 - a_{33}y_2 = 0. \end{cases} \tag{17}$$

For each pair of matched points between the two images, we obtain two such equations.

As A is defined up to a scale factor, we must impose a constraint over the coefficients. A simple condition is $a_{33} = 1$.

We thus obtain a linear system in eight unknowns a_{ij} (excluding a_{33}), and we can solve these equations using mean square techniques or Kalman filtering, as described in the appendix. This yields matrix A, from which the motion parameters and the plane equation can be deduced. The advantage of using the Kalman filter is that A can be estimated recursively (see Sect. 9) while taking into account pixel uncertainty.

Using Eq. 6, a similar technique can be applied to a line.

8. CONDITIONING OF THE PROBLEM

If we are to use this method with real data, we must study its sensitivity to noise, i.e.

- sensitivity of the estimation of A in the presence of noisy pixel data.
- sensitivity of the estimation of the physical parameters from A when A is noisy.

8.1. Estimation of A

Little is to be said about this part of the problem. The estimation of A implies the inversion of an 8×8 matrix (if we use a simple least-squares technique). The problem is thus conditioned by the proximity of a singularity of this matrix, which is regular if and only if the points used are projectively independent. Therefore, the less triplets of points are aligned and the more distant the points are from each other, the better the problem will be conditioned.

8.2. Estimation of the Physical Parameters from A

The main step in estimating the physical parameters from A is the singular value decomposition. The SVD is quite stable with respect to perturbations, as is demonstrated by two classic theorems on the stability of singular elements with respect to perturbations, that can be found, mathematically stated, in Ref. 14. We prefer to give only their English interpretation.

The first theorem states that singular values are stable under a perturbation in the sense that a variation of ε of the coefficients of A induces an equivalent variation of the singular values.

The second theorem states that this same variation of ε of the coefficients of A induces a variation ε/δ of the singular subspaces, δ measuring the isolation of the singular values.

It follows from the previous theorems that the problem will be all the more stable as we will be far from the degenerate configurations, i.e.

- translation orthogonal to the plane
- no translation.

This property has been checked experimentally on synthetic data in a typical case when we suppose that we are looking at a frontal plane, with the camera rotating 30 degrees around a vertical axis, and translating horizontally. A uniform noise is added to A, and Fig. 5 shows the error on the angle of rotation as a function of the direction of the

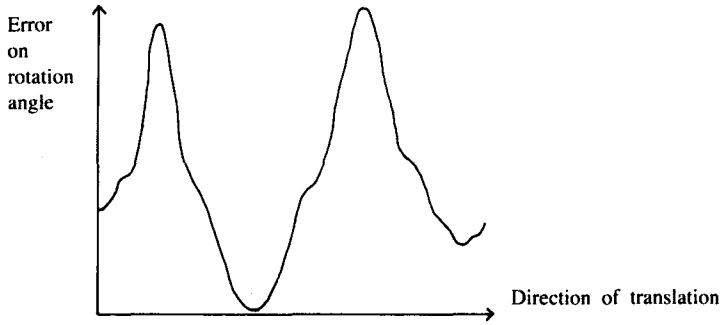


Fig. 5. Variation of the stability of the solution on a noisy matrix A as a function of the direction of translation: the two peaks correspond to the normal to the plane.

translation in the horizontal plane: we can notice two peaks, corresponding to the direction of the normal to the plane.

9. FINDING STRUCTURE AND MOTION BY PREDICTION AND VERIFICATION

We now have all the tools for explaining the world in terms of planes. The method that we propose is inspired from the one used in Refs. 15–17 and has two main steps:

- Hypothesis formation: find n geometric primitives, points or lines, in image 1 and image 2 and suppose they are the images of n coplanar primitives. In order to estimate $\boldsymbol{A}$, we need $n \geq 4$, with at least four primitives projectively independent. Notice that any four pairs of primitives provide a solution, and only the way the different hypotheses propagate discriminates the right hypotheses from the wrong.
- Hypothesis verification: we can then use the knowledge of matrix $\boldsymbol{A}$ as a constraint to find further matches. Indeed, if m_1 (resp. $\boldsymbol{n}_1$) is a point (resp. line) in the first image which is the image of a point (resp. line) in the same plane as the first n geometric primitives matched so far, then there exists in general a point m_2 (resp. line $\boldsymbol{n}_2$) in the second image such that $m_2 = \boldsymbol{A}m_1$ (resp. $\boldsymbol{n}_1 = \boldsymbol{A}'\boldsymbol{n}_2$). Knowing $\boldsymbol{A}$ allows us to restrict considerably the number of potential matches of m_1 (resp. $\boldsymbol{n}_1$), and in turn each further match can be used to refine the estimation of $\boldsymbol{A}$ recursively using the Kalman filtering technique, described in the appendix, and therefore the estimation of the physical parameters. The criterion used for matching is the Mahalanobis distance between a predicted point and a potential matching point, which is an euclidean distance weighted by the covariances both on the measurement of image points and on the transformation $\boldsymbol{A}$.

  If no match can be found, then this means that the current hypothesis is not correct and another one must be tried.

Notice that, in the general case, the problem is simplified when the observed scene contains several planar surfaces and we have found one: we now know the motion between the two positions of the camera, so that we fall in the case of stereovision rather than pure motion. We therefore can use the epipolar constraint to simplify the initial correspondence problem in the following way: matching *one* pair of lines between the two images can provide *two* pairs of matching points, by taking two points on the line in the

first image and matching them with the intersections of the line in the second image and the epipolar lines of the two points. It is of no use to take more than two points since three aligned points are not projectively independent.

9.1. Updating the World Model

Since our goal is to establish a 3-D description of the environment in which a mobile vehicle moves, two key problems are to obtain explicit information about the uncertainty in the world model used by the robot and to be able to update this information by combining a large number of measures from different sensors. These ideas are fully developed in Refs. 18 and 19.

In accordance with this approach, the updating of matrix A is made through the use of a Kalman filter: using an estimate of the desired parameters along with the covariance matrix of those elements, we can obtain a new estimate with its covariance matrix. We are thus able to relate the uncertainty of the world model to the pixel uncertainty, and to incorporate any previous information on the value or the precision of a parameter via the initial estimate.

The possibility of incorporating *a priori* information is of great value for finding a strategy of image analysis: commonsense knowledge about the world can be used to guide the system and help it converge more rapidly. Since we know, for example, that horizontal planes such as a floor and a ceiling are likely to be found in the image at a known height, why not first look for them, refining at the same time the estimation of the motion parameters? The initial estimates may also be the result of previous computations, the motion parameters being approximately known through proprioceptive sensors or because a plane has already been detected during the analysis of the last pair of images.

Let us be a little more specific about using *a priori* knowledge on an example: if we want to specify that the axis of rotation u is vertical, then we will give an *a priori* value of the rotation meeting that condition $u = (001)$, and we want to specify that the x and y coordinates must remain small. It follows from that requirement that the covariance matrix on the axis of rotation will be:

$$S_{\text{vert}} = \begin{pmatrix} \varepsilon & o & o \\ 0 & \varepsilon & 0 \\ 0 & 0 & K \end{pmatrix}$$

where $\varepsilon \ll K$. Using Eq. (3), we can derive a covariance matrix on A from the covariance matrices on the physical parameters (by linearizing to the first order), and this covariance matrix will contain the information on the verticality of the axis, thus inhibiting the variation of the horizontal coordinates of the axis.

Similarly, information concerning the normal to the plane or the translation can be injected in the Kalman filter.

10. RESULTS

The first thing to check is that we can discriminate the points lying in a plane from all the other points. Figure 6 shows two photographs of a scene containing two main vertical

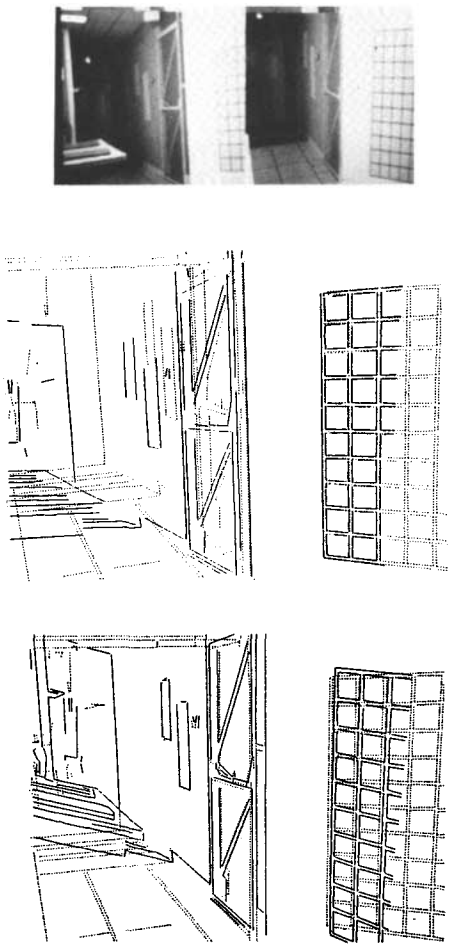


Fig. 6. The knowledge of A allows discrimination of the points belonging to the planes.

planes. It then shows the superpositions of predicted segments on the second image, in the two cases of propagation. In both cases, we can see that the superposition is very good for the segments of the corresponding plane, and bad for the other segments.

Let us now show that the main structures of the environment can be found. Figure 7 shows the photographs of a corridor, and the three planes that are found (the floor and the two walls): the crosses at the endpoints of the segments show the matched segments.

Finally, we have considered a stereo pair of the calibration grid. We therefore know accurately the relative position of the cameras, so that we can compare it to the motion obtained by matching corresponding segments. We show in Fig. 8 the superposition of the transformed first image onto the second image, as well as the (very accurate) numerical results. The right solution is obviously the second one.

11. CALIBRATION

We have also used the previous approach to calibrate our camera system. The calibration process can be decomposed into two steps:

- estimation of the intrinsic parameters of each camera: we assume this has already been done and refer the reader to Ref. 12.
- estimation of the extrinsic parameters, i.e. the relative positions of the cameras. This can be easily done by looking at a planar rectangular grid of known size (see Fig. 9), using our paradigm. The points at the intersection of the horizontal and vertical lines are extracted from the images of the grid, and are then matched between the images. We thus obtain two solutions for each pair of cameras. We can find the right

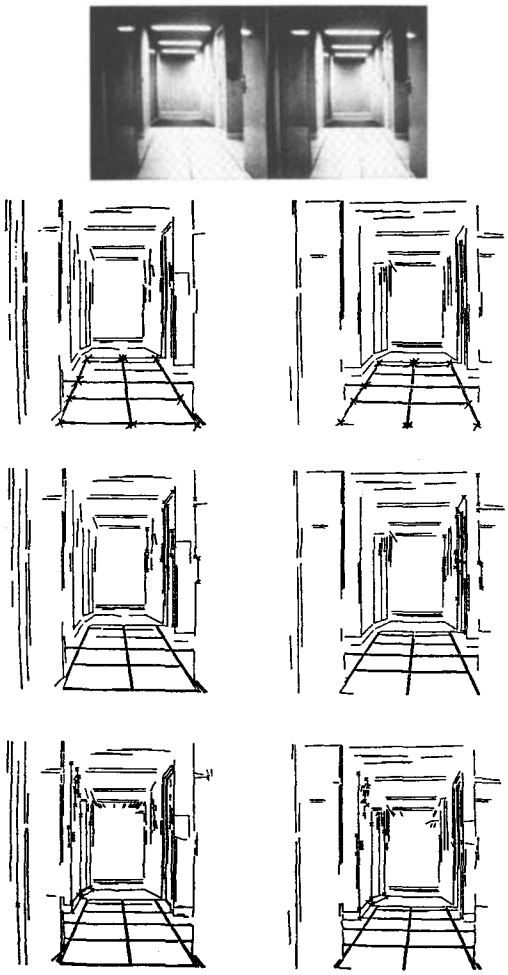
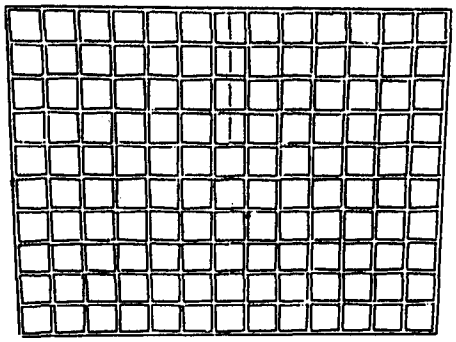


Fig. 7. In these images of a corridor, we were able to find the main structures (segments ended by crosses).

Superposition of the transformed first image and the second.



Two solutions of motion from planarity.

Angle of rotation in degrees	Axis of rotation			Direction of translation			Normal to the plane		
2.977515	0.332	−0.023	0.943	0.007	0.017	−0.143	0.404	−0.913	−0.051
9.773	0.864	0.348	0.365	0.082	0.032	−0.996	0.065	−0.127	−0.020

Motion from calibration.

Angle of rotation in degrees: 9.705
Axis of rotation: 0.865 0.338 0.369
Translation: −22.732 47.577 7.384

Fig. 8. Comparison of estimated and calibration results.

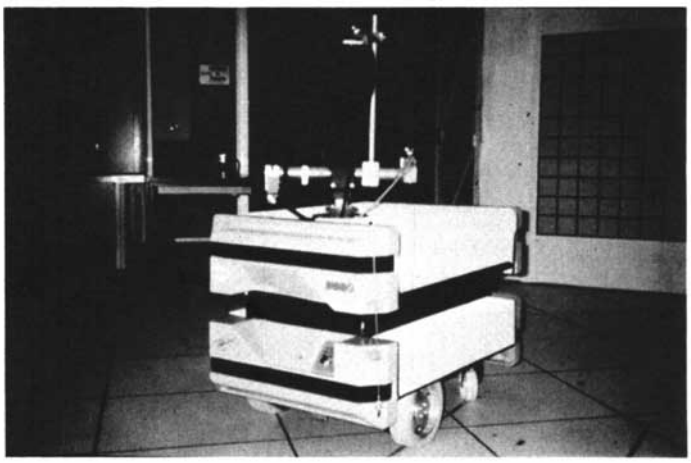


Fig. 9. The calibration grid and the trinocular system.

solution, either by using the three cameras if we are calibrating a trinocular system, or by using the knowledge we have about the angles of the edges on the grid. The scaling factor is then recovered using the knowledge of the absolute size of the grid.

This provides a fairly easy way to calibrate a camera system and thus obtain the epipolar geometry. It has, moreover, proved to be very accurate, and has been successfully used in stereo experiments.

12. CONCLUSION

We have shown, in this article, how a hypothesis on the kinds of structure present in the environment can be used to constrain the matching process in motion analysis. We have pushed the idea of a piecewise planar environment quite far using ideas from projective geometry and combining them with nonlinear filtering techniques and strategies for testing and verifying hypotheses.

This approach may perhaps be extended to other types of surfaces, like quadrics.

REFERENCES

1. A. R. Saint-Vincent, "Perception et modélisation de l'environnement d'un robot mobile: une approche par stéréovision," Thèse de l'Université Paul Sabatier, Toulouse, Novembre 1986.
2. Monique Thonnat, "Semantic interpretation of 3-D stereo data: finding the main structures," in *Proc. Int. Conf. on Pattern Recognition,* Oct. 1986, pp. 1051–1054.
3. G. Lorette and M. Gaudaire, "Methodes de reconnaissance d'images utilisables en l'absence de modèles d'objets et de référentiel unique," in *Actes 3^e congrès AFCET-INRIA Reconnaissance des Formes et Intelligence Artificielle,* Nancy, 1981, pp. 349–351.
4. A. Waxman and S. Ullman, "Surface structure and 3-D motion from image flow: planar surface in motion," *Int. J. of Robotics Research* **4** (1985) 95–108.
5. M. Subbarao and A. Waxman, "On the uniqueness of image flow solutions for planar surfaces in motion," *Computer Vision, Graphics and Image Processing* **36** (1986) 208–228.
6. G. Tziritas, "Gradient-based optical flow estimation," *Traitement du Signal* **3:1** (1986) 3–11.
7. S. Maybank, "The angular velocity associated with the optical flow field due to a rigid moving body," *Proc. R. Soc. Lond.,* **A401** (1985) 317–326.
8. Josef Krames, "Zur Ermittlung eines Objektes aus zwei Perspektiven," *Monatsh. Math. Physik* **49** (1940) 327–354.
9. R. Tsai and T. S. Huang, "Estimating three-dimensional motion parameters of a rigid planar patch, II: singular value decomposition", *IEEE Trans. on Acoustics, Speech and Signal Processing* **30** (1982).
10. H. C. Longuet-Higgins, "The reconstruction of a plane surface from two perspective projections," *Proc. Royal Society, London* **227** (1986) 399–410.
11. Francis Lustman, "Vision stéréoscopique et perception du mouvement en vision artificielle," Thèse de l'université de Paris-Sud, Dec. 1987.
12. O. D. Faugeras and G. Toscani, "Camera calibration for 3-D machine vision," in *Proc. Int. Workshop on Machine Vision and Machine Intelligence,* Tokyo, Japan, Feb. 1987, pp. 240–247.
13. J. G. Semple and G. T. Kneebone, *Algebraic Projective Geometry,* Oxford University Press, London, 1952.
14. G. H. Golub and C. F. Van Loan, *Matrix Computations,* Johns Hopkins University Press, 1983.

15. N. Ayache and O. D. Faugeras, "Hyper: a new approach for the recognition and positioning of two-dimensional objects," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **8:1** (1986) 44–54.
16. O. D. Faugeras and M. Hebert, "The representation, recognition, and locating of 3-D shapes from range data," *The Int. J. of Robotics Research* **5:3** (1986) 27–52.
17. N. Ayache and B. Faverjon, "Efficient registration of stereo images by matching graph descriptions of edge segments," *The Int. J. of Computer Vision* **1:2** (1987).
18. O. D. Faugeras, N. Ayache, and B. Faverjon, "Building visual maps by combining noisy stereo measurements," in *Proc. Int. Conf. on Robotics and Automation*, San Francisco, CA, Apr. 1986, pp. 1433–1438.
19. N. Ayache and O. D. Faugeras, "Maintaining representations of the environment of a mobile robot," in *Int. Symposium on Robotics Research*, Santa-Cruz, California, Aug. 1987.
20. A. M. Jazwinsky, *Stochastic Processes and Filtering Theory*, Academic Press, 1970.

APPENDIX: KALMAN FILTER EQUATIONS

We use the Kalman filter as a general tool for computer vision problems. The interested reader is referred to Ref. 20 for a general and detailed presentation of the Kalman filter, and to Ref. 19 for some applications.

Let us summarize what is useful for our application.

A.1. The Problem

We are confronted with the estimation of an unknown parameter $a \in \mathbb{R}^n$ given a set of k not necessarily linear equations of the form

$$f_i(x_i, a) = 0 \tag{18}$$

where $x_i \in \mathbb{R}^m$ and f_i is a function from $\mathbb{R}^m \times \mathbb{R}^n$ into $\mathbb{R}^p$. The vector x_i represents some random measurements in the sense that we only measure an estimate $\hat{x}_i$ of them, such that

$$\hat{x}_i = x_i + v_i \tag{19}$$

where v_i is a random error. The only assumption we make on v_i is that its mean is zero, its covariance is known, and that it is a white noise:

$$E[v_i] = 0$$
$$E[v_i v_i^t] = \Lambda_i \geq 0$$
$$E[v_i v_j^t] = 0 \ \forall \ i \neq j .$$

These assumptions are reasonable. If the estimator is biased, it is possible to subtract its mean to get an unbiased one. If we do not know the covariance of the error (or some other confidence measure on it), the estimator is meaningless. If two measurements $\hat{x}_i$ and $\hat{x}_j$ are correlated, we take the concatenation of them $\hat{x}_k = (\hat{x}_i, \hat{x}_j)$ and the concatenated vector function $f_k = [f_i^t, f_j^t]^t$. The problem is to find the optimal estimate $\hat{a}$ of a given the functions f_i and the measurements $\hat{x}_i$.

A.2. Linearizing the Equations

The most powerful tools developed in parameter estimation are for linear systems. Before using complicated nonlinear optimization techniques, it is worthwhile to try applying the linear tools to a linearized version of our equations. This is the Extended Kalman Filtering approach that we now develop.

For each nonlinear equation $f_i(x_i, a) = 0$ we need to know an estimate $\hat{a}_{i-1}$ of the solution a, and again a measure S_i of the confidence we have in this estimate[a]. We assume that $\hat{a}_{i-1}$ is given by

$$\hat{a}_{i-1} = a + w_i \tag{20}$$

where w_i is a random error. The only assumptions we make on w_i are the same as for v_i, i.e.

$$E[w_i] = 0$$

$$E[w_i \; w_i^t] = S_i \geq 0$$

where S_i is a given positive matrix. Here again, no assumption of Gaussianness is required.

Having an estimate $\hat{a}_{i-1}$ of the solution, the equations are linearized by a first order Taylor expansion around $(\hat{x}_i, \hat{a}_{i-1})$:

$$f_i(x_i, a) = 0 \approx f_i(\hat{x}_i, \hat{a}_{i-1}) + \frac{\widehat{\partial f_i}}{\partial x}(x_i - \hat{x}_i) + \frac{\widehat{\partial f_i}}{\partial a}(a - \hat{a}_{i-1}) \tag{21}$$

where the derivatives $\widehat{\partial f_i/\partial x}$ and $\widehat{\partial f_i/\partial a}$ are estimated at $(\hat{x}_i, \hat{a}_{i-1})$:

Equation 21 can be rewritten as:

$$y_i = M_i a + u_i \tag{22}$$

where

$$y_i = -f_i(\hat{x}_i, \hat{a}_{i-1}) + \frac{\widehat{\partial f_i}}{\partial a}\hat{a}_{i-1}$$

$$M_i = \frac{\widehat{\partial f_i}}{\partial a}$$

$$u_i = \frac{\widehat{\partial f_i}}{\partial x}(x_i - \hat{x}_i).$$

[a]In practice, we shall see that only an initial estimate $(\hat{a}_0, S_0)$ of a is required prior to the first measurement $\hat{x}_1$, while the next ones $(\hat{a}_i, S_i)$ are provided automatically by the Kalman filter itself.

Equation 22 is now a linear measurement equation, where y_i is the new measurement, M_i is the linear transformation, u_i is the random measurement error. Both y_i and M_i are readily computed from the actual measurement $\hat{x}_i$, the estimate $\hat{a}_{i-1}$ of a, the function f_i and its first derivative. The second-order statistics of u_i are derived easily from those of v_i:

$$E[u_i] = 0$$
$$W_i \triangleq E[u_i u_i^t] = \frac{\partial \widehat{f_i}}{\partial x} \Lambda_i \frac{\partial \widehat{f_i^t}}{\partial x}.$$

A.3. Recursive Kalman Filter

When no Gaussianness is assumed on the previous random errors u_i, v_i and w_i, the Kalman filter equations provide the best (minimum variance) linear unbiased estimate of a. This means that among the estimators which seek $\hat{a}_k$ as a linear combination of the measurements $\{y_i\}$, it is the one which minimizes the expected error norm

$$E[(\hat{a}_k - a)^t(\hat{a}_k - a)]$$

while verifying

$$E[\hat{a}_k] = a.$$

The recursive equations of the Kalman filter which provide a new estimate $(\hat{a}_i, S_i)$ of a from $(\hat{a}_{i-1}, S_{i-1})$ are the following ones[20]:

$$\hat{a}_i = \hat{a}_{i-1} + K_i(y_i - M_i\hat{a}_{i-1}) \tag{23}$$

$$K_i = S_{i-1}M_i^t(W_i + M_iS_{i-1}M_i^t)^{-1} \tag{24}$$

$$S_i = (I - K_iM_i)S_{i-1}$$

or equivalently

$$S_i^{-1} = S_{i-1}^{-1} + M_i^tW_i^{-1}M_i. \tag{26}$$

One can see that the previously estimated parameter $\hat{a}_{i-1}$ is corrected by an amount proportional to the current error $y_i - M_i\,\hat{a}_{i-1}$ called the innovation. The proportionality factor K_i is called the Kalman gain. At the end of the process, $\hat{a}_k$ is the final estimate and S_k represents the covariance of the estimation error:

$$S_k = E[(\hat{a}_k - a)(\hat{a}_k - a)^t].$$

The recursive process is initialized by $\hat{a}_0$, an initial estimate of a, and S_0, its error covariance matrix. Actually, the criterion minimized by the final estimate $\hat{a}_k$ is:

$$C = (a - \hat{a}_0)' S_0^{-1}(a - \hat{a}_0) + \sum_{i=1}^{k} (y_i - M_i a)' W_i^{-1}(y_i - M_i a). \tag{27}$$

It is interesting to note that the first term of Eq. 27 measures the squared distance of a from an initial estimate weighted by its covariance matrix, while the second term is nothing else than the classical least-square criterion, i.e. the sum of the squared measurement errors weighted by their covariance matrices. Indeed, initializing the process with an arbitrary $\hat{a}_0$ and $S_0^{-1} = 0$, criterion 27 provides the classical least-square estimate $\hat{a}_k$ obtained from the measurements only, while the initial estimate does not play any role.

The enormous advantage of such a recursive solution is that if we decide, after a set of k measurements $\{x_i\}$, to stop the measures, we only have to keep $\hat{a}_k$ and S_k as the whole memory of the measurement process. If we decide later to take into account additional measurements, we simply have to initialize $\hat{a}_0 \equiv \hat{a}_k$ and $S_0 \equiv S_k$ and to process the new measurements to obtain exactly the same solution as if we had processed all the measurements together.

A.4. Gaussian Assumption

Up to now, we did not introduce any Gaussian assumption on the random measurement errors $v_i = x_i - \hat{x}_i$ of Eq. 19 and on the prior estimate error $w_0 = a - \hat{a}_0$ of Eq. 20. However, in practice, these errors usually come from a sum of independent random processes, which tend toward a Gaussian process (Central Limit Theorem). If we actually identify v_i and w_0 with Gaussian processes, i.e.

$$v_i \equiv N(0, \Lambda_i)$$
$$w_0 \equiv N(0, S_0),$$

then, it follows that the noise u_i in Eq. 22 is also Gaussian, i.e. $u_i \equiv N(0, W_i)$ and that all the successive estimates provided by the recursive Kalman filter are also Gaussian:

$$\hat{a}_k = N(a, S_k).$$

Moreover, in this case, the Kalman filter provides the best (minimum variance) unbiased estimate $\hat{a}_k$ among all, even nonlinear, filters. This estimate $\hat{a}_k$ is also the maximum likelihood estimator of a. This comes from the fact that in the Gaussian case, the solution is the conditional mean $\hat{a}_k = E[a/y_1, \ldots, y_k]$ which both minimizes the variance and maximizes the likelihood while being expressed as a linear combination of the measurements y_i. Therefore in this case, the minimum variance and minimum variance linear estimates are the same, namely the estimate $\hat{a}_k$ provided by the Kalman filter[20].

In conclusion, in the Gaussian case, the Kalman filter provides the best estimate with the advantage of preserving Gaussianness of all the implied random variables, which means that no information on the probability density functions of the parameters is lost while keeping only their mean and covariance matrix.

A.5. Rejecting Outlier Measurements

At iteration i, we have an estimate $\hat{a}_{i-1}$ and an attached covariance matrix S_{i-1} for parameter a. We also have a noisy measurement $(\hat{x}_i, \Lambda_i)$ of x_i and we want to test the plausibility of this measurement with respect to the equation $f_i(x_i, a) = 0$.

If we consider again a first order expansion of $f_i(x_i, a)$ around $(\hat{x}_i, \hat{a}_{i-1})$(Eq. 21), considering that $(\hat{x}_i - x_i)$ and $(\hat{a}_{i-1} - a)$ are independent centered Gaussian processes, we see that $f_i(\hat{x}_i, \hat{a}_{i-1})$ is also (up to linear approximation) a centered Gaussian process whose mean and covariance are given by:

$$E[f_i(\hat{x}_i, \hat{a}_{i-1})] = 0$$

$$Q_i = E[f_i(\hat{x}_i, \hat{a}_{i-1}) f_i(\hat{x}_i, \hat{a}_{i-1})^t] = \frac{\widehat{\partial f_i}}{\partial x} \Lambda_i \frac{\widehat{\partial f_i^t}}{\partial x} + \frac{\widehat{\partial f_i}}{\partial a} S_{i-1} \frac{\widehat{\partial f_i^t}}{\partial a}.$$

Therefore, if the rank of Q_i is q, the generalized Mahalanobis distance:

$$d(\hat{x}_i, \hat{a}_{i-1}) = f_i(\hat{x}_i, \hat{a}_{i-1})^t Q_i^{-1} f_i(\hat{x}_i, \hat{a}_{i-1}) \tag{28}$$

has a χ^2 distribution with q degrees of freedom[b].

Looking at a χ^2 distribution table, it is therefore possible to reject an outlier measurement $\hat{x}_i$ at a 95% confidence rate by setting an appropriate threshold ϵ on the Mahalanobis distance, and by keeping only those measurements $\hat{x}_i$ which verify:

$$d(\hat{x}_i, \hat{a}_{i-1}) < \epsilon. \tag{29}$$

[b] If $q < p$ = the size of the measurement vector f_i, Q_i^{-1} is the pseudo-inverse of Q_i.

Olivier Faugeras is a Professor in Computer Science at Ecole Polytechnique, Paris, a Senior Lecturer at Ecole Normale Supérieure, in Paris, and the Ecole Polytechnique Fédérale, in Lausanne where he teaches computer vision. He is Director of the Vision and Robotics Laboratory at INRIA, Rocquencourt. His current research interests are in computer vision, artificial intelligence and mobile robots.

Francis Lustman received his engineering degree at Ecole Polytechnique in 1983 and his Ph.D. degree in computer science (Université d'Orsay, France) in 1987. From 1984 to 1988 he has been working at INRIA (Rocquencourt, France) under the direction of Professor Faugeras.