Rectification of images for binocular and trinocular stereovision

Nicholas Ayache, C. Hansen

INRIA

UNITÉ DE RECHERCHE
INRIA-ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél.:(1) 39 63 55 11

Rapports de Recherche

N° 860

RECTIFICATION OF IMAGES FOR BINOCULAR AND TRINOCULAR STEREOVISION

Nicholas AYACHE
Charles HANSEN

JUIN 1988

RR_0860

Rectification of Images for Binocular and Trinocular Stereovision

—

Nicholas Ayache Charles Hansen

October 1987

INRIA
BP 105
78153 Le Chesnay Cédex

Abstract

We present a technique for calibrating and rectifying in a very *efficient* and *simple* manner pairs or triplets of images taken for binocular or trinocular stereovision systems. After the rectification of images, epipolar lines are parallel to the axes of the image coordinate frames: therefore, potential matches between two or three images satisfy simpler relations, allowing for simpler and more efficient matching algorithms. Experimental results obtained with a binocular and a trinocular stereovision system are presented and complexity analysis is provided.

Nous présentons une technique pour calibrer et rectifier de manière efficace et simple des couples ou des triplets d'images stéréoscopiques. Après rectification, les droites épipolaires sont parallèles aux axes des repères images: ainsi, les appariements potentiels vérifient des contraintes plus simples, ce qui permet d'accélérer les algorithmes de mise en correspondance. Nous présentons des résultats expérimentaux et précisons la complexité algorithmique de la rectification.

Key-words: Rectification, Stereovision, Binocular, Trinocular, Calibration.

*This work was partially supported by esprit project $P940$.

PAPIER RECUPERÉ ET RECYCLÉ

Contents

Rectification of Images for Binocular and Trinocular Stereovision

—

Nicholas Ayache   Charles Hansen

October 1987

INRIA
BP 105
78153 Le Chesnay Cédex

Abstract

We present a technique for calibrating and rectifying in a very *efficient* and *simple* manner pairs or triplets of images taken for binocular or trinocular stereovision systems. After the rectification of images, epipolar lines are parallel to the axes of the image coordinate frames: therefore, potential matches between two or three images satisfy simpler relations, allowing for simpler and more efficient matching algorithms. Experimental results obtained with a binocular and a trinocular stereovision system are presented and complexity analysis is provided.

Nous présentons une technique pour calibrer et rectifier de manière efficace et simple des couples ou des triplets d'images stéréoscopiques. Après rectification, les droites épipolaires sont parallèles aux axes des repères images: ainsi, les appariements potentiels vérifient des contraintes plus simples, ce qui permet d'accélérer les algorithmes de mise en correspondance. Nous présentons des résultats expérimentaux et précisons la complexité algorithmique de la rectification.

Key-words: Rectification, Stereovision, Binocular, Trinocular, Calibration.

---

*This work was partially supported by esprit project $P940$.

Contents

1 Introduction

We present in this paper a simple formalism for the calibration of binocular and trinocular stereovision systems which includes the rectification of images for enforcing horizontal and/or vertical epipolar lines which leads to much more efficient stereomatching algorithms.

This rectification could be avoided if images were taken with identical cameras placed in very specific relative positions. In practice, it is impossible, and such assumptions lead to large approximations for the localization of epipolar lines and inaccurate 3D reconstructions.

Our approach is tied to but different from the ones developed by [1,2,3,4,5,6,7] in the sense that we insist on computing explicitely **neither** the intrinsic parameters of each camera (focal length, piercing point, horizontal and vertical image units...), **neither** their extrinsic parameters (rotation and translation with respect to an absolute reference frame), but only the perspective matrices used to relate, in projective coordinates, 3D scene points to their corresponding 2D image points.

First, we show how to estimate the perspective matrices through simple least squares procedures. Second we describe the epipolar geometry in the general case. Then we show how to rectify images at a minimal cost to obtain the simplest possible epipolar geometry. This cost is the storage of two 3x3 matrices in the binocular case or three such matrices in the trinocular case, and then of 6 multiplications, 6 additions and two divisions per rectified point. Finally, we indicate how 3D reconstruction is performed from the matches obtained on the rectified images. Experimental results obtained with binocular and trinocular stereovision systems are presented.

2 Image Modelling

Each camera is modelled by its optical center C and its image plane $\mathcal{P}$ (cf. figure 1). This is the classical pinhole model. A point P in the observed scene is projected on point I of the camera retina. Point I is the intersection of the straight line PC with the image plane $\mathcal{P}$.

The relationship between P and I is modelled as a linear transformation in projective coordinates. If we denote $I^* = (U, V, S)^t$ the projective coordinates of I and $(x, y, z)^t$ the coordinates of P, the following relation holds:

$$I^* = \begin{pmatrix} U \\ V \\ S \end{pmatrix} = T \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

where T is a 3×4 matrix usually called the *perspective matrix* of the considered camera.

If P is in the focal plane of the camera, (i.e. if the straight line CP is parallel to the image plane $\mathcal{P}$), then $S = 0$ and the coordinates $(u, v)^t$ of I are no longer defined. In the general case

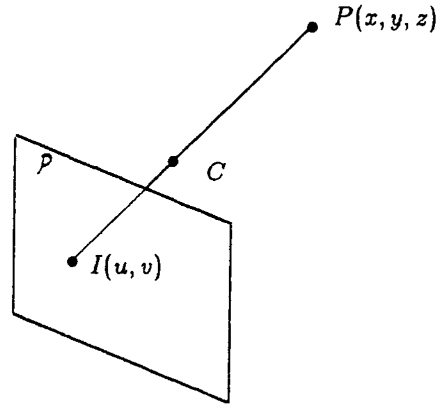

Figure 1: Camera Modelling

$S \neq 0$ and the image coordinates of I (usually expressed in *pixels*) are given by:

$$I = \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} U/S \\ V/S \end{pmatrix}$$

3 Determining the perspective matrix T

3.1 Principle

In the experiments conducted in our laboratory [1,2], T is obtained by analysing a calibration pattern which is a grid painted on a planar surface. The 3D position of the intersection points of the grid are well known in an absolute 3D coordinate frame and the grid is observed from several well defined different positions.

T is a matrix of dimension 3×4, but it is defined up to a scale factor, and one needs a constraint to specify T uniquely. The simplest constraint[1] consists in assuming that $t_{34} \neq 0$, then enforcing

$$t_{34} = 1$$

Each time an image point $I = (u, v)^t$ is matched with its corresponding scene point $P = (x, y, z)^t$, this provides the following two linear equations on the eleven unknowns remaining for determining T:

$$\begin{aligned} P^t t_1 + t_{14} - u(P^t t_3 + 1) &= 0 \\ P^t t_2 + t_{24} - v(P^t t_3 + 1) &= 0 \end{aligned} \tag{1}$$

[1] on the discussion of this constraint, see [1,2].

where t_{jk} is the element of rank (j,k) in T, and t_j is the 3-vector obtained from the first 3 elements of the j^{th} row of T:

$$t_j = (t_{j1}, t_{j2}, t_{j3})^t$$

In theory, six non coplanar points are sufficient for determining T uniquely [8]. In practice, several dozens of points are available, allowing for a global or recursive least squares estimation of T (see the Appendix).

4 General epipolar geometry

We now assume that we are dealing with at least two cameras, called camera 1 and camera 2, and we compute the relations between them.

4.1 Epipolar lines

Given a point I_1 in image 1 (cf. figure 2), we look for its corresponding point I_2 in image 2. It appears that I_2 belongs necessary to a straight line of image 2 entirely defined by the coordinates of I_1 and the relative geometry of the two cameras, called the *epipolar line* attached to I_1.

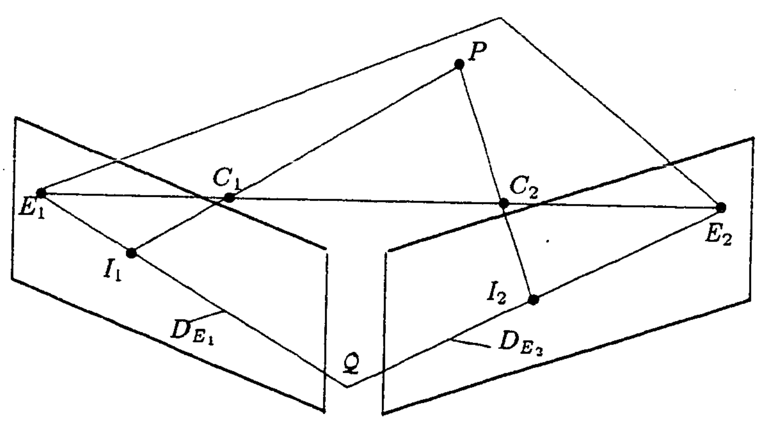

Figure 2: General epipolar geometry

In effect, the set of physical points P whose image corresponds to I_1 is the straight line C_1I_1 from which we exlude C_1. The image of this straight line in camera 2 is the epipolar line attached to I_1. The problem is perfectly symmetrical, and if one considers the plane Q defined by I_1, C_1 and C_2, it is clear that this plane intersects the image planes along two straight lines which are *conjugated* epipolar lines D_{12} and D_{21}. Any point I_1 of D_{12} has its potential corresponding points on D_{21} and vice-versa.

To compute analytically the equation of the epipolar lines, we need first to compute from each matrix T_i the optical center of the cameras and the inverse image of an arbitrary image point.

4.2 Determining optical centers

The 3D coordinates $(x_{C_i}, y_{C_i}, z_{C_i})$ of the optical center C_i of camera i (modelled by the perspective matrix T_i) are obtained by solving:

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = T_i \begin{pmatrix} x_{C_i} \\ y_{C_i} \\ z_{C_i} \\ 1 \end{pmatrix}$$

which is a system of three linear equations in the three unknowns $(x_{C_i}, y_{C_i}, z_{C_i})$.

4.3 Computing inverse images

We need to compute the straight line D which is the inverse image in the scene of a given image point I. This straight line D is composed of 3D scene points P having the same image I. If we look at figure 2 we see that D is simply the straight line defined by I and C_i.

To determine D analytically, let us re-write the system of equations 1 which relates point I_i to points P in the form:

$$\begin{aligned} (t_1^i - u_1\, t_3^i)^t P + t_{14}^i - u_1\, t_{34}^i &= 0 \\ (t_2^i - v_1\, t_3^i)^t P + t_{24}^i - v_1\, t_{34}^i &= 0 \end{aligned}$$

where the i index in t_j^i refers to camera i.

These equations are nothing else than the equations of two planes whose intersection defines D. A vector n colinear to D is the cross-product of the normals to the planes:

$$n = (t_1^i - u_1\, t_3^i) \times (t_2^i - v_1\, t_3^i)$$

which yields:

$$n = u_1\, t_2^i \times t_3^i + v_1\, t_3^i \times t_1^i + t_1^i \times t_2^i$$

which can be written

$$n = N_i I_i^* \tag{2}$$

with

$$N_i = [t_2^i \times t_3^i \quad t_3^i \times t_1^i \quad t_1^i \times t_2^i]$$

The parametric equation of the line $C_1 I_1$ is therefore given by

$$P = C_i + \lambda\, n$$

, where n is given by the previous equation and where λ is a real number.

4.4 Parametric equation of epipolar lines

It is now easy to compute the parametric equation of the epipolar line D_{21} in image 2 corresponding to the image point I_1 of coordinates (u_1, v_1) in image 1, because D_{21} is simply the image of the line $C_1 I_1$ by camera 2. Therefore D_{21} is composed of points I_2 whose projective coordinates satisfy:

$$I_2^* = T_2 \begin{pmatrix} C_1 + \lambda\, n \\ 1 \end{pmatrix}$$

If we denote

$$F_2^* = T_2'\, n \tag{3}$$

where T_2' is the 3x3 sub-matrix obtained from T_2 by suppressing its last column, and

$$E_2^* = T_2 \begin{pmatrix} C_1 \\ 1 \end{pmatrix} \tag{4}$$

then we get the parametric equation of the epipolar line D_{21} *in projective coordinates:*

$$I_2^* = E_2^* + \lambda\, F_2^*$$

Therefore, the parametric equation of D_{21} *in image coordinates* is:

$$u_2 = \frac{U_{E_2} + \lambda\, U_{F_2}}{S_{E_2} + \lambda\, S_{F_2}} \tag{5}$$

$$v_2 = \frac{V_{E_2} + \lambda\, V_{F_2}}{S_{E_2} + \lambda\, S_{F_2}} \tag{6}$$

From these equations, it is easy to see that the epipolar lines form a pencil of lines going through an epipolar center E_2 which is the image of C_1 in camera 2. One can also notice that a vector colinear to the epipolar line D_{21} is obtained by differentiation of equations 5 and 6 with respect to λ. This yields

$$\begin{pmatrix} \Delta u_2 \\ \Delta v_2 \end{pmatrix} = \begin{pmatrix} U_{F_2} S_{E_2} - U_{E_2} S_{F_2} \\ V_{F_2} S_{E_2} - V_{E_2} S_{F_2} \end{pmatrix} \tag{7}$$

When $S_{E_2} = 0$, this means that the epipolar center E_2 is rejected to infinity. In this case, the direction of the epipolar lines becomes independent of the coordinates (u_1, v_1) of I_1, and one can see from equation 7 that in this case all epipolar lines are parallel to the vector:

$$\begin{pmatrix} \Delta u_2 \\ \Delta v_2 \end{pmatrix} = \begin{pmatrix} U_{E_2} \\ V_{E_2} \end{pmatrix}$$

We now present a technique which provides new images for which the epipolar lines form such a pencil of parallel lines.

5 Rectification of two images

5.1 Getting horizontal epipolar lines

As we have just seen, in the particular case where the image planes $\mathcal{P}_1$ et $\mathcal{P}_2$ are coplanar and parallel to the vector C_1C_2 defined by the optical centers, then the epipolar centers are rejected to infinity and the epipolar lines form a pencil of parallel lines. If in addition the image coordinate frames are judiciously defined it is possible that the epipolar line attached to a point (u'_1, v'_1) in image 1 be the line $v'_2 = v'_1$ in image 2. We are then in the situation depicted by figure 3.

We show in this section that it is always possible to apply to each image a transformation which is linear in projective coordinates to obtain conjugated horizontal epipolar lines.

For doing this, let us consider again figure 3 where we have represented the optical centers C_i and the image planes $\mathcal{P}_i$ of each camera. The principle of the rectification is the definition of two new perpective matrices M and N which respectively define the same optical centers C_1 and C_2 as T_1 and T_2 but with a new common image plane $\mathcal{P}'$ parallel to C_1C_2. The rectification is the function which computes the new coordinates (u'_i, v'_i) from the old ones (u_i, v_i) for each image i.

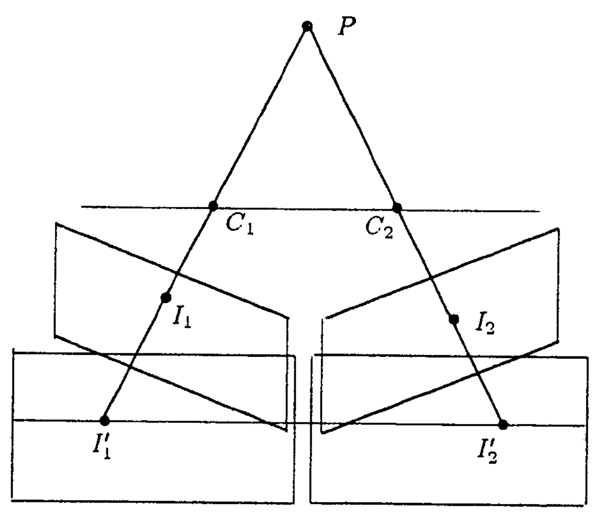

Figure 3: Rectification of two images

5.2 New perspective matrices

Let us describe the constraints holding on the new perspective matrices M and N.

1. The optical centers of M and N remain respectively C_1 and C_2. (this is necessary for having a univoque correspondance between the image points I_i and I'_i respectively before and after rectification).

2. The focal plane of M must be the same as the focal plane of N (this is necessay for having parallel epipolar lines in the rectified images).

3. For any point P *not* in this focal plane, the image points I_1' and I_2' respectively computed with M and N are such that $v_1' = v_2'$ (this is to simplify, as much as possible, the computation of the epipolar lines).

Let us now translate these constraints for the matrices M and N. For doing this, we denote.

$$M = \begin{pmatrix} m_1^t & m_{14} \\ m_2^t & m_{24} \\ m_3^t & m_{34} \end{pmatrix}; \qquad N = \begin{pmatrix} n_1^t & n_{14} \\ n_2^t & n_{24} \\ n_3^t & n_{34} \end{pmatrix}$$

Therefore, one must to estimate 24 parameters. But we have seen that the perspective matrices were defined up to a scale factor. It is therefore possible to choose for instance[2]:

$$m_{34} = n_{34} = ||C_1 \times C_2||^2 \tag{8}$$

There remains 22 parameters to compute to define M and N completely.

Constraint 2 implies that

$$\forall P \quad m_3^t P + ||C_1 \times C_2||^2 = 0 \Longleftrightarrow n_3^t P + ||C_1 \times C_2||^2 = 0$$

Therefore

$$m_3 = n_3 \tag{9}$$

Moreover, contraint 3 implies:

$$\forall P \,|\, m_3^t P + ||C_1 \times C_2||^2 \neq 0, \quad \frac{m_2^t P + m_{24}}{m_3^t P + ||C_1 \times C_2||^2} = \frac{n_2^t P + n_{24}}{n_3^t P + ||C_1 \times C_2||^2}$$

But, as $m_3 = n_3$, this yields:

$$m_2 = n_2 \text{ and } m_{24} = n_{24} \tag{10}$$

Finally, constraint 1 can be written as:

$$M \begin{pmatrix} C_1 \\ 1 \end{pmatrix} = N \begin{pmatrix} C_2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{11}$$

Equations 8 to 11 are summarized by the search of M and N of the form:

$$M = \begin{pmatrix} m_1^t & m_{14} \\ m_2^t & m_{24} \\ m_3^t & m_{34} \end{pmatrix} \qquad N = \begin{pmatrix} n_1^t & n_{14} \\ m_2^t & m_{24} \\ m_3^t & m_{34} \end{pmatrix}$$

[2]assuming that $||C_1 \times C_2|| \neq 0$.

with the system of constraints

$$
\begin{aligned}
m_1^t C_1 + m_{14} &= 0 \\
m_2^t C_1 + m_{24} &= 0 \\
m_2^t C_2 + m_{24} &= 0 \\
n_1^t C_2 + n_{14} &= 0 \\
m_3^t C_1 + \|C_1 \times C_2\|^2 &= 0 \\
m_3^t C_2 + \|C_1 \times C_2\|^2 &= 0 \\
m_{34} - \|C_1 \times C_2\|^2 &= 0
\end{aligned} \tag{12}
$$

In conclusion, equations 12 express 7 linear equations on the 16 remaining parameters defining M and N. Thus, there remain 9 degrees of freedom which correspond to some degrees of freedom still available to definitely choose the orientation and the distance of plane $\mathcal{P}'$ as well as the coordinate frames in the new images.

Let us fix in the simplest manner the remaining degrees of freedom. For instance, let us assume that

$$m_{14} = m_{24} = n_{14} = 0$$

Therefore, it is clear that the following properties must be satisfied:

1. m_1 must be orthogonal to C_1,
2. n_1 must be orthogonal to C_2,
3. m_2 must be orthogonal to C_1 and C_2,
4. m_3 must be orthogonal to $C_1 - C_2$.

To satisfy property 3, we choose

$$m_2 = (C_1 \times C_2)$$

Then, to satisfy property 1, we choose m_1 orthogonal to C_1 and also, to avoid a degenerate perspective matrix, m_1 orthogonal to m_2, which leads to

$$m_1 = (C_1 \times C_2) \times C_1$$

The same reasoning with property 1 leads to

$$n_1 = (C_1 \times C_2) \times C_2$$

Finally, to satisfy property 4, we choose m_3 as the cross product of $C_1 - C_2$ with a vector u such that $m_3^t C_1 + \|C_1 \times C_2\|^2 = 0$ which yields $u = C_1 \times C_2$ and:

$$m_3 = (C_1 - C_2) \times (C_1 \times C_2)$$

Therefore, matrices M and N are defined as follows:

$$M = \begin{pmatrix} ((C_1 \times C_2) \times C_1)^t & 0 \\ (C_1 \times C_2)^t & 0 \\ ((C_1 - C_2) \times (C_1 \times C_2))^t & \|C_1 \times C_2\|^2 \end{pmatrix} \tag{13}$$

$$N = \begin{pmatrix} ((C_1 \times C_2) \times C_2)^t & 0 \\ (C_1 \times C_2)^t & 0 \\ ((C_1 - C_2) \times (C_1 \times C_2))^t & \|C_1 \times C_2\|^2 \end{pmatrix} \tag{14}$$

5.3 Rectifying two images

As we saw in section 4.3 an image point $I_1(u_1, v_1)$ in image 1 comes from a 3D point $P(x, y, z)$ lying on the 3D straight line D defined by I_1C_1. The parametric equation of D is

$$P = C_1 + \lambda\, n$$

where n, a vector colinear to D, is given by equation 2:

$$n = u_1\, t_2^1 \wedge t_3^1 + v_1\, t_3^1 \wedge t_1^1 + t_1^1 \wedge t_2^1$$

The projective coordinates of the new image I_1' of P are computed as follows:

$$I_1'^* = \begin{pmatrix} U_1' \\ V_1' \\ S_1' \end{pmatrix} = M \begin{pmatrix} C_1 + n \\ 1 \end{pmatrix}$$

But C_1 is the optical center of M, which means that $M\,(C_1, 1)^t = 0$. Therefore, the computation of I_1' is simplified as:

$$I_1'^* = \begin{pmatrix} U_1' \\ V_1' \\ S_1' \end{pmatrix} = M'\, n$$

where M' is simply the 3×3 matrix obtained from M by deleting its 4^{th} column. As n is computed by an affine transformation from (u_1, v_1), one can compute

$$R_1 = \begin{pmatrix} ((C_1 \times C_2) \times C_1)^t \\ (C_1 \times C_2)^t \\ ((C_1 - C_2) \times (C_1 \times C_2))^t \end{pmatrix} [t_2^1 \times t_3^1 \quad t_3^1 \times t_1^1 \quad t_1^1 \times t_2^1]$$

a 3×3 matrix to obtain

$$\begin{pmatrix} U_1' \\ V_1' \\ S_1' \end{pmatrix} = R_1 \begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} \tag{15}$$

By a perfectly symmetric reasoning, we get:

$$R_2 = \begin{pmatrix} ((C_1 \times C_2) \times C_2)^t \\ (C_1 \times C_2)^t \\ ((C_1 - C_2) \times (C_1 \times C_2))^t \end{pmatrix} [t_2^2 \times t_3^2 \quad t_3^2 \times t_1^2 \quad t_1^2 \times t_2^2]$$

and we have:

$$\begin{pmatrix} U_2' \\ V_2' \\ S_2' \end{pmatrix} = R_2 \begin{pmatrix} u_2 \\ v_2 \\ 1 \end{pmatrix} \tag{16}$$

Therefore, the rectification of images 1 and 2 is reduced to the application of the two linear tranformations in projective coordinates expressed by the equations 15 and 16. After rectification, we have parallel horizontal epipolar lines, as desired, with the nice relationship

$$v_1' = v_2'$$

6 Rectification of three images

6.1 Getting horizontal and vertical epipolar lines

For three cameras, it is possible to rectify the images to get horizontal epipolar lines between images 1 and 2, and vertical epipolar lines between images 1 and 3. For doing this, the image planes $\mathcal{P}_1$ $\mathcal{P}_2$ and $\mathcal{P}_3$ must be coplanar and parallel to the plane defined by the optical centers $C_1C_2C_3$. If, in addition, the image coordinate frames are judiciously defined it is possible that the epipolar line attached to a point (u_1', v_1') in image 1 be the line $v_2' = v_1'$ in image 2 and the line $u_3' = u_1'$ in image 3. Moreover, it is possible to obtain a very simple relationship between images 2 and 3 of the form $u_2' = v_3'$. We are then in the situation depicted by figure 4. For doing this, one must follow the same reasoning as with two cameras.

6.2 New perspective matrices

Let us again precisely specify the constraints on the new perspective matrices M, N and Q:

1. The optical centers of M, N and Q remain respectively C_1, C_2 and C_3. (this is necessary for having a univoque correspondance between the image points I_i and I_i', $i = 1, 2, 3$ respectively before and after rectification).

2. The focal plane $\mathcal{F}$ of M, N and Q must be the same (this is necessary for having parallel epipolar lines in the rectified images).

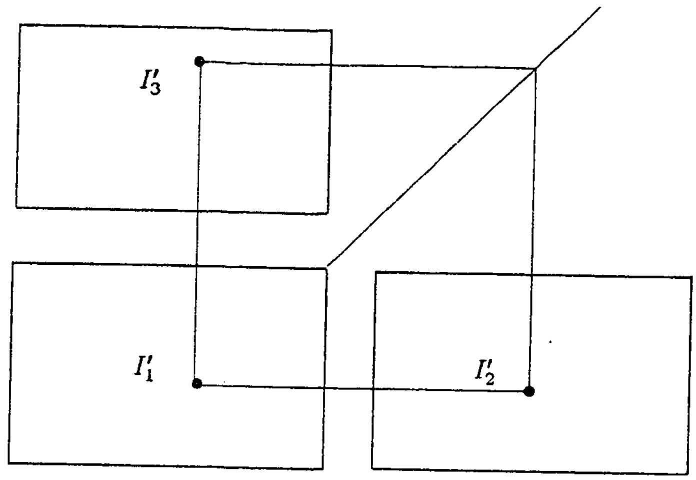

Figure 4: After the rectification of three images

3. For any point P *not* in the focal plane, the image points I'_1, I'_2 and I'_3 respectively computed with M, N and Q are such that

   - $v'_1 = v'_2$
   - $u'_1 = u'_3$
   - $u'_2 = v'_3$

Let us now translate these constraints for the matrices M, N and Q. For doing this, let us denote M and N as before and:

$$Q = \begin{pmatrix} q_1^t & q_{14} \\ q_2^t & q_{24} \\ q_3^t & q_{34} \end{pmatrix}$$

Therefore one has to estimate 36 parameters. But we have seen that the perspective matrices were defined up to a scale factor. It is therefore possible to choose for instance[3]:

$$m_{34} = n_{34} = q_{34} = -(C_1, C_2, C_3) \tag{17}$$

where (C_1, C_2, C_3) is the triple product

$$(C_1, C_2, C_3) = (C_1 \times C_2) \cdot C_3$$

There remains 33 parameters to compute to define M, N and Q completely.

[3]assuming that $(C_1, C_2, C_3) \neq 0$.

Constraint 2 implies that

$$\forall P \quad m_3^t\, P - (C_1, C_2, C_3) = 0 \Longleftrightarrow n_3^t\, P - (C_1, C_2, C_3) = 0 \Longleftrightarrow q_3^t\, P - (C_1, C_2, C_3) = 0$$

Therefore

$$m_3 = n_3 = q_3 \tag{18}$$

Moreover, constraint 3 implies, as before:

$$m_2 = n_2 \text{ and } m_{24} = n_{24} \tag{19}$$

$$m_1 = q_1 \text{ and } m_{14} = q_{14} \tag{20}$$

$$n_1 = q_2 \text{ and } n_{14} = q_{24} \tag{21}$$

Finally, constraint 1 can be written as:

$$M \begin{pmatrix} C_1 \\ 1 \end{pmatrix} = N \begin{pmatrix} C_2 \\ 1 \end{pmatrix} = Q \begin{pmatrix} C_3 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \tag{22}$$

Equations 17 to 22 are summarized by the search of M, N and Q of the form:

$$M = \begin{pmatrix} m_1^t & m_{14} \\ m_2^t & m_{24} \\ m_3^t & -(C_1, C_2, C_3) \end{pmatrix}; \quad N = \begin{pmatrix} n_1^t & n_{14} \\ m_2^t & m_{24} \\ m_3^t & -(C_1, C_2, C_3) \end{pmatrix}; \quad Q = \begin{pmatrix} m_1^t & m_{14} \\ n_1^t & n_{14} \\ m_3^t & -(C_1, C_2, C_3) \end{pmatrix}$$

satisfying the 10 linear constraints

$$\begin{aligned} m_1^t\, C_1 + m_{14} &= 0 \\ m_1^t\, C_3 + m_{14} &= 0 \\ m_2^t\, C_1 + m_{24} &= 0 \\ m_2^t\, C_2 + m_{24} &= 0 \\ n_1^t\, C_2 + n_{14} &= 0 \\ n_1^t\, C_3 + n_{14} &= 0 \\ m_3^t\, C_1 - (C_1, C_2, C_3) &= 0 \\ m_3^t\, C_2 - (C_1, C_2, C_3) &= 0 \\ m_3^t\, C_3 - (C_1, C_2, C_3) &= 0 \\ m_{34} - (C_1, C_2, C_3) &= 0 \end{aligned} \tag{23}$$

In conclusion, equations 23 express 10 linear equations on the 16 remaining parameters defining M N and Q. There remain 6 degrees of freedom which correspond to some degrees of freedom still available to definitely choose the distance of plane $\mathcal{P}'$ as well as some of the parameters of the coordinate frames in the new images.

Let us fix in the most simple way the remaining degrees of freedom. For instance, let us assume that

$$m_{14} = m_{24} = n_{14} = 0$$

Therefore, it is clear that the following properties must be satisfied:

1. m_1 must be orthogonal to C_1 and C_3,
2. m_2 must be orthogonal to C_1 and C_2,
3. n_1 must be orthogonal to C_2 and C_3,
4. m_3 must be orthogonal to $C_1 - C_2$ and $C_1 - C_3$.

To satisfy properties 1,2 and 3, we choose

$$m_1 = (C_3 \times C_1)$$

$$m_2 = (C_1 \times C_2)$$

$$n_1 = (C_2 \times C_3)$$

Then, to satisfy property 4, we choose

$$m_3 = (C_1 - C_2) \times (C_1 - C_3) = C_1 \times C_2 + C_2 \times C_3 + C_3 \times C_1$$

for which we can easily verify that $m_3^t C_1 = m_3^t C_2 = m_3^t C_3 = (C_1, C_2, C_3)$.

Therefore, matrices M, N and Q have the following values:

$$M = \begin{pmatrix} (C_3 \times C_1)^t & 0 \\ (C_1 \times C_2)^t & 0 \\ (C_1 \times C_2 + C_2 \times C_3 + C_3 \times C_1)^t & -(C_1, C_2, C_3) \end{pmatrix} \tag{24}$$

$$N = \begin{pmatrix} (C_2 \times C_3)^t & 0 \\ (C_1 \times C_2)^t & 0 \\ (C_1 \times C_2 + C_2 \times C_3 + C_3 \times C_1)^t & -(C_1, C_2, C_3) \end{pmatrix} \tag{25}$$

$$Q = \begin{pmatrix} (C_3 \times C_1)^t & 0 \\ (C_2 \times C_3)^t & 0 \\ (C_1 \times C_2 + C_2 \times C_3 + C_3 \times C_1)^t & -(C_1, C_2, C_3) \end{pmatrix} \tag{26}$$

One can notice that we get the same result as with two cameras for matrices M and N (except for a minor difference in the last line) when identifying $C_3 = C_1 \times C_2$ in the equations 13 and 14.

6.3 Rectifying three images

It is performed in exactly the same way as before with three 3x3 rectification matrices called R_1, R_2 and R_3. where

$$R_i = \begin{pmatrix} (C_{i-1} \times C_i)^t \\ (C_i \times C_{i+1})^t \\ (C_1 \times C_2 + C_2 \times C_3 + C_3 \times C_1)^t \end{pmatrix} [t_2^i \times t_3^i \quad t_3^i \times t_1^i \quad t_1^i \times t_2^i]$$

where $i + 1 = 1$ if $i = 3$ and $i - 1 = 3$ if $i = 1$.

After the rectification of the images we have, as desired, the nice relationships:

$$\begin{aligned} v_2' &= v_1' \\ u_3' &= u_1' \\ v_3' &= u_2' \end{aligned} \tag{27}$$

which was illustrated by figure 4.

7 Algorithmic complexity

The rectification of l images ($l = 2$ or 3) requires the storage of l 3x3 matrices, i.e. 9 l parameters. Then it requires 6 multiplications, 6 additions and 2 divisions per rectified point.

As the rectification process is a linear transformation in projective space, it preserves straight lines: therefore it is sufficient to apply it to the endoints of the linear segments of a polygonal approximation to get the endpoints of the segments of the rectified polygonal approximation. This is very useful for our stereovision algorithms [9,10,11] which actually deal with linear segments.

8 Intrinsic rectification

The matrices M, N and Q used for the rectification of the images depend on the choice of the origin O of the absolute 3D coordinate frame of the scene. The question is how to make the matrices independent of the origin O?

An answer is the following: it is sufficient to change the origin O of the scene coordinate frame into a point O' which is intrinsically defined by the relative geometry of the cameras themselves. For instance, O' could be the point which is at a minimum distance from the optical axes of the cameras.

Let us call D_i the optical axis of camera i, and let us define D_i as the set of 3D scene points whose image is a given point I_m chosen in the middle of the image plane. We remember that the parametric equation of D_i is given by:

$$OP = OC_i + \lambda n_i$$

where n_i is the 3-vector computed from the coordinates (u_m, v_m) of I_m by equation 2:

$$n_i = u_m\, t_2^i \times t_3^i + v_m\, t_3^i \times t_1^i + t_1^i \times t_2^i$$

The distance of O' from D_i is given by

$$d_i = \|O'H_i\|$$

where $O'H_i$ is computed as:

$$O'H_i = A_i\, O'C_i = A_i\, OC_i - A_i\, OO'$$

with A_i the 3×3 matrix:

$$A_i = I + (n_i n_i^t)/\|n_i\|^2$$

and I the 3x3 identity matrix. The proof is given in the Appendix.

Computing O' is then equivalent to minimizing

$$\mathcal{C} = \sum_{i=1}^{k} d_i^2$$

(for $k = 2$ or 3), which is obtained as the least square solution of the linear system of equations:

$$A\, OO' = b$$

where

$$A = \begin{pmatrix} A_1 \\ \vdots \\ A_k \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} A_1 OC_1 \\ \vdots \\ A_k OC_k \end{pmatrix}$$

The least square solution is given by

$$OO' = (A^t A)^{-1} A^t b$$

assuming that $A^t A$ has rank 3.

Once O' is found, the scene coordinates of any point P are changed into:

$$O'P = OP - OO'$$

and the matrices T_i $(i = 1, \ldots, k)$ are recomputed for the new coordinate frame, yielding new coordinates for the optical centers

$$O'C_i = OC_i - OO'$$

and new rectification matrices R_i, which are now intrinsically defined by the relative geometry of the cameras.

9 3D Reconstruction

Once two points have been matched between original (or rectified) images i (i = 1,2 and possibly 3), one can use matrices T_i's (resp. M, N and possibly Q) to compute the 3D coordinates of the corresponding scene point P.

Actually, if we rewrite the system of equations 1 for cameras i, one gets a new system of 4 (or possibly 6) equations in the three unknowns (x, y, z) of the P coordinates.

$$\begin{aligned}(t_1^i - u_i t_3^i)^t P + t_{14}^i - u_i t_{34}^i &= 0 \\ (t_2^i - v_i t_3^i)^t P + t_{24}^i - v_i t_{34}^i &= 0\end{aligned}$$

equations in which the index i of t_j^i refers to camera i and where (u_i, v_i) is replaced by (u_i', v_i') if we deal with the rectified images.

In theory these equations are not independent, because the image points are supposed to satisfy the epipolar constraints, and only three independent equations can be extracted from the system. Nevertheless, quantization errors and numerical limited accuracy both vote for a least-squares solution of the entire system. This is detailed in the Appendix.

10 Experimental Results

The reprojection technique presented in this paper has been successfully implemented for both binocular and trinocular stereo systems. We have tested this algorithm on a variety of indoor scenes and have found the accuracy of the rectified image to be within 1 pixel for 512 x 512 images. Since the rectification takes points (pixels) from the image plane and reprojects them onto a new plane, the rectified points are computed with, arbitrarily fine, sub-pixel accuracy in the rectified space.

Our stereo programs, both 2 and 3 camera [9,10], are feature-based where the features are linear segments. After obtaining a pair or triplet of images, edge points are extracted and chains of connected edges points are built and approximated by linear segments. These segments are the primitives for our stereo matching algorithms and thus are what we rectify. It should be noted that our rectification technique is general in that the rectification is not restricted to segments but can be applied to other types of features or to an entire intensity image thereby improving the matching speeds of different types of stereo programs.

Figure 5 shows a stereo pair of extracted linear segments from a typical room scene. We rectify the segments with the algorithm for two cameras presented in section 5. The rectified image pair is shown in figure 6. Note that the epipolar lines form the rows of the images and epipolar conjugates correspond to the same row number.

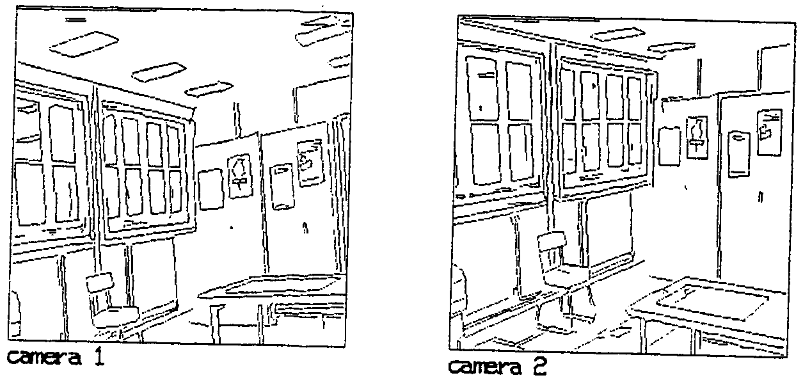

Figure 5: Stereo Pair of Room Scene

We have also tested the rectification algorithm for three cameras. Figure 7 again shows a typical room scene but with a calibrated triplet of cameras. Figure 8 shows the rectified image triplet. Notice that the rectification plane is different in the three camera case than with two cameras. This is because the plane is tilted to include the optical center of the third camera. This results in a slightly more distorted rectified image for some of the cameras, yet the conjugate epipolar lines still maintain the desired relationships: $v_2' = v_1'$, $u_3' = u_1'$ and $v_3' = u_2'$.

The computational expense for rectification depends on the complexity of the image. On a SUN-3/50 workstation, the computational time for one point is 1.53 millisecond. Thus, for an edge representation of a typical scene, such as figure 5 which contains approximately 400 segments, the computational time is about 0.12s on a SUN-3 workstation. We have tested our algorithm on more complex scenes, and as one would expect, we found that the computational time increases linearly with the number of segments. It should be pointed out that this relatively inexpensive operation greatly improves the speed of stereo matching. Since the speed of reliable matching is the highest cost in stereo programs, one can easily justify the expense of rectification. In actual experiments, the time of the stereo matching process was reduced by more than a factor of 2 if the images are rectified before matching [4,11].

11 Conclusions

In this paper, we have presented a general and efficient method for the rectification of images used in computational stereo. The method can be used for two or three cameras to enforce simple

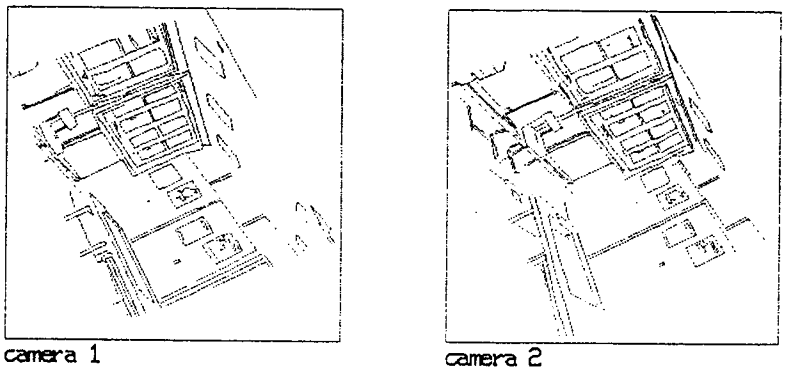

Figure 6: Rectified Stereo Pair of Room Scene

constraints on the conjugated epipolar lines in the images.

The presented technique does not rely on difficult, and often inaccurate, mechanical alignment of the cameras. Also, it does not require the explicit computation of the usually called *intrinsic* and *extrinsic* parameters of the cameras, but only the simple computation of the perpective matrix of each camera.

After rectification, the epipolar lines are parallel and aligned with the image coordinate frame. Furthermore, conjugate epipolar lines have corresponding row/column numbers. This technique greatly improves the speed of stereo matching by eliminating the need to compute the corresponding epipolar lines in the other image.

Experiments have been presented which have shown the method to be both effective and efficient.

12 Acknowledgements

O.D. Faugeras and G. Toscani pioneered the formalism we use for calibration, and the originality of this work comes from the rectification technique. Discussions with both of them and also with Francis Lustman who developed a different rectification technique (using intrinsic parameters of each camera) were very stimulating.

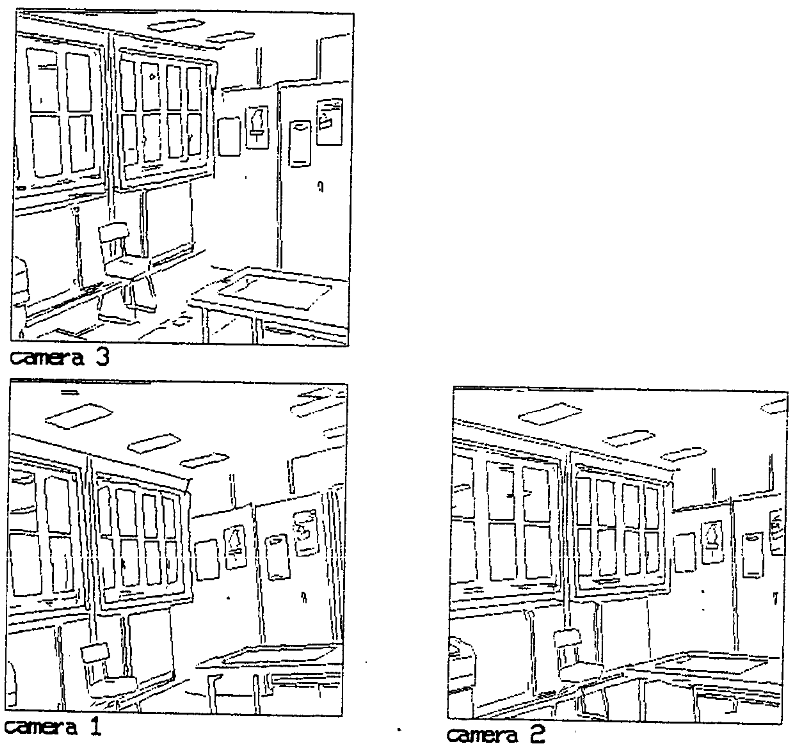

Figure 7: Triplet of Images of Room Scene

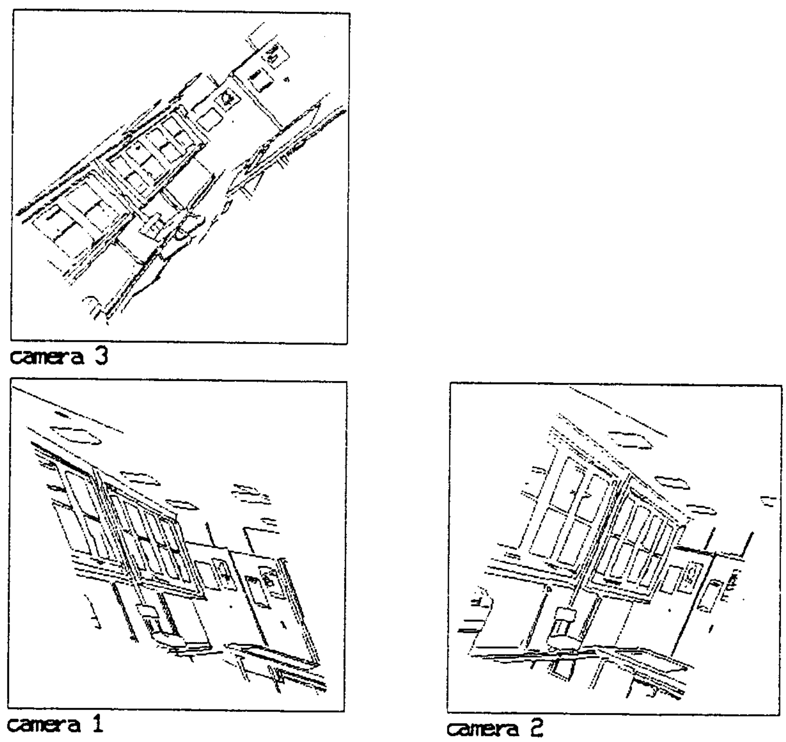

Figure 8: Rectified Triplet of Room Scene

A Estimation of T

A.1 Global least-squares estimation of T

We denote by a the sought parameter vector:

$$a = \begin{pmatrix} t_1 \\ t_{14} \\ t_2 \\ t_{24} \\ t_3 \end{pmatrix}$$

If the successively measured 3D and image points are denoted by P_i and I_i ($i = 1, \ldots, n$) respectively, this provides the following system of equations:

$$Aa = b$$

with

$$A = \begin{pmatrix} A_1 \\ \vdots \\ A_n \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

for which

$$A_i = \begin{pmatrix} \bar{P}_i^t & 0_{14} & -u_i P_i^t \\ 0_{14} & \bar{P}_i^t & -v_i P_i^t \end{pmatrix} \quad \text{and} \quad b_i = \begin{pmatrix} u_i \\ v_i \end{pmatrix}$$

with $\bar{P}_i^t = (x_i, y_i, z_i, 1)$ and 0_{14} is the null 1×4 matrix.

The least square solution is given by

$$a = (A^t A)^{-1} A^t b$$

assuming that $A^t A$ has rank 11.

A.2 Recursive least-squares estimation of T

It is possible to compute a recursive solution of the previous system, and also to take into account the uncertainty on the measurement process. For doing this, one consider a as a state vector and the measurement equations are of the form $A_i a = b_i + w_i$ where w_i is related to the noise on the position of the image point (u_i, v_i) and on the position of the scene point P_i. For additional details, one can refer to [12,13,14].

B Distance of O' to D

We compute here the distance d of a point O' to the straight line D given by a parametric equation of the form:

$$OP = OC + \lambda n$$

Let us call H the orthogonal projection of O' on D. Clearly,

$$d = ||O'H||$$

As H belongs to D, $O'H$ can be written as:

$$O'H = OH - OO' = OC + \lambda\, n - OO' = O'C + \lambda\, n$$

and

$$||O'H||^2 = \lambda^2\, ||n||^2 + 2\lambda\, n \cdot O'C + ||O'C||^2$$

As H is the point which minimizes the distance to D, the value of λ is such that

$$\frac{\partial ||O'H||^2}{\partial \lambda} = 0$$

This yields

$$\lambda = \frac{O'C \cdot n}{||n||^2}\, n$$

and therefore

$$O'H = O'C + (O'C \cdot u)u$$

where

$$u = \frac{1}{||n||} n$$

Finally, $O'H$ can be written as

$$O'H = [I + uu^t]O'C$$

where I is the 3x3 identity matrix.

C 3D reconstruction

C.1 Global least-squares estimation of P

For n cameras, one sets

$$A\, a = b$$

with $a = (x, y, z)^t$ and

$$A = \begin{pmatrix} A_1 \\ \vdots \\ A_n \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

where[4]

$$A_i = \begin{pmatrix} t_1^i - u_i t_3^i)^t \\ (t_2^i - v_i t_3^i)^t \end{pmatrix} \quad \text{and} \quad b_i = \begin{pmatrix} u_i t_{34}^i - t_{14}^i \\ v_i t_{34}^i - t_{24}^i \end{pmatrix}$$

The least-squares solution is then given by:

$$a = (A^t A)^{-1} A^t b$$

assuming $A^t A$ has rank 3.

C.2 Recursive least-squares estimation of P

Again, it is possible to compute a recursive solution of the previous system, and also to take into account the uncertainty on the measurement For additional details, one can refer to [13,14].

[4]one must change (u_i, v_i) into (u_i', v_i') if images are rectified.

References

[1] G. Toscani. *Système de Calibration optique et perception du mouvement en vision artificielle.* PhD thesis, Paris-Orsay, 1987.

[2] O.D. Faugeras and G. Toscani. The calibration problem for stereo. In *Proceedings CVPR '86, Miami Beach, Florida*, pages 15–20, IEEE, 1986.

[3] R. Tsai. An efficient and accurate camera calibration technique for 3d machine vision. In *Proc. International Conference on Computer Vision and Pattern Recognition*, pages 364–374, IEEE, June 1986. Miami Beach, Florida.

[4] F. Lustman. *Vision stéréoscopique et perception du mouvement en vision artificielle.* PhD thesis, Paris-Orsay, 1987.

[5] T. M. Strat. Recovering the camera parameters for a transformation matrix. In *Proc. Image Understanding Workshop*, pages 264–271, October 1984.

[6] B. Caprile and V. Torre. The epipolar transform. 1988. to appear.

[7] G. Gaillat. On the rectification of images. 1988. to appear.

[8] O.D. Faugeras. *Artificial 3D Vision.* To appear, 1988.

[9] N. Ayache and B. Faverjon. Efficient registration of stereo images by matching graph descriptions of edge segments. *The International Journal of Computer Vision*, 1(2), April 1987.

[10] N. Ayache and F. Lustman. Fast and reliable passive trinocular stereovision. In *Proc. First International Conference on Computer Vision*, pages 422–427, IEEE, June 1987. London, U.K.

[11] N. Ayache and F. Lustman. Trinocular stereovision, recent results. In *Proc. International Joint Conference on Artificial Intelligence*, August 1987. Milano, Italy.

[12] N. Ayache. Construction et fusion de représentations visuelles tridimensionnelles; applications à la robotique mobile. Thèse d'Etat, Université de Paris-Sud, Orsay, May 1988. INRIA Internal Report.

[13] N. Ayache and O.D. Faugeras. Building, registrating and fusing noisy visual maps. In *Proc. First International Conference on Computer Vision*, pages 73–82, IEEE, June 1987. London, U.K., also an INRIA Internal Report 596, 1986.

[14] N. Ayache and O.D. Faugeras. Maintaining representations of the environment of a mobile robot. In *International Symposium on Robotics Research*, August 1987. Santa-Cruz, California.

Imprimé en France
par
l'Institut National de Recherche en Informatique et en Automatique