Maintaining Representations of the Environment of a Mobile Robot

Nicholas Ayache, Olivier Faugeras

Maintaining Representations of the Environment of a Mobile Robot

NICHOLAS AYACHE AND OLIVIER D. FAUGERAS

Abstract—**In this paper we describe our current ideas related to the problem of building and updating 3-D representation of the environment of a mobile robot that uses passive Vision as its main sensory modality. Our basic tenet is that we want to represent both geometry and uncertainty. We first motivate our approach by defining the problems we are trying to solve and give some simple didactic examples. We then present the tool that we think is extremely well-adapted to solving most of these problems: the extended Kalman filter (EKF). We discuss the notions of minimal geometric representations for 3-D lines, planes, and rigid motions. We show how the EKF and the representations can be combined to provide solutions for some of the problems listed at the beginning of the paper, and give a number of experimental results on real data.**

I. Introduction

IN THE last few years, Computer Vision has gone extensively into the area of three-dimensional (3-D) analysis from a variety of sensing modalities such as stereo, motion, range finders, and sonars. A book that brings together some of this recent work is [24].

Most of these sensing modalities start from pixels which are then converted into 3-D structures. A characteristic of this work as compared to previous work (like in image restoration, for example) where images were the starting and the ending point is that noise in the measurements is, of course, still present but, contrary to what has happened in the past, it has to be taken into account all the way from pixels to 3-D geometry.

Another aspect of the work on 3-D follows from the observation that if noise is present, it has to be evaluated, i.e., we need models of sensor noise (sensor being taken here in the broad sense of sensory modality), and reduced. This reduction can be obtained in many ways. The most important ones are as follows:

- First, the case of one sensor in a fixed position: it can repeat its measurements and thus maybe obtain better estimations.
- Second, the case of a sensor that can be moved around: given its measurements in a given position, what is the best way to move in order to reduce the uncertainty and increase the knowledge of the environment in a way that is compatible with the task at hand.
- Third, is the case of several different sensors that have to combine their measurements in a meaningful fashion.

Interesting work related to those issues has already emerged which is not reported in [24]. In the area of robust estimation procedures and models of sensors noise, Hager and Mintz [22] and McKendall and Mintz [27] have started to pave the ground. Bolle and Cooper [12] have developed maximum likelihood techniques to combine range data to estimate object positions. Darmon [16] applies the Kalman filter formalism to the detection of moving objects in sequences of images. Durrant-Whyte [18], in his Ph.D. dissertation has conducted a thorough investigation of the problems posed by multi-sensory systems. Applications to the navigation of a mobile robot have been discussed by Crowley [15], Smith and Cheeseman [32], and Matthies and Shafer [28]. The problem of combining stereo views has been attacked by Ayache and Faugeras [3], [4], [19], Porril *et al.* [30], and Kriegman [25]. It also appears that the linearization paradigm extensively used in this paper has been already used in the photogrammetry field [26].

Several problems related to these preliminary studies need more attention. Modeling sensor noise in general and more specifically visual sensor noise appears to us an area where considerable progress can be achieved; relating sensor noise to geometric uncertainty and the corresponding problem of representing geometric information with an eye toward describing not only the geometry but also the uncertainty on this geometry are key problems to be investigated further as is the problem of combining uncertain geometric information produced by different sensors.

II. What are the Problems that we are Trying to Solve

We have been focusing on a number of problems arising in connection with a robot moving in an indoor environment and using passive vision and proprioceptive sensory modalities such as odometry. Our mid-term goals are to incrementally build on the robot an increasing set of sensing and reasoning capabilities such as:

- build local 3-D descriptions of the environment,
- use the descriptions to update or compute motion descriptions where the motion is either the robot's motion or others,
- fuse the local descriptions of neighboring places into more global, coherent, and accurate ones,
- "discover" interesting geometric relations in these descriptions,
- "discover" semantic entities and exhibit "intelligent" behavior.

We describe how we understand each of these capabilities and what are the underlying difficulties.

Manuscript received April 5, 1988; revised May 2, 1989. This work was partially supported by the Esprit Project P940.

The authors are with INRIA—Rocquencourt, Domaine de Voluceau, Rocquencourt, B. P. 105—78153, Le Chesney Cedex, France.

IEEE Log Number 8930460

1042-296X/89/1200-0804$01.00

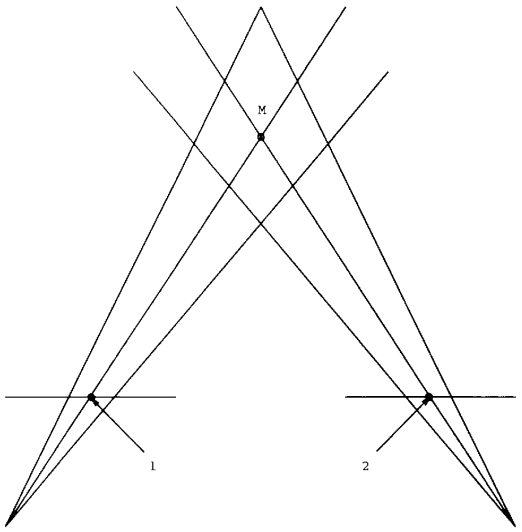


Fig. 1. Effect of pixel noise on 3-D reconstruction.

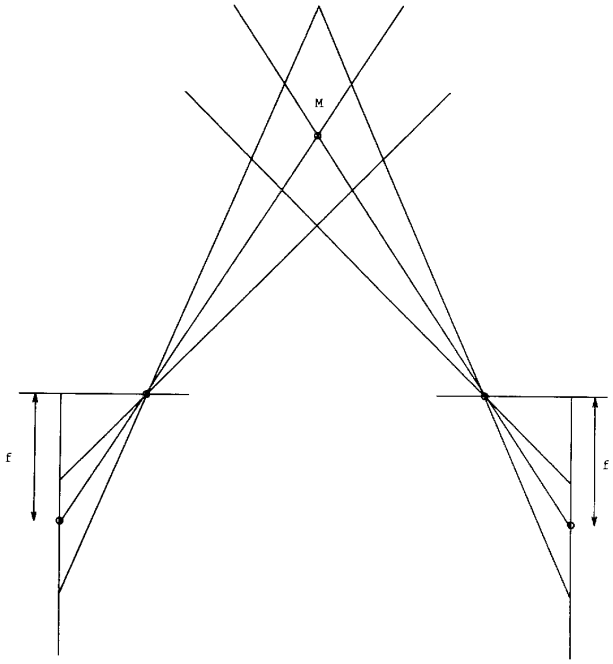


Fig. 2. Effect of calibration errors on 3-D reconstruction.

A. Build Local 3-D Descriptions of the Environment

Up until now, our main source of 3-D information has been Stereo [5], [9] even though we have made considerable progress toward the use of structure from motion as well [21]. In any case, the problems are very similar for both sensing modalities and we concentrate on Stereo. As announced in the Introduction, our main concern is to track uncertainty all the way from pixel noise to geometric descriptions. Fig. 1 shows, for example, that in a Stereo system, if pixels positions are imperfectly known, then the corresponding 3-D point varies in an area with a quite anisotropic diamond shape. This is a clear example of a relation between pixel uncertainty and geometric (the position of point M) uncertainty. Another source of uncertainty in Stereo is the calibration uncertainty. In a stereo rig, intrinsic parameters of the cameras such as focal length, and extrinsic parameters such as relative position and orientation of the cameras have to be calculated. Fig. 2 shows the effect on the reconstruction of a point M of an uncertainty on the focal lengths of the two cameras. Again, M varies in a diamond-like shape. Of course, this source of uncertainty adds to the previous pixel uncertainty.

Another example of the propagation of uncertainty is given in Fig. 3 where pixels in left and right images are grouped into line segments: pixel uncertainty is converted into 2-D line uncertainty. Line segments are then matched and used to reconstruct 3-D line segments: 2-D line uncertainty and calibration uncertainty are converted into 3-D uncertainty.

Yet another set of examples of this kind of propagation is shown in Fig. 4 where coplanar and cocylindrical line segments are grouped together; again, the question is, what is the uncertainty on the plane or on the cylinder? (the uncertainty on the position of the lines, plane, and cylinder is represented on the picture in a symbolic manner by ellipses).

From these examples, we see that the main problem that needs to be solved in order to build local 3-D descriptions of the environment is how geometric uncertainty propagates

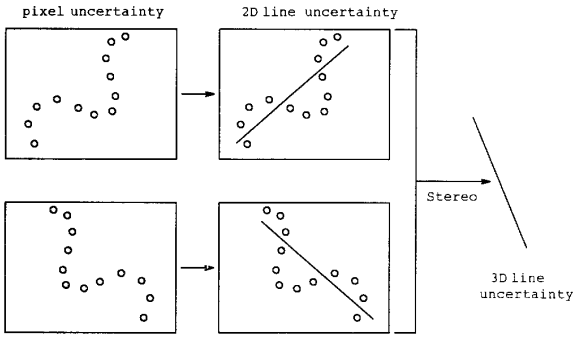


Fig. 3. From pixel uncertainty to 3-D line uncertainty.

when we build more complex primitives from simpler ones. This, in turn, generates two questions:

1) How do we represent geometric primitives?
2) How do we represent uncertainty on these primitives?

B. Update Position and Motion Information

Fig. 5 shows a measurement of a physical point made in two positions 1 and 2 of a mobile vehicle. In position 1, it "sees" M with some uncertainty represented by the ellipse around it. In position 2, it "sees" P with another uncertainty. Assuming that the displacement between 1 and 2 is exactly known, it is possible to express P and M in the same coordinate system. If the displacement estimate is wrong, as it is in Fig. 5, the two zones of uncertainty do not intersect and it is very unlikely that the observer will realize that the points M and P are instances of the same physical point. If we now take into account the uncertainty on the displacement (assuming that we can estimate it) we have Fig. 6 where the combination of

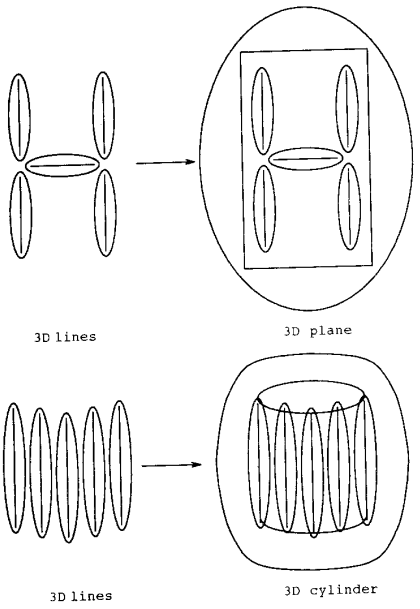


Fig. 4. From 3-D line uncertainty to 3-D surface uncertainty.

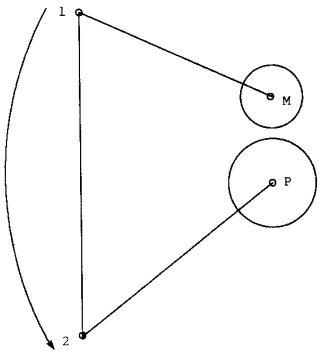


Fig. 5. Measuring a point in two positions (wrong displacement estimation).

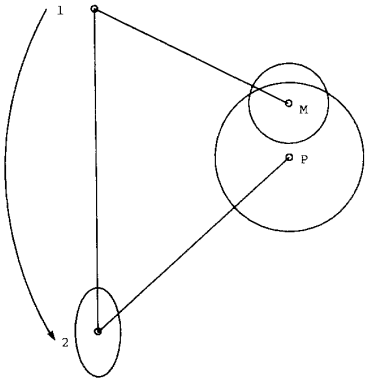


Fig. 6. Measuring a point in two positions (displacement and uncertainty estimation).

displacement uncertainty and measurement uncertainty produces a larger ellipse around P which intersects the one around M: the observer can now infer that the probability of M and P being the same physical point is quite high and use the two measurements to obtain a better estimate of the displacement

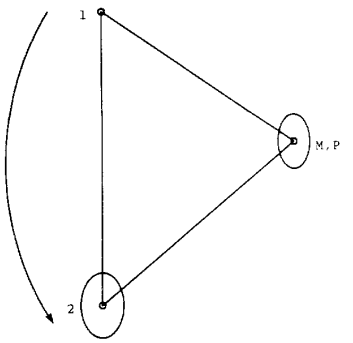


Fig. 7. Improving the estimation of the points position.

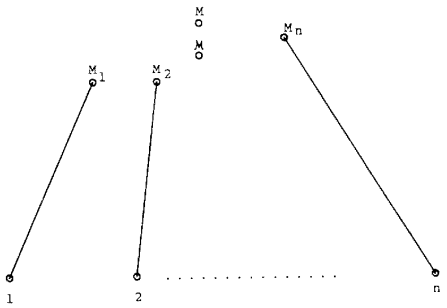


Fig. 8. Fusing n points measured from different positions.

and reduce its uncertainty. We explain how to do this in Section V. The measurements can also be used to produce better estimates of the positions (Fig. 7). This is related to what we call geometric fusion.

C. Fusing Geometric Entities

Fig. 8 shows a slightly more general case than what is depicted in Fig. 7. The mobile vehicle has measured the physical point M in n positions numbered from 1 and n. Each measurement yields a point M_i, $i = 1, \cdots, n$ and some uncertainty in the coordinate system attached to the robot. Displacement uncertainty is also available. Using the ideas described in Section V, we can improve the estimates of the displacements and reduce their uncertainty by discovering that points $M_1, \cdots, M_n$ are all instanciations of the same point. We can also use this observation to reduce the uncertainty on, let us say M_1, by combining the n measurements and produce a point $\mathfrak{M}$, fusion of $M_1, \cdots, M_n$, as well as its related uncertainty. The points $M_1, \cdots, M_n$ can then be erased from the representation of the environment, they can be forgotten. What remains is the point $\mathfrak{M}$ expressed in the coordinate system attached to position 1, for example, and the displacement from 1 to 2, 2 to 3, etc...., which allows us to express $\mathfrak{M}$ in the other coordinate systems.

Fusing geometric entities is therefore the key to "intelligent" forgetting which, in turn, prevents the representation of the environment from growing too large.

D. Discovering "Interesting" Geometric Relations

Using this approach also allows us to characterize the likelihood that a given geometric relation exists between a

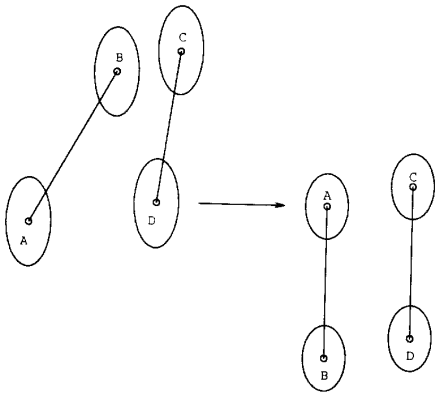


Fig. 9. Discovering that AB and CD are parallel.

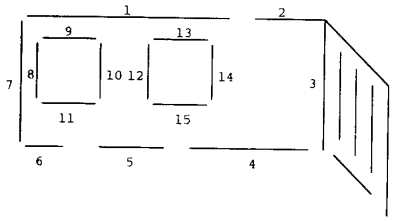


Fig. 10. Hypothesizing walls, windows, and doors.

number of geometric entities and to use this information to obtain better estimates of these entities and reduce their uncertainty. For example, as shown in Fig. 9, segments AB and CD which have uncertainty attached to their endpoints have a high likelihood to be parallel. Assuming that they are, we can update their position (they become more parallel) and reduce the uncertainty of their endpoints. The same reasoning can be used, for the relation "to be perpendicular."

E. Discovering Semantic Entities

Fig. 10 shows the kind of "semantic" grouping that is of interest to us in the context of a mobile robot moving indoors, to combine geometry and some *a priori* description of the environment. The line segments numbered from 1 to 15 are found, using the ideas described in Section II-D, to be coplanar with a high probability; the corresponding plane is found to be vertical with a very high probability which can be deduced from the geometric uncertainty of the line segments. This observation can then be used to infer that the plane has a high probability to be a wall. If we also observe that segments 8 to 11 and 12 to 15 form approximately two rectangles this can be used to infer that they have a high probability to be parts of a window or a door.

III. What is the Tool that we are Using

In this section, we introduce the Extended Kalman Filter (EKF) formalism which is applied in Sections IV and V to solve the problems we have just listed in Section II.

A. Unifying the Problems

In all of these previously listed problems, we are confronted with the estimation of an unknown parameter $\boldsymbol{a} \in R^n$ given a set of k possibly nonlinear equations of the form

$$\boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{a}) = \boldsymbol{0} \tag{1}$$

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

$$\hat{\boldsymbol{x}}_i = \boldsymbol{x}_i + \boldsymbol{v}_i \tag{2}$$

where $\boldsymbol{v}_i$ is a random error. The only assumption we make on $\boldsymbol{v}_i$ is that its mean is zero, its covariance is known, and that it is a white noise

$$E[\boldsymbol{v}_i] = 0$$

$$E[\boldsymbol{v}_i \boldsymbol{v}_i^t] = \boldsymbol{\Lambda}_i \geq \boldsymbol{0}$$

$$E[\boldsymbol{v}_i \boldsymbol{v}_j^t] = 0 \qquad \forall i \neq j.$$

These assumptions are reasonable. If the estimator is biased, it is often possible to subtract its mean to get an unbiased one. If we do not know the covariance of the error (or at least an upper bound of it), the estimator is meaningless. If two measurements $\hat{\boldsymbol{x}}_i$ and $\hat{\boldsymbol{x}}_j$ are correlated, we take the concatenation of them $\hat{\boldsymbol{x}}_k = (\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{x}}_j)$ and the concatenated vector function $\boldsymbol{f}_k = [\boldsymbol{f}_i^t, \boldsymbol{f}_j^t]^t$. The problem is to find the optimal estimate $\hat{\boldsymbol{a}}$ of $\boldsymbol{a}$ given the function $\boldsymbol{f}_i$ and the measurements $\hat{\boldsymbol{x}}_i$.

B. Linearizing the Equations

The most powerful tools developed in parameter estimation are for linear systems. We decided to apply these tools to a linearized version of our equations. This is the EKF approach that we now develop.

For each nonlinear equation $\boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{a}) = \boldsymbol{0}$ we need to know an estimate $\hat{\boldsymbol{a}}_{i-1}$ of the sougth parameter $\boldsymbol{a}$, and again a measure $\boldsymbol{S}_i$ of the confidence we have in this estimate.[1] Actually, we model probalistically the current estimate $\hat{\boldsymbol{a}}_{i-1}$ of $\boldsymbol{a}$ by assuming that

$$\hat{\boldsymbol{a}}_{i-1} = \boldsymbol{a} + \boldsymbol{w}_i \tag{3}$$

where $\boldsymbol{w}_i$ is a random error. The only assumptions we make on $\boldsymbol{w}_i$ are the same as for $\boldsymbol{v}_i$, i.e.,

$$E[\boldsymbol{w}_i] = 0$$

$$E[\boldsymbol{w}_i \boldsymbol{w}_i^t] = \boldsymbol{S}_i \geq \boldsymbol{0}$$

where $\boldsymbol{S}_i$ is a given non-negative matrix. Here again, no assumption of gaussianness is required.

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

$$\boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{a}) = \boldsymbol{0} \approx \boldsymbol{f}_i(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{a}}_{i-1}) + \frac{\hat{\partial \boldsymbol{f}_i}}{\partial \boldsymbol{x}} (\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i) + \frac{\partial \boldsymbol{f}_i}{\partial \boldsymbol{a}} (\boldsymbol{a} - \hat{\boldsymbol{a}}_{i-1}) \tag{4}$$

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

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

Equation (4) can be rewritten as

$$y_i = M_i a + u_i \tag{5}$$

where

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

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

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

Equation (5) 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

$$E[u_i] = 0$$

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

C. 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 a_k as a linear combination of the measurements $\{y_i\}$, it is the one which minimizes the expected error norm squared

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

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 as follows [23]:

$$\hat{a}_i = \hat{a}_{i-1} + K_i(y_i - M_i\hat{a}_{i-1}) \tag{6}$$

$$K_i = S_{i-1}M_i^t(W_i + M_iS_{i-1}M_i^t)^{-1} \tag{7}$$

$$S_i = (I - K_iM_i)S_{i-1} \tag{8}$$

or equivalently

$$S_i^{-1} = S_{i-1}^{-1} + M_i^tW_i^{-1}M_i. \tag{9}$$

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

$$C = (a - \hat{a}_0)^tS_0^{-1}(a - \hat{a}_0) + \sum_{i=1}^{k}(y_i - M_ia)^tW_i^{-1}(y_i - M_ia). \tag{10}$$

It is interesting to note that the first term of (10) measures the squared distance of a from an initial estimate, weighted by its covariance matrix, while the second term is nothing else but the classical least square criterion, i.e., the sum of the squared measurement errors weighted by the covariance matrices. Indeed, initializing the process with an arbitrary $\hat{a}_0$ and $S_0^{-1} = \mathbf{0}$, criterion (10) 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 $\{\hat{x}_i\}$, to stop the measures, we only have to keep $\hat{a}_k$ and $\hat{S}_k$ as the whole memory of the measurement process. If we decide later to take into account additional measurements, we simple have to initialize $\hat{a}_0 \sim \hat{a}_k$ and $S_0 \sim S_k$ and to process the new measurements to obtain exactly the same solution as if we had processed all the measurements together.

D. Gaussian Assumption

Up to now, we did not introduce any Gaussian assumptions on the random measurement errors $v_i = x_i - \hat{x}_i$ of (2) and on the prior estimate error $w_0 = a - \hat{a}_0$ of (3). 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.,

$$v_i \sim N(\mathbf{0}, \Lambda_i)$$

$$w_0 \sim N(\mathbf{0}, S_0)$$

then, it follows that the noise u_i in (5) is also Gaussian, i.e.,

$$u_i \sim N(\mathbf{0}, W_i)$$

and that all the successive estimates provided by the recursive Kalman filter are also Gaussian, with mean a and covariance S_k

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

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 $\hat{a}_k = E[a/y_1, \cdots, y_k]$ 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 [23].

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.

E. Rejecting Outlier Measurements

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

If we consider again a first-order expansion of $\boldsymbol{f}_i(\boldsymbol{x}_i, \boldsymbol{a})$ around $(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{a}}_{i-1})$ (4), considering that $(\hat{\boldsymbol{x}}_i - \boldsymbol{x}_i)$ and $(\hat{\boldsymbol{a}}_{i-1} - \boldsymbol{a})$ are independent centered Gaussian processes, we see that $\boldsymbol{f}_i(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{a}}_{i-1})$ is also (up to a linear approximation) a centered Gaussian process whose mean and covariance are given by

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

$$\boldsymbol{Q}_i = E[\boldsymbol{f}_i(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{a}}_{i-1})\boldsymbol{f}_i(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{a}}_{i-1})^t] = \frac{\hat{\partial \boldsymbol{f}_i}}{\partial \boldsymbol{x}} \boldsymbol{\Lambda}_i \frac{\hat{\partial \boldsymbol{f}_i^t}}{\partial \boldsymbol{x}} + \frac{\hat{\partial \boldsymbol{f}_i}}{\partial \boldsymbol{a}} \boldsymbol{S}_{i-1} \frac{\hat{\partial \boldsymbol{f}_i^t}}{\partial \boldsymbol{a}}.$$

Therefore, if the rank of $\boldsymbol{Q}_i$ is q, the generalized Mahalanobis distance

$$d(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{a}}_{i-1}) = [\boldsymbol{f}_i(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{a}}_{i-1})]^t \boldsymbol{Q}_i^{-1} [\boldsymbol{f}_i(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{a}}_{i-1})] \quad (11)$$

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

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

$$d(\hat{\boldsymbol{x}}_i, \hat{\boldsymbol{a}}_{i-1}) < \epsilon. \quad (12)$$

We shall see in the experimental section at the end of this paper how this formalism can be used in practice, and how well it fits with reality.

IV. Geometric Representations

In this section, we give the details of the geometric representations that we have found useful at various stages of our work. It is first important to note that we have been dealing so far only with points, lines, and planes, i.e., with affine geometric entities. This may appear to be quite a restriction on the type of environments that we can cope with. This is indeed the case but there are a number of reasons why we think that our approach is quite reasonable.

1) The obvious one is that for the kind of environment that our mobile robot moves into, these primitives are very likely to cover most of the geometric features of importance.

2) A second reason is that more complicated curved features can be first approximated with affine primitives which are then grouped into more complicated nonaffine primitives.

3) A third reason is that we believe that the techniques we have developed for representing and combining uncertainty of affine primitives are generic and directly applicable to nonaffine primitives.

Let us now discuss specifically lines, planes, and rigid displacements.

[2] If $q < p$ = the size of the measurement vector $\boldsymbol{f}_i$, $\boldsymbol{Q}_i^{-1}$ is the pseudo-inverse of $\boldsymbol{Q}_i$.

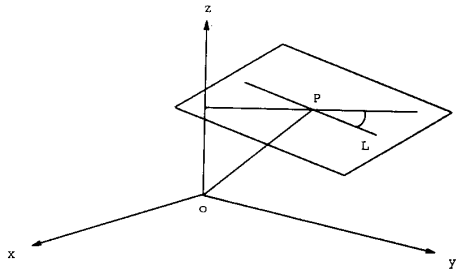


Fig. 11. A possible 3-D line representation.

A. Line Segments

The 3-D segments that we deal with are usually constructed from stereo [4], [9]. Their endpoints may be quite unreliable, even though they can be of some use from time to time, and we largely depend on the infinite lines supporting those line segments.

We concentrate here on how to represent 3-D lines. The obvious representation we mention here only for pedagogical reasons, is by two points; this representation is six-dimensional and, as we will see next, not minimal. Another way to represent a line is to choose a point on it (three parameters), and a unit vector defining its direction (two parameters). The corresponding representation is five-dimensional and, again, not minimal. In fact, the set of affine 3-D lines is a manifold of dimension 4 for which we will exhibit later an atlas of class C^∞.[3] This implies that a minimal representation of a straight line has four parameters.

One such representation can be obtained by considering the normal to the line from the origin (if the line goes through the origin it is the same as a vector line and can be defined by two parameters only). The point of intersection between the normal and the line is represented by three parameters. If we now consider (see Fig. 11) the plane normal at P to OP, the line is in that plane and can be defined by one more parameter, its angle with an arbitrary direction, for example, the line defined by P, and one of the axis of coordinates (in Fig. 11, the z axis). Of course, when the line is parallel to the xy plane this direction is not defined and we must use either the x or the y axis. This brings up an interesting point, namely, that a global minimal representation for affine lines, i.e., one which can be used for all such lines, does not exist. We must choose the representation as a function of the line orientation. Mathematically, this means that the manifold of the affine straight lines cannot be defined with only one map. This is quite common and is also true for affine planes and rotations of R^3, as will be shown next.

The previous representation for a line is not in fact the one

[3] Grossly speaking, a manifold of dimension d is a set that can be defined locally by d parameters. When the functions that transform one set of parameters into another are p times differentiable, the manifold is said to be of class C^p. For more details, see [14].

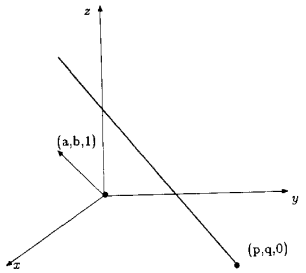


Fig. 12. A better 3-D line representation.

we have been using. In effect, the parameters involved in the previous representation are usually combined in a highly nonlinear manner in the measurement equations expressing geometric relationships between geometric entities (cf. next section), which is not good for the extended Kalman filtering approach. Also, the angular parameter must be assigned some fixed bounds (for instance $]0, \pi[$), which might cause some problems during a recursive evaluation with the Kalman filter. This latter constraint also appears in the representation recently proposed by Roberts [31].

Therefore, we prefer the following representation in which the retained parameters are usually combined linearly in the measurement equations, and are not constrained to any bounded interval. This representation considers a line (not perpendicular to the z axis) as the intersection of a plane parallel to the y axis, and a plane parallel to the x axis

$$\begin{cases} x = az + p \\ y = bz + q. \end{cases} \tag{13}$$

The intersection is represented by the four-dimensional vector $\boldsymbol{L} = [a, b, p, q]^T$ which has the following geometric interpretation (see Fig. 12): The direction of the line is that of the vector $[a, b, 1]^T$, and the point of intersection of the line with the xy plane has coordinates p and q. Since the last coordinate of the direction vector is equal to 1, the line cannot be perpendicular to the z axis or parallel to the xy plane. If we want, and we do in practice, represent such lines, we must choose another representation, for example

$$\begin{cases} y = ax + p \\ z = bx + q \end{cases} \tag{14}$$

which cannot represent lines parallel to the yz plane, or perpendicular to the x axis, or

$$\begin{cases} z = ay + p \\ x = by + q \end{cases} \tag{15}$$

which excludes lines parallel to the zx plane.

Each representation defines a one-to-one mapping between R^4 and a subset (in fact an open subset) of the set of affine 3-D lines and it can be shown that these three mappings define on this set a structure of C^∞ manifold for which they form an atlas. In practice, this means the representation is not exactly four-dimensional, but is made of the four numbers a, b, p, and q and an integer i taking the values 1, 2, and 3 to indicate which map 13, 14, or 15 we are currently using.

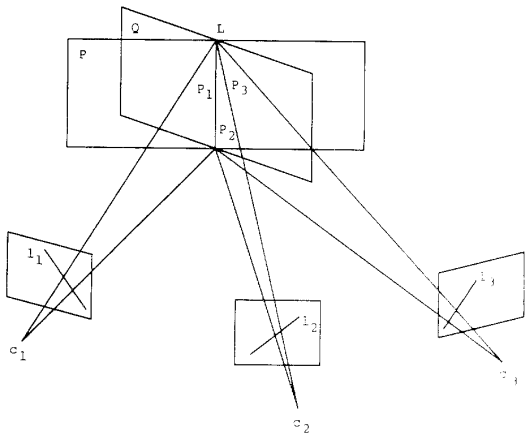


Fig. 13. Reconstruction of 3-D lines.

The fact that the set of affine 3-D lines has been given a structure of C^∞ manifold implies that the a', b', p', q' of a given representation are C^∞ functions of the a, b, p, q of another representation for all lines for which the two representations are well defined (for example, all lines not parallel to the xy and yz planes). The representation of a line also includes a 4×4 covariance matrix Δ_L on the vector $\boldsymbol{L}$.

It is interesting at this stage to trace the computation of this covariance matrix all the way from pixel to 3-D. In order to do this, we must briefly explain how 3-D lines are computed in our current Stereo system [9]. We use three cameras as indicated in Fig. 13. In theory, the three planes defined by the 2-D lines l_1, l_2, and l_3 and the optical centers C_1, C_2, and C_3 belong to the same pencil and intersect along the 3-D line L. In practice they do not because of noise, and we have to find the "best" line satisfying the measurements, i.e., l_1, l_2, and l_3. This can be done by using the idea of pencil of planes, described more fully in [21]. We assume that in the coordinate system attached to camera 1, for example, the equation of the ith plane P_i, $i = 1, 2, 3$, is given by

$$u_i x + v_i y + w_i z + r_i = 0$$

where the three four-vectors $\boldsymbol{P}_i = [u_i, v_i, w_i, r_i]^T$ are known, as well as their covariance matrix Λ_{P_i} (we show later how to compute them). If we use representation (13) for the 3-D line, it is represented as the intersection of the two planes P of the equation $x = az + p$ and Q of the equation $y = bz + q$. Writing that the five planes P, Q, and P_i, $i = 1, 2, 3$, form a pencil allows us to write six equations

$$\begin{cases} w_i + au_i + bv_i = 0, \\ r_i + pu_i + qu_i = 0, \end{cases} \qquad i = 1, 2, 3$$

in the four unknowns a, b, p, and q.

We can apply directly the Kalman formalism to these measurement equations and choose $\boldsymbol{a} = [a, b, p, q]^T$, and x_i as the four-vector $\boldsymbol{P}_i$. We can therefore simply compute an estimate $\hat{\boldsymbol{a}}$ of $\boldsymbol{a}$ and its covariance matrix $\Lambda_{\hat{a}}$ from the $\boldsymbol{P}_i$'s and Λ_{P_i}'s.

Let us now show how we can compute the $\boldsymbol{P}_i$'s and Λ_{P_i}'s. Each line l_i, $i = 1, 2, 3$, is obtained by fitting a straight line to

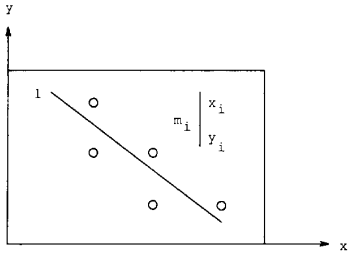


Fig. 14. 2-D line approximation.

a set of edge pixels which have been detected using a modified version of the Canny edge detector [13], [17]. Looking at Fig. 14, let $x \cos\theta + y \sin\theta - p = 0$ be the equation of the line l which is fit to the edge pixels m_i of coordinates x_i, y_i, $(0 \leq \theta < 2\pi, \rho \geq 0)$. We assume that the measured edge pixels are independent and corrupted by a gaussian isotropic noise and take the parameter $\boldsymbol{a}$ equal to $[\theta, \rho]^T$ and the measurement $\boldsymbol{x}$ as the vector $[x, y]^T$. The measurement equation is therefore

$$f(\boldsymbol{x}, \boldsymbol{a}) = x \cos\theta + y \sin\theta - \rho.$$

Applying the EKF formalism to the n-edge pixels forming the line provides the best estimate $\hat{\boldsymbol{a}}$ of the line parameters and its covariance matrix. Having done this for all three cameras, it is easy to deduce the equations of the three planes P_i and the covariance matrices on their coefficients.

B. Planes

Planes can receive pretty much the same treatment as lines. A plane is defined by three parameters, and this is minimal. A possible representation is the representation by the normal $\hat{\boldsymbol{n}}$ (a unit norm vector), and the distance d to the origin. The problem with this representation is that it is not unique since $(-\boldsymbol{n}, -d)$ represents the same plane. It is possible to fix that problem by assuming that one component of $\boldsymbol{n}$, say n_z, is positive, i.e., we consider planes not parallel to the z axis. For these planes we must choose another convention, for example, that n_z is positive. Again, this works well for planes not parallel to the x axis. The third possible representation is to assume n_y positive which excludes planes parallel to the y axis.

So, we have three one-to-one mappings of open subsets of the product $S_2 \times R$, where S_2 is the usual Gaussian sphere into open subsets of the set of planes

$$(\boldsymbol{n}, d),\ n_z > 0 \rightarrow \text{planes not parallel to } Oz$$

$$(\boldsymbol{n}, d),\ n_x > 0 \rightarrow \text{planes not parallel to } Ox$$

$$(\boldsymbol{n}, d),\ n_y > 0 \rightarrow \text{planes not parallel to } Oy.$$

It is easy to show that these three mappings define on the set of 3-D planes a structure of C^∞ manifold of dimension 3.

One practical disadvantage of the previous representations is that the normal $\boldsymbol{n}$ is constrained to lie on the unit sphere S_2, i.e., it must satisfy the constraint $\|\boldsymbol{n}\| = 1$. A possibly simpler representation is obtained by considering the mapping from R^3 to the set of 3-D planes defined by

$$p_1 : (a, b, c) \rightarrow ax + by + z + c = 0. \tag{16}$$

This can represent all planes except those parallel of Oz and it is a one-to-one continuous mapping from R^3 to the open subset of the set of 3-D planes constituted of the planes not parallel to the z axis. In order to obtain all possible planes, we must also consider the mappings

$$p_2 : (a, b, c) \rightarrow x + ay + bz + c = 0 \tag{17}$$

$$p_3 : (a, b, c) \rightarrow bx + y + az + c = 0. \tag{18}$$

p_2 (respectively, p_3) excludes planes to the x axis (respectively, the y axis). It is easy to show that p_1, p_2, p_3 also define on the set of 3-D planes a structure of C^∞ manifold of dimension 3.

C. Rigid Displacements

In a previous paper [4], [6] we have proposed the use of the exponential representation of rotations. This is the same as saying that a rotation is defined by its axis $\boldsymbol{u}$ (a unit vector) and its angle θ. The vector $\boldsymbol{r} = \theta\boldsymbol{u}$ can be used to represent the rotation and we have

$$\boldsymbol{R} = e^{\boldsymbol{H}}$$

where $\boldsymbol{H}$ is an antisymmetric matrix representing the cross product with the vector $\boldsymbol{r}$ (i.e., $\boldsymbol{Hx} = \boldsymbol{r} \times \boldsymbol{x}$, for all $\boldsymbol{x}$). In this case, the rotation is represented by the three coordinates of $\boldsymbol{r}$, i.e., by three independent numbers. There are several other possible representations for rotations, the most widely known being the one with orthogonal matrices or quaternions. Their main disadvantage is that an orthogonal matrix is defined by nine numbers subject to six quadratic constraints, whereas a quaternion is defined by four numbers subject to one quadratic constraint. These constraints are not easy to deal with in the EKF formalism and, moreover, these two representations are more costly than the exponential one.

Let us see how we can define a structure of manifold on the set of rotations using this representation. If we allow θ to vary over the semi-open interval $[0, 2\pi[$, the vector $\boldsymbol{r}$ can vary in the open ball $B(0, 2\pi)$ of R^3 of radius 2π. But the mapping f: $B(0, 2\pi)$ into the set of rotations is not one to one because (u, π) and $(-u, \pi)$ represent the same rotation. To enforce uniqueness we can assume that one of the coordinates, for example u_z, of the rotation axis u is positive. We can then represent uniquely the open subset of the set of rotations for which the axis is not perpendicular to the z axis, and has a positive component along the axis, and the mapping is continuous. If we consider the open set of rotations defined by (u, θ), $u_z < 0$, we have another one-to-one continuous mapping. With these two mappings, we cannot represent rotations with an axis perpendicular to the z axis. In order to obtain all possible rotations, we have to introduce the other four mappings defined by (u, θ) and $u_x > 0$ (respectively, $u_x < 0$, $u_y > 0$, $u_y < 0$) which represent rotations with an axis not perpendicular to the x axis (respectively, the y axis). We are sill missing the null vector, i.e., we have no representation for the null rotation, the identity matrix. In order to include it, we have to add a seventh map by considering for example the rotations defined by the "small" open ball $B(0, \epsilon)$ where ϵ

must be smaller than π. These seven mappings define on the set of rotations a structure of C^∞ manifold of dimension 3.[4]

It is interesting that in all three cases (3-D lines, planes, and rotations), unique global representation does not exist and that we must deal with at least three local mappings.

It is now instructive to study how the group of rigid displacements operates on the representations for lines and planes.

1) Applying Rigid Displacement to Lines: The easiest way to derive how representation (13) changes under rotation and translation is by considering that the line is defined by two points M_1 and M_2 of coordinates (x_1, y_1, z_1) and (x_2, y_2, z_2). It is then easy to verify that

$$a=\frac{x_2-x_1}{z_2-z_1} \quad b=\frac{y_2-y_1}{z_2-z_1}$$

$$p=\frac{z_2x_1-z_1x_2}{z_2-z_1} \quad q=\frac{z_2y_1-z_1y_2}{z_2-z_1}.$$

Introducing the vector $M_1M_2 = [A, B, C]^T$, we have $a = A/C$, and $b = B/C$. a and b are therefore only sensitive to rotation

$$M_1M_2 \rightarrow \boldsymbol{R}\, M_1M_2$$

$$\begin{bmatrix} A \\ B \\ C \end{bmatrix} \rightarrow \begin{bmatrix} A' \\ B' \\ C' \end{bmatrix} = \boldsymbol{R} \begin{bmatrix} A \\ B \\ C \end{bmatrix}$$

$$\begin{bmatrix} a \\ b \end{bmatrix} \rightarrow \begin{bmatrix} a' \\ b' \end{bmatrix} = \begin{bmatrix} A'/C' \\ B'/C' \end{bmatrix}.$$

This yields

$$a'=\frac{\boldsymbol{r}_1 \cdot \boldsymbol{m}}{\boldsymbol{r}_3 \cdot \boldsymbol{m}} \qquad b'=\frac{\boldsymbol{r}_2 \cdot \boldsymbol{m}}{\boldsymbol{r}_3 \cdot \boldsymbol{m}}$$

where $\boldsymbol{m} = [a, b, 1]^T$, and the $\boldsymbol{r}_i$'s are the row vectors of matrix $\boldsymbol{R}$. This is true only if $\boldsymbol{r}_3 \cdot \boldsymbol{m} \neq 0$; if $\boldsymbol{r}_3 \cdot \boldsymbol{m} = 0$, the transformed line is perpendicular to the z axis and representation (14) or (15) must be used.

To treat the case of p and q, let us introduce $P = p(z_2 - z_1) = pC$ and $Q = q(z_2 - z_1) = qC$. It is easy to show that

$$\begin{bmatrix} P \\ Q \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix} OM_1 \times OM_2 = H(OM_1 \times OM_2).$$

This allows us to study how P and Q change under rotation and translation

$$OM_1 \rightarrow \boldsymbol{R}\, OM_1 + t \qquad OM_2 \rightarrow \boldsymbol{R}\, OM_2 + t.$$

Therefore

$$\begin{bmatrix} P \\ Q \end{bmatrix} \rightarrow \begin{bmatrix} P' \\ Q' \end{bmatrix} = \boldsymbol{H}(\boldsymbol{R}(OM_1 \times OM_2) + t \times \boldsymbol{R}\, M_1M_2).$$

Using the previous notations, $M_1M_2 = [A, B, C]^t$, and $OM_1 \times OM_2 = [Q, -P, X]^t$ where X is unknown. But noticing that $M_1M_2 \cdot (OM_1 \times OM_2) = 0$ we have

$$AQ - BP + CX = 0$$

and therefore

$$X=\frac{BP-AQ}{C}=bP-aQ.$$

C is not equal to 0 since by definition, the line is not perpendicular to the z axis. Putting everything together

$$\begin{bmatrix} P' \\ Q' \end{bmatrix} = CH\left(\boldsymbol{R}\begin{pmatrix} q \\ -p \\ bp-aq \end{pmatrix} + t \times \boldsymbol{Rm}\right).$$

Finally

$$\begin{bmatrix} p' \\ q' \end{bmatrix} = \begin{bmatrix} P'/C' \\ Q'/C' \end{bmatrix} = \frac{C}{C'}\, \boldsymbol{H}\left(\boldsymbol{R}\begin{pmatrix} q \\ -p \\ bp-aq \end{pmatrix} + t \times \boldsymbol{Rm}\right)$$

and we know from the previous derivation that $C/C' = 1/\boldsymbol{r}_3 \cdot \boldsymbol{M}$, therefore

$$\begin{bmatrix} p' \\ q' \end{bmatrix} = \frac{1}{\boldsymbol{r}_3 \cdot \boldsymbol{m}}\, \boldsymbol{H}(\boldsymbol{Rp} + t \times \boldsymbol{Rm})$$

where we have taken $\boldsymbol{p} = [q, -p, bp - aq]^t$.

2) Applying Rigid Displacements to Planes: Given a plane represented by its normal n and its distance to the origin d, if we apply to it a rotation along an axis going through the origin represented by a matrix $\boldsymbol{R}$ followed by a translation represented by a vector $\boldsymbol{t}$, the new plane is represented by $\boldsymbol{Rn}$ and $d - \boldsymbol{t} \cdot \boldsymbol{Rn}$ [20].

This allows us to compute how the representation (16), for example, is transformed by the rigid displacement. From the previous observation:

$$\begin{pmatrix} a \\ b \\ 1 \end{pmatrix} \rightarrow \boldsymbol{R}\begin{pmatrix} a \\ b \\ 1 \end{pmatrix} \quad \text{and} \quad c \rightarrow c - t \cdot \boldsymbol{R}\begin{pmatrix} a \\ b \\ 1 \end{pmatrix}.$$

Introducing the three row vectors $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, $\boldsymbol{r}_3$ of matrix $\boldsymbol{R}$, we have, assuming that $\boldsymbol{r}_3 \cdot \boldsymbol{m} \neq 0$.

$$a'=\frac{\boldsymbol{r}_1 \cdot \boldsymbol{m}}{\boldsymbol{r}_3 \cdot \boldsymbol{m}} \qquad b'=\frac{\boldsymbol{r}_2 \cdot \boldsymbol{m}}{\boldsymbol{r}_3 \cdot \boldsymbol{m}} \qquad c'=\frac{c - t \cdot \boldsymbol{Rm}}{\boldsymbol{r}_3 \cdot \boldsymbol{m}}$$

if $\boldsymbol{r}_3 \cdot \boldsymbol{m} = 0$, this means that we cannot use the same representation for the transformed plane since it is parallel to the z axis, therefore, we must choose the representation (17) or (18).

V. Registration, Motion, and Fusion of Visual Maps

In this section we show how to solve the problems listed in Section II within the formalism and the representations detailed in Sections III and IV.

A. Initial Assumptions

We are given two visual maps $\mathcal{V}$ and $\mathcal{V}'$, each of them attached to a coordinate reference frame $\mathcal{F}$ and $\mathcal{F}'$ (see Fig. 15).

[4] In [10] and [11] one can find an atlas of rotations with only four maps.

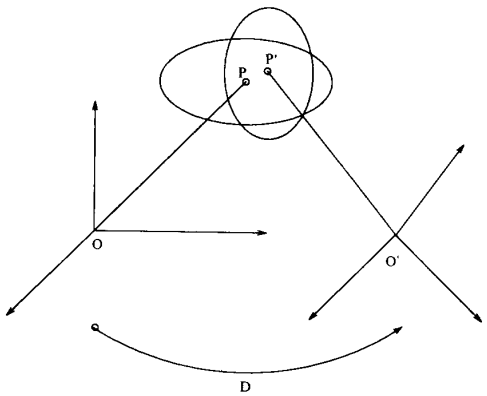


Fig. 15. The general registration motion fusion problem.

TABLE I
RELATIONS BETWEEN THE PRIMITIVES

Relations	Points	Lines	Planes
Points	$\equiv$	$\subset$	$\subset$
Lines		$\equiv \parallel \perp$	$\subset \parallel \perp$
Planes			$\equiv \parallel \perp$

Each visual map $\mathcal{V}$ is composed of primitives $\mathcal{P}$, described by a parameter vector $\boldsymbol{P}$. We have an estimate $\hat{\boldsymbol{P}}_0$ of $\boldsymbol{P}$ and an error covariance matrix $\boldsymbol{W}_{P_0}$.

The coordinate frames $\mathcal{F}$ and $\mathcal{F}'$ are related by a rigid displacement $\mathcal{D}$ such that each point M' of $\mathcal{F}'$ is related to a point M of $\mathcal{F}$ by the relation

$$O'M' = \boldsymbol{R}\, OM + \boldsymbol{t}$$

where $\boldsymbol{R}$ is the rotation matrix and $\boldsymbol{t}$ the translation vector of the displacement $\mathcal{D}$. We also have an estimate $\hat{\boldsymbol{D}}_0$ of $\boldsymbol{D}$, with an error covariance matrix $\boldsymbol{W}_{D_0}$.

B. Defining Geometric Relations

We define a set of geometric relations between the primitive $\mathcal{P}$ and $\mathcal{P}'$ of two visual maps $\mathcal{V}$ and $\mathcal{V}'$. These relations are given in Table I.

The list of relations/primitives is not exhaustive but only demonstrative. The relation "identical" expresses the fact that the primitives $\mathcal{P}$ and $\mathcal{P}'$ represented in $\mathcal{V}$ and $\mathcal{V}'$ actually describe the same physical primitive. The relation "included" expresses that $\mathcal{P}$ describes a physical primitive which is part of the physical primitive described by $\mathcal{P}'$. The relations "parallel" and "orthogonal" are interpreted in a similar fashion.

Each geometric relation can be expressed by a vector equation of the form

$$f_i(\boldsymbol{P}, \boldsymbol{P}', \boldsymbol{D}) = \boldsymbol{0}. \tag{19}$$

C. Expressing Geometric Relations

We rewrite (19) for the geometric relations of Table I. We denote by $\bar{\boldsymbol{P}}$ the parameters of the primitive $\bar{\mathcal{P}} = \boldsymbol{D}(\mathcal{P})$, the image of $\mathcal{P}$ by the rigid displacement $\boldsymbol{D}$. The computation of $\bar{\boldsymbol{P}}$ from $\boldsymbol{P}$ is, in the case of points, $\overline{OM} = \boldsymbol{R}\, OM + \boldsymbol{t}$. The case of lines and planes was detailed in the previous section. The measurement equations are as follows:

Point-Point:

$$\text{relation } \equiv : O'M' - \overline{OM} = \boldsymbol{0}.$$

Point-Line: assuming the line is not orthogonal to the z axis:

$$\text{relation } \subset : \begin{cases} \bar{x} - a'\bar{z} - p' = 0 \\ \bar{y} - b'\bar{z} - q' = 0. \end{cases}$$

Point-Plane: assuming the plane is not parallel to the z axis:

$$\text{relation } \subset : a'\bar{x} + b'\bar{y} + \bar{z} + c' = 0.$$

Line-Line: assuming the two lines are not orthogonal to the z axis:

$$\text{relation } \equiv : (a', b', c', d')^t - (\bar{a}, \bar{b}, \bar{c}, \bar{d})^t = \boldsymbol{0}$$

$$\text{relation } \parallel : (a', b')^t - (\bar{a}, \bar{b})^t = \boldsymbol{0}$$

$$\text{relation } \perp : a'\bar{a} + b'\bar{b} + 1 = 0.$$

Line-Plane: assuming the line is not orthogonal and the plane not parallel to the z axis:

$$\text{relation } \subset : \begin{cases} a'\bar{a} + b'\bar{b} + 1 = 0 \\ a'\bar{p} + b'\bar{q} + c' = 0 \end{cases}$$

$$\text{relation } \parallel : a'\bar{a} + b'\bar{b} + 1 = 0$$

$$\text{relation } \perp : (a', b')^t - (\bar{a}, \bar{b})^t = \boldsymbol{0}.$$

Plane-Plane: assuming the plane is not parallel to the z axis:

$$\text{relation } \equiv : (a', b', c')^t - (\bar{a}, \bar{b}, \bar{c})^t = \boldsymbol{0}$$

$$\text{relation } \parallel : (a', b')^t - (\bar{a}, \bar{b})^t = \boldsymbol{0}$$

$$\text{relation } \perp : a'\bar{a} + b'\bar{b} + 1 = 0.$$

This approach should be compared to that of [29].

D. Registration

1) Principle: The registration (or matching) of two primitives $\mathcal{P}$ and $\mathcal{P}'$ consists in detecting that their parameters $\boldsymbol{P}$ and $\boldsymbol{P}'$ verify (19) for one of the above listed geometric relations, with respect to the current noisy estimates $(\hat{\boldsymbol{P}}_0, \boldsymbol{W}_{P_0})$, $(\hat{\boldsymbol{P}}'_0, \boldsymbol{W}_{P'_0})$, and $(\hat{\boldsymbol{D}}_0, \boldsymbol{W}_{D_0})$ of $\boldsymbol{P}$, $\boldsymbol{P}'$, and $\boldsymbol{D}$.

This "detection" is done by computing between each pair of primitive the generalized Mahalanohis distance given by (11), and by matching a pair of primitives each time the χ^2 acceptance test given by inequality (12) is verified, i.e., when

$$d(\hat{\boldsymbol{P}}_0, \boldsymbol{W}_{P_0}, \hat{\boldsymbol{P}}'_0, \boldsymbol{W}_{P'_0}, \hat{\boldsymbol{D}}_0, \boldsymbol{W}_{D_0}) < \epsilon. \tag{20}$$

2) Reliability: The above-described registration procedure detects what would be called *plausible* matches between geometric primitives. When the uncertainty attached to the primitives parameters is large, it may happen that a plausible match is false. In order to improve the reliability of the procedure, one can use a strategy (inspired by [16]) which starts by registrating primitives whose parameters have a small covariance matrix, or primitives which can be matched

unambiguously. Such a strategy is exemplified in the experimental results section of this paper and also in another paper [8].

3) Efficiency: In order to avoid a $O(n^2)$ complexity algorithm, it is of course possible to use additional control structures to select a subset of candidate primitives for each test. For instance, to test the relation "$\equiv$" between points or lines, bucketing techniques can be used with efficiency (see for instance [1], [7]).

E. Motion

Having registered two primitives $\mathcal{P}$ and $\mathcal{P}'$, the motion problem consists in reducing the uncertainty on the *motion parameters* $\boldsymbol{D}$ while taking into account the uncertainty on the parameters $\boldsymbol{P}$, $\boldsymbol{P}'$, and $\boldsymbol{D}$.

This is done by setting $\boldsymbol{a} = \boldsymbol{D}$ and $\boldsymbol{x} = (\boldsymbol{P}, \boldsymbol{P}')^t$, and by using the relation equation (19) as a measurement equation (1)

$$f_i(\boldsymbol{x}, \boldsymbol{a}) \equiv f_i((\boldsymbol{P}, \boldsymbol{P}'), \boldsymbol{D}) = \boldsymbol{0}.$$

Starting from the initial estimate $\hat{\boldsymbol{a}}_0 = \hat{\boldsymbol{D}}_0$, $\boldsymbol{S}_0 = \boldsymbol{W}_{D_0}$, and using the measurement $\hat{\boldsymbol{x}}_1 = (\hat{\boldsymbol{P}}_0, \boldsymbol{P}_0')^t$) with

$$\boldsymbol{W}_1 = \begin{pmatrix} \boldsymbol{W}_{P_0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{W}_{P_0'} \end{pmatrix}$$

one applies the EKF formalism to obtain a new estimate $\hat{\boldsymbol{a}}_1$ of the motion with a reduced covariance matrix $\boldsymbol{S}_1 < \boldsymbol{S}_0$. (In the sense $\boldsymbol{S}_0 - \boldsymbol{S}_1$ is nonnegative).

This process is recursively repeated: at iteration i, if a new pair of primitives can be registered with the new motion estimate $(\hat{\boldsymbol{a}}_{i-1}, \boldsymbol{S}_{i-1})$, the additional measurement equations they bring lead to a new better estimate $\hat{\boldsymbol{a}}_i$ of the motion with a still reduced covariance matrix $\boldsymbol{S}_i$. This process ends after the matching of k primitives with a final estimate $(\hat{\boldsymbol{a}}_k, \boldsymbol{S}_k)$ of the motion parameter $\boldsymbol{D}$.

F. Fusion

1) General Fusion: The fusion problem is exactly the dual of the motion problem, as it consists, after the registration of two primitives, in reducing the uncertainty on the *primitive parameters* $\boldsymbol{P}$ and $\boldsymbol{P}'$ while taking into account the uncertainty on the parameters $\boldsymbol{P}$, $\boldsymbol{P}'$, and $\boldsymbol{D}$.

This is done by "switching the attention," i.e., by choosing $\boldsymbol{a} = (\boldsymbol{P}, \boldsymbol{P}')^t$ and $\boldsymbol{x} = \boldsymbol{D}$ while using again the relation equation (19) as a measurement equation (1)

$$f_i(\boldsymbol{x}, \boldsymbol{a}) \equiv f_i(\boldsymbol{D}, (\boldsymbol{P}, \boldsymbol{P}')) = 0.$$

The initial estimate is taken as $\hat{\boldsymbol{a}}_0 = (\hat{\boldsymbol{P}}_0, \hat{\boldsymbol{P}}_0')^t$ and

$$\boldsymbol{S}_0 = \begin{pmatrix} \boldsymbol{W}_{P_0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{W}_{P_0'} \end{pmatrix}$$

and one uses the measurement $\hat{\boldsymbol{x}}_1 = \hat{\boldsymbol{D}}_0$ with $\boldsymbol{W}_1 = \boldsymbol{W}_{D_0}$ to apply the EKF formalism and obtain a new estimate $\hat{\boldsymbol{a}}_1$ of the primitive parameters with a reduced covariance matrix $\boldsymbol{S}_1 < \boldsymbol{S}_0$.

If additional relations hold between these primitives and other ones, the same treatment allows for a further reduction in their parameters uncertainty, and therefore a more accurate estimation of the primitive parameters.

2) Forgetting Primitives: After the treatment of a constraint, the parameters $\boldsymbol{P}_1$ and $\boldsymbol{P}_1'$ of the primitives are usually correlated, which means that the covariance matrix

$$\text{cov}\,(\hat{\boldsymbol{P}}_1, \hat{\boldsymbol{P}}_1') = \begin{pmatrix} \boldsymbol{W}_{P_1} & \boldsymbol{W}_{P_1 P_1'} \\ \boldsymbol{W}_{P_1' P_1} & \boldsymbol{W}_{P_1} \end{pmatrix}$$

contains $\boldsymbol{W}_{P_1' P_1} = \boldsymbol{W}_{P_1 P_1'}^t \neq 0$.

Therefore, it is no longer possible to treat independently $\mathcal{P}$ and $\mathcal{P}'$ in successive measurement equations. One has to consider them as a new primitive, either by keeping only one of them, or the union of them.

For instance if one updates the parameters of $\mathcal{P}'$ with those of an "identical" primitive $\mathcal{P}$ observed in a previous visual map, one keeps only the updated parameters of $\mathcal{P}'$ in the new map, with their covariance matrix $\boldsymbol{W}_{p'}$, forgetting the previous parameters $\mathcal{P}$ after having used them.

On the other hand, if one updates the parameters of two lines by detecting that they are orthogonal, one keeps the new primitive formed by the union of the updated two lines, with the corresponding covariance matrix. One must use this kind of relation carefully, in order to control the size of the state parameter $\boldsymbol{a}$.

3) Autofusion: In the special case where $\mathcal{V} \equiv \mathcal{V}'$, all primitives come from the same visual map, and the motion parameters vanish as they correspond to the identity transform and are perfectly known.

Nevertheless, one can still detect the previous geometric relations between pairs of primitives $\mathcal{P}$ and $\mathcal{P}'$, and use them to reduce the uncertainty on the primitives parameters.

VI. Experimental Results

The basic principles presented in this paper were tested on a variety of synthetic and real data. The interested reader can find registration and motion results with real points and lines in [3], registration and fusion results with synthetic and real points and lines in [2], and results on the building of global 3-D maps from passive stereovision in [9]. In this paper we only present results of the motion estimation from two 3-D maps from passive stereovision in [9]. In this paper we only present results of the motion estimation from two 3-D maps, the fusion of several inaccurate 3-D maps, and the detection of colinearity within a single 3-D map (what we called "autofusion"). In each of these examples, the 3-D map is made of 3-D lines.

A. Registration and Motion

Fig. 16 shows the edges of a triplet of images taken by the mobile robot in a first position. From these edges, the trinocular stereovision system computes a set of 3-D segments. Each 3-D segment is represented by the parameters (a, b, p, q) of the 3-D line supporting it and by the error covariance computed—as explained in Section IV-B—from the uncertainty on the edge points in the three images (we took an isotropic Gaussian density function of covariance 1 pixel around each edge point). Each 3-D line is bounded by two endpoints obtained from the endpoints measured in the three images which are projected on the reconstructed 3-D line.

Fig. 16. Triplet of images taken in position 1.

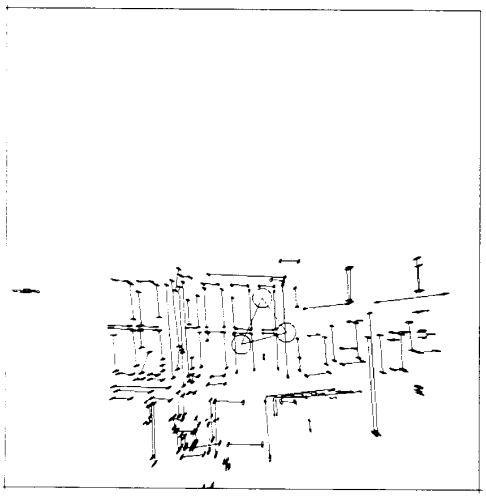


Fig. 17. Front view of reconstructed 3-D lines.

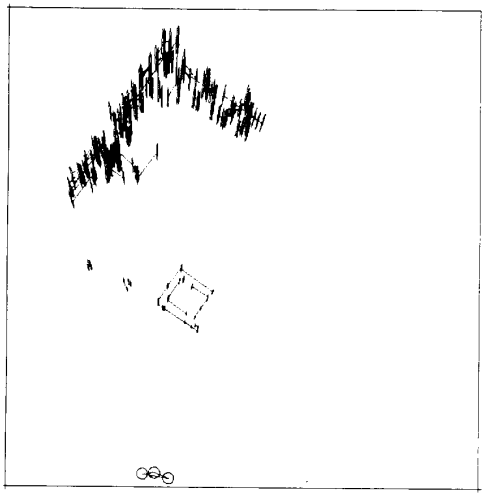


Fig. 18. Top view of reconstructed 3-D lines.

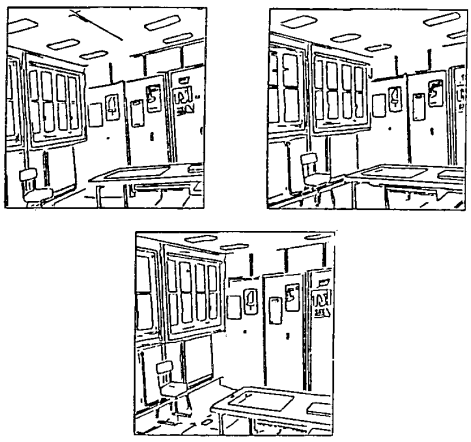


Fig. 19. Triplet of images taken in position 2.

We show in Figs. 17 and 18, respectively, the horizontal and vertical projections of the reconstructed 3-D segments. We also show the uncertainty attached to the reconstructed 3-D lines by showing the uncertainty it produces on the coordinates of their endpoints. The 95-percent confidence regions of the endpoints positions are ellipsoids whose projections are the ellipses shown in Figs. 17 and 18. One can see the anisotropic distribution of the uncertainty on the three coordinates of the points and its variation as a function of their position relative to the cameras (the projections of the three optical centers of the cameras correspond to the vertices of the triangle located grossly in the middle of the front view and at the bottom of the top view. Also, the circles around these vertices have been given an arbitrary radius of 20 cm to allow the reader to estimate the uncertainty attached to the other primitives).

The robot now moves a little, a new triplet of images is taken (Fig. 19) and another set of 3-D lines is computed. Initially, the robot is given a very crude estimate of its motion between the two views. Applying this crude estimate to the 3-D lines obtained in position 1, and projecting them in one of the images obtained in position 2 (the image of camera 3), one obtains the crude superimposition observed in Fig. 20. Solid lines are the transformed 3-D segments computed in position 1, while the dotted lines are the 2-D segments observed in position 2.

We now ask the system to discover the relation "$\equiv$" between the 3-D lines (see Section V-C) reconstructed in position 1 and 2, given the initial crude motion estimate and its uncertainty. The program takes each 3-D line in position 1, applies the noisy current motion estimate to place it in the 3-D map obtained in position 2 with a new covariance matrix (combining the initial uncertainty with the motion uncertainty), and computes its Mahalanobis distance (11) to all the other lines of position 2 (see Section V-D).

The program detects a match each time a pair of lines passes the χ^2 test of (12). If a line in position 1 can be matched to several lines in position 2, this is an *ambiguous* match, and nothing is done. On the other hand, each time an *unambiguous* match is found, the parameters of the motion are updated

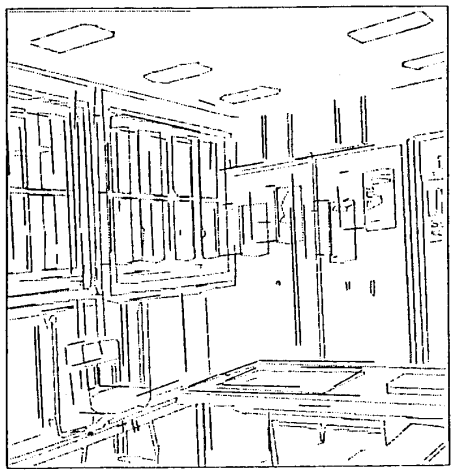


Fig. 20. Superimposition of 3-D segments of position 1 with 2-D edges of position 2 (crude initial motion estimate).

Fig. 21. Superimposition of 3-D segments of position 1 with 2-D edges of position 2 (final motion estimate).

as it is explained in Section V-E. As the uncertainty on motion decreases after each new match, some previously ambiguous matches can now become unambiguous. Therefore, the entire matching process is repeated until no more lines can be matched (three iterations in this example). The final estimate of the motion is very accurate as can be seen in Fig. 21 where the obtained superimposition is now almost perfect.

Applying exactly the same technique to a set of six triplets of views taken during the motion of the robot (Figs. 22–27), the system was able to build a global 3-D map of the room shown in Fig. 28 where rotating segments at the bottom right are the computed successive robot positions. Fig. 29 gives a hand sketched semantic interpretation of this global map.

B. Registration, Motion, and Fusion

In this experiment, the robot is looking from four different positions at a regular pattern (Figs. 30 and 31) formed by vertical lines floating in front of horizontal lines, and builds in each position a local 3-D map. Exactly the same technique as in the previous example was used to register each successive

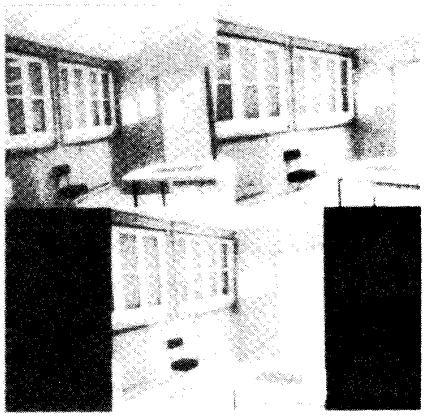


Fig. 22. First triplet of laboratory images.

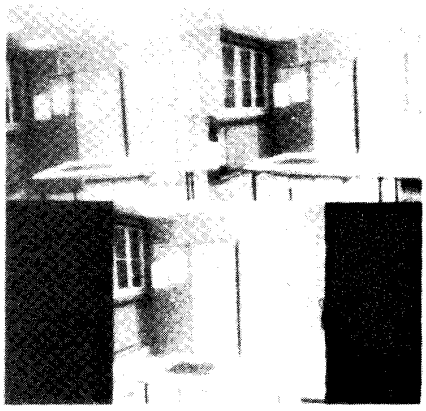


Fig. 23. Second triplet of laboratory images.

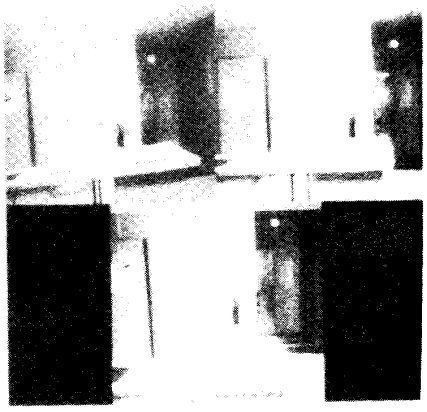


Fig. 24. Third triplet of laboratory images.

local 3-D map, and put all of them in a single absolute reference frame. Fig. 32 shows the resulting 3-D map before fusion. Fusion is achieved by discovering the relation "$\equiv$" computed between lines in the global 3-D map, and taking into account the uncertainty on the 3-D lines due to their reconstruction and to the successive motion estimations. Fusion yields a reduction from 1808 to 650 segments and improves accuracy, as can be seen by looking at the front and

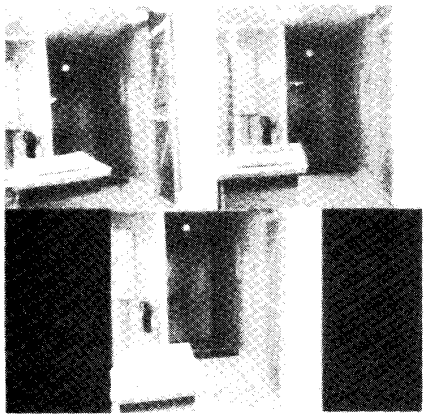


Fig. 25. Fourth triplet of laboratory images.

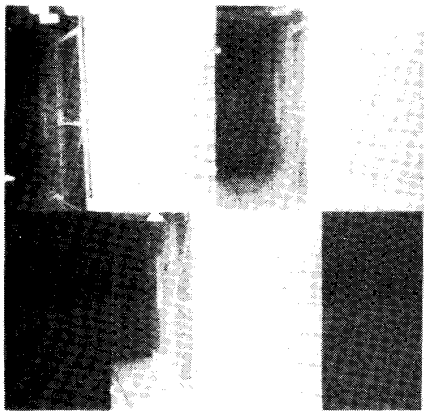


Fig. 26. Fifth triplet of laboratory images.

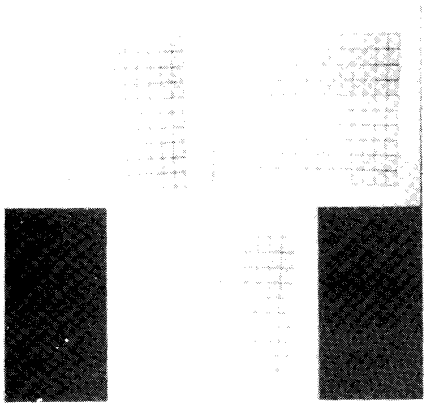


Fig. 27. Sixth triplet of laboratory images.

top view of the reconstructed 3-D pattern after fusion (Fig. 33).

C. Detecting Colinearity in Space

In this experiment, the robot is looking at the regular pattern only once. We show in Fig. 34 the vertical and horizontal projections of the initially reconstructed 3-D segments. We also show in Fig. 35 the uncertainty attached to reconstructed 3-D lines by showing the uncertainty it produces on the

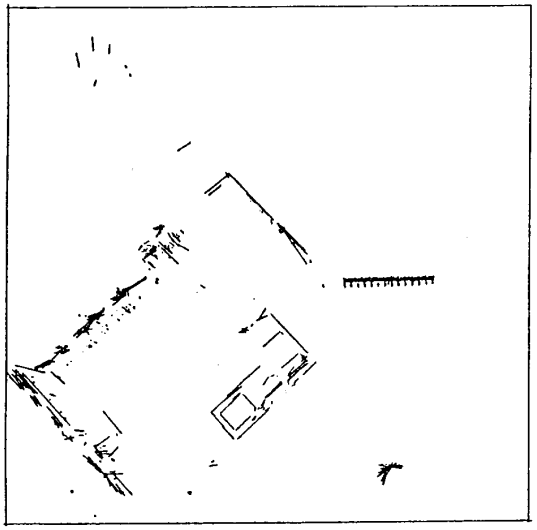


Fig. 28. Top view of a global 3-D map of the room computed from six local 3-D maps.

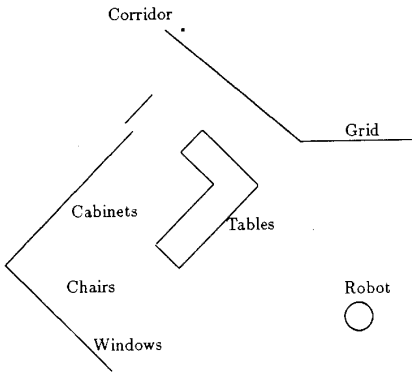


Fig. 29. Semantic interpretation of the previous global 3-D map.

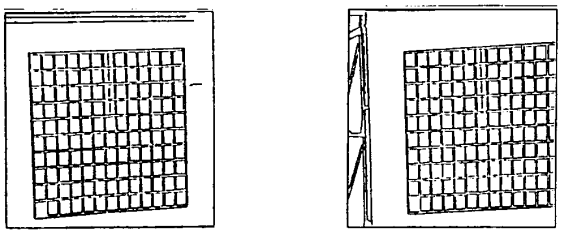


Fig. 30. A regular grid observed from position 1 and 2.

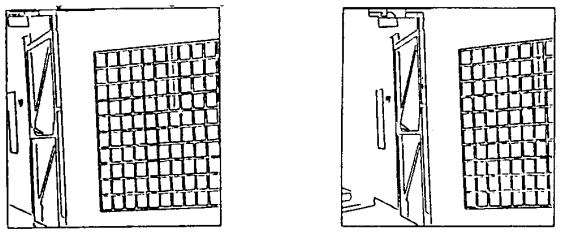


Fig. 31. A regular grid observed from positions 3 and 4.

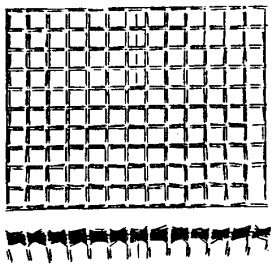


Fig. 32. Front an top view of reconstructed 3-D lines before identical lines are detected.

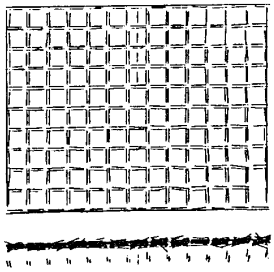


Fig. 33. Front and top view of 3-D lines after the fusion of identical lines.

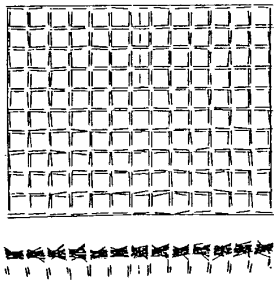


Fig. 34. Front and top view of reconstructed 3-D lines, before colinearity is detected.

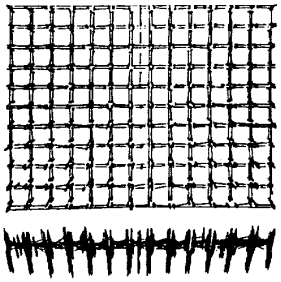


Fig. 35. Initial uncertainty attached to 3-D lines endpoints.

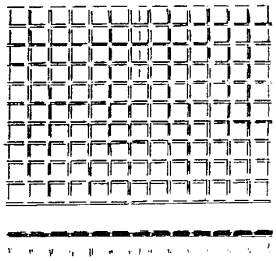


Fig. 36. Front and top view of 3-D lines when colinearity is discovered and enforced.

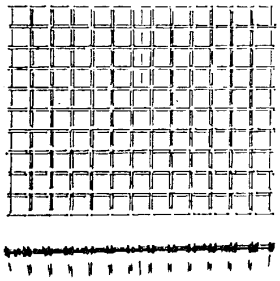


Fig. 37. Uncertainty attached to 3-D lines endpoints after the fusion of colinear segments.

coordinates of their endpoints (in the same way as in the first experiment).

We now ask the system to discover the relation "$\equiv$" between the 3-D lines (see Section V-C). The program takes a first 3-D line, computes its Mahalanobis distance (11) to all the other lines of the scene, and accepts the first line which passes the χ^2 test of (12) (see Section V-D). The two lines are fused using the technique of Section V-F and one keeps only the parameters of the optimal line representing both of them with an updated covariance matrix. The remaining lines are now compared to this new virtual line still with the Mahalanobis distance of (11) but with the new updated covariance matrix, while the χ^2 test of (12) remains unchanged. This process is repeated until no more lines can be matched with the first one, and then repeated with all the remaining unmatched lines.

The result is a reduced set of virtual lines on which the endpoints of the original segments have been projected, as shown in Fig. 36. The uncertainty on the line parameters has been greatly reduced: Fig. 37 shows the resulting uncertainty on the lines endpoints, which agrees very well with the reality.

VII. Conclusion

In this paper we have proposed a methodology for building and maintaining a geometric representation of the environment of a mobile robot. This methodology has the following salient features:

Representation: 1) We use geometric primitives to describe the environment and rigid displacements to describe the motion. These entities are described with a minimal number of parameters. 2) Uncertainty is modeled by a probability density function of these parameters. 3) Relationships between geometric entities are represented by algebraic equations on their parameters.

Algorithms: detecting geometric relationships, computing or updating the parameters of geometric entities (both for primitives and displacements) is done by recursive prediction-and-verification algorithms including the Extended Kalman Filter. These algorithms, better detailed in [8], [10], and [11], take into account prior knowledge to compute and propagate uncertainties.

Finally, the experimental results showed that the major approximations we made (linearization of the algebraic equations, second-order approximation of the probability density functions) were valid in a number of practical cases. Of course, a lot of theoritical and experimental work is still necessary to extend the approach to a wider class of problems for which such approximations cannot be made. This might be a good direction for future research.

Acknowledgment

The authors want to thank N. Gaudechoux for her precious help in the preparation of this paper and the reviewers for their helpful comments.

References

[1] N. Ayache and O. D. Faugeras, "Hyper: A new approach for the recognition and positioning of two-dimensional objects," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 1, pp. 44–54, Jan. 1986.

[2] ——, "Building a consistent 3d representation of a mobile robot environment by combining multiple stereo views," in *Proc. Int. Joint Conf. on Artificial Intelligence* (Milano, Italy, Aug. 1987).

[3] ——, "Building, registrating and fusing noisy visual maps," in *Proc. Int. Conf. on Computer Vision* (London, UK, June 1987), pp. 73–82. Also an INRIA Int. Rep. 596, 1986.

[4] ——, "Maintaining representations of the environment of a mobile robot," in *Int. Symp. on Robotics Research* (Santa Cruz, CA, Aug. 1987), pp. 337–350.

[5] N. Ayache and B. Faverjon, "Efficient registration of stereo images by matching graph descriptions of edge segments," *Int. J. Computer Vision*, vol. 1, no. 2, Apr. 1987.

[6] N. Ayache and O. D. Faugeras, "Building, registrating and fusing noisy visual maps," *Int. J. Robotics Res.*, vol. 7, no. 6, pp. 45–65, Dec. 1988 (Special Issue on Sensor Data Fusion).

[7] N. Ayache, O. D. Faugeras, and B. Faverjon, "Matching depth maps obtained by passive stereovision," in *Proc. 3rd Workshop on Computer Vision: Representation and Control*, pp. 197–204, Oct. 1985.

[8] N. Ayache, O. D. Faugeras, F. Lustman, and Z. Zhang, "Visual navigation of a mobile robot," in *IEEE Int. Workshop on Intelligent Robots and Systems (IROS'88)* (Tokyo, Japan, Oct. 1988).

[9] N. Ayache and F. Lustman, "Fast and reliable passive trinocular stereovision," in *Proc. Int. Conf. on Computer Vision* (London, UK, June 1987), pp. 422–427.

[10] 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 Int. Rep.

[11] ——, *Vision Stéréoscopique et Perception Multisensorielle—Application à la Robotique Mobile.* Inter-Editions, 1989. English translation will be available from MIT Press in 1990.

[12] R. M. Bolle and D. B. Cooper, "On optimally combining pieces of information, with application to estimating 3D complex-object position from range data," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 619–638, 1986.

[13] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, no. 6, pp. 679–698, 1986.

[14] M. P. do Carmo, *Differential Geometry of Curves and Surfaces.* Englewood Cliffs, NJ: Prentice Hall, 1976.

[15] J. L. Crowley, "Representation and maintenance of a composite surface model," in *Proc. Int. Conf. on Robotics and Automation* (San Francisco, CA, Apr. 1986), pp. 1455–1462.

[16] C. Darmon, "A new recursive method to detect moving objects in a sequence of images," in *Proc. IEEE Conf. on Pattern Recognition and Image Processing*, pp. 259–261, June 1982.

[17] R. Deriche, "Using Canny's criteria to derive an optimal edge detector recursively implemented," *Int. J. Computer Vision*, vol. 2, Apr. 1987.

[18] H. F. Durrant-Whyte, "Consistent integration and propagation of disparate sensor observations," in *Proc. Int. Conf. on Robotics and Automation* (San Francisco, CA, Apr. 1986), pp. 1464–1469.

[19] 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.

[20] O. D. Faugeras and M. Hébert, "The representation, recognition, and locating of 3d objects," *Int. J. Robotics Res.*, vol. 5, no. 3, pp. 27–52, 1986.

[21] O. D. Faugeras, F. Lustman, and G. Toscani, "Motion and structure from motion from point and line matches," in *Proc. Int. Conf. on Computer Vision* (London, UK, June 1987), pp. 25–34.

[22] Hager and M. Mintz, "Estimation procedures for robust sensor control, in the integration of sensing with actuation to form a robust intelligent control system," GRASPLAB Rep. 97, Dep. Comput. Informat. Sci., Moore School of Elec. Eng. Univ. of Pennsylvania, Mar. 1987.

[23] A. M. Jazwinsky, *Stochastic Processes and Filtering Theory.* New York, NY: Academic Press, 1970.

[24] T. Kanade, *Three-Dimensional Machine Vision.* New York, NY: 1987.

[25] D. J. Kriegman, E. Triendl, and T. O. Binford, "A mobile robot: Sensing, planning, and locomotion," in *Proc. Int. Conf. on Robotics and Automation* (Raleigh, NC 1987), pp. 402–408.

[26] E. M. Mikhail, *Observations and Least Squares.* University Press of America, 1976.

[27] McKendall and M. Mintz, "Models of sensor noise and optimal algorithms for estimation and quantization in vision systems," GRASPLAB Rep. 97, Dep. Comput. Informat. Sci., Moore School of Elec. Eng., Univ. of Pennsylvania, Mar. 1987.

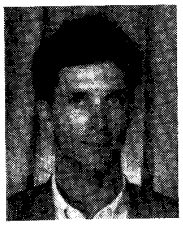
[28] L. Matthies and S. A. Shafer, "Error modelling in stereo navigation," *IEEE J. Robotics Automat.*, vol. RA-3, no. 3, pp. 239–248, 1987.

[29] J. L. Mundy, "Reasoning about 3-d space with algebraic deduction," in O. D. Faugeras and G. Giralt, Eds., *Robotics Research, The Third International Symposium.* Cambridge, MA: MIT Press, 1986, pp. 117–124.

[30] J. Porrill *et al.*, "Optimal combination and constraints for geometrical sensor data," Mar. 1987, to be published.

[31] K. Roberts, "A new representation for a line," in *Proc. Int. Conf. on Computer Vision and Pattern Recognition*, pp. 635–640, 1988.

[32] R. C. Smith and P. Cheeseman, "On the representation and estimation of spatial uncertainty," *Int. J. Robotics Res.*, vol. 5, no. 4, pp. 56–68, 1987.



Nicholas Ayache was born in Paris, France, on November 1, 1958. He graduated from the Ecole Nationale Supérieure des Mines in 1980, received the M.S. degree from the University of California, Los Angeles in 1981, and the Docteur Ingénieur and Docteur d'Etat degrees in computer science from the University of Paris XI in 1983 and 1988, respectively.

He is currently the Research Director of the Medical Images and Robotics Project *Epidaure* at INRIA (Institut National de Recherche en Informatique et en Automatique, Le Chesnay, France), and an Associate Professor at the University of Paris XI and Ecole Centrale. His research interests are in Computer Vision, Robotics, Artificial Intelligence, and Medical Applications. He is the author of the book *Stereovision and Multisensor Perception.*

Olivier D. Faugeras is Director of Research at INRIA (National Institute for Research in Computer Science and Control Theory) where he leads the Robotics and Computer Vision Project. His research interest include Computer Vision, Robotics, Shapes Representation, Computational Geometry, and Architectures for Vision. He is also lecturer in Applied Mathematics at the Ecole Polytechnique in Palaiseau where he teaches Computer Vision and Computational Geometry. He also teaches Computer Vision and Robotics in the "Magistère de Mathématique et Informatique" at the Ecole Normale Supérieure de la rue d'Ulm and at l'Ecole Polytechnique Féderérale of Lausanne in Switzerland. He is Associated Editor of several international scientific journals including: IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, *International Journal of Computer Vision, Internation Journal of Robotics Research, Pattern Recognition Letters, Signal Processing, and Robotics and Autonomous Systems.* In April 1989, he received from the French Science Academy the "Institut de France—Fondation Fiat" prize for his work in Vision and Robotics.