Artificial 3D Vision

Olivier D. Faugeras
INRIA
Domaine de Voluceau-Rocquencourt
BP 105 - 78153 Le Chesnay Cedex - France

EXTENDED ABSTRACT

1 Introduction

In this talk, we discuss the following problem: suppose we have a mobile robot moving around in an otherwise unknown man made environment. is it possible to let the thing wander around and build incrementally, using passive Vision, a three dimensional representation which:

1. does not grow too large even if many measurements are accumulated (capability of "intelligently forgetting").
2. is accurate even though the motion of the robot is not accurately known (converges in time toward the "real world" description)?

We have made some progress toward the solution of this problem whose applications to the field of Robotics should be obvious.

This presentation is in three parts. in the first part we somewhat detail the basic assumptions and techniques that have allowed us to come up with efficient solutions to the problem of building a local 3D map from Stereo Vision and Structure from Motion. In the second part, we show how the motion of the vehicle can be computed accurately by Visual Motion based techniques. In the third part we present a purely geometric approach to the problem of combining several viewpoints into a single surface and volume representation of the environment. In the fourth part, we present a solution to the same problem that takes into account the Uncertainty in the visual measurements and the motion of the robot.

In our work, we have followed two guiding lights. The first is Geometry: a Vision System is mostly a geometric engine and ours deals directly with geometric entities which are constructed from the visual inputs. The second guiding light is that of Uncertainty. Uncertainty is always present in the real world and cannot be engineered away. Therefore it must be present explicitely in the representations which are manipulated by the system.

2 Building 3D maps from passive Vision

We exploit the fact that in a man made environment many polyhedral or closely polyhedral objects are present. Therefore, it is natural to adopt a representation that is based on linear geometric primitives: points, lines, and planes. The line primitives are computed immediatly from the input images and then used as tokens in the Stereo and Motion Matchers that compute 3D Geometry and Motion. We have developed several Stereo matchers which all use the basic paradigm of Hypothesis Prediction and Testing to match 2D line segments [5,6] and differ by the number of cameras inputs they use. The fastest, most reliable and accurate matcher is the one using three cameras.

They all use as their basic principle the idea of working not on the images themselves to establish the matches but on symbolic representations of these images. These symbolic representations are neighborhood graphs of line segments extracted from the images. The matchers can accomodate any camera geometry thanks to a powerful calibration technique described in [11] from which the epipolar geometry can be easily computed. Thanks to this, they

---

[4] This work was partially supported by Esprit Project P940

can operate in a purely geometrical setting which increases their speed and reliability.

3 Obtaining a more accurate estimation of the robot motion from passive Vision

Some information about the robot motion is available from, for example, inert ial guidance and/or odometry. This information is usually not accurate enough and passive Vision is an efficient way of obtaining a more accurate estimation. Inversely, if the robot is not moving or moving with an accurately known motion, then passive Vision can be used to estimate the motion of a mobile object passing by, using exactly the same methods. We have developed a number of techniques for solving this problem which are all based on a combination of three ingredients: Hypothesis Prediction and Testing to establish the token matches, representing and manipulating Uncertainty to evaluate their likelihood, and Geometry to provide the framework.

3.1 Motion from 2D points and lines

This set of techniques exposed in |7j allows us to obtain the motion of a camera by matching points in the retinas at two positions at times t and $t + At$, or lines in the retinas at two positions at times t, (+ At_1, and $t + At_2$. It therefore uses the same primitives as the ones which are used by the Stereo matcher and can be computed at Video rates. It turns out that, contrarily to what has been reported in the literature, this token matching based approach for the estimation of Motion is quite robust to noise and provides very accurate results.

3.2 Motion from Stereo

Supposing that we have computed 3D by Stereo at times t and t ■+ At, by matching 3D primitives such as points, lines, and planes between the two representations, then we can reliably compute the motion of the robot. We use the same ingredients as before: Hypothesis Prediction and Testing, Uncertainty manipulation, and Geometry. This work is a generalization of some previous work (4,10] and is described in more details in [2].

4 Combining several viewpoints: a geometric approach

The problem which is tackled here is that of producing from the previous Geometric and Kinematic data a coherent description which is both:

1. surface oriented: to allow recognition / manipulation.
2. volume oriented: to allow trajectory planning / manipulation.

In order to achieve this we use our previously computed estimates of motion to express all our local 3D data (or a subset of it) in the same coordinate system. We then compute a volumetric tesselation of 3D space that is intrinsic to the measurements. This part uses heavily some recent results in Computational geometry [12]. We then exploit a simple visibility property to distinguish the tessels which are empty (their reunion provides the volume representation of free space) from those which are not (their border provides the surface representation of the objects/obstacles). This work is described in more details in [8j. It solves the problem of building a coherent 3D description of the environment but does not completely satisfies the first requirement that we set up at the beginning. Indeed, if the number of measurements increases, so does the complexity of the representation (as measured for example by the number of texels). A solution to this problem is provided in the next Section.

5 Combining several viewpoints: a probabilistic approach

If we not only represent the geometry of our primitives, but also the uncertainty that is attached to them we end up with a slightly different picture from the one of the previous Section. Our representation of the environment is made of a number of uncertain geometric primitives attached to coordinate frames related by uncertain rigid motions. The more we move the robot and make visual measurements, the more we increase the number of geometric primitives in the representation. This is clearly unsatisfactory and we must provide the system with means of "forgetting intelligently". This is achieved by exploiting the following simple idea: if a physical line segment S, for example the edge of a desk, is detected in, let us say, two positions 1 and 2 of the mobile robot, it is then present as segment S_1 in the coordinate system attached to position 1 and as segment S_2 in the coordinate system attached to position 2. Since we can relate by a rigid motion positions 1 and 2, by applying the correct transformation to S_2, the physical segment S is represented by two segments 5, and S'_2 in the coordinate system attached to position 1. By comparing the representations of S_1 and S_2 in terms of both their geometry and their uncertainty, our system is capable of discovering that these two segments are very likely to be instances of the same physical segment. it then proceeds to merge them into a single segment whose representation is a combination of the representations of S_1 and S_2. It has therefore "forgotten" the two instances of S and merged them into a single "better" one, better in the sense that it is more accurate than either one. We have therefore reached two goals:

1. the accuracy of the description increases with the number of measurements.
2. its size does not grow out of control since we are capable of fusing pieces of it which are instances of the same physical event.

This part of our work is described in more details in [9,2,1,3].

6 Conclusions

The system that we have sketched in this Abstract has been fully implemented and is used to drive a mobile robot designed at INRIA. This robot is used by a variety of users as a testbed for Vision algorithms. In a near future it will be equipped with video rate capabilities for computing the line segments which are at the basis of our 3D representations. All the matching algorithms that we use are very highly parallel, and this is an obvious way of improving the response time performances of the system. This approach is followed in a joint effort with a number of european academic and industrial partners. We think that the most important contribution of our system is to provide a proof of existence that Vision theory and practice have now reached a stage where fascinating questions such as the relation of shape and function, the relation of language and our perception of the 3D world, and many others can be reasonably asked without confining them to solipsistic systems.

References

[1] N. AYACHE and O.D. FAUGERAS. Building a consistent 3d representation of a mobile robot environment by combining multiple stereo views. In *Proceedings IJCAI '87, Milan,* August 1987.

[2] N. AYACHE and O.D. FAUGERAS. Building, registrating and fusing noisy visual maps. In *Proceedings ICCV '87, London,* IEEE, June 1987. also an INRIA Internal Report 596, 1986.

[3] N. AYACHE and O.D. FAUGERAS. Geometry and uncertainty in visual maps making. In R.C. BOLLES, editor, *Robotics Research, The Fourth International Symposium,* MIT Press, 1987.

[4] N. AYACHE and O.D. FAUGERAS. Hyper: a new approach for the recognition and positioning of two-dimensional objects. *IEEE Transactions on PAMI,* 8(1):44 54, January 1986.

[5] N. AYACHE and B. FAVERJON. Efficient registration of stereo images by matching graph descriptions of edge segments. *International Journal of computer Vision,* 1(2), April 1987.

[6] N. AYACHE and F. LUSTMAN. Trinocular stereovision: recent results. In *Proceedings IJCAI '87, Milan,* August 1987.

[7] O. FAUGERAS, F. LUSTMAN, and G. TOSCANI. Motion and structure from motion from point and line matches. In *Proceedings ICCV '87, London,* IEEE, June 1987.

[8] O.D. FAUGERAS. Artificial 3d vision. In Michael Brady, editor, *Robotics Research,* MIT Press, 1988.

[9] O.D. FAUGERAS, N. AYACHE, and B. FAVERJON. Building visual maps by combining noisy stereo measurements. In *Proceedings of the Conference on Robotics and automation,* IEEE, April 1986.

[10] O.D. FAUGERAS and M. HEBERT. The representation, recognition, and locating of 3d shapes from range data. *The International Journal of Robotics Research,* 5, No 3:27-52, 1986.

[11] O.D. FAUGERAS and G. TOSCANI. The calibration problem for stereo. In *Proceedings CVPR '86, Miami Beach, Florida,* pages 15-20, IEEE, 1986.

[12] F. PREPARATA and SHAMOS M. *Computational Geometry.* Springer-Verlag, New-York, 1985.