Trinocular Stereovision : Recent Results*

Nicholas AYACHE and Francis LUSTMAN

INRIA
Domaine de Voluceau-Rocquencourt
BP 105 - 78153 Le Chesnay Cedex - France

Abstract

We present an original approach to build rapidly and reliably a 3D description of the environment of a mobile robot by means of passive stereovision using three cameras.

This technique has been successfully applied to many indoor scenes and has proved faster than a previously developed binocular stereo technique, while providing more reliable and more accurate results. Moreover, the algorithm is highly parallelisable and has indeed been parallelised, thus highly increasing its speed.

Results showing the construction of the 3D visual map of a complex indoor scene are included.

1 Introduction

Stereovision is a technique to build a three dimensional description of a scene observed from several viewpoints. It is quoted as passive if no additional lighting of the scene, for instance by a laser beam, is required. So defined, passive stereovision happens to be very attractive for many applications in robotics, including 3D object recognition and localisation as well as 3D navigation of mobile robots

Most of the research on passive stereovision has been devoted to binocular vision for which two cameras are observing the same scene from two slightly different viewpoints (see [1] [2] and included referciKeii). Presently, following Yachida [3], an increasing number of stud--- are now concerned with trinocular vision in which a third camera is used to simplify the stereovision problem. Detailed references can be found in [4].

We introduce in this article an original method for trinocular stereovision which has the following attractive features:

- Flexibility, it allows for arbitrary positions of three different cameras. Calibration is obtained by a simple automatic procedure.
- Speed: 200 line segments (corresponding to approximatively 3000 image points) are matched in about 2s
- Reliability: typically less than 2 percent of the found matches are erroneous. Moreover noise or occluding boundaries which are often mismatched by conventional binocular systems, are reliably checked by trinocular systems, providing safer reconstruction results.
- Accuracy: the use of a third camera provides an additional measurement which increases the reconstruction accuracy of each 3D segment.

The scheme of the method is the following:

- Preprocessing: a graph based description of a polygonal approximation of the contours is extracted from each image and the images are rectified so that the epipolar geometry, becoming simpler, less computations are needed.
- Hypotheses Prediction: triplets of potential matches are derived from the previously constructed graphs by simple geometric verifications.
- Hypotheses Validation: local consistency checks are performed to remove erroneous matches.

2 Geometry of Trinocular Stereovision

Figure 1 shows the geometry of trinocular stereovision. We have three cameras, modelled by an optical center C, and an image plane P, Given a physical point A_j its image on camera i is defined as the intersection of the straight line *AC,* with the image plane Pi. Let us denote a,, i =1,2,3 the image of *A* on camera t. $a_{1t}a_2,a_3$ are called homologous image points.

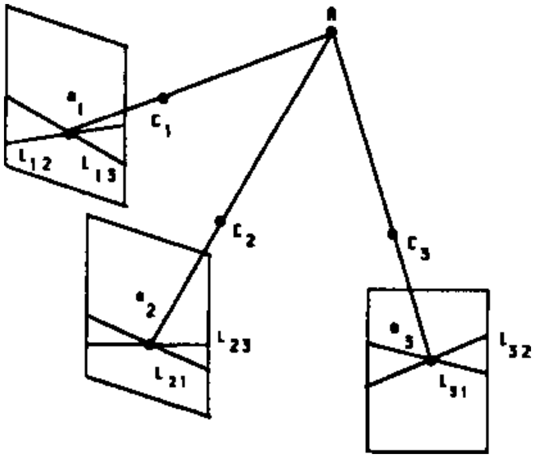


Figure 1: Geometry of trinocular stereovision

Given a pair (i,j) of cameras and a physical point A, the epipolar plane P_{ij} is defined by the triplet of points (C_i, A, C_j). The intersection of this epipolar plane with camera i is the epipolar line L_{ij}, while its intersection with camera j is the epipolar line L_{ji}. L_{ij} and L_{ji} are called conjugated epipolar lines. Any point a_i on L_{ij} (resp. a_j on L_{ji}) has its homologous image point a_j on L_{ji} (resp. a_i on L_{ij}). Therefore, using two cameras, the search for homologous image points is a search along conjugated epipolar lines.

Any triplet (a_1, a_2, a_3) of homologous image points is such that a_1 lies at the intersection of the epipolar lines L_{12} and L_{13} defined by the two other image points a_2 and a_3. Therefore the search for homologous image points between two images can now be reduced to a simple verification at a precise location in the third image. For instance checking that (a_1, a_2) form a pair of homologous image points consists in verifying the presence of a_3 at the intersection of L_{31} and L_{32}.

---

*This work was partially supported by Esprit Project P940

3 Calibration and Reprojection

Faugeras and Toscani [5] have defined a calibration procedure through which the parameters of the perspective transformations attached to each camera are automatically obtained by looking at a reference pattern in a sequence of known positions. From the knowledge of these transformations, it is straightforward to compute the epipolar line of any point of a given retina onto another retina (see appendix of [2] for more details).

Though very simple (6 multiplications and 4 additions), this computation of the epipolar lines and the subsequent intersection computations that are needed during the matching phase are done so many times that it is a substantial saving in computational cost to have the simplest possible epipolar geometry: parallel epipolar lines, horizontal if possible. This can be obtained **without loss of generality** by reprojecting the segments of each image onto a plane parallel to the plane containing the three optical centers. Figure 2 illustrates the principle of reprojection. In this plane, we can furthermore choose the three coordinate systems of the retinas: if we choose the coordinate systems of figure 3, the epipolar line of point $M(x_0, y_0)$ is $y = y_0$ in image 2 and it is a horizontal line (though a little more difficult to compute) in image 3 (cf figure 3). This simple transformation divided the computing time by a factor of 2.

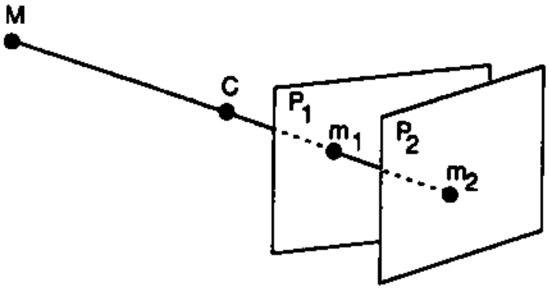


Figure 2: Reprojection

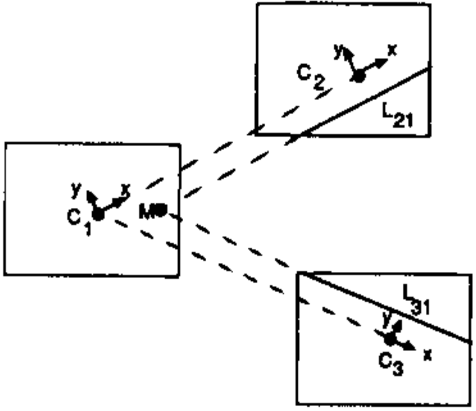


Figure 3. Choice of the coordinate systems on the reprojection plane

4 Structuring the image

Another operation worth optimizing is accessing the segments lying within a given region of the image and having a given orientation. We therefore compute buckets, i-e we superimpose a virtual grid composed of square windows on the image and compute, for each window, the list of segments intersecting it. This structure is computed in linear time with respect to the number of segments. Moreover, at the expense of a unique preliminary sorting of the segments orientations (done with algorithmic complexity O(nLogn)), buckets are actually filled with segments sorted by increasing orientation.

Accessing a segment in a given region of the image and having a certain orientation is then reduced to a dichotomy search in each bucket of the region.

5 Matching of homologous segments

The algorithm works as follows:

```
for each segment $S_1$ of image 1 do
    consider the segments $S_2$ of image 2 intersecting the epipolar line $L_{21}$ corresponding to the middle $a_1$ of $S_1$ within a given disparity interval
    if $S_2$'s features (angle, length, gradient) are compatible with $S_1$'s then do
        compute the intersection $a_3$ of the epipolar lines $L_{31}$ and $L_{32}$ of $a_1$ and $a_2$ in image 3, and compute the predicted orientation of a matching segment;
        if a segment $S_3$ lying within a given neighbourhood of $a_3$ whose orientation is compatible with the predicted orientation and whose features are compatible with $S_1$ and $S_2$ then
            keep $(S_1, S_2, S_3)$ as a potential matching triplet;
```

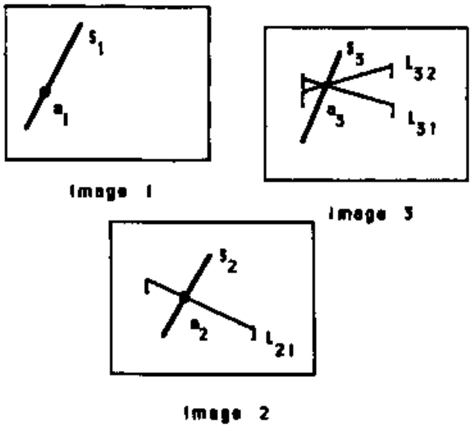


Figure 4: Matching of homologous segments

Discussion of the constraints enforced on each segment feature and of the meaning of *neighborhood*, as well as a description of the validation procedure can be found in [4].

6 Experimental results and discussion

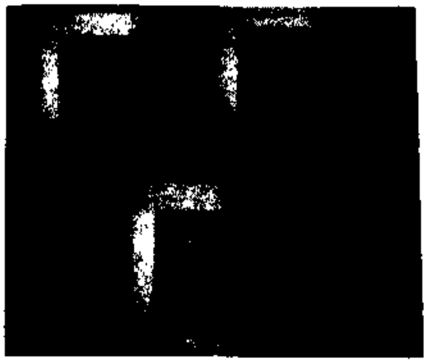
This stereo-matching technique has been tested on a number of indoor scenes. We only present the following typical results.

Three images of a room are taken simultaneously with our previously calibrated three camera system. Two triplets are shown in figures 5 and 6. From these 512x512 pixels images, edge points are extracted and chains of connected edge points are built and approximated by a set of linear segments, oriented with respect to the contrast sign across the segment. This preprocessing is independent of the stereo-matching program, and we believe that it could be done in a fraction of a second on dedicated hardware.

The computing times are shown in the captions. To compare with point matching algorithms, a matched segment corresponds to an average of 15 matched points. These times do not include the validation phase: at the present time, this phase takes about the same time as the matching process but we have serious hopes to be able to shorten that time.

In general the number of errors is less than 2 percent of the total number of matches. Figures 5 and 6 show the stereo-matching results.

The algorithm is highly parallel: preprocessing can be done independently over each of the three images and the outer loop of the matching algorithm can be split in whatever number of processors we can get We implemented the algorithm on a Sequent decaprocessor machine, so that we were able to split it straightforwardly over 9 processors and divide the computing time almost by 9 with respect to the sequential version.



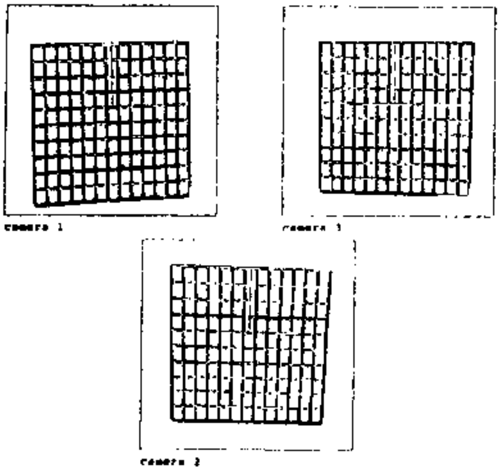


Figure 5· Triplet of the calibration grid, angular position 0 · about 600 segments per image, preprocessing 0 35s per image, 580 matched segments in 6 7s

Finally, 3D segments are reconstructed using the technique developed in [6] to reconstruct a 3D line from its projections on several retinas. Figure 7 shows the view from above of the reconstructed room (obtained from a panoramic view including figures 5 and 6, but notice the horizontal line correponding to the projection of the grid of figure 5 on the right-hand side of figure 7 and the room corner of figure 6 on the left-hand side of figure 7).

References

[1] N AYACHE and B FAVERJON A fast stereo vision matcher based on prediction and recursive verification of hypotheses In *Proceedings of Third Workshop on Computer Vision. Representation and Control*, pages 27-37, IEEE, october 1986 Shorter version in CVPR '85, San Francisco, 1985

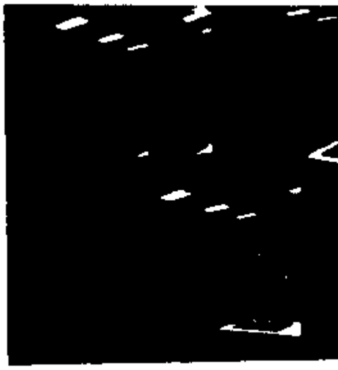
[2] N AYACHE and B FAVERJON Efficient registration of stereo images by matching graph descriptions of edge segments *International Journal of computer Vision*, 1(2), april 1987

[3] M YACHIDA 3d data acquisition by multiple views In *Robotics Research. the Third International Symposium*, pages 11-18, MIT Press, Cambridge, Mass , 1986

[4] N AYACHE and F LUSTMAN Fast and reliable passive trinocular stereovision In *Proceedings of ICCV '87, London, UK*, IEEE, June 1987

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

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



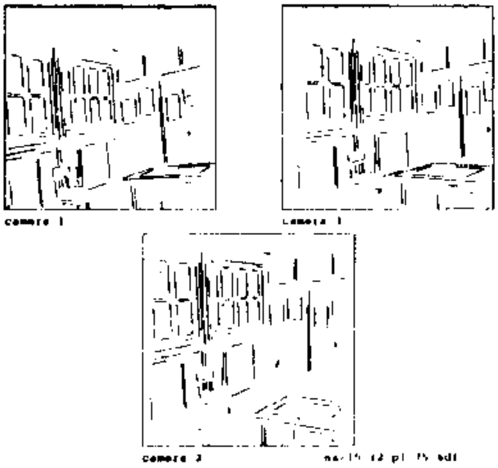


Figure 6: Triplet at angular position 75 . about 350 segments per image, preprocessing 0.25s per image, 200 matched segments in 2s

Figure 7: 3D map of the room: view from above