A method for the 3D reconstruction of indoor scenes from monocular images

Paolo Olivieri[1], *Maurizio Gatti*[1], *Marco Straforini*[2] *and Vincent Torre*[2]

[1] Dipartimento di Informatica e Scienza dell'Informazione
[2] Dipartimento di Fisica
Università di Genova, Italy

Abstract. The recovery of the 3D structure of indoor scenes from a single image is an important goal of machine vision Therefore, a simple and reliable solution to this problem will have a great influence on many tasks in robotics, such as the autonomous navigation of a mobile vehicle in indoor environments.

This communication describes the recovery, in a reliable and robust way, of the 3D structure of a corridor and of obstacles from a sequence of images obtained by a T.V. camera moving through the corridor. The obtained 3D information can be used to extract the free space in the viewed scene in order to plan the trajectory of a mobile vehicle. This application is being worked on at the moment and the results will be illustrated in a future communication.

1 The recovery of a line-drawing

Fig. 1A illustrates an image of polyhedral objects on a table. Using standard routines for edge detection it is possible to obtain the edge map illustrated in Fig. 1B. It is useful to extract straight segments from this map and identify junctions. These features allow the construction of a line-drawing from which it is simple to obtain 3D information (see [H1] and [B2]).

The procedure for the recovery of the line-drawing is fully described in previous works (see [C1]). Fig. 1C shows the results of the first step of the algorithm for the extraction of the line-drawing. In this elaboration segments are fused together and the junctions are identified by the symbols L, T, Y and X. In order to obtain a fully-connected line-drawing, i.e. one whose segments have both ends belonging to identified junctions, it is possible to delete unconnected segments recursively. When this procedure is applied to the line-drawing of Fig. 1C, the fully connected line-drawing of Fig. 1D is obtained.

The algorithm used for the extraction of a line-drawing illustrated in Fig. 1 does not fully exploit the geometrical constraint present in the scene. The algorithm is rather general and can also be used for images of rounded objects or with complex surfaces. Many images of indoor scenes, such as that of Fig. 2A, can be usefully analysed by exploiting geometrical properties of the scene. By assuming that the viewed scene belongs to a Legoland world, where objects have planar surfaces, with either parallel or perpendicular edges, it is possible to make the algorithm for the recovery of line-drawing efficient and robust. Fig. 2B reproduces the polygonal approximation of the edge map obtained from Fig. 2A. Figs. 2C and 2D reproduce, respectively, the line-drawing obtained by using the algorithm previously described and the algorithm making use of the assumption of a Legoland world. It is evident that the line-drawing of Fig. 2D is more accurate and its segments and junctions are more correct.

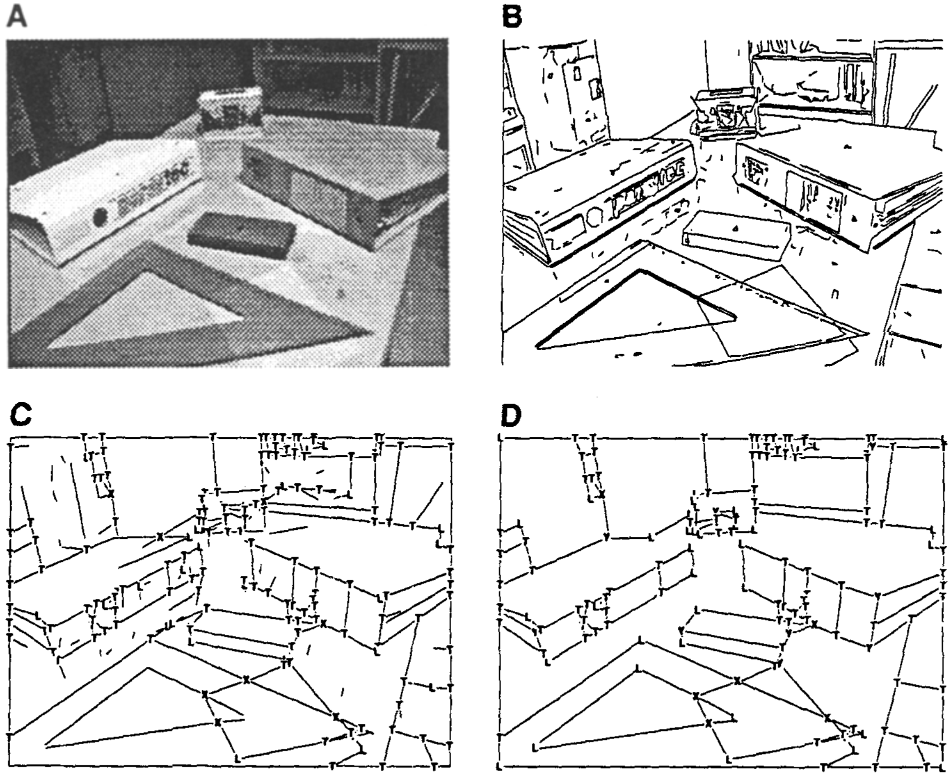

Fig. 1. The recovery of a line-drawing. A: an image of 512x512 pixels acquired with a Panasonic camera and digitalised with a FG100 Imaging Technology board. B: the segments obtained with a polygonal approximation of the edges extracted with a Canny filter. C: the line-drawing with labelled junctions (L, T, Y and X). The thresholds used to merge the segments are: $\pm 5°$ (collinearity), 8 pixels (adjacent), 50 pixels (distance). D: the final line-drawing after the recursive deletion of unconnected segments. The L junctions are detected if two vertices are closer than 7 pixels.

2 Extraction of polygons

Using the line-drawing of Fig. 2D, it is possible to extract maximal simple polygons (see [S1]), which are the perspective projection on the image of planar surfaces with similar attitude in space. Each simple polygon may be labelled with a different orientation; this depends on the attitude in space of the projected planar surfaces. Simple polygons in images of scenes belonging to Legoland can have, at most, three different orientations. The fig. 3B shows the polygons extracted from the line-drawing obtained from the image 3A; the three different textures correspond to horizontal, vertical and planar surfaces, white regions correspond to complex polygons.

3 Detection of the dimensions of the corridor

The 3D structure of viewed scenes is described by simply using 3D boxes, the largest box corresponding to the empty corridor and other boxes representing different objects

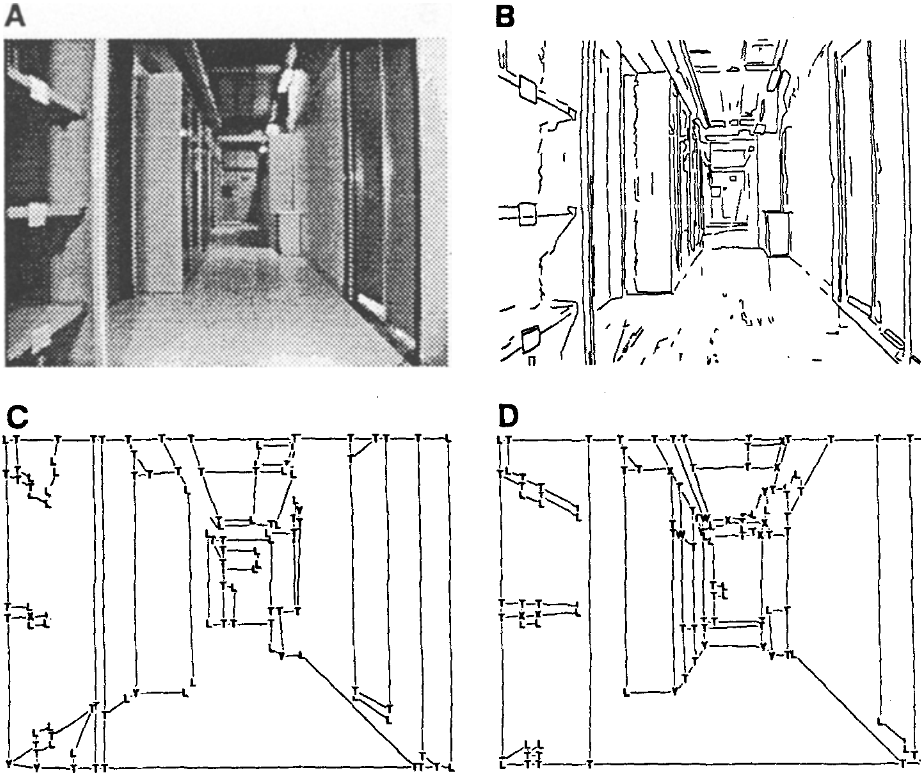

Fig. 2. The recovery of a line-drawing from an image of Legoland. A: an image of a corridor at the Dipartimento di Fisica. The viewed scene can be described as belonging to a Legoland world, where objects' boundaries are straight lines mutually parallel or orthogonal. B: the segments map. C: the line-drawing obtained with the procedure explained in the text and illustrated in Fig. 1. D: the line-drawing obtained making use of the assumption of a Legoland world. The parameters used are the same as the ones in Fig. 1.

or obstacles. The algorithm able to extract this information can be divided into three main steps:

1. identification, on the image, of the bottom end of the corridor. (see Fig. 3C).
2. identification on the image of lines separating the floor and walls, and those separating the ceiling and walls (see Fig. 3D).
3. validation of the consistency of the first two steps.

By assuming that the distance from the floor of the optical center of the viewing camera is known, it is possible to make an absolute estimate of the side of the box in Fig. 3E and F. The image of Fig. 3A was acquired with an objective having a focal length of 8 mm and the T.V. camera placed at 115 cm from the floor. The estimate of 195 cm for the width of the corridor (the true value is 200 cm) can be obtained by using simple trigonometry.

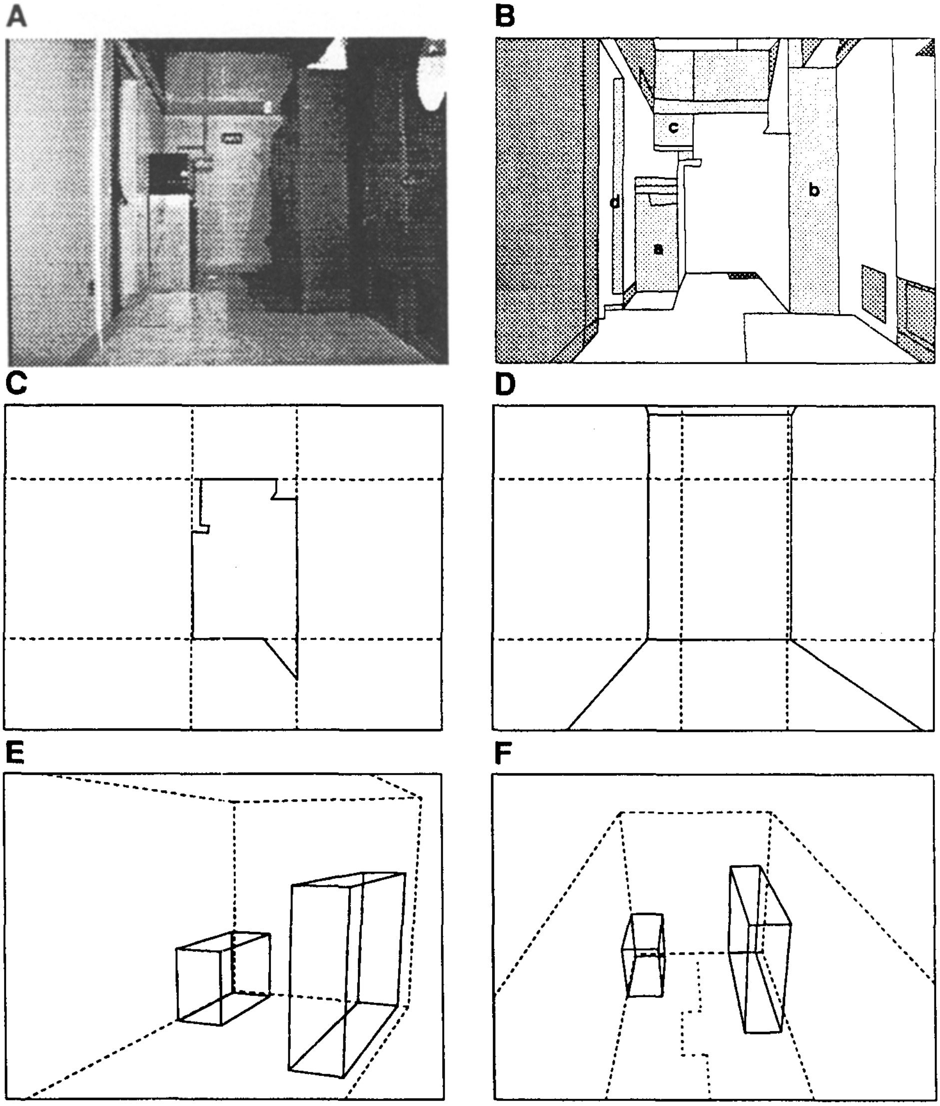

Fig. 3. The recovery of the 3D structure. A: an image of a corridor at the Dipartimento di Fisica. B: maximal simple polygons. Polygons a, b, c and d are candidates to be the front panel of obstacles. Polygons c and d are rejected because they are too high (polygon c) or outside the frame of the corridor (polygon d). C: the detection of the boundaries of the bottom end. D: the detection of the lines separating floor, walls and ceiling. E, F: the 3D structure of the corridor, represented with two different perspective projections. The broken line in D is a collision-free trajectory for a mobile vehicle.

4 Detection of obstacles

When the largest 3D box corresponding to the empty corridor has been detected it is useful to detect and localize other objects or obstacles, such as filing cabinets and drawers. The algorithm for the detection of these boxes is divided into four steps:

1. detection of polygons, which are good candidates for the frontal panel of the obstacle for example polygons a, b, c and d in Fig. 3B.
2. validation of the candidates.
3. a 3D box is associated with each validated polygon using a procedure which is very similar to that used in constructing the 3D box associated with the empty corridor.
4. the consistency of the global 3D structure of the scene is checked, that is to say all obstacles must be inside the corridor.

Figs. 3E and F reproduce two views of the 3D structure of the scene of image 3A. It is evident that the global 3D structure of viewed corridors is well described by the boxes illustrated

Conclusion

The algorithm described in this paper seems to be efficient for the recovery of the 3D structure of indoor scenes from one or a sequence of images. The proposed algorithm produced good results for different corridors under a variety of lighting and complexity. Similar procedures can be used in order to determine the presence of, and locate, other rectangular objects, such as cabinets, boxes, tables, ... Therefore, when a sequence of many images is available it is possible to obtain an accurate and robust 3D description of the scene by exploiting geometrical properties of Legoland and by using a simple Kalman filter.

Acknowledgements

We wish to thank Dr. M. Campani, Dr. E. De Micheli and Dr. A. Verri for helpful suggestions on the manuscript. Cristina Rosati typed the manuscript and Clive Prestt checked the English. This work was partially supported by grants from the EEC (ESPRIT II VOILA), E.B.R.A. Insight Project 3001, EEC BRAIN Project No. 88300446/JU1, Progetto Finalizzato Trasporti PROMETHEUS, Progetto Finalizzato Robotica, Agenzia Spaziale Italiana (ASI).

References

[B2] Barrow, H.G, Tenenbaum, J.M.: Interpreting line-drawings as three-dimensional surfaces. Artif. Intell. **17** (1981) 75–116

[C1] Coelho C., Straforini M., Campani M.: A fast and precise method to extract vanishing points, SPIE's International Symposia on Applications in Optical Science and Engineering, Boston 1990.

[H1] Haralick, R.M.: Using perspective transformation in scene analysis. Comput. Graphics Image Process **13** (1980) 191–221

[S1] Straforini, M., Coelho, C., Campani, M., Torre V.: The recovery and understanding of a line drawing from indoor scenes. PAMI in the press (1991)