MVA '92 IAPR Workshop on Machine Vision Applications Dec. 7-9,1992, Tokyo

A MONOCULAR TECHNIQUE FOR THE RECONSTRUCTION OF 3D SHAPE USING A COOPERATING ROBOT ARM.

C.harles Allen, Roy Booth
Robotics Research Group
Newcastle University England NE1 7RU

ABSTRACT

A robust method for the reconstruction of 3D object shape is described based on the inverse perspective method [1,2], which uses a single camera mounted on a commercial manipulator arm for accurate 3D positioning and orientation of the sensor head in 3D space. Preliminary measurements are presented and the reconstruction theory used by the technique.

INTRODUCTION

The accurate modelling of the workspace for a robot or mobile platform is a necessary requirement for the successful navigation of autonomous vehicles in the advanced robotics application areas [3].

Techniques are currently being developed using a number of approaches. Makela [4] uses laser bar code reading from markers placed in the workspace, whilst others [5] use ultra-sound sensors borne on the vehicle to detect the location of obstacles and static features in the path of the vehicle. This latter method is useful for on-line collision avoidance but clearly is unsuitable for accurate modelling of the workspace. Stereo vision systems are under development at many institutions. In the UK two noteworthy projects are at Stratchclyde and Sheffield Universities [5,6]. Stereo, whilst a valid model for understanding the human visual system only offers a 2 1/2D sketch representation, and has several well documented problems with mismatching and the need for special epipolar geometries.

In 1990/91 a new technique was reported by Booth et al [2], which offered a more robust method of scene/object reconstruction based on region growing 2D object segmentation, combined with polygon fitting of the regions. The polygon vertices were then used as the primary features for matching between camera frames from differing views, backed up with the knowledge that each vertex was attached to a common polygon. This wireframe representation was particularly effective at representing 2D images from office or industrial scenes. See Figure 1a and 1b below:

Figure 1. A sample office scene used for testing 3D reconstruction algorithm with 3 cameras.

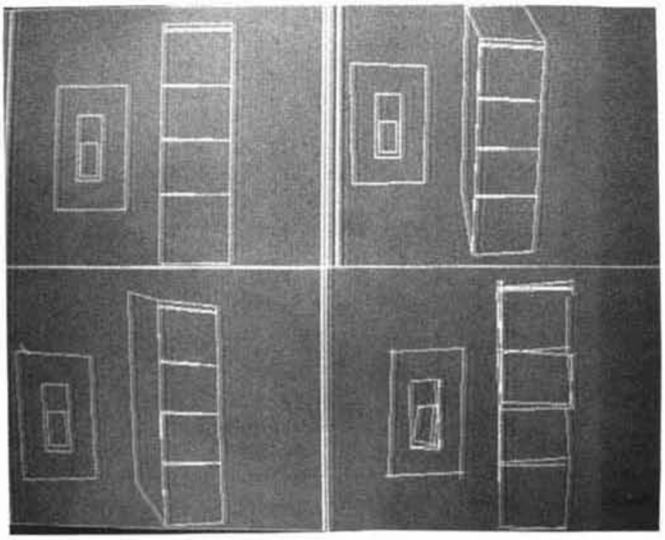


Figure 1b. Polygon representation of the office scene shown from three views (Top-left, Top-right and Bottom-left).

The fourth quadrant picture shown in Figure 1b shows the reconstructed wireframe in 3D coordinate axes. A multi-transputer processing system was constructed in 1990/91 based on this method which used 13 cooperating transputers for image grabbing and analysis of images from three cameras rigidly mounted on the mobile robot frame. Two-D processing was achieved in a cycle time of 2.5 s with code written in 3L parallel-C. This processing involved grabbing 512x512x8-bit images, performing a smoothing algorithm to average grey level intensity locally to nearest neighbours reducing the image size to a smoothed 256x256x8bit digitised representation. This was performed simultaneously with three T-801 based frame grabber cards. The processed images were then passed to a second set of T-805 cards which constructed polygon representations in 2D using a recursive algorithm of vertex assignment much akin to that reported many years ago by Feng & Pavlides [7]. The scene analyser is shown in Figure 2 below mounted on a trike DC-servo driven mobile platform:

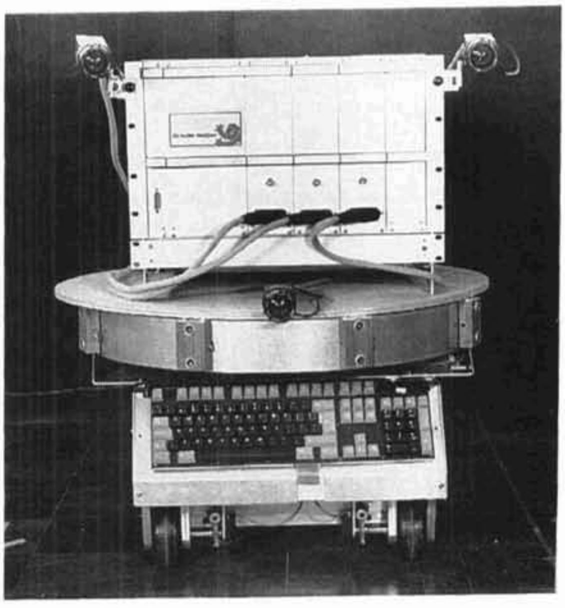
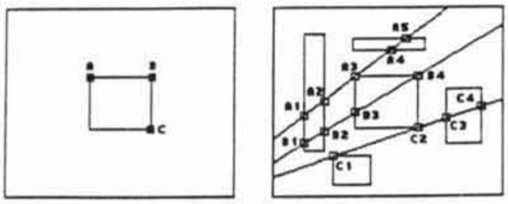


Figure 2. The Newcastle 3D wireframe analyser mounted on a trike AGV.

In testing the system it was found to create quite good models but that for more complex scenes the 3D matching often failed to converge. This was thought due to the inaccuracies in knowledge of the positioning and orientations of the three cameras. This was due to the calibration method adopted, which used a single illuminated source of Led lights mounted on a circle of radius 10 cm, which was placed at 16 positions before the sensor head. The circle of bright lights was then used to calculate the relative positioning of each camera with respect to each other. Additionally the depth position was entered to the analyser manually using a tape measure. In fact the self calibration method typically only produced camera positioning and orientation accuracies within +/- 4% which seriously degraded the ability of the 3D reconstruction algorithm which relies on an accurate knowledge of their positions. This unforeseen problem has now been tackled in a different way by replacing the rigid three camera sensor head with a single robot mounted camera, which can be oriented accurately in space using the kinematics of the robot itself. The following sections describe the 3D matching method applied to a cooperating robot, and some preliminary measurements showing a considerable improvement in reconstructions.

THE INVERSE PERSPECTIVE METHOD

The 2D processing briefly outlined in the introduction to this paper provides vertex lists and the interconnection topology i.e. edges for each segmented region in a scene. Each region is thus labelled by a characteristic intensity level which can be normally assigned to represent the sides of a real object before the camera. This region is then represented as a polygon as shown in Figure 3 for the case of two camera views. The left camera has detected a simple region shaped as a square with vertices A,B,C. The task of 3D reconstruction then requires that a match be found for each of these connected vertices in the right camera frame. Both cameras are known to be placed in space at defined world coordinates X_1, Y_1, Z_1 and at X_2, Y_2, Z_2, and with known orientations in world coordinates of o_1, a_1, n_1 and of o_2, a_2, n_2. In this notation the o unit vector is the orientation unit vector, a is the approach unit vector and n is the cross product of oxa. Thus for each camera a homogeneous transformation can be defined based on the kinematics of the robot manipulator and the world base coordinate origin at X_0, Y_0, Z_0.



	A1	A2	A3	A4	A5	C1	C2	C3	C4
B1	✓	—	—	—	—	—	—	—	—
B2	—	✓	—	—	—	—	—	—	—
B3	—	—	✓	—	—	—	—	—	—
B4	—	—	✓	—	—	—	✓	—	—

B=B4

Since B−A & B−C and B4−A3 & B4−C2

Figure 3. Connected corner list matching using linear 3D coordinate transformations from camera 1 to camera 2, positioned at known positions in space.

The Denavit-Hartenburg representation for each frame is thus defined by:

$$._oT^1 = \begin{vmatrix} n_x & o_x & a_x & p_x \\ n_y & o_y & a_y & p_y \\ n_z & o_z & a_z & p_z \\ 0 & 0 & 0 & 1 \end{vmatrix} \qquad (1)$$

The required function of inverse perspective reconstruction is to take each visual vertex cue from the first left camera, which will be known to be positioned at a pixel coordinate point in the image plane of X_s , Z_s and using the known geometry of the camera i.e. its optical magnification factor, the camera resolution in pixels squared and the angular field of view of the lens attachment to generate the ray from the actual point infront of the camera at an unknown distance d from the camera such that the ray intersects both points. The unknown coordinate points may be assigned as X_d , Y_d , Z_d in the robot's world coordinate frame.

This is illustrated in diagramatic form in Figure 4 below:

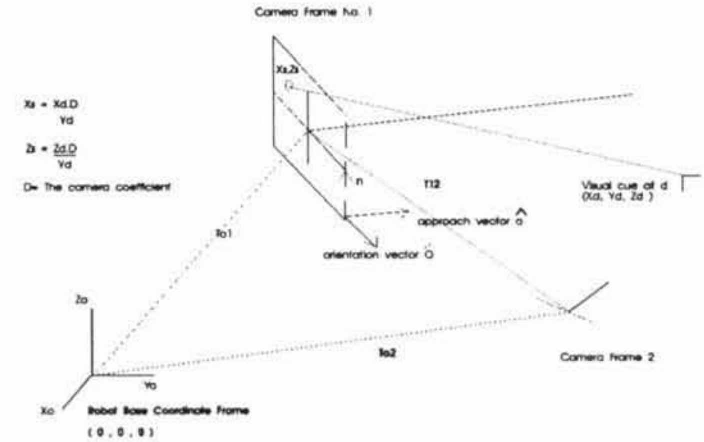


Figure 4. The reconstruction rays from an unknown vertex of a real object in the scene at point: X_d , Y_d , Z_d projected onto the camera image frame at a pixel coordinate X_s , Z_s .

The exact position of the visual cue in the camera frame depends on the camera geometry, and in particular the camera optical magnification factor or coefficient D is defined as:

$$D = \frac{R/2}{Tan\ (\theta/2)} \qquad (2)$$

where R = camera resolution in pixels and θ is the angular field of view of the camera lens. Both are known.

Then the point X_s , Z_s is related to the unkown source point of this cue at X_d , Y_d , Z_d by the simple relationships:

$$X_s = \frac{X_d \, . \, D}{Y_d} \ , \ Z_s = \frac{Z_d \, . \, D}{Y_d} \ , \ Y_d = d \qquad (3)$$

Eqn (3) defines the ray equation and it is known that the distant object cue is within a bound of the focus of the optical system. In this case the lens focal point is limited to within 10m from the image plane. The three camera positions are defined by the world coordinate position and end-effector (camera) orientation vectors **o,a,n** and these are controllable in world coordinates. Thus cameras are positioned to

collect images from any point up to 10m as shown in Figure 5.

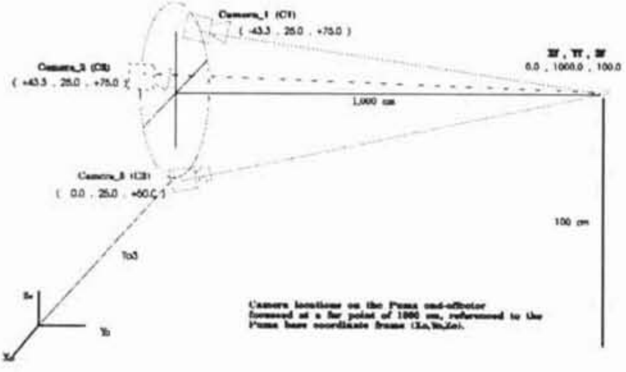


Figure 5. Specific robot positioning case for camera views at C1,C2 and C3 in a vertical circle of radius 50 cm, focussed at a far point of 1000 cm.

Given the required camera positions C1,C2 and C3 the inverse transformation can be computed directly from known data as:

$$T_{01}^{-1} = \left|\begin{array}{c|c} [n_1\ o_1\ a_1]^{\mathrm{T}} & \begin{array}{c} -p_1 . n_1 \\ -p_1 . o_1 \\ -p_1 . a_1 \end{array} \\ \hline 0 & 1 \end{array}\right| \quad (4)$$

This specific example is thus the inverse transformation of points in the image frame referenced to the world coordinate frame of the robot. It can be used for any of the C1,C2 or C3 camera points.

Then using equation (4) back into the ray relations (3) we can define the exact line equations as:

$$X_{\mathrm{d}} = \frac{d . {T_{01}}^{-1} . X_{\mathrm{s}}}{D} \quad Z_{\mathrm{d}} = \frac{d . {T_{01}}^{-1} . Z_{\mathrm{s}}}{D}, Y_{\mathrm{d}} = d \quad (5)$$

This line is then matched to visual cues from the second camera. Once matched, as illustrated in Figure 3, the actual depth d is known.

RESULTS

Analysis using this revised method was undertaken using a PUMA 560 robot, which can position to an accuracy of +/- 1mm for x-y-z space and with +/- 5' of arc in orientation accuracy. A working envelope as shown in Figure 5 was used i.e. cameras placed on a vertical circle 50 cm above the origin and of radius 50 cm. The image analyser was replaced with a single line of three transputer cards; the first for capture and image smoothing, the second for polygon fitting and the third for 3D analysis. The sweep time for the robot was 1.5s, and 2D processing took 2.5s per image, reduced by the pipeline to 1.25s. Matching took the T805 processor 10s on a more complex office scene.

CONCLUSIONS

This paper has described a novel approach to 3D object analysis and 3D scene reconstruction. Inaccuracies apparant in the first implementation using rigid cameras in a sensor head have been improved by using a robot manipulator to accurately position each camera in set positions. A ten-fold improvement in camera geometry accuracy has been achieved.

Secondly, the new method allows changes to the viewing angle by simply selecting robot end effector positions, and all parameters can be read directly from the robot controller.

REFERENCES

[1] R. Booth 3D scene reconstruction using passive cameras, PhD Thesis, Robotics Group, Newcastle University (1989).

[2] C.R. Allen, M.Farsi, R.Booth, On the reconstruction of 3D shape using wireframes, Proc. IEE Conf. Control'91, Edinburgh, March 1991,816-821(1991).

[3] H. Makela & K. Koshinen, Navigation of outdoor mobile robots using dead reckoning, Proc IARP'91, Pisa, Italy, IEEE , 1051-1056(1991).

[4] M. Brady et al, Sensor-based control of AGVs, Proc. DTI Conf on Domestic Robots, Newcastle-upon-Tyne,England,173-188(1989).

[5] Porrill & Pridmore, A transputer-based stereo analyser, SERC ACME Project (1991-93).

[6] J. Takeno & S. Hachiyama, New technology on stereo vision for mobile robots, Proc IARP'91, Pisa Italy,1383-1391(1991).

[7] H.Y.F.Feng & T. PavlidesThe generation of polygon outlines of objects from grey level pictures, IEEE Trans ECS, 22,427-439(1975).