Ground Plane Segmentation from Multiple Visual Cues

Bojian Liang and Nick Pears
Department of Computer Science,
University of York,
York, YO10 5DD, UK
email: bojian,nep@cs.york.ac.uk
Fax: +44-1904-432767

March 1, 2002

Abstract

An approach which uses multiple sources of visual information (or visual cues) to identify and segment the ground plane in indoor mobile robot visual navigation applications is presented. Information from color, contours and corners and their motion are applied, in conjunction with planar homography relations, to identify the navigable area of the ground, which may be textured or non-textured. We have developed new algorithms for both the computation of the homography, in which a highly stable two point method for pure translation is proposed, and the region growing. Also, a new method for applying the homography to measure the height of a visual feature to the ground using an uncalibrated camera is also developed. Regions are segmented by color and also by their sizes and geometric relation and these region boundarys are extracted as contours. By controlled manoeuvres of a mobile robot, the methods of coplanar feature grouping developed in this paper are not only applicable to corner correspondences but also to contours. This leads to robust, accurate segmentation of the ground plane from the other image regions. Results are presented which show the validity of the approach.

Keywords: Visually-guided robot navigation, Grouping and segmentation, Multiple cue systems.

1 Introduction

In this paper, we focus on the detection and segmentation of the ground plane for visual navigation of mobile robots in indoor environments. The fundamental assumption is that the floors are planar to some approximation. Apart from this basic requirement, we impose no further *environmental* restrictions and aim to be able to navigate robustly in a broad range of indoor scenarios. (Note, however, that since our robot has a blind area around its wheels, we also need an *initial position* assumption, such that the robot starts on the floor and can make a small motion, typically 0.1m, to initialise our ground plane detection algorithms before colliding with an obstacle.) Once the robot begins to manoeuvre, scene structure is automatically measured. In our approach, we apply the planar homography, **H**, and multiple visual cues for ground plane segmentation, which aim to improve performance and robustness in comparison to single visual cue system. The recovered **H** matrix, in conjunction with a cross ratio construct, is further applied to measure the height of a visual feature above the floor in terms of the height of the camera optical center. This provides a means to both detect the ground and obstacles using an uncalibrated camera. To make the approach robust, specific vehicle manoeuvres are applied to probe the structure of the scene and, in this way, the computation of the **H** matrix is greatly simplified and is more stable.

2 Outline of the ground plane segmentation procedure

Here we give a brief high level description of our algorithm for segmenting the ground plane.

1. Extract visual features in the image. In particular, the Plessey corner detector and Canny edge detectors are used.

2. Perform a color based region segmentation based on a quadtree split-merge algorithm and determine the boundary of each region.

3. The robot moves and all the corners, contours and regions are tracked to get the correspondence over one or more frames.

4. Feature tracks are used to determine whether the motion between a pair of frames is (approximately) pure translation. This determines whether H is computed from a general 4-point correspondences method or a more stable 2-point correspondence method for pure translation.

5. For pure translation, all available corner correspondences are used to get the vanishing point and we choose the available features (corners/edges/region boundary) from the region nearest to the camera (the 'seed region') for the computation of the horizon line and hence the homography (H matrix). It is not necessary that any corner correspondences exist within this region, as correspondences on the boundary of the region can be used. The orientation of the horizon line can check if this region can not be the ground plane, in which case the robot can take evasive action. (It does not, however, guarantee that it is the ground plane as it may be near-parallel. In future, virtual parallax checks need to be added to disambiguate such planes.)

6. If the seed region is deemed to be on the ground plane, the computed homography is used to check the boundary of all other regions to test whether they are coplanar with the ground plane.

Thus a ground plane segmentation consisting of several (possibly) different color regions is obtained, in any arbitrary topology (eg adjacent or not adjacent regions, compact or non-compact regions, etc). In the following sections, we describe this process in more detail. Since the corner and edge detectors used are standard, we start with region segmentation and go on to describe computation of the ground plane homography and ground plane region grouping.

3 Region Segmentation

To find the ground plane, regions in the image must be segmented and re-grouped into a co-planar set. We apply a split and merge technique for automatic region growing. In contrast to other approaches, our approach generates both non-textured and textured regions. The procedure is as follows:

- Split the image using the quadtree method on the basis of color difference. The variance of the color difference in the block is used to determine whether the block is divided further.
- Regions are classified by their area and this can be combined with the decomposition procedure. Each region is represented by the coordinates of the top-left corner, the mean color of the block and the dimensions of the block.
- Using the mean color of the largest block as the initial seed, find all blocks with similar color (the difference of mean color less than threshold) in the tree structure to form a list of 'similar color' blocks, then merge all geometrically adjacent blocks in the list. These blocks are marked on the block tree as one single region.
- Repeat the above procedure for the unmarked blocks in the tree, until no further blocks can be merged.

After grouping blocks with similar color properties, there remains some small (eg. less than 4 by 4 pixels) blocks. They may be adjacent to each other but with significant different colors and hence can only be merged by the geometric adjacency relation. Such regions are textured areas in the image. Some blocks merged in such a way may be non-coplanar but are separated by coplanarity checking at a later stage. This coplanarity checking can be done by both corners and contours inside the region, in addition to the region boundary itself.

4 Grouping coplanar features using homographies

4.1 Computation of the H Matrix

A planar homography (or plane to plane projectivity) defines relations of images of points on a planar surface at two view-points. Let $\mathbf{X}_i$ be a set of points which are coplanar in the 3D world. The images of $\mathbf{X}_i$ from two view-points are related by a plane to plane projectivity or homography, $\mathbf{H}$, such that,

$$\lambda \mathbf{x}_{i2} = \mathbf{H}\mathbf{x}_{i1}. \tag{1}$$

where λ is a scalar, $\mathbf{x}_{i1}$ and $\mathbf{x}_{i2}$ are homogenous image coordinates of the images of point $\mathbf{X}_i$, $\mathbf{H}$ is a 3 by 3 matrix representing the homography. As homogenous coordinates are defined up to a scale factor, the $\mathbf{H}$ matrix has only eight degrees of freedom. Once the $\mathbf{H}$ matrix of the ground plane from two viewpoints has been recovered, it can be used to check whether other feature points in the scene lie in the same plane and hence a coplanar point set can be constructed. Note, however, Eq-1 does not provide quantitative measurement for non-coplanar points. Early work on exploiting coplanar relations has been presented by Tsai and Huang [9], Longuet-Higgins [10] and Faugeras and Lustman [13].

The $\mathbf{H}$ matrix has eight degrees of freedom and it can be determined by standard linear methods. Four corresponding point pairs in general position (no three collinear) provide eight independent constraints and the solutions of the linear system defines the $\mathbf{H}$ matrix up to a scaling factor. When the number of point pairs is more than four, a standard least square method can be used, usually in conjunction with some form of sample consensus to reject outliers. (For pure translation the eigenvectors of the $\mathbf{H}$ matrix indicate the plane normal/horizon line and distinguish it from other planes.) However, we have noted that there are several disadvantages with using approaches which directly use four or more image correspondences and, as an alternative, we propose a horizon line

- vanishing point (2-point) method, which exhibits greater robustness when the robot undergoes pure translation. Consider two camera centered coordinate systems, frame 1 and frame 2, so that we can write

$$\mathbf{X}_2 \,=\, \mathbf{R}\,\mathbf{X}_1 \,+\, \mathbf{T}, \tag{2}$$

where $\mathbf{X}_1$ and $\mathbf{X}_2$ are the coordinates of the same 3D point, expressed in frames 1 and 2 respectively and where $\mathbf{R}$ and $\mathbf{T}$ are the rotation and the translation matrices encoding the relative position of the two coordinate systems. Now assume that $\mathbf{X}_1$ is a point on the plane defined by:

$$A\,X_1 \,+\, B\,Y_1 \,+\, C\,Z_1 \,+\, 1 \,=\, 0. \tag{3}$$

This is a plane which does not pass through the origin (i.e. the optical center of the camera) and $\mathbf{N} = (A, B, C)^T$ is the plane normal. Thus we have $\mathbf{N}^T\,\mathbf{X}_1 = -1$ and denoting $\mathbf{T} = k\mathbf{t}$, where k is a scalar and $\mathbf{t}$ is a unit vector, we have:

$$\begin{aligned} \mathbf{X}_2 \;&=\; \mathbf{R}\,\mathbf{X}_1 \,-\, k\mathbf{t}\,N^T\,\mathbf{X}_1 \\ &=\; (\mathbf{R} \,-\, k\mathbf{t}\,N^T)\,\mathbf{X}_1. \end{aligned} \tag{4}$$

The images of the scene point can be written as:

$$\begin{aligned} \mathbf{x}_2 \;&=\; \mathbf{P}(\mathbf{R} \,-\, k\mathbf{t}\,N^T)\mathbf{P}^{-1}\mathbf{x}_1 \\ &=\; \mathbf{H}\mathbf{x}_1. \end{aligned} \tag{5}$$

where $\mathbf{P}$ is the (unknown) camera model. For a pure translation, $\mathbf{R} = \mathbf{I}$, and so $\mathbf{H}$ has the form

$$\begin{aligned} \mathbf{H} \;&=\; \mathbf{P}(\mathbf{I} \,-\, k\mathbf{t}\,N^T)\mathbf{P}^{-1} \\ &=\; \mathbf{I} \,-\, k\mathbf{P}\mathbf{t}\,N^T\mathbf{P}^{-1}. \end{aligned} \tag{6}$$

We note that $\mathbf{Pt}$ is the vanishing point, $\mathbf{v}_p$, and $N^T\mathbf{P}^{-1}$ is the horizon line, ${\mathbf{v}_l}^T$, in the image. Thus, we have

$$\mathbf{H} = \mathbf{I} - k\mathbf{v}_p\mathbf{v}_l^T \tag{7}$$

As shown in Fig-1, two corresponding point pairs fully define the horizon line and the vanishing point. Given that we know the vanishing point and horizon line, scalar k can be recovered by substituting any one known corresponding point pair and thus the H matrix can be recovered. From 7 we have

$$\mathbf{x}_2 = \mathbf{x}_1 - k\mathbf{v}_p\mathbf{v}_l^T\mathbf{x}_1 \tag{8}$$

Since this equation is defined up to a scale factor we have

$$\lambda\mathbf{x}_2 = \mathbf{x}_1 - ks\mathbf{x}_t \tag{9}$$

where $s\mathbf{x}_t = \mathbf{v}_p\mathbf{v}_l^T\mathbf{x}_1 = [sx_t, sy_t, s]^T$. Normalising homogenous vector $\mathbf{x}_2$ gives

$$x_2 = \frac{x_1 - ksx_t}{1 - ks}, \quad y_2 = \frac{y_1 - ksy_t}{1 - ks} \tag{10}$$

Thus we have two estimates of the scalar k as

$$k_x = \frac{x_2 - x_1}{s(x_2 - x_t)}, \quad k_y = \frac{y_2 - y_1}{s(y_2 - y_t)} \tag{11}$$

Now suppose there are n ($n \geq 2$) sets of corresponding point pairs, indexed as ($0 \leq i < n$), then a least squares fit can be applied to obtain the scalar k, as

$$k = \frac{1}{2n} \sum_{i=0}^{n-1} (k_{x_i} + k_{y_i}) \tag{12}$$

Once k has been computed, H can be recovered by Eq-7. Compared with using 4 point correspondences to compute H, this approach generates a "well formed" H matrix. By this we mean that it encodes a motion of pure translation and its eigenvectors are the points on the horizon line. This is valuable in terms of 3D reconstruction relative to the ground plane.

In practice, the vanishing point can be computed by using all corner correspondences, not just those on the ground plane. Intersection of the two lines which join each pair of end points of the loci of the co-planar point pair is a point on the horizon line (see fig 1). These intersection points can generate the horizon line using robust approaches such as RANSAC.

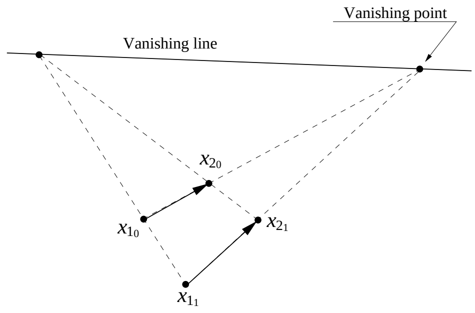


Figure 1: Two corresponding point pairs fully define the vanishing point and the horizon line.

The image region nearest to the camera is the best candidate region for the initial ground plane test and the features nearest to the camera can be used to compute the horizon line. If the nearest features are not two corresponding point pairs (corners) but image contours, the corresponding points can be defined by choosing a point on one contour, constructing a line passing through this point and the vanishing point, and finding the intersection of this line with the remaining contour.

4.2 Height above the ground plane

We note that the **H** matrix does not provide a quantitative measurement of how far a point is from the plane which defines the homography. This may be problematic in practice, since the assessment of a measurement error is necessary and the measurement of the height of a potential obstacle above the ground is a fundamental requirement to find the navigable region. Here, we show that, using an uncalibrated camera, this can be done under pure translation in terms of the height of the camera optical center.

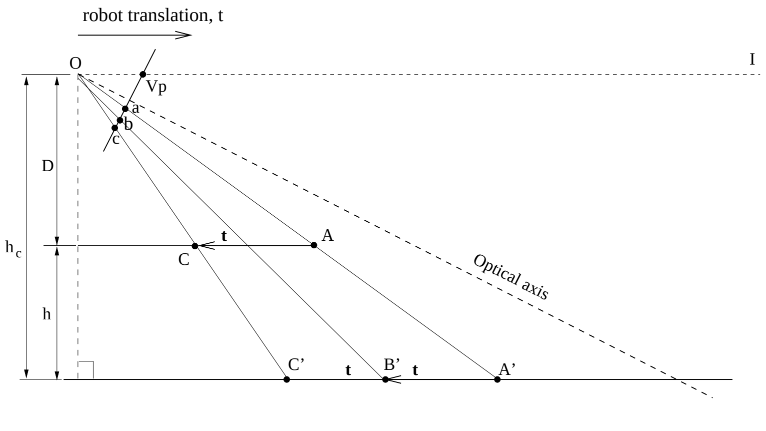


Figure 2: Computation of height of point A.

Our aim is to recover the height of corner point A shown in figure 2, when the robot undergoes pure (forward) translation, t (and thus the scene point translates t units towards the robot). Point A is the actual position of the corner point relative to the camera before the translation and point C is the position of the corner after the translation. Points A' and C' are the projections of these actual corner positions onto the ground plane. Points a and c are the image positions of the corner at positions A and C respectively and b is the predicted image position of the corner point, if the corner point were to lie in the ground plane. Image point b is computed from the recovered H matrix as $\mathbf{b} = \mathbf{Ha}$.

Now the height of the corner point relative to the height of the camera optical centre is

$$h_r = \frac{h}{h_c} = 1 - \frac{D}{h_c} \tag{13}$$

Using similar triangles, and denoting the distance between points x and y as $d(x, y)$, we note that:

$$\frac{D}{h_c} = \frac{d(OC)}{d(OC')} = \frac{d(AC)}{d(A'C')} \tag{14}$$

For pure translation, $d(A, C) = d(A', B')$, so that

$$h_r = 1 - \frac{d(A'B')}{d(A'C')} \tag{15}$$

Now, the four image points (a, b, c, V_p), where V_p is the vanishing point, and the corresponding four ground plane points (A', B', C', ∞) are collinear. The cross ratio for this set of points remains invariant under projection and so we can write:

$$\frac{d(A'B')}{d(A', C')} = \frac{d(a, b)\, d(c, V_p)}{d(a, c)\, d(b, V_p)} \tag{16}$$

Hence we can compute relative height as:

$$h_r = 1 - \frac{d(a, b)\, d(c, V_p)}{d(a, c)\, d(b, V_p)} \tag{17}$$

This can be interpreted as the height of point A units of height h_c.

Note that this approach only needs the ground plane homography, H, and the tracked image correspondences a and c of the feature to determine the height above the ground plane. By thresholding the measured height above the plane, the method can be used to check for ground plane points, which can be driven over, and for sufficiently high feature points which can be driven under. Note that this is achieved without camera calibration.

We note that a similar idea has been proposed by Criminisi *et. al* [14]. He proposed a method to compute the distance (refered to a common scaling factor) between a plane parallel to some some reference plane. However, we have removed the constraint of needing a known vanishing point of a reference direction from Criminisi's method and our method can be applied to compute the height from any isolated point to the reference plane.

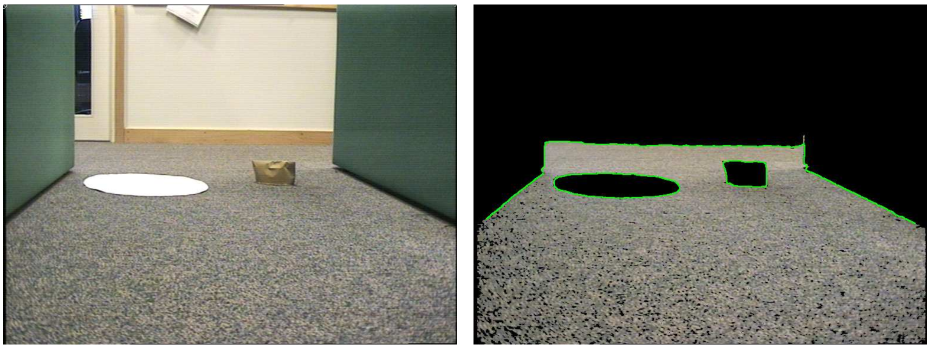
4.3 Ground plane segmentation

The region segmentation procedure, described in section 3, generates a list of regions whose boundaries are detected. The region nearest the camera is assumed to be the most likely candidate for a ground plane region and features within this region are tracked and used to compute the **H** matrix (the horizon line validates that the region is approximately the correct orientation). If validated, this region is used as the "ground region seed". The boundaries and feature points in the adjacent regions are then used to check whether it is coplanar with the ground region seed (by using the height measurement method described in the previous section). The ground region is thus grown by combining the adjacent coplanar regions.

5 Experimental results

In this section, experimental results validating our ground plane segmentation approach are presented. Image sequences were grabbed by a camera mounted on a mobile robot which moved in the pure translation mode.

In the experiment described here, we try to merge an elliptical coplanar patch (a piece of white paper on the ground) with the ground plane seed region and separate a box shaped non co-planar patch (an obstacle!) from the ground plane seed region (see Fig-3(a)). The region growing yielded a ground seed region which has two holes on it, one elliptical and one roughly rectangular (Fig-3(b)). The three contours on the region were automatically tracked (Fig-3(c)) and subsequently coplanarity checking was applied. The estimation of the height ratio of the ellipse and obstacle region boundaries are plotted in Fig-4. The ellipse region was then automatically merged to the ground plane as in Fig-3(d), as it is flush with the carpeted area shown in Fig-3(b).



(a). Raw data.

(b). Initial seed region.

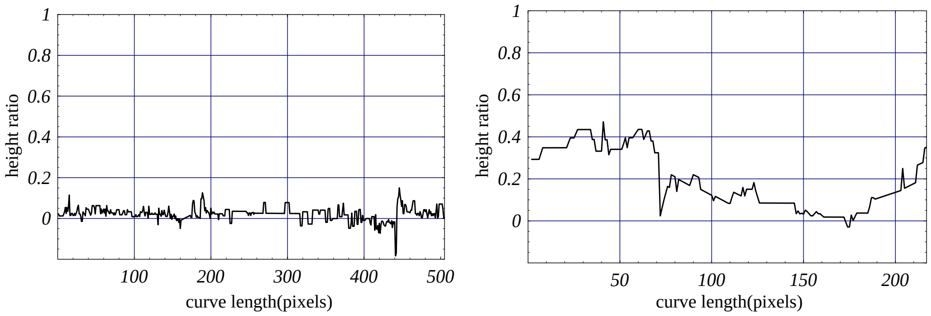
(c). Tracking the region boundarys contours.

(d). The ground region.

Figure 3: Separation of the obstacle by co-planarity checking.

6 Conclusions

We have presented a method of ground plane segmentation for mobile robot visual navigation applications, which employs multiple sources of visual information, in conjunction with planar homographyies. In particular, we illustrated how, for pure translation, a homography can be computed from just two pairs of corresponding corner features. We also showed how, for pure translation, we can determine the height of corner features above the ground plane using the recovered homography and a construct based on the cross ratio. This allows us to detect points which can be driven over, as their height is measured to be close to zero, and points which are sufficiently high to drive under. Our experimental results have shown the viability of the approach



(a). Estimation of height of the in-plane boundary.

(b).Estimation of height of the obstacle boundary.

Figure 4: Co-planarity checking for the contours inside the region.

over long image sequences, and we plan to expose our procedures to a wide range of scenarios to demonstrate its robustness.

References

[1] Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge, 2001.

[2] Guerrero, J.J and Sagues, C.: Navigation from uncalibrated monocular vision. In Proc. 3rd IFAC symposium on Intelligent Autonomous Vehicles. (1998) 210–215

[3] Santos, V., Sandini, G., Gurotto, F. and Garibaldi, S.: Divergent Stereo in Autonomous Navigation: From Bees to Robots. Int. Journal of Computer Vision (1995) 14:159–177

[4] Cipolla, R., Blake, A.: Surface Orientation and Time to Contact from Image Divergence and Deformation. In Proc. 2nd European Conf. on Computer Vision and Pattern Recognition, (1992) 761–764

[5] Coombs, D., Herman, M., Hong, T.H. and Nashman, M.: Real-time Obstacle Avoidance Using Central Flow Divergence and Peripheral Flow. Int. Journal of Robotics and Automation. (1998) 14(1):49–59

[6] Sinclair, D., Blake, A.: Quantitative planar region detection. Int. Journal of Computer Vision. (1996) 18(1):77–91

[7] Jain, A.K.: Color Distance and Geodesics in Color 3 Space. Journal of the Optical Society of America, (1972) 62:1287–1290.

[8] Hartley, R., Gupta, R. and Chang, H.: Stereo from Uncalibrated Cameras In Proc. IEEE Conf. on Computer Vision and Pattern Recognition, (1992) 761–764

[9] Tsai, R., Huang, T.: Estimating Three Dimensional Motion Parameters of A Rigid Planar Patch IEEE Trans. Acoustics, Speech and Signal Processing, vol. 29, no. 6, (1981) 1147–1152

[10] Longueit-Higgins, H.C.: The Reconstruction of A Plane Surface from Two Perspective Projections. In Proc. Royal Society London, B227, (1986) 339–410

[11] Irani, M., Anandan, P.: Parallax Geometry of Pairs of Points for 3D Scene Analysis. Computer Vision - ECCV96, (1996) 17–30

[12] Cheong,L.F., Fermuller, C. and Aloimonos, Y,: Spatiotemporal Representations for Visual Navigation. Computer Vision - ECCV96, (1996) 673–684

[13] Faugeras, O., Lustman, F.: Motion and Structure from Motion in A Piecewise Planar Environment. Int. Journ. Pattern Recognition and Artificial Intelligence. vol. 2, no.3, (1988). 485–508

[14] Criminisi. A., Reid. I and Zisserman. A.: Single View Metrology International Journal of computer Vision, vol.40, no.2, 2000, 123-148