Ground Plane Segmentation for Mobile Robot Visual Navigation

Nick Pears and Bojian Liang
Department of Computer Science
University of York
York, YO10 5DD, UK
email: nep@cs.york.ac.uk, bojian@cs.york.a.uk

Abstract

We describe a method of mobile robot monocular visual navigation, which uses multiple visual cues to detect and segment the ground plane in the robot's field of view. Corner points are tracked through an image sequence and grouped into coplanar regions using a method which we call an H-based tracker. The H-based tracker employs planar homographys and is initialised by 5-point planar projective invariants. This allows us to detect ground plane patches and the colour within such patches is subsequently modelled. These patches are grown by colour classification to give a ground plane segmentation, which is then used as an input to a new variant of the artificial potential field algorithm.

1 Introduction

In this paper we discuss visual navigation (using a standard CCD camera) for mobile robots in indoor environments. Our focus on indoor environments means that planar regions in the scene will be common. In particular, floors which are planar to some approximation is a fundamental assumption. Apart from this ground planarity requirement, we impose no further restrictions and ultimately aim to be able to navigate in a broad range of indoor scenes. This is a challenging problem, since, as the vehicle moves around, the various visual cues that aid navigation disappear and reappear in the robot's visible environment.

Various (isolated) visual cues have been employed to facilitate navigational functions with uncalibrated cameras. These include navigation down corridors both by using the focus of expansion of non-vertical scene lines [7] and wide field peripheral flow [9]. Other approaches have used time-to-contact from image divergence [3], a combination of central flow divergence and peripheral flow [4], and quantitative planar region detection using point correspondences [13]. Most of these techniques work in some types of scene, but will fail when a particular type of feature is not well supported within the image data. This has motivated us to use multi-cue systems where, initially, we are looking at corner points, colour and texture. Of these cues, corner points are the most fundamental and can be used to recover scene structure. This is because, unlike edge motion, which suffers from the aperture effect, it is possible to fully extract their motion in the image plane, across a sequence of images.

Our initial high level requirements for navigation are (i) to determine the region in the image that corresponds to the ground plane and (ii) to determine which parts of the ground plane are navigable. Parts of the ground plane are not navigable, simply because of the robot's finite dimensions. Ground plane regions must be excluded which have obstacles or walls that are less than a half the dimension of the robot in their neighbourhood. Other areas which have overhanging obstacles less than the robot's height must also be excluded. These requirements suggest that need to extract the ground-plane and reconstruct other environmental features and obstacles in terms of units of robot height and robot width. This paper focusses on the ground plane segmentation problem.

2 Review of corner based approaches

Ultimately we hope to use many types of visual cue to aid navigation. However, we believe that the structural information that can be extracted by tracking corner points should be central to our system. Therefore, we briefly review three methods which use corner tracking (or correspondences) to elicit structural information.

2.1 Navigation using F

Perhaps the most common approach used to track corner features through an image sequence is the so called "F-based tracker", where "**F**" stands for the *fundamental matrix*. The fundamental matrix models the epipolar geometry between two views taken by uncalibrated cameras and the F-based tracker is an iterative

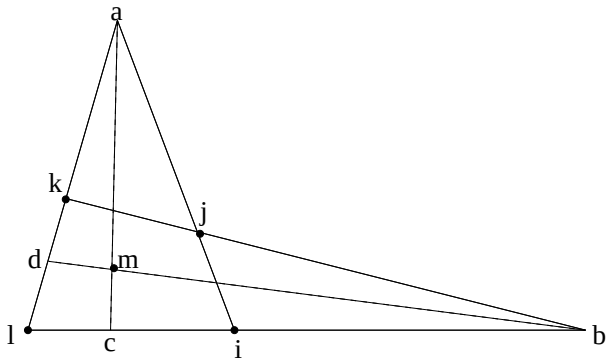

Figure 1: Projective construction for 5 point planar invariant

process which simultaneously estimates **F** and and the correspondences consistent with that **F**. Once **F** is estimated, it may be used to reconstruct 3D position of the points in the scene up to an ambiguity of a projective transformation [12]. Furthermore, if camera parameters are approximately known, the projective skew can be "unwound" to give a "quasi-Euclidean" structure which may be used for navigation purposes.

2.2 Navigation using invariants

Another approach used to detect coplanar points is the direct use of projective invariants, as exemplified by [13]. This uses the fact that if we have four collinear points in the scene, say a, b, c, d, then a ratio of ratios of distance (the cross ratio) is a projective invariant. This fact can be extended to five points in a general position on a plane, since, using projective constructions, we can get two sets of four collinear points, which are invariant if and only if the original five points are coplanar. Figure 1 shows this construction. The point groupings l, d, k, a and l, c, i, b give the two invariants as:

$$I_1 = \frac{d_{lk} d_{da}}{d_{la} d_{dk}}, \quad I_2 = \frac{d_{li} d_{cb}}{d_{lb} d_{ci}} \tag{1}$$

2.3 Navigation using H

Early work on exploiting coplanar relations has been presented by Tsai and Huang [14], Longuet-Higgins [11] and Faugeras and Lustman [5]. We summarise the coplanar relation as follows: If a set of corner features in the scene lie in a plane, and they are imaged from two viewpoints, then the corresponding points in the two images (separated by k frames) are related by a plane-to-plane projectivity or homography, H, such that:

$$\lambda \mathbf{x_i} = \mathbf{H} \mathbf{x_{i-k}} \tag{2}$$

where $\mathbf{x}$ represents a homogenous image coordinate $(x, y, 1)^T$, $\mathbf{H}$ is a 3 by 3 matrix representing the homography and λ is a scalar. Since this equation is valid up to a scale factor, $\mathbf{H}$ has only eight degrees of freedom, and it is normal practice to choose λ such that element h_{33} in $\mathbf{H}$ is set to unity. Eight degrees of freedom requires that we have four corresponding coplanar features in general position (no three collinear), since each pair of corresponding points then provides two independent constraints, and $\mathbf{H}$ can be determined by standard linear methods.

Equation 2 suggests a method of grouping corner features into coplanar sets. Namely, if we can select a set of four coplanar corresponding point pairs which are in a sufficiently general configuration in both images (each point is unique and no three are collinear), then $\mathbf{H}$ can be computed and used to check whether other points in the scene lie in the same plane.

3 Navigation using an H-based tracker

Due to the degeneracies and sensitivities to noise in the estimation of **F**, particularly in scenes with a single dominant plane (such as a ground plane), we aim to use primarily **H** relations to detect the ground plane and planar projective invariants to help bootstrap this process. We call our system an *H-based* tracker. In this section, we give a top down description of our algorithm, and the corresponding subsections describe each of the main stages in more detail.

H-based tracker algorithm
We first run an initialisation stage where we

1. Detect corners using a standard corner detector.
2. Track these points over n frames using a Kalman filter, with a standard motion model (of velocity) and cross correlation to determine matches.

This generates a reasonable disparity between corresponding corner points in frame 1 and frame n before attempting to estimate H. In subsequent frames we search for correspondences between frame i and frame $i - n$. Thus from frame $n + 1$, we run the H-based tracker. The key modification from the basic tracker used in the initialisation stage is that two process models are employed in the state prediction and data association stages of the tracker. The first stage is the standard motion model used in the initialisation stage. The correspondences generated from this allows bootstrapping of the ground plane **H** by testing a population of putative **H** matrices. This is a sample consensus approach similar to RANSAC [6], but the samples are

not selected randomly (see section 3.3). H matrices can then be used as a model to predict and associate measurements in an iterative manner. To summarise these steps we

1. Bootstrap the system by computing a population of putative **H** matrices for the corner points which have their vertical component of image motion in a downwards direction (i.e. are below the horizon line).

2. Select the dominant **H** model i.e. that which verifies the largest number of corner associations. The corners points that are verified are deemed *inliers*.

3. Recompute **H** by applying orthogonal least squares to the inliers.

4. Retest the data associations of corner points to tracks using the least squares estimate of **H** to get an updated set of inliers.

5. Iterate around the previous two steps until the number of inliers stabilises.

6. Check that the coplanar points extracted are ground plane points by computing the plane normal.

It is possible to remove all of these coplanar corners, and repeat the whole procedure to find further significant co-planar corner groupings in the scene. Indeed, it may be necessary to do this if we find that the dominant plane can not be the ground plane, due to the computed plane normal. In subsequent frames, we simply sample from the group of points that are deemed to be in the ground plane and choose a suitable selection of basis points to compute a new **H**. We now describe the steps in the algorithm in more detail.

3.1 The corner detector

The Plessey-Harris detector [8] is used as it is stable and gives sub-pixel accuracy. Since this corner detector well known and standard, it will not be described further here.

3.2 The tracker

Currently we are using a Kalman filter to track features through the initial image sequence. In environments with a low density of corner features, this works well since, for most tracks, there is a single feasible feature match within the validation gate. However, in scenarios with a high density of corner features, there are many cases where there are several well correlated features within the validation gate and the correct measurement to track association is not obvious. In future work we plan to use, a more sophisticated tracking tool, which is an extension of the simple (nearest neighbour) Kalman filter, and is known as the joint probabilistic data association filter (JPDAF). The key difference of the JPDAF compared with the KF is a more sophisticated data association mechanism, which models the probability of *joint* association events across all tracks [1].

3.3 Bootstrapping ground plane detection

To bootstrap the system, we wish to select four points in general position which lie on the ground plane. In addition to being on the ground plane, it is required that the correspondences should be positioned so that they give a reasonably accurate estimation of H. Thus, we have the following requirements for a bootstrap procedure

1. points within an image should not be too distant. (This ensures a good chance of coplanarity).

2. points within an image should not be too close (for good **H** estimation accuracy).

3. points within an image should not be near collinear (for good **H** estimation accuracy).

4. corresponding points across two images should have a large disparity in position (for good **H** estimation accuracy).

Bootstrap algorithm

Firstly we delete correspondences between the two frames where the image motion is below a threshold. Effectively, this removes points close to the epipole, and complies with requirement 4 above. With the remaining points, we construct a symmetric adjacency matrix where the entries give the (square of the) Euclidean separation of points *within* the latest frame. For each point in this matrix we

1. Find a neighbouring point which is closest to the selected point but above a minimum separation threshold.

2. From the remaining neighbouring points, find a third point which is above the minimum separation threshold for the two selected points, not near-collinear with those points and a minimum product of the separation from those points.

3. Find a fourth point above the minimum separation threshold for the three selected points, not near collinear with any of the three pairs of those three points, and a minimum product of the separation from those points.

4. Find a fifth point above the minimum separation threshold for the four selected points, not near-collinear with any of the four of the six pairs of points (the diagonal pairings are not checked), and a minimum product of the separation from those points.

5. We then remove groupings of five points that are not invariant over the two views. A grouping of five points is deemed to be non-coplanar if either of the two invariants from equation 1 changes by a threshold. i.e. we require $\delta I_1 < t_i$, $\delta I_2 < t_i$.

6. For each remaining grouping of five points, we remove the point correspondence pair with minimum disparity and use the remaining 4 points for a projective basis to compute an initial value for H. Thus we have a population of putative $\mathbf{H}$ matrices.

3.4 Data association in the tracker

A crucial part of the H-based tracker is generating an appropriate search window within which to make corner to track associations. To do this we assume that the measurement errors (dx_1, dy_1) at corner point (x_1, y_1) have zero mean, Gaussian distribution. We assume that the measurement errors in the x and y directions are independent, so that the covariance matrix $\mathbf{\Sigma}_1$ is diagonal:

$$\mathbf{\Sigma}_1 \quad = \quad \begin{bmatrix} \sigma_{x1}^2 & 0 \\ 0 & \sigma_{y1}^2 \end{bmatrix}. \tag{3}$$

Estimated corner measurement errors in image-1 are transferred to an estimate of the error in predicted position of the same corner in image-2 by:

$$\begin{bmatrix} dx_2 \\ dy_2 \end{bmatrix} \quad = \quad \mathbf{D} \begin{bmatrix} dx_1 \\ dy_1 \end{bmatrix} \tag{4}$$

where the Jacobian, D, is given by

$$\mathbf{D} \quad = \quad \begin{bmatrix} \frac{\partial x_2}{\partial x_1} & \frac{\partial x_2}{\partial y_1} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial y_1} \end{bmatrix} \tag{5}$$

and we have

$$\frac{\partial x_2}{\partial x_1} \quad = \quad \frac{(h_{11}h_{32} - h_{12}h_{31})y_1 + h_{11} - h_{13}h_{31}}{(h_{31}x_1 + h_{32}y_1 + 1)^2}, \tag{6}$$

and similar expressions for the other three partial derivatives. (Note that we are assuming that there are no errors in h_{ij}, which is not strictly true, although we expect these to be negligible if an accurate least squares estimate of $\mathbf{H}$ has been made.) The covariance matrix $\mathbf{\Sigma}_2$ of measurement error (dx_2, dy_2) is given by:

$$\mathbf{\Sigma}_2 \quad = \quad \mathbf{D}\,\mathbf{\Sigma}_1\,\mathbf{D}^T$$

The eigenvalues and eigenvectors of this covariance matrix define an equal probability density ellipse in which the prediction of a corner correspondence falls with a given probability. This defines a dynamic search area which is equivalent to searching within an integer number of standard deviations of the predicted corner position. Note that the size of the equal probability density ellipse depends on the position of the point in the (original) frame for a given mapping $\mathbf{H}$.

4 Combining corner and colour cues

In this section we describe our method of combining corner and colour cues to extract a first estimate of the navigable image region. At present, our implemented method is fairly basic, but our results illustrate how effective the general technique of combining cues can be. Our algorithm is as follows:

1. Corner points are tracked and classified as either *on the ground plane* or *off the ground plane*, using the H-based tracker described in previous sections.

2. Ground plane corner points are then grouped into one or more *ground plane patches*. These are collections of ground plane points where the distance to the nearest neighbour ground plane point within a patch is below a threshold.

3. A bounding polygon for these corner points defines an image region in which the colour space of the ground plane is modelled. (Currently we use a simple bounding ellipse in normalised colour space.)

4. Thus the region(s) classified as the ground plane (i.e. within the bounding polygons) can then be grown by classifying small image regions as either *ground plane colour* or *not ground plane colour*.

5 Robot motion commands from the extracted ground plane

To show that visual navigation is possible using extracted ground plane information, we have implemented a *wandering behaviour* on an experimental mobile robot. This is based on the idea of artificial potential fields (APF) [10]. The ideas which are new (to our

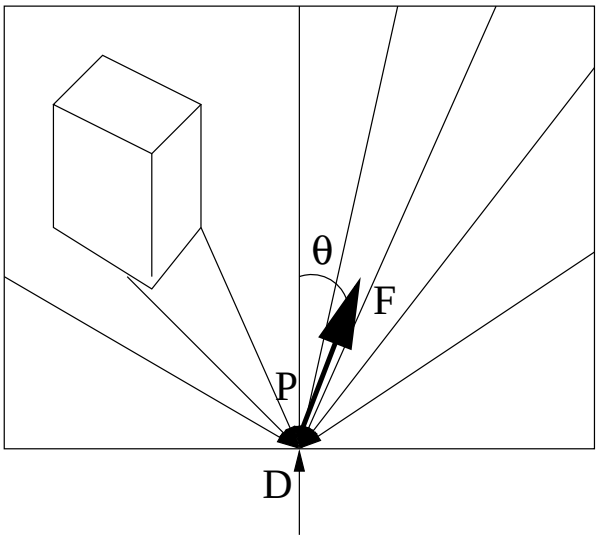

Figure 2: Image based APF algorithm

Figure 3: Tracked and grouped corners

knowledge) are (i) that we use image based information directly in the APF algorithm (ii) that we employ a fictional *driving force* to move the robot forward. Figure 2 shows this idea. A set of rays are cast out from the bottom of the image to the edges of the extracted ground plane. We generate a fictional force, which is in inverse square proportion to the length of the cast ray, and in the reverse direction of the cast ray i.e towards the *pivot* point P. We do a vector sum of these forces, and include the fictional driving force D, to produce a resultant vector, F. This will tend to point into open spaces and we can generate speed and steering instructions for the robot as

$$V = sign(F)k_V|F|, \kappa = -k_\theta\theta \tag{7}$$

where V is demand speed, κ is demand turning curvature and k_V, k_θ are constants. If the robot falls into a local minimum, a state sequencer turns it through 180 degrees and then returns it to wandering mode. This idea fits into Brook's concept of 'behavioural' robotics [2], but here we only intend to use it to test the reliability of our navigable ground plane detection algorithms. (Note that the small isolated misclassifications shown in figure 4 have to be removed for this technique to work. Our most recent work has shown how this can be done by combining additional cues, such as texture, in a probabilistic way.)

6 Results

Figures 3 and 4 illustrate the H-based tracker and colour region growing processes respectively. (For clarity, figure 3 only shows the strongest corner feature within a 20 by 20 pixel window.) The corners marked with a cross have been matched to previous positions, as shown by their trailing lines, and have been used to estimate the **H** matrix by orthogonal least squares. Other crosses, which are also inside the bounding polygon, are corners not used in the **H** matrix estimation, but whose correspondences lie within the matching ellipse associated with this **H** matrix, and so are deemed to lie on the same plane. Some of these small ellipses are overlaid on the image and it can be seen that the corners fall within their boundaries. All corner correspondences outside the bounding polygon failed the data association test defined in section 3.4. Again, some of the 'failed association' ellipses are shown on the bottom right of the obstacle. Once the bounding polygon has been extracted, the colour space of the ground plane is sampled, and a region growing algorithm expands the polygon to edges in the image where there is a change in colour. Figure 4 highlights the final ground plane region extracted from this technique. Notice how the ground plane detected can extend into regions there are no corners due to the texture gradient of the imaged carpet. Obviously, the technique works well in this particular case, because the ground plane has sufficient corner features, and the colour space of the ground plane is unimodal (i.e. homogenous). However, in further work we aim to develop a range of techniques, a selection of which can be automatically deployed depending on the image context.

Figure 4: Ground plane region extraction

7 Conclusions

We have described a method of mobile robot visual navigation, which aims to use multiple visual cues to improve the robustness of operation in indoor environments. It was argued that corner tracking should be central to the system, since the motion of corners can provide structural information. For initial ground plane detection, we have proposed a hybrid system which uses planar projective invariants to bootstrap the system, in conjunction with what we call an *H-based tracker* to track ground plane corners. Using colour cues in conjunction with corners, we have illustrated the potential of building robust visual navigation systems and we have shown how we aim to test this using a new variant on the *artificial potential field algorithm*.

References

[1] Y. Bar-Shalom and T. E. Fortmann. *Tracking and Data Association*. Academic Press Inc., Boston, USA., 1988.

[2] R.A. Brooks. A layered intelligent control system for a mobile robot. In *Third Intl. Symp. Robotics Research*, pages 365–372, 1986.

[3] R. Cipolla and A. Blake. Surface orientation and time to contact from image divergence and deformation. In *Proc. 2nd European Conf. on Computer Vision*, pages 187–202, 1992.

[4] T. H. Hong D. Coombs, M. Herman and M. Nashman. Real-time obstacle avoidance using central flow divergence and peripheral flow. *Int. Journal of Robotics and Automation*, 14(1):49–59, 1998.

[5] O. Faugeras and F. Lustman. Motion and structure from motion in a piecewise planar environment. *Int. Journ. Pattern Recognition and Artificial Intelligence*, 2(3):485–508, 1988.

[6] M. A. Fischler and R. C. Bolles. Random sample consensus: apardigm for model fitting with application to image analysis and automated cartography. *Commun. Assoc. Comp. Mach.*, 24:381–395, 1981.

[7] J.J Guerrero and C. Sagues. Navigation from uncalibrated monocular vision. In *Proc. 3rd IFAC Symposium on Intelligent Autonomous Vehicles*, pages 210–215, 1998.

[8] C. J. Harris and M. Stephens. A combined corner and edge detector. In *4th Alvey Vision Conference Manchester*, pages 147–151, 1988.

[9] F. Curotto J. Santos Victor, G. Sandini and S. Garibaldi. Divergent stereo in autonomous navigation: From bees to robots. *Int. Journal of Computer Vision*, 14:159–177, 1995.

[10] B. H. Krough. A generalised potential field approach to obstacle avoidance control. In *Robotics Research Conference Papers*, 1984.

[11] H. C. Longuet-Higgins. The reconstruction of a plane surface from two perspective projections. *Proc. Royal Society London*, B227:399–410, 1986.

[12] R. Gupta R. Hartley and T. Chang. Stereo from uncalibrated cameras. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 761–764, 1992.

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

[14] R. Tsai and T. Huang. Estimating three-dimensional motion parameters of a rigid planar patch. *IEEE Trans. Acoustics, Speech and Signal Processing*, 29(6):1147–1152, 1981.