IGR report

Robust, Adaptive Visual Navigation using Multiple Diverse Visual Cues

N. E. Pears, Dept. Computer Science, University of York, York, UK

1 Background/Context

This report summarises the activity and achievements of an 18 month (September 2000 - March 2002) fast-track grant, where around 62K was awarded by EPSRC. The majority of this money was used to fund a post doctoral RA, Dr Bojian Liang.

1.1 Background to the research area

Robust Computer Vision is becoming increasingly important as an enabling technology in a broad range of applications which include industrial automation, industrial inspection, medical diagnosis, human-computer interaction, surveillance, document processing, video database management and virtual reality. Although, our primary motivation is aimed at robotic visual navigation, the ability of computer vision systems to employ many diverse visual cues will be relevant to many of these application domains, allowing decision making processes within these applications to be maximally informed [4].

It is the fundamental aim of this research to attack robustness and adaptability issues within the domain of mobile robot visual navigation. This is a challenging application, since, as the vehicle moves around, the various visual cues that aid navigation disappear and reappear in the robot's visible environment.

The main objective is to develop new computer vision algorithms which employ multiple visual cues to reliably segment the ground plane in the robot's field of view [3]. The assumption is that the robot is equipped with a single uncalibrated camera and that many scenes will contain features which are mostly coplanar so that structure and motion methods based on estimation of fundamental matrices cannot be used. The algorithms should work in *any* indoor scenario, irrespective of feature content, provided the floor is flat to some approximation.

2 Key Advances

The key advances in this work are as follows

- Stable computation of an appropriate model for the image motion of the ground plane. In general, this motion is modelled by a homography matrix (H matrix) with eight degrees of freedom. In our work we have used standard methods (using four corresponding point pairs) for general motion, but we have developed more stable methods when the robot motion is automatically detected to be (near) pure translation. These methods include a corresponding points method (these may be two corners or two corresponding points on a contour) and a correlation based method. These are described in sections 2.1 and 2.2 respectively.

- Given a stable estimate of the image motion of the ground plane, we have developed methods which use this estimate to aid vehicle navigation. The most significant of these, described in section 2.3, is the measurement of the height of features (corners, contours, etc) above the ground plane, as this can be used detect and group ground plane segments. A further useful result we have generated is the measurement of planar rotation of the mobile robot (or at least its camera) using homographies. This is described in section 2.4.
- We have developed a new variant of a quadtree split-and-merge colour and texture segmentation algorithm. This is described in section 2.5.
- We have integrated our multi-cue structure and segmentation algorithms and tested them on a broad range of scenes to validate our new results. This is described in section 2.6.

2.1 Computation of an elation using point-to-point correspondences

If the motion of the camera is restricted to a translation, the plane to plane homography specialises to a transformation called an elation. We have developed a highly stable two-point method of determining this model for the cases when (near) pure translation is detected [2]. Compared to the general approach for determining a planar homography, this method has fewer degrees of freedom (four rather than eight) and we have been able to show that it can provide a more stable estimation of the ground plane image motion.

2.2 Computation of an elation without point-to-point correspondences

A method of recovering elations and hence detecting ground plane regions without point-to-point (eg corner) correspondences has been developed. Firstly we automatically detect for (near) pure translation, and we recover the vanishing point. We then change the pixel coordinate system so that it is centered on the vanishing point and perform an autocorrelation of the image along lines of constant θ (radial directions) when the image is described in coordinates $(\frac{1}{r}, \theta)$. The orientation of the vanishing line can be determined by fitting a sinusoid to the correlation results. The method is more widely applicable than corner based methods, since scenes often do not contain stable corner features. Even if corner features are available, this method gives more reliable results as much more of the image data is used, not just the parts of the image with corner features, which may be very sparse.

2.3 Measurement of height using planar homographies

A new method for applying planar homographies to measure the height of a visual feature above the ground using an uncalibrated camera has been developed [2]. This allows us to identify segmented image regions as belonging to the ground (near zero height) and therefore related to parts of the scene can be driven over. We can also detect image regions as obstacles that either need to be avoided or which are overhanging and are of sufficient height for the robot to drive under.

We have noted that Criminisi [5] has also proposed a method to compute the distance (up to a scale factor) between a plane parallel to a reference plane. This approach needs to know the vanishing line of the reference plane together with a vanishing point of a reference direction which is not parallel to the plane. This method also needs to know the intersection points of a line at the reference direction on the two planes and this is not always available in practice. These constraints make the approach less valuable for robot navigation in an unknown environment.

In comparison to Criminisi's work, we use the homography of a reference plane (the ground plane) and a corresponding point pair for the computation of this distance. This distance is referenced to the height of the camera optic center above the reference plane. Each point pair in our approach can be clearly identified and there is no ordering ambiguity in the computation of the cross ratio. We remove the constraint of a known vanishing point of a reference direction from Criminisi's method and our method can be applied to compute the height between any isolated point to the reference plane (i.e the ground). The height measured provides a useful quantitative measurement for planar feature grouping.

2.4 Vehicle rotation

A general planar motion includes both rotation and translation. In this case, the translation vector is always perpendicular to the plane normal and the axis of rotation (screw axis) is parallel to the plane normal. The H matrix of a planar motion generally has one real and two complex eigenvalues. We have shown that the real eigenvalue is unity and its corresponding eigenvector is the image of the intersection of the screw axis and the plane of motion. We have also shown that the two complex conjugate eigenvalues have unity amplitude and their phase angles are equal to the actual angle of rotation that the robot (or camera) undergoes. Thus we can measure the robot's rotation, independent of any camera calibration [2].

2.5 Image segmentation

To find the ground plane, regions in the image must be segmented and then re-grouped into co-planar sets. An incremental improvement on quadtree-based split-and-merge segmentation methods has been developed [1]. In contrast to other approaches, which generally are (non-textured) colour based segmentation approaches, our approach generates both non-textured and textured regions.

The technique splits the image using a quadtree decomposition based on colour moments. It then merges adjacent blocks based on difference of mean colour. After grouping blocks with similar colour properties, there remains some small (eg. less than 4 by 4 pixels) blocks. They may be adjacent to each other but with significant different colours 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.

A problem which was solved effectively was how to avoid exhaustive searching in the merge phase. This was done by run-length coding the tree location of the blocks in order to effectively determine their adjacency relations.

2.6 Integration and implementation

These achievements, when put together and implemented in a real robot navigation system, have allowed us to reliably detect and segment the ground plane in a mobile robot's field of view. The system works reliably in a diverse range of indoor scenes. The system that we have built is the most accurate and reliable ground plane segmentation system that I have seen, both in the Computer Vision and Robotics communities. Performance of the system can be viewed as mpeg videos on the project web page at: http://www-users.cs.york.ac.uk/ ˜ nep/research/vn.html

3 Project Plan Review

We have not fulfilled the original objective as presented in the grant proposal, namely the development of a general framework for multi-cue integration. I remain convinced that the use of multiple cues is the key to robust vision systems but I felt it was necessary to pursue sub-goals prior to achieving our longer term aim Thus, our generic long term aims were focussed on to more specific shorter term aims, which were more suitable to a project of length 18 months.

I realised that you need to work out how multiple visual cues can interact to yield pertinent task-orientated information for *specific* tasks before trying to build a *general* theoretical framework of how multiple cues can interact for any arbitrary robotic task which is supported by visual sensing. If this were not the case, one would not have a baseline with which to evaluate the performance of automatic cue integration systems.

Essentially, we chose what we saw as the most fundamental visual task which would be required to support a number of visual navigation functions, such as obstacle avoidance, localisation and mapping. This task was *detection and segmentation of ground plane regions* within the robot's field of view. Given that different scenes have different features (eg some have corners and some don't), it seemed appropriate that we should investigate methods using a variety of visual cues. Then we were able to manually investigate which cue or combination of cues would give the most reliable results in a wide range of indoor scenes which varied in the number and type of cues available. The final implementation of our system is capable of using corners, edges, contours of colour and texture regions, and the motion of all these features.

We have successfully shown how the use of many feature types can give more robust performance in the specific case of ground plane segmentation, and now our knowledge of this specific case will allow us to generalise and experiment with a number of different automatic cue integration frameworks such as Bayesian networks, Dempster-Shafer systems, Kalman filter estimators and so on.

4 Research Impact and Benefits to Society

The performance of our ground plane segmentation system is excellent, as witnessed in the mpeg videos on our website. This is due to the high stability of the methods employed to estimate the model of the image motion of the ground plane, and the excellent performance of the image segmentation method and ground plane region grouping method. Unfortunately, we were not able to present these video results at our latest conference submission (IEEE Int. Conf. on Robotics and Automation, Washington, USA) as Dr Liang's visa did not arrive on time. However, we are confident that these video results will have significant impact at our next conference presentation in August 2002 (2nd Int. Conf. on Image and Graphics), and we plan to write one further conference paper on this work (we will submit to CVPR03).

I believe that there is relevance of this work to researchers interested in biologically-inspired robotic vision systems, as it is widely believed that animate vision systems exploit the integration of multiple visual cues. Researchers involved with the psychophysics of human vision have shown how different visual cues dominate in different task-environment scenarios. I have made contact with four European research institutions with a view to working on biologically-inspired vision architectures.

Recently, a number of domestic robots have come on the market, with British companies being among the leading competitors. Such robots include those which vacuum the floor and those which mow the lawn. A system which can reliably and effectively segment the ground plane from scenes has obvious benefits in these applications and I have recently been given a contact within one of these companies and I will request support for further development of this work (perhaps as a case studentship).

5 Explanation of Expenditure

The only significant variance in the original spending plans was in the use of the equipment money allocated for the SGI R12K O2 workstation (7.05K allocated). It came to our attention that we could purchase an SGI R10K O2 workstation for significantly less money (2.35K) with only a small compromise in performance relative to the R12K. With the remaining equipment money we purchased a laptop with network card (3.19K), which we could piggy back on our mobile robot and allow us to collect large amounts of video data to disk, which was analysed off line using the O2. Although data can be collected directly onto the O2, the use of a laptop allows data to be collected without moving the workstation. This is essential to collect a wide variety of scene data with which to test our algorithms. The laptop also functioned as a general machine for conference presentations, preparation of research papers and so on. We also purchased a standard video camera and framegrabber (1.36K) on the equipment budget so that we could video our mobile robot in action as the vehicle avoids obstacles using its ground plane segmentation algorithms.

6 Further Research and Dissemination Activities

We have not yet disseminated a significant result of our work, namely computation of an elation using correlation, and so we plan to publish one more conference paper. We will aim for CVPR 2003. In addition, I believe that the project has yielded sufficient new results, which have been validated over a number of long video sequences, to publish two top level journal papers, which, if successful, will bring the publication count for this project to 7 papers. In particular we plan to submit a paper to *The International Journal of Computer Vision*, thus disseminating our results to the computer vision community and a paper with a different emphasis to *IEEE Transactions on Robotics and Automation* thus disseminating our results to the robotics community. The planned time-scale for these is submission before the start of the next academic year (October 2002).

Once these journal papers are submitted, I plan to submit a proposal to the EPSRC to continue with this work in collaboration with Dr Huosheng Hu (University of Essex), where we will focus on a general multi-cue integration framework. This framework will run on PC, whilst the cues themselves will be extracted using a lower level layer of hardware (FP-GAs). If successful with this application, we hope to re-employ Dr Bojian Liang (currently employed on the DAME project within our department), who was the key member of staff which made this project a success.

References

1. Liang, B. and N. E. Pears, 2002, "Ground Plane Segmentation from Multiple Visual Cues", to appear in *Proc. Second Int. Conf. on Image and Graphics (SPIE series)*, August 16-18, Hefei, China.

2. Liang, B. and N. E. Pears, 2002, "Visual Navigation Using Planar Homographies", in *Proc. 2002 IEEE Int. Conf. on Robotics and Automation*, Washington, USA.

3. Pears N. E. and B. Liang, 2001, 'Ground plane segmentation for mobile robot visual navigation', in *Proceedings of the 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp 1513-1518.

4. Pears N. E., J. Austin, and J. McDermid, 2001, 'Multi-cue vision, novel architectures and high integrity concepts for dependable robots', *Proc. 1st IARP/IEEE-RAS Joint Workshop on the Technical Challenge for Dependable Robots in Human Environments*, paper IV-3

5. Criminisi A, I Reid and A. Zisserman, 2000, 'Single View Metrology', *International Journal of Computer Vision*, vol 40, number 2, pp 125-148.