

Interpreting the Structure of Single Images by Learning from Examples

Osian Haines

Department of Computer Science, University of Bristol, UK

Advisor: Andrew Calway

PhD thesis defended 3rd October 2013, University of Bristol

Received 1st March 2015; accepted 14th May 2015

Abstract

An important problem in computer vision is the interpretation of the content of a single image. A particularly challenging case is the extraction of the underlying 3D structure apparent in an image, where there is ambiguity due to the lack of depth information. Nevertheless, knowledge of the regular and predictable nature of the 3D world imposes constraints upon images, which can be used to recover basic structural information, by taking advantage of the fact that some 3D configurations are more likely than others. Our work in this area was inspired by the human visual system, which appears to have little difficulty in interpreting complex scenes from only a single viewpoint. Humans are thought to rely heavily on learned prior knowledge for this, and so we took a machine learning approach, and learned the relationship between appearance and scene structure from training examples.

Our work investigated this challenging area by focusing on the task of plane detection, which is important since planes are a ubiquitous feature of human-made environments. We developed a new plane detection method, which works by learning from labelled training data, and can find planes and estimate their orientation. This is done from a single image, without relying on explicit geometric information, nor requiring depth. This was the first work to show that accurate plane orientations could be recovered from single images without geometric structure such as vanishing lines.

The work comprised of two main parts: a plane recognition step to classify whether a given image region corresponds to a plane or not, and if so to estimate its orientation with respect to the camera; and a segmentation stage, which uses this to divide an image into regions corresponding to distinct planar surfaces.

Plane recognition was achieved by describing an image region using gradient and colour descriptors centred about salient points, and classifying against pre-labelled training data. Orientations of planar regions were then estimated using multivariate regression (learning was done with relevance vector machines [5]). For this we obtained good results compared to labelled ground truth obtained from a location independent of our training data – namely a mean classification accuracy of 92%, and a median orientation error of 11° . A histogram of errors is shown in Fig. 1, and example results are shown in Fig. 2 (see also [2]).

Correspondence to: <osian.haines@bristol.ac.uk>

Recommended for acceptance by Jorge Bernal

DOI <http://dx.doi.org/10.5565/rev/elcvia.729>

ELCVIA ISSN:1577-5097

Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

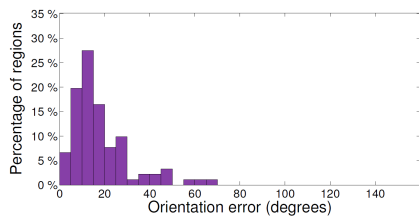


Figure 1: Errors in orientation estimation, for individual planar regions



Figure 2: Examples of correctly classified plane/nonplane regions, showing estimates of 3D orientation

This recognition algorithm formed the core of our plane detector, by applying it repeatedly to overlapping image regions to estimate plane likelihood across the image. This was used to initialise a Markov random field based segmentation, to first separate planar from non-planar regions, then to further segment the planar regions into individual planes with different orientations [1]. When evaluating on independent test data, 82% of points in the images were assigned to regions of the correct class, and the median orientation error for correctly detected planes was 16° . We show some example results in Fig. 3; many more can be seen in [2].

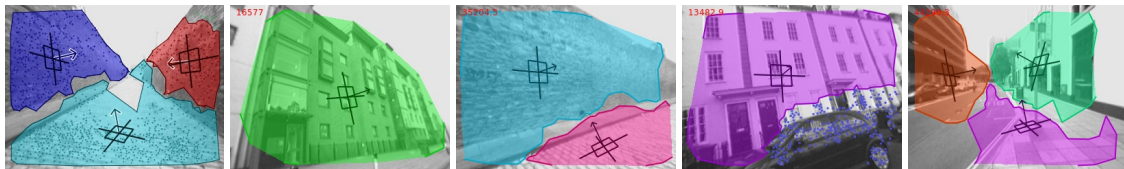
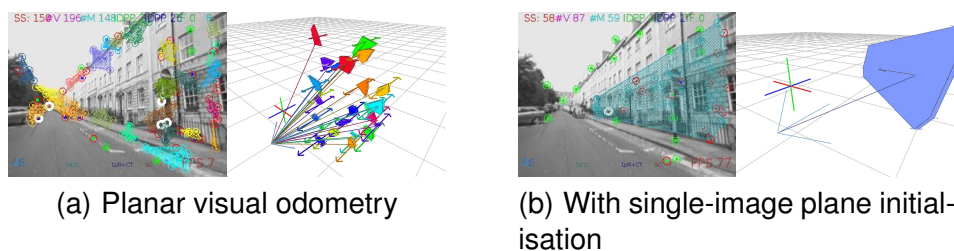


Figure 3: Examples of plane detection from a single image

Finally, we demonstrated an application of this plane detection algorithm to monocular visual odometry, i.e. the localisation of a moving camera in an unknown environment. This was done by enhancing a plane-based visual odometry system [4] – which previously required sufficient parallax before being able to find planes geometrically – to detect and initialise planes from one frame [3]. As shown in Fig. 4, while the baseline algorithm needed to investigate many candidate locations, our version allowed large planar regions to be quickly initialised in the 3D map from only one frame. This enabled rapid extraction of structured maps while exploring, and allowed the visual odometry to run at an increased frame-rate.



(a) Planar visual odometry

(b) With single-image plane initialisation

Figure 4: Using single image plane detection with a visual odometry system allows large planes to be initialised quickly (b), compared to searching based on geometry alone (a). On the right are corresponding 3D views.

References

- [1] O. Haines and A. Calway. Detecting planes and estimating their orientation from a single image. In *Proc. British Machine Vision Conf.*, 2012.

- [2] O. Haines and A. Calway. Recognising planes in a single image. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2014.
- [3] O. Haines, J Martínez-Carranza, and A Calway. Visual mapping using learned structural priors. In *Proc. IEEE Int. Conf. Robotics and Automation*, 2013.
- [4] J. Martínez-Carranza and A. Calway. Efficient visual odometry using a structure-driven temporal map. In *Proc. IEEE Int. Conf. Robotics and Automation*, 2012.
- [5] M.E. Tipping. Sparse bayesian learning and the relevance vector machine. *The Journal of Machine Learning Research*, 2001.