Detecting planes and estimating their orientation from a single image

Osian Haines
www.cs.bris.ac.uk/~haines

Andrew Calway
www.cs.bris.ac.uk/~andrew

Visual Information Laboratory
University of Bristol

We propose an algorithm to detect planes in a single image of an outdoor urban scene, capable of identifying multiple distinct planes, and estimating their orientation. Planar surfaces are ubiquitous in man-made environments, and convey useful information about scene structure; detecting them from a single image can be very useful since it requires no parallax, and can be done when only one image is available. Previous methods have shown how this can be used for estimating camera placement, wide-baseline matching, and creating simple 3D reconstructions. Such methods fall into two broad categories: those that explicitly use the geometric properties of the image, such as vanishing lines [6] or texture [3], and those that use machine learning techniques to relate appearance to structure [5] – our method is motivated by recent examples of the latter.

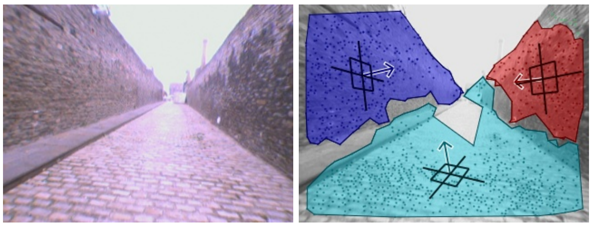
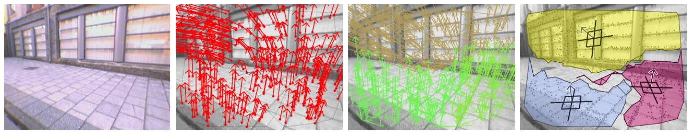


Figure 1: The detection of planes, and estimating their 3D orientation

Our goal is to group the salient points in an image into planar and non-planar regions, and give each plane an orientation estimate with respect to the camera – see figure 1. The method is inspired by humans' apparent ability to interpret scenes easily without explicit reference to geometric information, based on our prior visual experience; so we take a machine learning approach, trained on data gathered in an urban environment, to learn the relationship between appearance and structure. From a single image, we detect planes by sweeping a plane estimation algorithm [4] over the image, which is able to classify an image region as being planar or not, and estimate its orientation with respect to the camera. This is used to build an estimate of planarity and orientation at each of a set of salient points, which are segmented using Markov random fields to find distinct planar surfaces in the image.

The plane estimation algorithm is trained using data extracted from a set of manually marked-up images, in which the planar and non-planar regions are labelled according to their class, and plane ground-truth orientation is sepcified by marking four points corresponding to the corners of a rectangle lying on the plane in 3D. For each region, salient points are detected, which are described using gradient and colour feature descriptors – to compactly represent whole image regions we use the bag of words model, so features are quantised to two codebooks (for gradient and colour descriptors) created by clustering a set of example features. The two bags of words are combined using Orthogonal Non-negative Matrix Factorisation [2] (a variant of Latent Semantic Analysis), which measures the occurrence of latent topics for each region. The histogram of topics for an image region is combined with spatial distribution information from the 2D points to form a spatiogram [1] – a generalisation of a histogram containing the mean and covariance for each bin. These spatiograms are used to train Relevance Vector Machines (RVM), to classify new regions (plane or non-plane) and regress their orientation (normal vector in $\mathbb{R}^3$). Such a method was previously shown by Haines and Calway [4] to accurately identify planar regions and predict their orientation, but is not able to actually detect planes in an image, since their extent is unknown.

In order to detect planes, we sweep a window over the image, centred on each salient point in turn, applying the plane estimation at each location. The plane estimates from these overlapping windows are accumulated at the points, using the median plane probability from the RVM, and the geometric median of normal estimates, from all regions in which the point lies. The process results in an estimate of the probability of belonging to a plane, and the orientation, at each point (see figure 2(b)); while this appears to reflect the underlying structure it is not sufficient for



(a) (b) (c) (d)

Figure 2: For an image (a), we sweep a classifier over the image to obtain a point-wise local plane estimate (b). This is segmented into distinct regions (c), from which the final plane detections are derived (d)

plane detection since it does not tell us where the individual planes are; but it allows us to segment the points into distinct planes.

Segmentation happens in two stages, first to separate planes from non-planes, then into planes of different orientations. These are formulated as Markov random fields (MRF), on a graph formed from the salient points and the edges of a Delaunay triangulation, which are optimised using Iterative Conditional Modes. The MRF for separating planes from non-planes sets each point to be the most likely class – plane or not – given its observed probability and neighbours. Then, to segment the planar regions according to their orientation, we assume there are a finite number of planar surfaces, whose orientations correspond to the modes in a density estimate of all the observed normals, which are found with mean shift. Segmentation then consists of deciding to which of these discrete planes each point should belong; figure 2(c) shows the result.

Finally, the extracted planar segments are run through the plane estimation algorithm once more, to get a planarity and orientation estimate based only on the image area in question (figure 2(d)).

Our results, illustrated in figure 3, show that we can successfully detect planes, and estimate their orientation, in a variety of scenes – being able to separate planes from non-planar surroundings, and successfully disambiguate planes of different orientations. The mean classification error for individual points was 88%, and the mean orientation error for detected regions was 18.3°, compared to ground truth.

[1] Birchfield and Rangarajan. Spatiograms versus histograms for region-based tracking. In *CVPR*, 2005.

[2] Choi. Algorithms for orthogonal nonnegative matrix factorization. In *IJCNN*, 2008.

[3] Gårding. Direct estimation of shape from texture. *TPAMI*, 1993.

[4] Haines and Calway. Estimating planar structure in single images by learning from examples. *ICPRAM*, 2012.

[5] Hoiem, Efros, and Hebert. Recovering surface layout from an image. *IJCV*, 2007.

[6] Košecká and Zhang. Extraction, matching, and pose recovery based on dominant rectangular structures. *CVIU*, 2005.

Figure 3: Example results. Planar points are enclosed in coloured regions, displaying their orientation. Final image shows a failure case