An Accurate Method for Line Detection and Manhattan Frame Estimation

Ron Tal and James H. Elder

York University
{rontal,jelder}@yorku.ca

Abstract. We address the problem of estimating the rotation of a camera relative to the canonical frame of an urban scene, from a single image. Solutions generally rely on the so-called 'Manhattan World' assumption [1] that the major structures in the scene conform to three orthogonal principal directions. This can be expressed as a generative model in which the dense gradient map of the image is explained by a mixture of the three principal directions and a background process [2]. It has recently been shown that using sparse oriented edges rather than the dense gradient map leads to substantial gains in both accuracy and speed[3]. Here we explore whether further gains can be made by basing inference on even sparser extended lines. Standard Houghing techniques suffer from quantization errors and noise that make line extraction unreliable. Here we introduce a probabilistic line extraction technique that eliminates these problems through two innovations. First, we accurately propagate edge uncertainty from the image to the Hough map through a bivariate normal kernel that uses natural image statistics, resulting in a non-stationary 'soft-voting' technique. Second, we eliminate multiple responses to the same line by updating the Hough map dynamically as each line is extracted. We evaluate the method on a standard benchmark dataset [3], showing that the resulting line representation supports reliable estimation of the Manhattan frame, bettering the accuracy of previous edge-based methods by a factor of 2 and the gradient-based Manhattan World method by a factor of 5.

1 Introduction

The problem of single-view 3D reconstruction is of great practical interest, with potential applications in 3D mapping, 2D-to-3D film conversion, and robot navigation. The problem is in general ill-posed, but indoor scenes and outdoor urban scenes contain many regularities that can be exploited. One of these is the so-called 'Manhattan World' assumption [1] that the oriented structure of the scene is dominated by three mutually orthogonal 3D directions comprising the Manhattan frame. This constraint has been used in a number of interesting recent algorithms for single-view 3D reconstruction [4,5], and can in principle be used to recover, up to a scale factor, the overall 3D structure of a scene conforming to the Manhattan assumption. Thus accurate and reliable estimation of the Manhattan frame, i.e., the 3D rotation of the Manhattan scene relative to the

J.-I. Park and J. Kim (Eds.): ACCV 2012 Workshops, Part II, LNCS 7729, pp. 580–593, 2013.

camera, is a crucial problem in single-view 3D reconstruction, and is the problem we address here.

Coughlan and Yuille [2] devised a generative framework for the problem in which the dense gradient map is explained by a mixture of the three principal Manhattan directions and a random background process. A maximum-likelihood technique is then used to estimate the rotation of the Manhattan frame relative to the camera. Later, Denis et al. [3] adapted this framework to use sparse oriented edges rather than the dense gradient map. In using edges, they reduced the size of the input space by a factor of 10, making the algorithm much more efficient. Further, they argued that by eliminating redundancy, the naive Bayes model on which the framework is based becomes more accurate.

In this work, we take the next logical step and explore whether accuracy can be further improved by using a relatively small number of accurately extracted lines. This additional reduction in redundancy further improves the accuracy of the naive Bayes model. Moreover, since the Manhattan structure in the scene is typically organized as line segments rather than randomly positioned edges, this approach should improve the discrimination between Manhattan and non-Manhattan structures, and lead to more accurate estimation.

Our contribution is twofold:(1) An improved line extraction framework that uses more accurate, adaptive probabilistic kernels, and (2) a maximum likelihood method for estimating the pose of the camera with respect to the Manhattan frame. We evaluate the performance of our contributions on a standard benchmark dataset [3] and show that our method performs better than leading methods from the literature.

The standard method for extended line extraction is to histogram edges into Hough maps. However, this standard Hough approach is sensitive to bin resolution and suffers from noise and quantization error. To overcome these problems, we introduce a probabilistic line extraction technique that relies upon two innovations. First, we introduce a novel kernel-based voting scheme that accurately propagates observation uncertainty from the image domain, using natural statistics, into Hough space. Second, we eliminate multiple redundant detections by selecting lines in a greedy fashion and probabilistically subtracting the evidence for detected lines from the Hough map.

To relate these lines to the Manhattan frame, we adapt the Manhattan World mixture model to use line observations, replacing the image-based observation model relating gradient position and orientation to vanishing point locations, with an observation model on the Gauss sphere. Finally, we adapt the maximum likelihood technique used by Denis et al. [3] to our new observation model.

2 Overview of the Hough Transform

The standard Hough transform [6,7] uses the normal parametrization of a line in polar coordinates,

$$\rho = xcos(\theta) \ + \ ysin(\theta), \tag{1}$$

where ρ represents the signed distance of the line from the origin (the principal point) and $\theta \in (0, \pi]$ represents the angle of the normal relative to the image

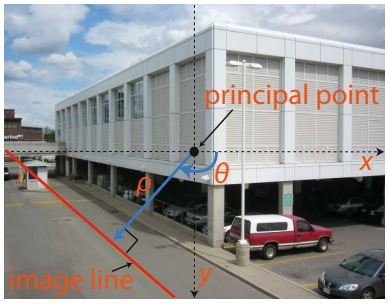


Fig. 1. The parametric representation of a line in the Hough transform

frame (ρ is positive if the point on the line closest to the origin lies in the lower half of the image, negative if it lies in the upper half: Fig. 1). The Hough transform involves the computation of a histogram $H(\rho, \theta)$ representing the image evidence for all possible lines passing through the image. Detection of an edge in the image can be taken as evidence for the existence of a line passing through the edge. If we neglect noise in the estimated position (x, y) of the edge, then the edge is consistent with the one-dimensional sinusoidal family of lines satisfying (1), and bins of $H(\rho, \theta)$ intersected by this sinusoid are incremented [7]. However, edge detection normally provides an estimate of edge orientation as well as its position, and it has long been recognized [8,9] that this orientation information can be used to limit the family of possible lines to a section of the sinusoid.

If edges were not noisy, the standard Hough transform would work well. Each oriented edge would map to a single bin in the Hough map, and all edges associated with the same line would map to the same bin. Due to noise in both edge position and orientation, edges associated with the same line will in practice map to different bins. This means that peaks in the Hough map will in general be displaced from the correct (ρ, θ) values, some peaks will be missing, and there will be many false peaks that do not correspond to real lines in the image. Further, the exact nature of these errors depends strongly on the histogram resolution: coarse resolutions lead to mislocalization and missed lines (due to merging of neighbouring lines), while fine resolutions lead to many false positives.

These problems have motivated research in probabilistic Houghing methods that map each edge to a smooth local distribution over the Hough parameters. The Hough map thus becomes a summation of overlapping smooth kernels, resulting in fewer false peaks. Since kernel dimensions are defined in absolute (ρ, θ) units, the number of false positives becomes independent of histogram resolution, permitting fine resolutions that allow precise localization. See [10] for a good survey of early probabilistic methods.

Stephens[11] modeled the parameters (position and orientation) of each edge as independent when conditioned upon line parameters. The probability of each edge given a hypothesized line is modeled as a mixture of a normal distribution, reflecting generation of the edge by the line, and a uniform distribution, reflecting

generation by some independent background process. A probabilistic Hough map is then computed by summing the log-likelihood over all edge observations. The proposed method is extremely computationally intensive, as it requires that the log likelihood be computed for every edge observation at every cell of the Hough map. No results on real data are reported.

Using only edge position, Kiryati & Bruckstein [12] reported a probabilistic Houghing method based on a bivariate normal model of edge position error, but demonstrated only limited results (detection of a single line from a single image). Li and Xie [13] detectd line endpoints as seed features and then paired these with randomly-selected edge points, using first-order error analysis to propagate uncertainty into the Hough map as a bivariate normal kernel. They also reported only limited, qualitative results.

Rather than using random edge pairs, Fernandes and Oliveira [14] formed local clusters of approximately collinear edges, computed a least-squares fit of a line to the cluster, estimated the uncertainty in the slope-intercept parameters of the line in the image, and propagated this uncertainty to the (ρ, θ) variables of the Hough map using first-order linear uncertainty propagation [15]. Each cluster thus contributes an oriented Gaussian kernel to the Hough map. The authors reported qualitative results on a wide array of images.

Barinova et al.[16] adopted an approach somewhat similar to Stephens'[11], modeling edge observations as conditionally independent, but explicitly introducing hidden association variables that tie edge observations to hypothesized lines. While they used both edge orientation and location information to generate soft votes in the Hough map, they employed a heuristically-derived fixed kernel rather than explicitly modelling error propagation from the image to the Hough domain. Perhaps the most interesting component of their approach is a greedy method for selecting associations once the initial Hough map is constructed. This results in a dynamically changing Hough map that avoids heuristic post-processing such as histogram smoothing or non-maxima suppression. This principle is implemented using probabilistic vote subtraction.

The probabilistic line detection technique we propose is in part inspired by these prior efforts. We use a kernel-voting technique to generate a smooth Hough map. As in [13,14], we employ first-order error propagation to determine the kernel in the Hough domain, and use a kernel subtraction technique related to the method of [16] to eliminate false positives without the need for post-processing steps like non-maximum suppression and smoothing.

However, our method also has several advantages in terms of effectiveness and efficiency. First, unlike [12], we use both edge position *and* orientation information, as both provide valuable constraints. Second, we avoid seed identification, clustering and line-fitting pre-processing steps [13,14], which increase the complexity of implementation and the number of parameters that must be tuned. Rather we propagate each observed edge independently, using learned image statistics and first-order error propagation. Unlike the fixed kernel of Barinova et al.[16], this leads to a voting kernel that varies dramatically over Hough space, better capturing the variation in uncertainty with edge parameters. We show that

these kernels can easily be precomputed and stored, making the voting process efficient. Unlike [11,16] we use a normal model rather than a mixture model for each conditional observation, and accumulate the sum of the likelihood as opposed to the log likelihood, so that the Hough map encodes the estimated length of hypothesized lines.

3 Proposed Framework

3.1 Propagating Observation Uncertainty to the Hough Domain

Edge detectors provide both position and orientation information. Thus, it makes sense to model the uncertainty of each observation with respect to both. That way, each edge can contribute a single bivariate normal (BVN) kernel to the Hough map (Fig. 2).

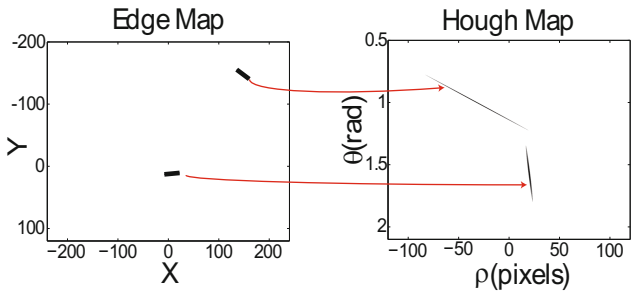
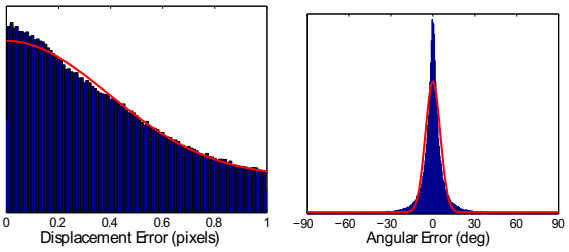


Fig. 2. Each edge votes according to a BVN kernel

We detect edges using a multiscale edge detector[17] that provides accurate sub-pixel edge position and orientation estimates, though our analysis is applicable to any edge detection algorithm such as Canny[18]. We model the uncertainty of edge observation by examining the statistics of image edges with respect to ground-truth lines[3]. Fig. 3 shows empirical densities for the displacement and orientation deviation of edges within one pixel of each line. We model each as a mixture of a normal distribution generated by the line and a uniform distribution generated by a background process. We assume isotropic displacement error, so that the data in Fig. 3 provide sufficient statistics to estimate the three key uncertainties σ_x, σ_y and σ_θ, i.e., the space constants of the normal models for horizontal and vertical displacement and angular error. The maximum likelihood estimates are: $\sigma_x = \sigma_y = 0.49$ pixels and $\sigma_\theta = 5.3$ deg. For comparison, we have also computed the same statistics for the popular Canny edge detector (MATLAB implementation with default parameters) and the space constants are: $\sigma_x = \sigma_y = 1.3$ pixels and $\sigma_\theta = 3.9$ deg.



(a) (b)

Fig. 3. Edge observation uncertainty with respect to ground-truth lines. Red curves show fit of two-component normal+uniform mixture model. (a) Distance of edge elements from ground-truth line. (b) Angular deviation of edge element from ground-truth line.

Approximating the deviations in these three dimensions as independent, we define the covariance of uncertainty of edge observations in the image domain as

$$C_I = \begin{bmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_x^2 & 0 \\ 0 & 0 & \sigma_\theta^2 \end{bmatrix}, \tag{2}$$

where we use σ_x for both horizontal and vertical displacement. Using linear propagation of uncertainty[15], the covariance of the corresponding parameters in the Hough domain can be computed as

$$C_h(x, y, \theta) = \nabla \boldsymbol{P}_h C_I \nabla \boldsymbol{P}_h^T \tag{3}$$

where $\nabla \boldsymbol{P}_h$ is the Jacobian of the parameter vector with respect to the observation vector

$$\nabla \boldsymbol{P}_h = \begin{bmatrix} \frac{\partial \rho}{\partial x} & \frac{\partial \rho}{\partial y} & \frac{\partial \rho}{\partial \theta} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} & \frac{\partial \theta}{\partial \theta} \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) & \frac{\partial \rho}{\partial \theta} \\ 0 & 0 & 1 \end{bmatrix}$$

and

$$\frac{\partial \rho}{\partial \theta} = -x \sin(\theta) + y \cos(\theta). \tag{4}$$

Thus, (3) becomes

$$C_h(x, y, \theta) = \begin{bmatrix} \sigma_x^2 + \left(\frac{\partial \rho}{\partial \theta}\right)^2 \sigma_\theta^2 & \frac{\partial \rho}{\partial \theta} \sigma_\theta^2 \\ \frac{\partial \rho}{\partial \theta} \sigma_\theta^2 & \sigma_\theta^2 \end{bmatrix}. \tag{5}$$

The vote each edge observation contributes at each location in the Hough map is described by the BVN distribution

$$\text{Vote}_i(\rho,\theta|x_i,y_i,\theta_i) \;=\; \frac{1}{2\pi|C_h|}\exp\left(-\frac{1}{2}\begin{bmatrix}\rho-\rho_i\\ \theta-\theta_i\end{bmatrix}^T C_h^{-1}\begin{bmatrix}\rho-\rho_i\\ \theta-\theta_i\end{bmatrix}\right), \qquad (6)$$

where ρ_i and θ_i correspond to a direct mapping of the i^{th} edge onto Hough space. The Hough map $H(\rho,\theta)$ is the sum of these votes at each location due to all observations

$$H(\rho,\theta) \;=\; \sum_i \text{Vote}_i(\rho,\theta|x_i,y_i,\theta_i). \qquad (7)$$

Note that while σ_x and σ_θ are constants, $\partial\rho/\partial\theta$ varies over the image: for our images, $\partial\rho/\partial\theta$ ranges from roughly -400 to 400. (Note that this dramatic variation underlines the importance of modelling the first-order error propagation rather than using a fixed kernel.) We sample $\partial\rho/\partial\theta$ in increments of 1 over this range, computing and storing the voting kernel given by (6) for each value, truncating at 3σ. Then when constructing the Hough map, for each edge observation (x_i,y_i,θ_i), $\partial\rho/\partial\theta$ is computed, the nearest stored kernel is selected, shifted to the corresponding (ρ_i,θ_i) location in the Hough map, and added to $H(\rho,\theta)$.

3.2 Peak Selection

After a Hough map is constructed, the peaks that correspond to dominant line features need to be identified. Ideally, image lines correspond to the modes of H. However, due to the limited sample of edges in the image, spurious local maxima will occur despite the kernel voting scheme, especially within the neighborhood of true peaks. Typically, this problem is addressed by either post-hoc smoothing [19] or non-maxima suppression. However, both approaches involve the ad-hoc selection of parameters, and in practice we find neither method works well.

Inspired by the recent dynamic updating approach of [16], we propose a simple, iterative, greedy technique, in which the kernel contributions of edges that can be associated with a detected line are subtracted from the Hough map. At each iteration, the global maximum of H is determined and is added to the list of selected lines. All edges that fall within a 3σ bound in terms of both distance from the line and angular deviation are identified. The probabilistic kernels for all of the identified edges are then subtracted from H (Fig. 4). This procedure is repeated until $\max(H)$ is smaller than a specified threshold which is a fraction of the global maximum of the Hough map. Using 5-fold cross-validation on the training images, we found that a threshold of 0.25, resulting in detection of 36.5 lines per image on average, yields minimum error for our Manhattan frame estimation application.

3.3 Manhattan Frame Estimation

We adapt the probabilistic mixture model used by Coughlan and Yuille [1,2] and Denis et al.[3] to observations consisting of extended lines. Each observed

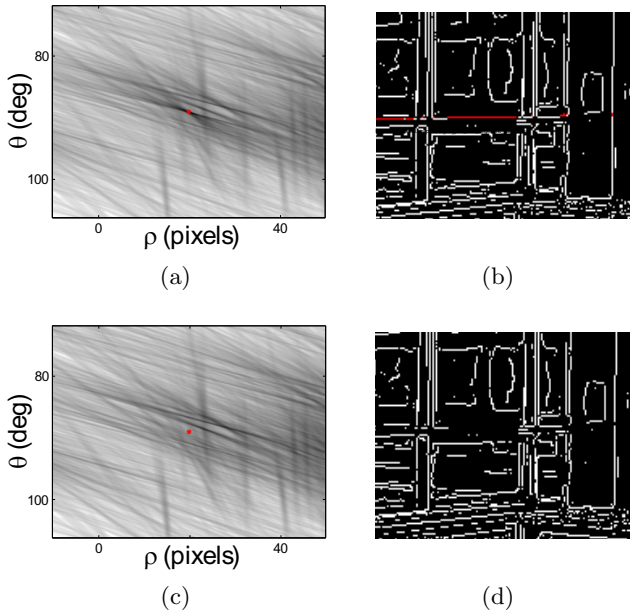


Fig. 4. Iterative peak detection with vote removal (a) Hough map with global maxima shown in red. (b) Corresponding edge map. Edges associated with the detected peak are shown in red. (c) Updated Hough map after kernel subtraction (d) Residual edge map.

line $\boldsymbol{\ell}_i$ is assumed to be generated by a latent Manhattan variable $m_{\boldsymbol{\ell}_i}$ that can assume values representing four classes of structure: vertical, horizontal(1), horizontal(2), background. If we knew the Euler angles Ψ describing the rotation of the camera with respect to the Manhattan frame of the scene, we could compute the probability of observing the line as a mixture of the four possible causes

$$p(\boldsymbol{\ell}_i|\Psi) \;=\; \sum_{m_{\boldsymbol{\ell}_i}} p(\boldsymbol{\ell}_i|\Psi, m_{\boldsymbol{\ell}_i}) p(m_{\boldsymbol{\ell}_i}) \tag{8}$$

where $p(m_{\boldsymbol{\ell}_i})$ is the prior probability for the latent variable $m_{\boldsymbol{\ell}_i}$ and $p(\boldsymbol{\ell}_i|\Psi, m_{\boldsymbol{\ell}_i})$ is the probability of a line conditioned on $m_{\boldsymbol{\ell}_i}$. Next, we provide a statistical model for $p(\boldsymbol{\ell}_i|\Psi)$.

Previous gradient- and edge-based methods [2,3] used an error model that relates localized, oriented image features to a set of vanishing points. Lines, however, are infinite in length and so cannot be related to vanishing points in the same way. Instead, it is natural to employ a Gauss sphere representation [20] (Fig. 5(a)). A line in the image plane, together with the optical centre of the camera, form an interpretation-plane that can be represented by its normal vector $\boldsymbol{\ell}$. In a noise-free world, all lines conforming to the same 3D orientation will produce interpretation plane normals that are coplanar. Thus, the error of

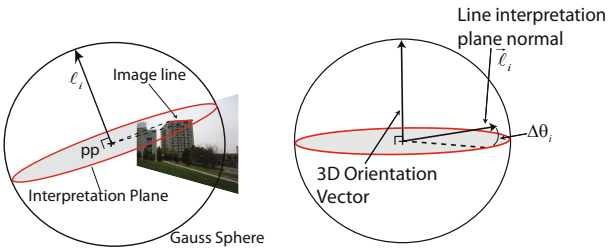


Fig. 5. Modeling uncertainty on the unit circle. (a) A line detected in the image can be represented in the Gauss sphere by its interpretation plane normal. (b) Error model for a line in the Gauss sphere.

an estimated line with respect to its true 3D orientation is given by the angular deviation of the interpretation-plane normal for the line from this common plane. The normal of this plane corresponds to the 3D orientation of the lines. Denis et al. [3] found that a Gauss-sphere error model was less accurate than an image model for edge primitives. However, line primitives have the potential to be much more accurate, as they integrate over many edges, and we find that for lines, the Gauss-sphere model works well.

We develop a statistical model for $p(\boldsymbol{\ell}_i|\Psi, m_{\boldsymbol{\ell}_i})$ by considering hand-labeled ground-truth lines from the database[3]. The error $\Delta\theta_i$ is the angular deviation of the interpretation plane normal for a line $\boldsymbol{\ell}$ from the plane normal to the 3D Manhattan orientation vector (Fig. 5(b)). A histogram of $\Delta\theta_i$ indicates that a Laplace model for $p(\boldsymbol{\ell}_i|\Psi, m_{\boldsymbol{\ell}_i})$ is suitable (Fig. 6). Thus,

$$p(\boldsymbol{\ell}_i|\Psi, m_{\boldsymbol{\ell}_i}) \;=\; \frac{1}{2b} e^{\frac{-|\Delta\theta_{m_{\boldsymbol{\ell}_i}}|}{b}} \tag{9}$$

when $m_{\boldsymbol{\ell}_i}$ is not background, where $b = 0.80$ deg for horizontal lines and $b = 0.57$ deg for vertical lines.

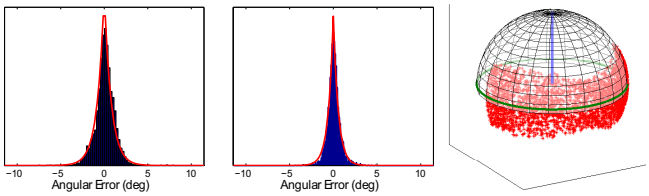


Fig. 6. Distribution of line interpretation plane normals with respect to their associated (a) horizontal and (b) vertical Manhattan direction. Maximum likelihood Laplace model shown in red. (c) Sample of uniform distribution of lines observable in the image.

We make the assumption that lines produced by the background process are uniformly distributed over the Gauss Sphere. However, to be observable, a section of the line must fall within the subtense of the image. This leads to a distribution that is roughly uniform over a rectangular region of the angular coordinates of the Gauss Sphere (Fig. 6(c)). For the camera parameters associated with the images in the YorkUrbanDB database, this leads to the uniform likelihood $p(\boldsymbol{\ell}_i|m_{\boldsymbol{\ell}_i} = B) = \frac{1}{0.68\pi}$.

The association priors $p(m_{\boldsymbol{\ell}_i})$ may depend on the line extraction algorithm and its parameters, and thus cannot be determined using the ground-truth data. The association priors are therefore learned independently for each line extraction algorithm and parameter setting, by maximizing the likelihood of the mixture model over the training data.

The next stage is to estimate the 3D orientation of the Manhattan frame given a set of line observations $\mathbb{L} = \{\boldsymbol{\ell}_i\}_{1\ldots N}$ and a model for their association. The likelihood of the relative rotation of the Manhattan frame with respect to the camera is given by the product of the probabilities of observing all lines

$$p(\mathbb{L}|\Psi) = \prod_i p(\boldsymbol{\ell}_i|\Psi) \tag{10}$$

so that

$$\hat{\Psi} = \arg\max_{\Psi} \sum_i \log p(\boldsymbol{\ell}_i|\Psi). \tag{11}$$

Kosecka and Zhang [21] and Schindler and Dellaert [22] have used the EM algorithm to estimate $\hat{\Psi}$. This works well when 3D orientations are independent. However, satisfying the Manhattan constraint (orthogonality of the 3 Manhattan directions) makes the M-step a constrained optimization with no closed-form solution. Thus applying EM requires gradient descent on each M-step, increasing computational cost and the risk of missing the global optimum. For these reasons, Denis et al.[3] found empirically that a search based on the standard iterative BFGS quasi-Newton algorithm[23] performs better than EM on the Manhattan problem, and we apply the same method here.

4 Results

4.1 Line Detection

We evaluate our extended line detection algorithm on the standard, publicly-available YorkUrbanDB test dataset [3]. We define a correct detection as an extended line passing within 3 pixels of both endpoints of a ground truth line segment in the dataset. We compute a 1:1 greedy bipartite match between lines and ground truth segments: additional lines matching the same segment are considered false positives. Fig. 7 shows that precision-recall performance for our method compares favourably to a recent state-of-the-art method for line detection [16]. Although one may be tempted to ask whether this improved accuracy

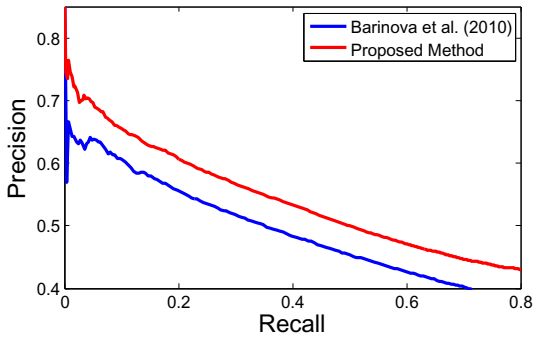


Fig. 7. Quantitative evaluation on the York UrbanDB test dataset (51 images) showing Precision recall plot for the proposed extended line detection method vs. Barinova et al.[16].

should be attributed to the improved localization and detection of edges, we find that when supplying our line detection algorithm with edge estimates obtained from the Canny edge detector used by [16], our proposed method still outperform theirs. Our experiments show that in fact the superior performance of our method is due primarily to its lower susceptibility to the multiple response problem, likely deriving from a faithful representation of the underlying statistics and more accurate propagation of uncertainty to the Hough domain.

4.2 Manhattan Frame Estimation

We evaluate our probabilistic line-based method for Manhattan Frame estimation on the YorkUrbanDB benchmark database [3], comparing against previous gradient-based [2] and edge-based [3] methods, as well as a standard non-probabilistic Hough method implemented in-house. We note that while algorithms for vanishing point detection have been published more recently (e.g., [24,25]), these have not specifically addressed the problem of estimating the Manhattan frame, and our own attempts to adapt these methods to this problem have led to poor results (average error of 19 deg).

The parameters for the line-based method were selected using a process of cross-validation, in which the training set was divided using a random 50-50 split: the Hough transform parameters were tuned and the association priors were learned on one subset and evaluated on the other. This process was repeated over 20 trials, and the configuration that produced the lowest mean error on the training data was used for final evaluation on the test set. With optimized parameters, our probabilistic line-based method uses 36.5 lines per image, on average, to estimate the Manhattan frame.

We measure performance by the angle between the estimated Manhattan frame and ground truth. Evaluation on the test set (Fig. 8) shows that our probabilistic Hough method, incorporating accurate non-stationary kernels and

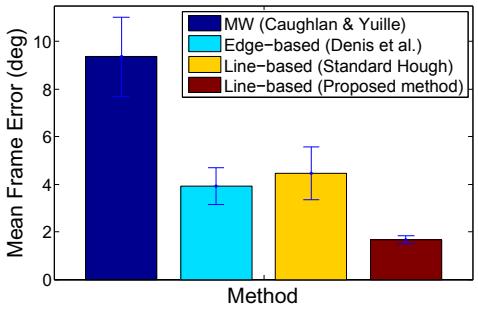


Fig. 8. Mean error of estimated camera pose relative to ground truth

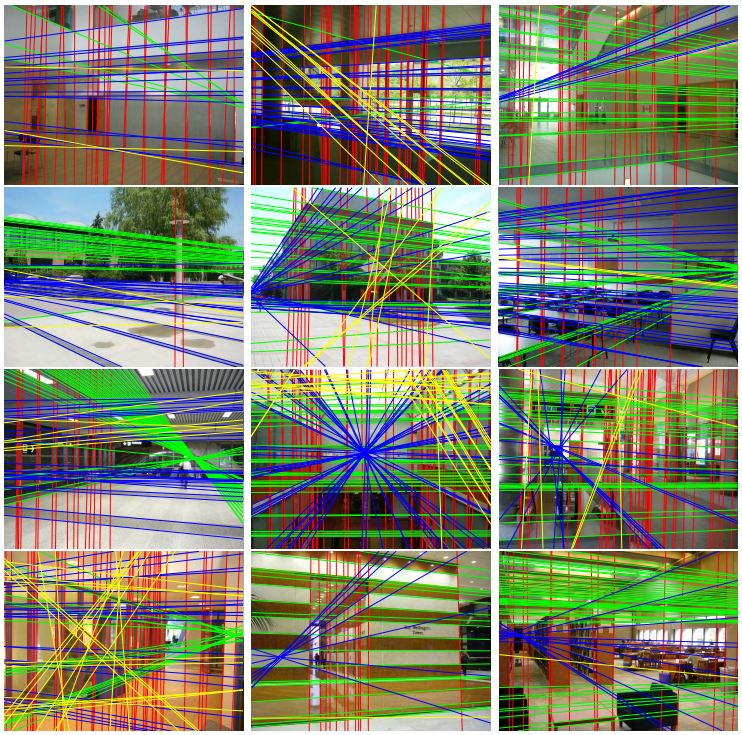


Fig. 9. Examples of automatically detected lines and association with Manhattan frame. All lines used to estimate the Manhattan frame are shown. Lines are colour coded according to the most probable Manhattan cause (red: vertical, green & blue: horizontal, yellow: background).

dynamic kernel subtraction, increases accuracy by a factor of more than 2 over prior methods, achieving an average frame error of 1.7 degrees. We provide a wide range of images to demonstrate the effectiveness of our contribution in both accurate line detection and robust Manhattan frame estimation in Fig. 9.

At roughly 8 s/image, our probabilistic line detection method is much faster than the original gradient-based Manhattan World method [2] (22 s/image), and comparable to the edge-based method of Denis et al. [3] (5 s/image). (All times based upon a 3.0 GHz Intel Core Duo CPU.) Running time is dominated by the time required to add and subtract kernels from the Hough map: these operations could be greatly accelerated if mapped to the GPU.

5 Conclusion

We have developed a novel probabilistic, dynamic line-based method for estimating the Manhattan frame of a scene, and shown it to be more than twice as accurate as the state-of-the-art, using a standard publicly-available benchmark dataset. Importantly, we have shown that standard line extraction methods are insufficient: achieving this improvement depends crucially on novel methods for a) accurate propagation of uncertainty in edge detection to the Hough domain, and b) dynamic updating of the Hough map to reflect inferred edge-line associations.

The superiority of our line-based method over the prior edge-based state-of-the-art [3] likely reflects a) greater accuracy of the naive Bayes mixture model for the less redundant and noisy line primitives and b) inferential leverage deriving from the tendency for orientation energy to be organized more as lines in the Manhattan structures and more as texture in the background.

References

1. Coughlan, J.M., Yuille, A.L.: Manhattan world: Compass direction from a single image by Bayesian inference. In: International Conference on Computer Vision (1999)
2. Coughlan, J.M., Yuille, A.L.: Manhattan world: Orientation and outlier detection by Bayesian inference. Neural Computation 15, 1063–1088 (2003)
3. Denis, P., Elder, J.H., Estrada, F.J.: Efficient Edge-Based Methods for Estimating Manhattan Frames in Urban Imagery. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 197–210. Springer, Heidelberg (2008)
4. Yu, S.X., Zhang, H., Malik, J.: Inferring spatial layout from a single image via depth-ordered grouping. In: CVPR Workshop (2008)
5. Hedau, V., Hoiem, D., Forsyth, D.: Recovering the spatial layout of cluttered rooms. In: ICCV 2009, pp. 1849–1856 (2009)
6. Hough, P.V.C.: Method and Means for Recognizing Complex Patterns. U. S. Patent 3, 069, 654 (1962)
7. Duda, R.O., Hart, P.E.: Use of the Hough transformation to detect lines and curves in pictures. Communications of the ACM 1, 11–15 (1972)
8. O'Gorman, F., Clowes, M.: Finding picture edges through collinearity of feature points. IEEE Transactions on Computers C-25, 449–456 (1976)
9. Ballard, D.: Generalizing the hough transform to detect arbitrary shapes. Pattern Recognition 13, 111–122 (1981)

10. Klviinen, H., Hirvonen, P., Xu, L., Oja, E.: Probabilistic and non-probabilistic Hough transforms: overview and comparisons. Image and Vision Computing 13, 239–252 (1995)
11. Stephens, R.: Probabilistic approach to the Hough transform. Image and Vision Computing 9, 66–71 (1991)
12. Kiryati, N., Bruckstein, A.M.: Heteroscedastic Hough transform (HtHT): an efficient method for robust line fitting in the 'errors in the variables' problems. Comput. Vis. Image Underst. 78, 69–83 (2000)
13. Li, Q., Xie, Y.: Randomised Hough transform with error propagation for line and circle detection. Pattern Analysis and Applications 6, 55–64 (2003)
14. Fernandes, L.A.F., Oliveira, M.M.: Real-time line detection through and improved Hough transform voting scheme. Pattern Recognition 41, 299–314 (2008)
15. Bevington, P.R.: Data Reduction and Error Analysis for the Physical Sciences. McGraw-Hill (1969)
16. Barinova, O., Lempitsky, V., Kohli, P.: On detection of multiple object instances using Hough transforms, pp. 2233–2240 (2010)
17. Elder, J.H., Zucker, S.W.: Local scale control for edge detection and blur estimation. Transactions on Pattern Analysis and Machine Intelligence 20, 699–716 (1998)
18. Canny, J.: A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8, 679–698 (1986)
19. Bishop, C.M.: Pattern Recognition and Machine Learning, 1st edn. Springer (2006)
20. Barnard, S.T.: Interpreting perspective images. Artificial Intelligence 21, 435–462 (1983)
21. Košecká, J., Zhang, W.: Video Compass. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part IV. LNCS, vol. 2353, pp. 476–490. Springer, Heidelberg (2002)
22. Schindler, G., Dellaert, F.: Atlanta world: An expectation maximization framework for simultaneous low-level edge grouping and camera calibration in complex man-made environments. In: CVPR 2004, pp. 203–209 (2004)
23. Avriel, M.: Nonlinear Programming: Analysis and Methods. Prentice Hall (2003)
24. Tardif, J.P.: Non-iterative approach for fast and accurate vanishing point detection. In: ICCV 2009 (2009)
25. Barinova, O., Lempitsky, V., Tretiak, E., Kohli, P.: Geometric Image Parsing in Man-Made Environments. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part II. LNCS, vol. 6312, pp. 57–70. Springer, Heidelberg (2010)