Manhattan World: Orientation and Outlier Detection by Bayesian Inference

James M. Coughlan A.L. Yuille *

Smith-Kettlewell Eye Research Institute

2318 Fillmore St.

San Francisco, CA 94115

coughlan@ski.org, yuille@ski.org. Tel 415 345-2146 (2144). Fax 415 345-8455.

Abstract

This paper argues that many visual scenes are based on a "Manhattan" three-dimensional grid which imposes regularities on the image statistics. We construct a Bayesian model which implements this assumption and estimates the viewer orientation relative to the Manhattan grid. For many images, these estimates are good approximations to the viewer orientation (as estimated manually by the authors). These estimates also make it easy to detect outlier structures which are unaligned to the grid. To determine the applicability of the Manhattan world model we implement a null hypothesis model which assumes that the image statistics are independent of any three dimensional scene structure. We then use the log-

likelihood ratio test to determine whether an image satisfies the Manhattan world assumption. Our results show that if an image is estimated to be Manhattan then the Bayesian model's estimates of viewer direction are almost always accurate (according to our manual estimates), and vice versa.

* A.L. Yuille's new address is: Prof. Alan L. Yuille, Department Statistics and Psychology, UCLA, 7461D Franz Hall, Los Angeles, CA 90095-1563, Tel: 310 267-5383, Fax: 310 206-5895, Email: yuille@stat.ucla.edu.

1 Introduction

In recent years, there has been growing interest in the statistics of natural images (see Huang and Mumford 19990 for a recent review). Much of the interest in these statistics lies in their usefulness for quantitatively describing the regularities of images. Image statistics are also useful, however, for solving visual inference problems. They can be used to design statistical edge detectors (Konishi, Yuille, Coughlan, Zhu 1999, 2003), to determine statistical image invariants (Chen, Belhumeur, Jacobs 2000), and to determine semantic categories (Oliva and Torrabla 2001). It is plausible that the receptive fields of neurons, adapt to these statistics and exploit them for making inferences about the world, see (Balboa and Grzywacz 2000).

This paper investigates the Manhattan world assumption (Coughlan and Yuille 1999, 2000) where statistical image regularities arise from the geometrical structure of the scene being viewed. It assumes that the scene has a natural cartesian x, y, z coordinate system and the image statistics will be determined by the alignment of the viewer with respect to this system. The Manhattan world assumption

is plausible for indoor and outdoor city scenes. But, as we will show, it also applies to some scenes in the country and even to some paintings.

Informal evidence that human observers use a form of the Manhattan world assumption is provided by the Ames room illusion, see Figure (1), where the observers appear to erroneously make this assumption, thereby perceiving the sizes of objects in the room to be grotesquely distorted. The Ames room is actually constructed in a shape that strongly *violates* the Manhattan assumption but human observers, and our model (see top row of Figure (18)), interpret the room as if it had a Cartesian structure.

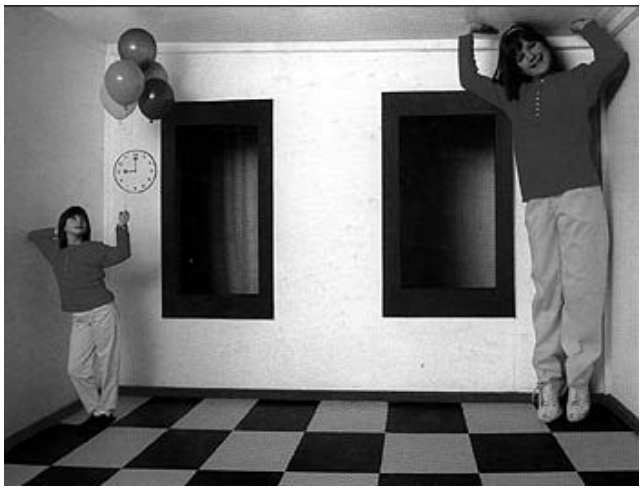

Figure 1: The Ames room, a geometrically distorted room that nevertheless appears rectangular from a special viewpoint. Despite appearances, the two people are the same size.

In particular, we demonstrate a Bayesian statistical model that exploits the Manhattan-world assumption to determine the orientation of the viewer relative to the grid. This enables us to to determine the orientation of the viewer in a scene, indoor or outdoor, from a single image. It might, for example, be used as part of the "reptilian layer" of a vision system in accordance with the active

vision philosophy (Blake and Yuille 1992). It gives a calibration method that is an alternative to well established techniques in computer vision based on calibration patterns (Faugeras 1993) or motion flow (Hartley and Zisserman 2000). It is related to methods for calibration by estimating vanishing points, see review in (Deutscher, Isard, and MacCormack 2002) but these methods require multiple images or manual labelling of detected lines. The Bayesian model also allows us to detect outlier edges which are not aligned to the dominant structures in the scene and which may simplify object detection.

We evaluate our model by comparing its estimates of the viewer orientation with estimates made manually by the authors (see (Deutscher, Isard, and MacCormack 2002) for recent comparisons of Manhattan-type algorithms to alternative methods and groundtruth). This is demonstrated by figures in the text (enabling the reader to make his/her own judgement). In addition, we construct a null hypothesis model which assumes that the image statistics are independent of the three-dimensional scene structure. This enables us to determine if an image obeys the Manhattan world assumption by comparing the evidence of the Manhattan and the null hypothesis model.

This paper is organized as follows. In Sections (2) and (3) we describe the geometry of the problem and the connection between three-dimensional structures in the scene and the corresponding properties of the image that are measured. Section (4) describes our statistical model of images. Section (5) describes the full Bayesian model of Manhattan world and Section (6) shows experimental results applying this model to a range of images and uses the null hypothesis model to estimate if an image is Manhattan. In Section (7) we describe the application of the Manhattan model to finding outlier objects unaligned to the Manhattan

grid. Section (8) presents a consistency check of the accuracy of the prior model. Finally, Section (9) summarizes the paper.

2 The Geometry of the Problem

There has been an enormous amount of work studying the geometry of computer vision (Faugeras 1993, Mundy and Zisserman 1992). Techniques from projective geometry have been applied to finding the vanishing points (Brillault-O'Mahony 1991, Lutton, Maître, Lopez-Krahe 1994); see (Shufelt 1999) for a recent review and analysis of these techniques. This work, however, has typically proceeded through the stages of edge detection, Hough transforms, and finally the calculation of the geometry. Alternatively, a sequence of images over time can be used to estimate the geometry, see for example (Torr and Zisserman 1998). In this paper, we demonstrate that accurate results can be obtained from a single image directly without the need for techniques such as edge detection and Hough transforms, by exploiting the statistical regularities of scenes.

For completeness, we give the basic geometry. The camera orientation is defined by a set of camera axes aligned to the camera body which has been rotated relative to the Manhattan xyz axis system. We define the camera axes by three unit vectors $\vec{a}$, $\vec{b}$ and $\vec{c}$. $\vec{a}$ and $\vec{b}$ specify the horizontal and vertical directions, respectively, along the film plane in xyz coordinates. $\vec{c}$ is the unit vector that points along the line of sight of the camera, so that $\vec{a} \times \vec{b} = -\vec{c}$. The projection from a 3-d point $\vec{r} = (x, y, z)$ to 2-d film coordinates $\vec{u} = (u, v)$ (centered at the physical center of the film) is given by:

$$u = \frac{f\vec{r} \cdot \vec{a}}{\vec{r} \cdot \vec{c}}, v = \frac{f\vec{r} \cdot \vec{b}}{\vec{r} \cdot \vec{c}}. \tag{1}$$

where f is the focal length of the camera. Here the film coordinates are chosen such that the u-axis is aligned to $\vec{a}$ and the v-axis is aligned to $\vec{b}$.

The camera orientation relative to the Manhattan xyz axis system may be specified by three Euler angles α, β, and γ. We can think of starting with the vectors $\vec{c}$, $\vec{a}$ and $\vec{b}$ aligned to the $x, -y, z$ axes and applying three successive rotations to the coordinate frame defined by these vectors (i.e. active transformations rather than passive coordinate transformations). The first angle is the *compass angle* (or *azimuth*) α, which rotates the camera about the z axis, yielding a transformed coordinate system $x'y'z'$. Next, the camera is rotated around the y' axis by the *elevation angle* β, yielding the next coordinate system $x''y''z''$. This has the effect of elevating the line of sight from the xy plane. Finally, a *twist angle* γ applies a rotation γ about the x'' axis, producing a twist in the plane of the film. We use $\vec{\Psi}$ to denote all three angles (α, β, γ) of the camera orientation. (In previous work we made the assumption that $\beta = \gamma = 0$ and only allowed the compass angle α to vary, which was a reasonable approximation for many images in our database.)

To derive expressions for $\vec{a}$, $\vec{b}$ and $\vec{c}$ as functions of $\vec{\Psi}$ we follow a standard procedure (Mathews and Walker 1970) of constructing rotation matrices $R_x(\gamma), R_y(\beta), R_z(\alpha)$ and defining the overall rotation matrix $R(\vec{\Psi}) = R_{x''}(\gamma)R_{y'}(\beta)R_z(\alpha)$ $= R_z(\alpha)R_y(\beta)R_x(\gamma)$. The resulting expression is $R(\vec{\Psi})$ given by

$$\begin{pmatrix} \cos\alpha\cos\beta & -\sin\alpha\cos\gamma + \cos\alpha\sin\beta\sin\gamma & \sin\alpha\sin\gamma + \cos\alpha\sin\beta\cos\gamma \\ \sin\alpha\cos\beta & \cos\alpha\cos\gamma + \sin\alpha\sin\beta\sin\gamma & -\cos\alpha\sin\gamma + \sin\alpha\sin\beta\cos\gamma \\ -\sin\beta & \cos\beta\sin\gamma & \cos\beta\cos\gamma \end{pmatrix} \tag{2}$$

We then obtain that $\vec{a} = R(\vec{\Psi})(0,-1,0)^T$, $\vec{b} = R(\vec{\Psi})(0,0,1)^T$ and $\vec{c} = R(\vec{\Psi})(1,0,0)^T$.

3 Mapping Manhattan Lines to Image Gradients

Straight lines in the x, y, z directions in the Manhattan world project to straight lines in the image plane. Our Bayesian model, see section (5), will use estimates of the orientations of these lines on the image plane, provided by image gradient information, to find the camera orientation which is most consistent with these orientation measurements. We now derive the orientation in the image plane of an x, y, or z line projected at any pixel location (u, v) as a function of camera orientation $\vec{\Psi}$. The calculation proceeds by first calculating the x, y, z vanishing point positions in the image plane as a function of $\vec{\Psi}$. The resulting (u, v) coordinates for the x, y, z vanishing points are $(fa_x/c_x, fb_x/c_x)$, $(fa_y/c_y, fb_y/c_y)$ and $(fa_z/c_z, fb_z/c_z)$, respectively.

Next, it is a straightforward calculation to show that a point in the image at $\vec{u} = (u, v)$ with intensity gradient direction $(\cos\theta_x, \sin\theta_x)$ is *consistent with an x line in the sense that it points to the vanishing point* if $\tan\theta_x = (uc_x - fa_x)/(fb_x - vc_x)$. This calculation is for an ideal edge for which the intensity gradient direction $(\cos\theta_x, \sin\theta_x)$ points exactly perpendicularly to the x vanishing point; our Bayesian model will exploit the fact that the orientation of true intensity gradients fluctuates about the ideal direction (see Section (5)) by modeling these fluctuations statistically. Observe also that this equation is unaffected by adding $\pm\pi$ to θ_x and so it does not depend on the polarity of the edge. We get similar expressions for y and z lines: $\tan\theta_y = (uc_y - fa_y)/(fb_y - vc_y)$ for intensity gradient

direction $(\cos\theta_y, \sin\theta_y)$ and $\tan\theta_z = (uc_z - fa_z)/(fb_z - vc_z)$ for intensity gradient direction $(\cos\theta_z, \sin\theta_z)$. See Figure 2 for an illustration of this geometry.

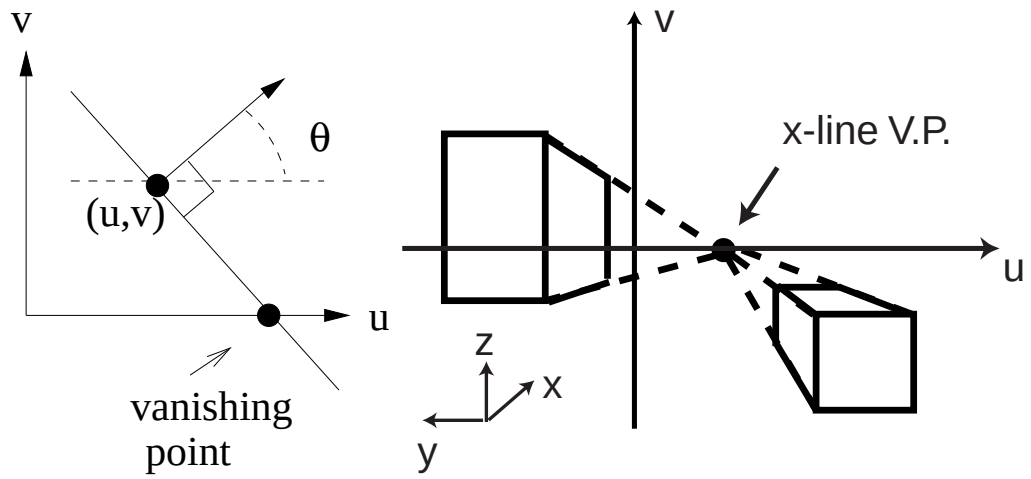

Figure 2: Examples of projection. Left Panel: Geometry of an x line projected onto the (u, v) image plane. θ is the normal orientation of the line in the image. Right Panel: the vanishing point due to two square boxes aligned to the Manhattan grid. In both cases the camera is assumed to point in a horizontal direction, and so the x vanishing point lies on the u axis.

4 P_{on} and P_{off}: Characterizing Edges Statistically

A key element of our approach is that we do not use a binary edge map. Such edge maps make premature decisions based on too little information. The poor quality of some of the images we used – underexposed and overexposed – makes edge detection particularly difficult. Our algorithm showed significant decrease in performance when we adapted it to run on edge maps unless the edge detection threshold is varied from image to image.

Instead we use the power of statistics. Following work by Konishi *et al.* (Konishi, Yuille, Coughlan, Zhu 1999, 2003), we determine probabilities $P_{on}(E_{\vec{u}})$ and $P_{off}(E_{\vec{u}})$ for the probabilities of the response $E_{\vec{u}}$ of an edge filter at position $\vec{u}$ in the image *conditioned on whether we are on or off an edge*. These distributions were learned by Konishi *et al* for the Sowerby image database which contain one hundred presegmented images (see (Konishi, Yuille, Coughlan, Zhu 1999, 2003) for the similarity of these statistics from image to image). The more different P_{on} is from P_{off} then the easier edge detection becomes, see Figure 3. A suitable measure of difference is the Chernoff Information (Cover and Thomas 1991) $C(P_{on}, P_{off}) = -\min_{0 \leq \lambda \leq 1} \log \sum_y P_{on}^{\lambda}(y) P_{off}^{1-\lambda}(y)$. This is motivated by theoretical studies of the detectability of edge contours (Yuille, Coughlan, Wu, Zhu 2001). Moreover, empirical studies (Konishi, Yuille, Coughlan, Zhu 1999, 2003) showed that the Chernoff Information for this task correlates strongly with other performance measures based on the ROC curve. Konishi *et al* tested a variety of different edge filters and ranked them by their effectiveness based on their Chernoff information. For this project, we chose a very simple edge detector $\left|\vec{\nabla} G_{\sigma=1} * I\right|$– the magnitude of the gradient of the grayscale image I filtered by a Gaussian $G_{\sigma=1}$ with standard deviation $\sigma = 1$ pixel units – which has a Chernoff of 0.26 nats or 0.37 bits (1 bit = $\log_e 2$ nats ≈ 0.69 nats). More effective edge detectors are available – for example, the gradient at multiple scales using color has a Chernoff of 0.51 nats or 0.74 bits. But we do not need these more sophisticated detectors.

We extend the work of Konishi *et al* by putting probability distributions on how accurately the edge filter gradient estimates the true perpendicular direction of the edge, see figure (4). These were learned for this dataset by measuring the true orientations of straight-line edges and comparing them to those estimated

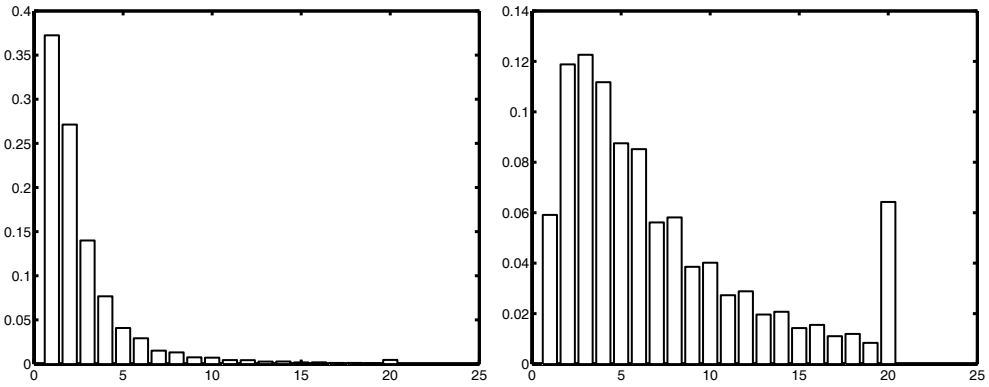

Figure 3: $P_{off}(y)$ (left) and $P_{on}(y)$(right), the empirical histograms of edge responses off and on edges, respectively. Here the response $y = \left|\vec{\nabla} I\right|$ is quantized to take 20 values and is shown on the horizontal axis. Note that the peak of $P_{off}(y)$ occurs at a lower edge response than the peak of $P_{on}(y)$. These distributions were very consistent for a range of images.

from the image gradients.

This gives us distributions on the magnitude and direction of the intensity gradient $P_{on}(\vec{E}_{\vec{u}}|\theta), P_{off}(\vec{E}_{\vec{u}})$, where $\vec{E}_{\vec{u}} = (E_{\vec{u}}, \phi_{\vec{u}})$, θ is the true normal orientation of the edge, and $\phi_{\vec{u}}$ is the gradient direction measured at point $\vec{u}$. We make a *factorization assumption* that $P_{on}(\vec{E}_{\vec{u}}|\theta) = P_{on}(E_{\vec{u}})P_{ang}(\phi_{\vec{u}} - \theta)$ and $P_{off}(\vec{E}_{\vec{u}}) = P_{off}(E_{\vec{u}})U(\phi_{\vec{u}})$. $P_{ang}(.)$ (with argument evaluated modulo 2π and normalized to 1 over the range 0 to 2π) is based on experimental data, see Figure 4, and is peaked about 0 and π. In practice, we use a simple box function model: $P_{ang}(\delta\theta) = (1-\epsilon)/4\tau$ if $\delta\theta$ is within angle τ of 0 or π, and $\epsilon/(2\pi - 4\tau)$ otherwise (i.e. the chance of an angular error greater than $\pm\tau$ is ϵ). In our experiments $\epsilon = 0.1$ and $\tau = 6°$. By contrast, $U(.) = 1/2\pi$ is the uniform distribution.

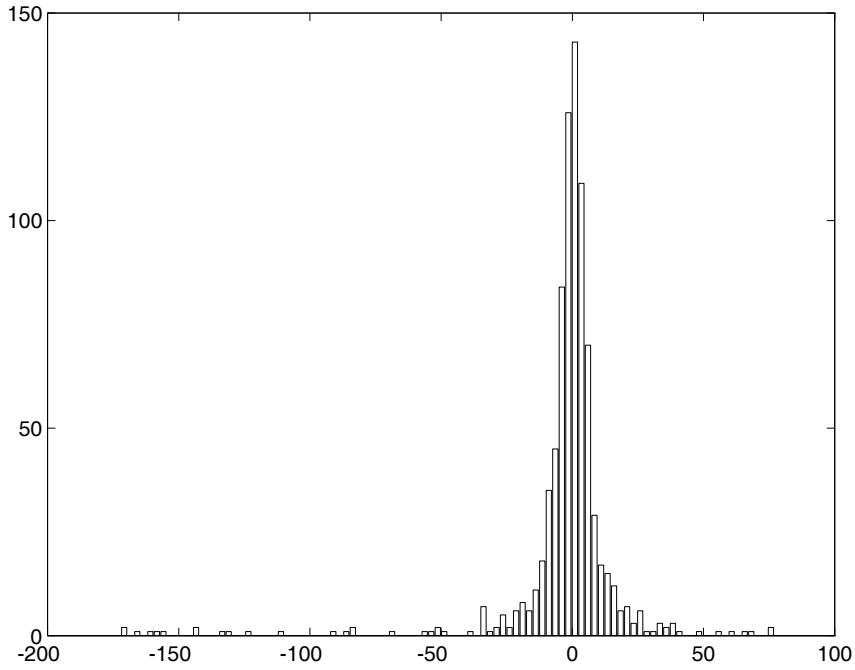

Figure 4: Distribution $P_{ang}(.)$ of edge orientation error (displayed as an unnormalized histogram, modulo 180°). Observe the strong peak at 0°, indicating that the image gradient direction at an edge is usually very close to the true normal orientation of the edge. We modelled this distribution using a simple box function.

5 Bayesian Model

We now devise a Bayesian model which combines knowledge of the three-dimensional geometry of the Manhattan world with statistical knowledge of edges in images. The model assumes that, while the majority of pixels in the image convey no information about camera orientation, most of the pixels with high edge responses arise from the presence of x, y, z lines in the three-dimensional scene. The edge orientations measured at these pixels provide constraints on the camera angle, and although the constraining evidence from any single pixel is weak, the Bayesian model allows us to pool the evidence over all pixels (both on and off edges), yielding a sharp posterior distribution on the camera orientation. An important feature of the Bayesian model is that *it does not force us to decide prematurely which pixels are on and off* (or whether an on pixel is due to x, y or z), *but allows us to sum over all possible interpretations of each pixel.*

5.1 Evidence at one pixel

The image data $\vec{E}_{\vec{u}}$ at pixel $\vec{u}$ is explained by one of five models $m_{\vec{u}}$: $m_{\vec{u}} = 1, 2, 3$ mean the data is generated by an edge due to an x, y, z line, respectively, in the scene; $m_{\vec{u}} = 4$ means the data is generated by a random edge (not due to an x, y, z line); and $m_{\vec{u}} = 5$ means the pixel is off-edge. The prior probability $P(m_{\vec{u}})$ of each of the edge models was estimated empirically to be $0.02, 0.02, 0.02, 0.04, 0.9$ for $m_{\vec{u}} = 1, 2, \ldots, 5$, see section (8).

Using the factorization assumption mentioned before, we assume the probability of the image data $\vec{E}_{\vec{u}}$ has two factors, one for the magnitude of the edge

strength and another for the edge direction:

$$P(\vec{E}_{\vec{u}}|m_{\vec{u}}, \vec{\Psi}, \vec{u}) = P(E_{\vec{u}}|m_{\vec{u}})P(\phi_{\vec{u}}|m_{\vec{u}}, \vec{\Psi}, \vec{u}) \tag{3}$$

where $P(E_{\vec{u}}|m_{\vec{u}})$ equals $P_{off}(E_{\vec{u}})$ if $m_{\vec{u}} = 5$ or $P_{on}(E_{\vec{u}})$ if $m_{\vec{u}} \neq 5$. Also, $P(\phi_{\vec{u}}|m_{\vec{u}}, \vec{\Psi}, \vec{u})$ equals $P_{ang}(\phi_{\vec{u}} - \theta(\vec{\Psi}, m_{\vec{u}}, \vec{u}))$ if $m_{\vec{u}} = 1, 2, 3$ or $U(\phi_{\vec{u}})$ if $m_{\vec{u}} = 4, 5$. Here $\theta(\vec{\Psi}, m_{\vec{u}}, \vec{u}))$ is the predicted normal orientation of lines determined by the equation $\tan\theta_x = (uc_x - fa_x)/(fb_x - vc_x)$ for x lines, $\tan\theta_y = (uc_y - fa_y)/(fb_y - vc_y)$ for y lines, and $\tan\theta_z = (uc_z - fa_z)/(fb_z - vc_z)$ for z lines. The structure of the probability distribution for all the variables relevant to one pixel (i.e. $\vec{E}_{\vec{u}}, m_{\vec{u}}, \vec{\Psi}$) is graphically depicted in the Bayes net shown in Figure 5.

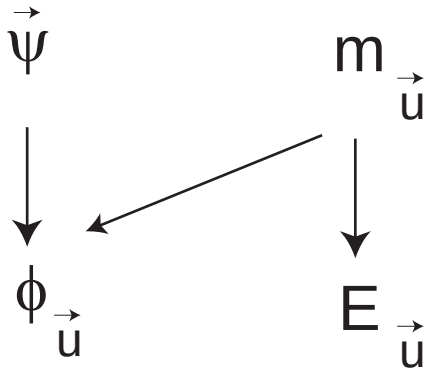

Figure 5: Bayes net for all variables pertaining to a single pixel. The net represents the structure of the joint probability of the image gradient direction $\phi_{\vec{u}}$ and magnitude $E_{\vec{u}}$ at pixel $\vec{u}$, the assignment variable $m_{\vec{u}}$ at that pixel and the camera orientation $\vec{\Psi}$. (The dependence of $\phi_{\vec{u}}$ on the pixel location $\vec{u}$ is not shown since we assume $\vec{u}$ is known.)

In summary, the edge strength probability is modeled by P_{on} for models 1 through 4 and by P_{off} for model 5. For models 1,2 and 3 the edge orientation is modeled by a distribution which is peaked about the appropriate orientation of

an x, y, z line predicted by the compass angle at pixel location $\vec{u}$; for models 4 and 5 the edge orientation is assumed to be uniformly distributed from 0 through 2π.

Rather than decide on a particular model at each pixel, we *marginalize* over all five possible models (i.e. creating a mixture model):

$$P(\vec{E}_{\vec{u}}|\vec{\Psi}, \vec{u}) = \sum_{m_{\vec{u}}=1}^{5} P(\vec{E}_{\vec{u}}|m_{\vec{u}}, \vec{\Psi}, \vec{u})P(m_{\vec{u}}) \tag{4}$$

In this way we can determine evidence about the camera orientation $\vec{\Psi}$ at each pixel without knowing which of the five model categories the pixel belongs to.

5.2 Evidence over all pixels: finding the MAP

To combine evidence over all pixels in the image, denoted by $\{\vec{E}_{\vec{u}}\}$, we assume that the image data is conditionally independent across all pixels, given the camera orientation $\vec{\Psi}$:

$$P(\{\vec{E}_{\vec{u}}\}|\vec{\Psi}) = \prod_{\vec{u}} P(\vec{E}_{\vec{u}}|\vec{\Psi}, \vec{u}) \tag{5}$$

Conditional independence is a key assumption of the Manhattan model (depicted in the Bayes net in Figure 6). By neglecting coupling of image gradients at neighboring pixels, the conditional independence assumption makes an approximation that yields a model for which MAP inference is tractable (see equation 6). Note also that the conditional independence assumption provides a way of combining evidence across pixels without an explicit grouping process (e.g. by which pixels could be grouped into straight line segments). Conditional independence is, of course, an approximation of the form used in many statistical inference algorithms. Indeed, there exist theoretical studies (Yuille, Coughlan, Wu, Zhu 2001) which prove that, for certain types of problem, approximate models can give results almost as good as the correct models. We did implement a Manhattan model

which included spatial interactions, but the results did not improve significantly and the model was far slower to implement.

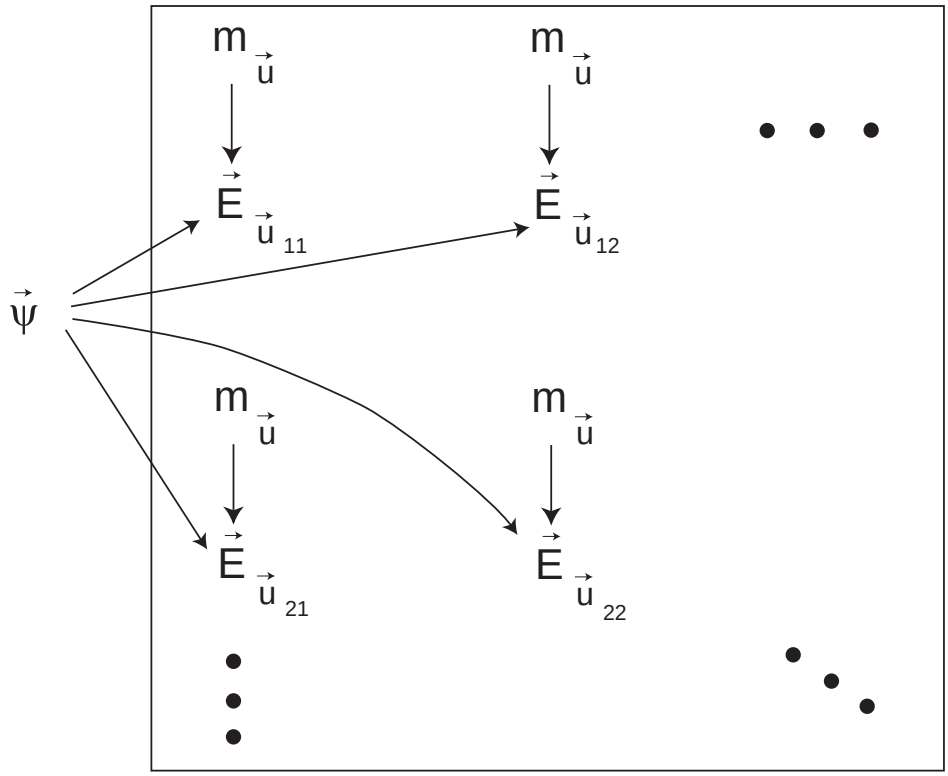

Figure 6: Bayes net for all variables in the Manhattan model. The net represents the structure of the joint probability of the image gradient vector $\vec{E}_{\vec{u}_{ij}}$ at pixel $\vec{u}_{ij}$ (the pixel at row i and column j), the assignment variable $m_{\vec{u}_{ij}}$ at that pixel and the camera orientation $\vec{\Psi}$. The box represents the entire image, with an image gradient vector $\vec{E}_{\vec{u}_{ij}}$ and assignment variable $m_{\vec{u}_{ij}}$ at each pixel location. The structure of the net graphically illustrates the assumption that the image gradient vectors $\vec{E}_{\vec{u}}$ are conditionally independent from pixel to pixel.

The posterior distribution on the camera orientation is thus given by: $\prod_{\vec{u}} P(\vec{E}_{\vec{u}}|\vec{\Psi}, \vec{u})P(\vec{\Psi})/Z$ where Z is a normalization factor and $P(\vec{\Psi})$ is a uniform prior on the camera orientation. To find the MAP (maximum a posterior) estimate, we need to maximize

the log posterior term (ignoring Z, which is independent of $\vec{\Psi}$):

$$\log[P(\{\vec{E}_{\vec{u}}\}|\vec{\Psi})P(\vec{\Psi})] = \log P(\vec{\Psi}) + \sum_{\vec{u}} \log[\sum_{m_{\vec{u}}=1}^{5} P(\vec{E}_{\vec{u}}|m_{\vec{u}}, \vec{\Psi}, \vec{u})P(m_{\vec{u}})] \qquad (6)$$

We denote the MAP estimate by $\vec{\Psi}^*$. To find the MAP, our algorithm evaluates the log posterior numerically for the compass direction for a quantized set of $\vec{\Psi}$ values. (The conditional independence assumption makes the form of the log posterior simple enough that it can be evaluated for any given $\vec{\Psi}$ value by summing over only 5 terms for each pixel.) One such set of quantized values that works for a range of images is given by searching over all combinations of α from $-45°$ to $45°$ in increments of $4°$, elevation β from $-40°$ to $40°$ in increments of $2°$, and twist γ from $-4°$ to $4°$ in increments of $2°$. In our preliminary work (Coughlan and Yuille 1999) we assumed that the camera was pointed horizontally, so we effectively set $\beta = 0$ and $\gamma = 0$ and searched for all α from $-45°$ to $45°$ in increments of $2°$.

A coarse-to-fine search strategy was employed to speed up the search for the MAP estimate, which succeeded for most images for which the true values of β and γ were close to 0. The first stage of the search was to find the best value of α from $45°$ to $45°$ in increments of $4°$, while setting $\beta = 0$ and $\gamma = 0$. The best value of α that was obtained, α_c, was used to initialize a medium-scale search, which searched over all (α, β, γ) of the form $(\alpha_c + i\Delta\alpha_m, j\Delta\beta_m, k\Delta\gamma_m)$, where $i, j, k \in \{-1, 0, 1\}$ and $\Delta\alpha_m = 2°, \Delta\beta_m = 5°$ and $\Delta\gamma_m = 5°$. The best Euler angles thus obtained, $(\alpha_m, \beta_m, \gamma_m)$, were then used to initialize a fine-scale search, which searched over all (α, β, γ) of the form $(\alpha_m, \beta_m + j\Delta\beta_f, \gamma_m + k\Delta\gamma_f)$, where $j, k \in \{-2, -1, 0, 1, 2\}$ and $\Delta\beta_f = 2.5°$ and $\Delta\gamma_f = 2.5°$.

We should mention the issues of algorithmic speed. At present the algorithm takes half a minute on a Pentium 3 using the coarse-to-fine search strategy on

images of size 640×480. Optimizing the code and subsampling the image will allow the algorithm to work significantly faster. Other techniques involve rejecting image pixels where the edge detector response is so low that there is no realistic chance of an edge being present. This would mean that at least 70% of the image pixels could be removed from the computation. We observe that the algorithm is entirely parallelizable. Stochastic gradient-descent techniques may also be employed for significant speed-ups (Deutscher, Isard and MacCormick 2002).

6 Experimental Results for Determining Manhattan Structure

Our model has been tested on four datasets of images: indoor scenes, outdoor city scenes, outdoor rural scenes and miscellaneous non-Manhattan scenes. Images from the first two datasets were taken by an unskilled photographer unfamiliar with the goals of the study; the outdoor rural scenes were obtained from a database of scenes of English countryside; and the non-Manhattan scenes were downloaded from the web.

We tested our model is two ways. Firstly, we compared the vanishing points estimated by the algorithm to manual estimates made by the authors. Secondly, we implemented a null hypothesis model and used a log-likelihood test to estimate whether an image does, or does not, obey the Manhattan world assumption. The null hypothesis model removes the image intensity depedence on any three-dimensional scene structure.

Our experiments show that the algorithms' estimates are usually close to the manual estimates for the first three domains, see section (6.1,6.2,6.3). Moreover,

images in these domains are almost always estimated as obeying the Manhattan world assumption while the miscellaneous images are typically estimated as not obeying the assumption, see Section (6.4).

6.1 Estimating Viewpoint for Indoor Scenes

On this dataset, the camera was held roughly horizontal but no special attempt was made to align it precisely. The camera was set on automatic so some images are over- or under- exposed.

A total of twenty-five images were tested. Since the camera was held roughly horizontal, we set $\beta = 0$ and $\gamma = 0$ and searched for the optimal value of α (the results are similar when searching simultaneously over all three camera angles α, β and γ). On twenty-three images, the angles estimated by the algorithm was within 5° of the manual estimate made by the authors. On two images, the orientation of the camera was far from horizontal and the estimation was poor. Examples of successes, demonstrating the range of images used, are shown in Figures 7,8. The log posteriors for typical images, plotted as a function of α, are shown in Figure 9 and are sharply peaked.

6.2 Estimating Viewpoint for Outdoor City Scenes

We next tested the accuracy of estimation on outdoor city scenes. Again we used twenty-five test images (taken by the same photographer as for the indoor scenes). In these scenes the vast majority of the results (twenty-two) were accurate up to 10° (with respect to the manual estimates made by the authors). On three of the images the angles were worse than 10°. Inspection of these images showed that the log posterior had multiple peaks, one peak corresponding to the true

Figure 7: Estimates of the compass angle and geometry obtained by our algorithm. The estimated orientations of the x and y lines are indicated by the black line segments drawn on the input image. At each point on a subgrid two such segments are drawn – one for x and one for y. Left panel: Observe how the x directions align with the wall on the right hand side and with features parallel to this wall. The y lines align with the wall on the left (and objects parallel to it). Right panel: Observe that the x, y directions align with the appropriate walls despite the poor quality of the image (i.e. under-exposed).

Figure 8: Another indoor (left panel) scene, its exterior (right panel), and another indoor scene (right panel). Same conventions as above. The vanishing points are estimated to within 5° (perfectly adequate for our purposes). Note poor quality of the indoor image (i.e. over-exposed). (Indoor 23,8 and Outdoor 12).

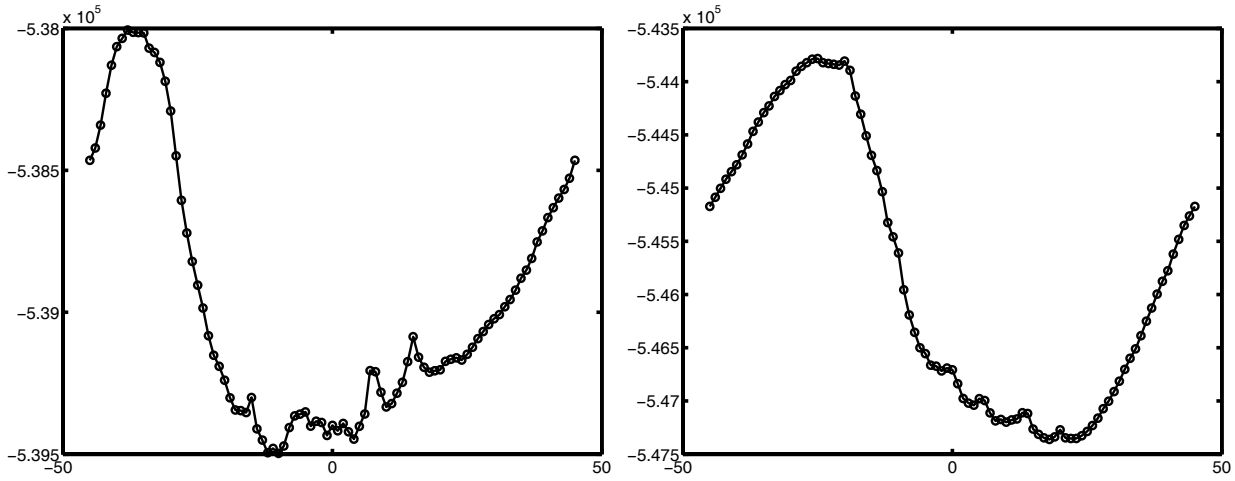

Figure 9: The log posteriors as a function of α (from $-45°$ to $45°$ along the horizontal axis) for images Indoor 17 (left) and Indoor 15 (right). These results are typical for both the indoor and outdoor dataset. (For these plots it was assumed that the camera is roughly horizontal so only the angle α needs to be varied.)

compass angle (to within 10°), as well as false peaks which were higher. The false peaks typically corresponded to the presence of structured objects in the scene (e.g. stairway railings) which did not align to Manhattan structure.

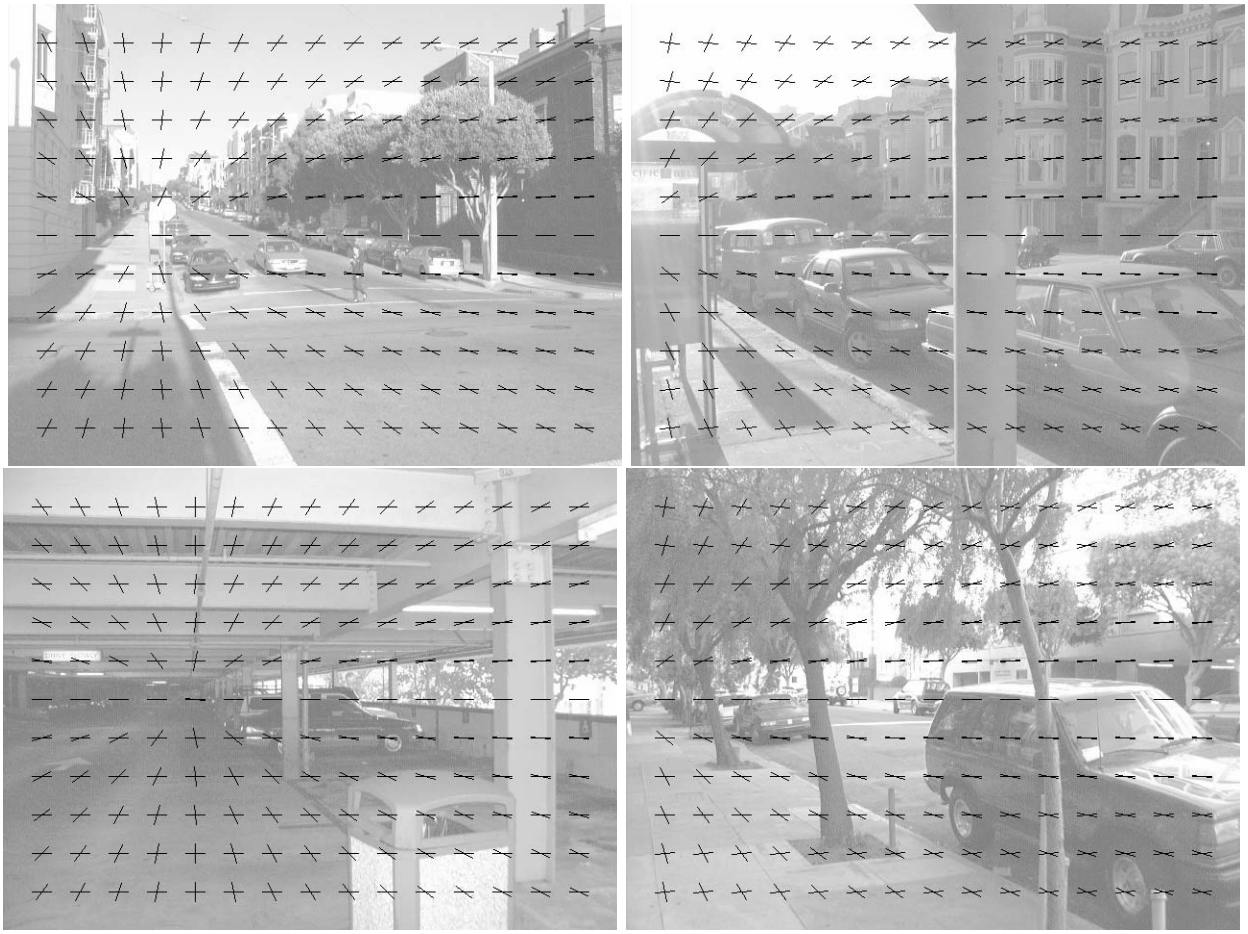

Figure 10: Results on four outdoor images. Same conventions as before. Observe the accuracy of the x, y projections in these varied scenes despite the poor quality of some of the images.

On twenty-two of the twenty-five images, however, the algorithm gave estimates accurate to 10° (compared with the authors' manual estimates). See Figure 10 for a representative set of images on which the algorithm was successful. We also demonstrate the algorithm on two aerial views of cities downloaded from

the web, see Figure 11. On these images we searched for all three camera angles α, β and γ simultaneously, which was necessary since the camera was tilted from the horizontal.

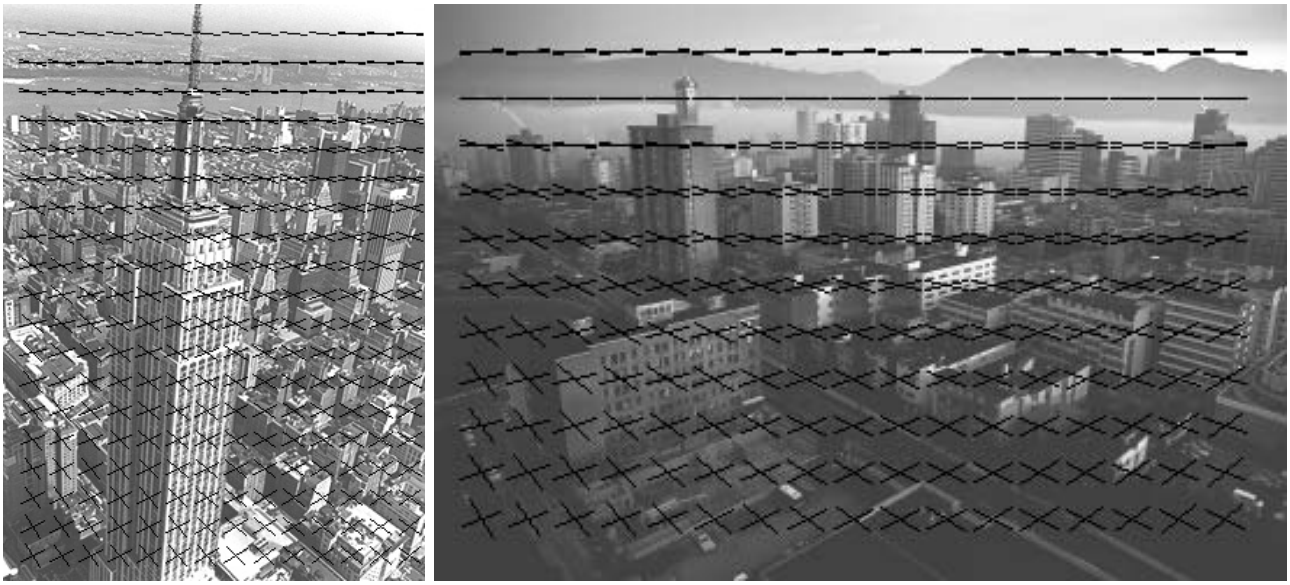

Figure 11: Aerial views of Manhattan, left and Vancouver, right. Note that the camera is tilted from the horizontal in both cases.

6.3 Estimating Viewpoint for Outdoor Rural Scenes

We also applied the Manhattan model to less structured scenes in the English countryside (see Coughlan and Yuille 2000 for a first report). Figure (12) shows two images of roads in rural scenes and two fields. These images come from the Sowerby database.

But some scenes, see Figure (13), contain so little Manhattan structure that the Manhattan model may base its inference on chance alignments of various parts of the scene.

The next four images, see Figure (14), were either downloaded from the web

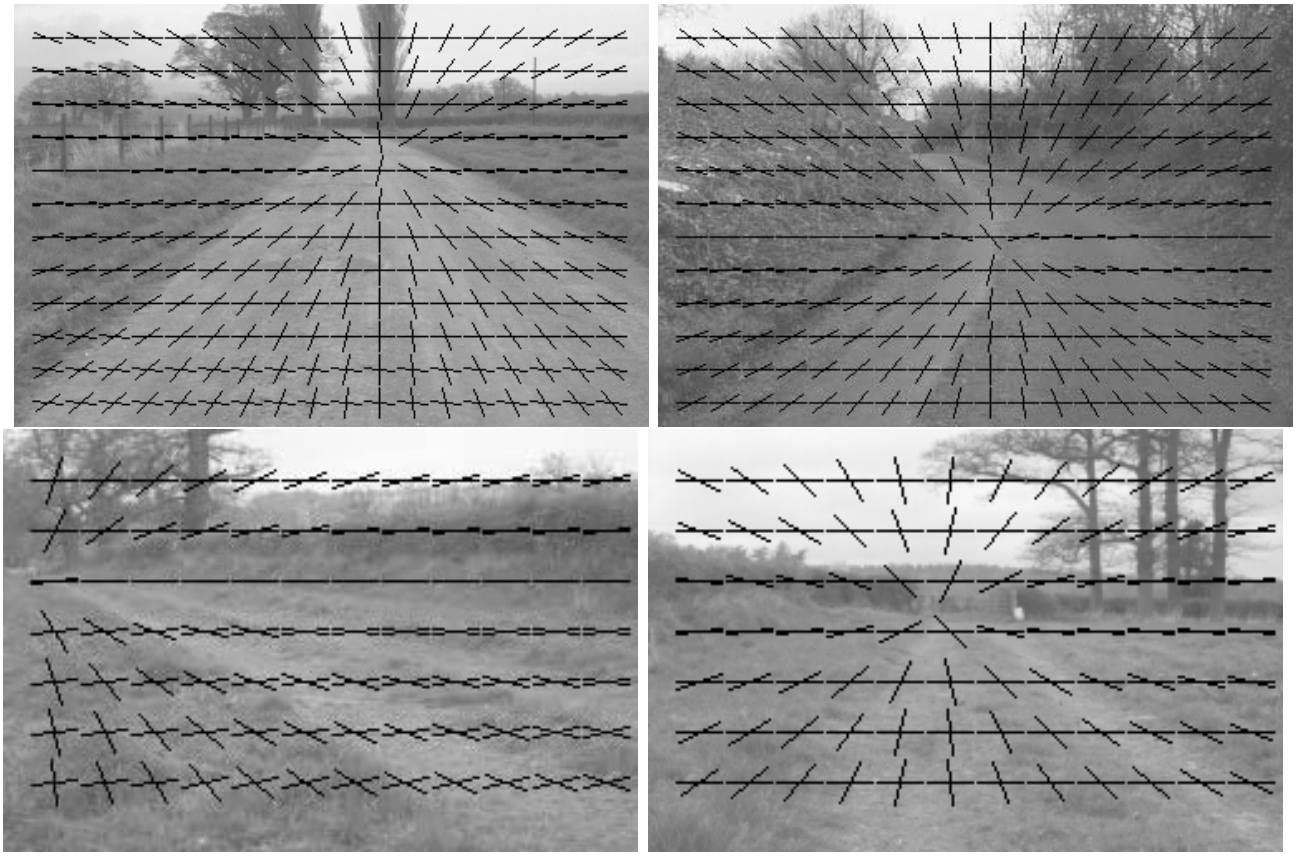

Figure 12: Results on rural images in England without strong Manhattan structure. Same conventions as before. Two images of roads in the countryside (left panels) and two images of fields (right panel).

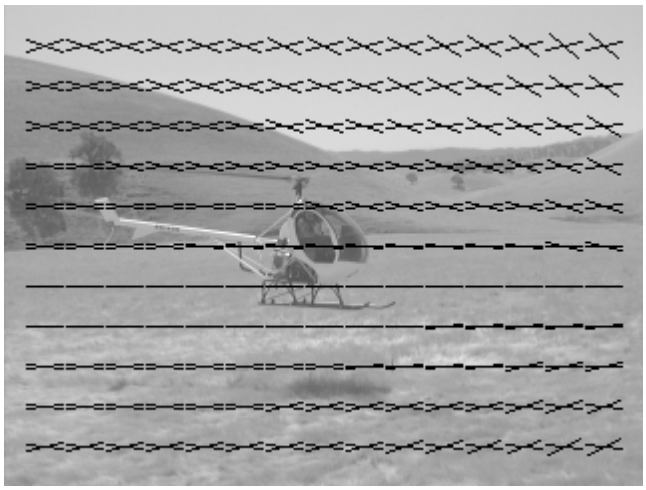

Figure 13: Example of a scene with low Manhattan structure. The hill silhouettes are mistakenly interpreted as x lines.

or digitized (the painting). These are the mid-west broccoli field, the Parthenon ruins, the painting of the French countryside, and the ski scene. On all of the images in this section, the Manhattan algorithm searched for all three camera angles α, β and γ simultaneously. In almost all of these cases, the Manhattan model makes reasonable inferences despite the absence of strong Manhattan structure, and in some cases despite the absence of strong straight-line edges.

6.4 Manhattan World and the Null Hypothesis

We now propose a test to determine whether an image obeys the Manhattan world assumption. Our previous results, see sections (6.1,6.2,6.3), show that on several image classes we can detect accurate vanishing points (with respect to our manual estimates) but there are many images, such as underwater images, for which the Manhattan world assumption is highly implausible.

We proceed by constructing a null hypothesis model and then use model selection to estimate whether an image obeys the Manhattan world assumption. The

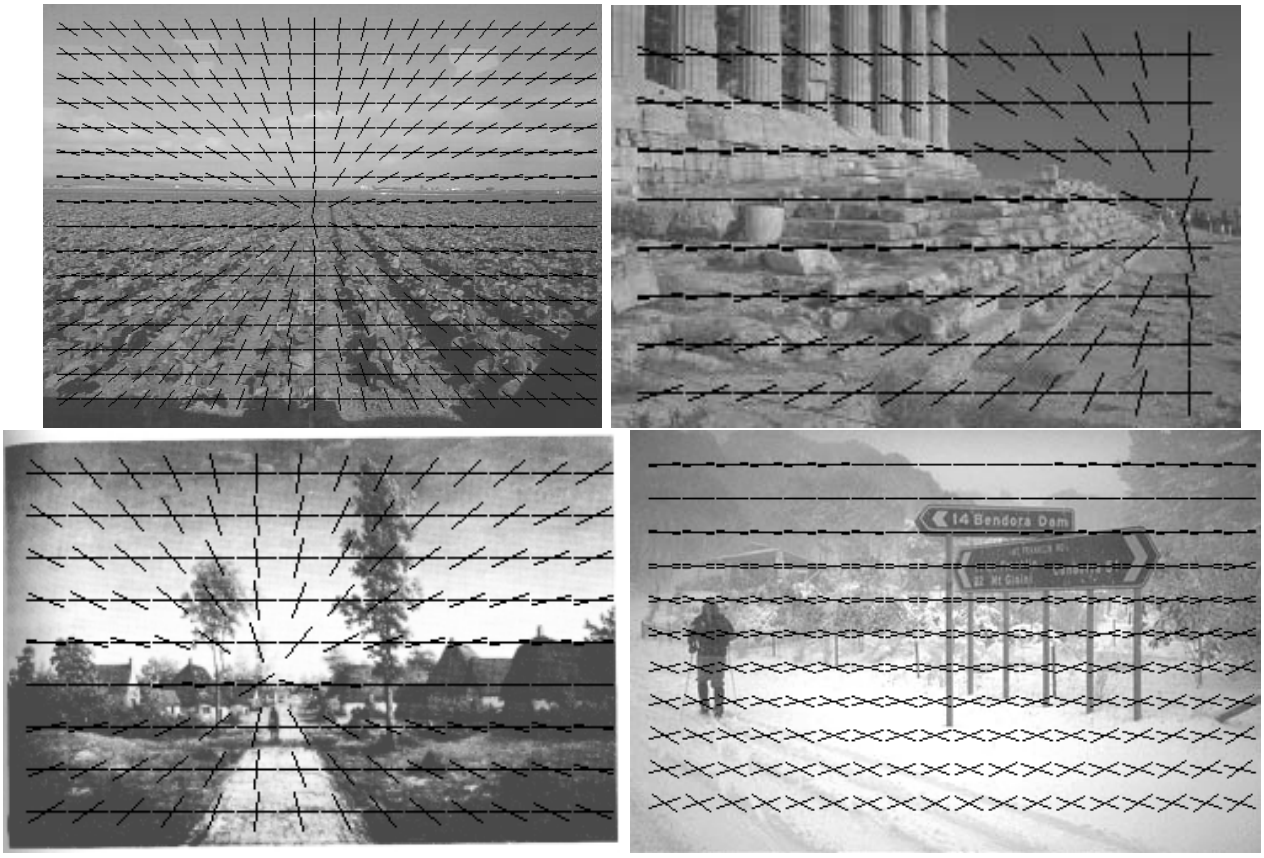

Figure 14: Results on an American mid-west broccoli field, the ruins of the Parthenon, a digitized painting of the French countryside and a ski scene.

null hypothesis model is constructed by modifying our Manhattan model of equation (3) by setting $P(\phi_{\vec{u}}|m_{\vec{u}}, \vec{\Psi}, \vec{u}) = U(\phi_{\vec{u}})$ where $U(.)$ is the uniform distribution. This give our null hypothesis model to be:

$$P_{null}(\{\vec{E}_{\vec{u}}\}) = \prod_{\vec{u}} [P(\text{edge})P_{on}(E_{\vec{u}}) + P(\text{not} - \text{edge})P_{off}(E_{\vec{u}})]U(\phi_{\vec{u}}), \qquad (7)$$

where $P(\text{edge}) = \sum_{i=1}^{4} P(m_i) = 0.1$ and $P(\text{not} - \text{edge}) = P(m_5) = 0.9$. In other words, we no longer distinguish between different types of edges and no longer assume that the image statistics reflect any three-dimensional scene structure.

To do model comparison for an image with statistics $\{\vec{E}_{\vec{u}}\}$, we compute the evidence $\log P_{manhat}(\{E_{\vec{u}}\}) = \log \sum_{\vec{(\Psi)}} P(\{\vec{E}_{\vec{u}}\}|\vec{\Psi})P(\vec{\Psi})$ of the Manhattan model and subtract the evidence $\log P_{null}(\{\vec{E}_{\vec{u}}\})$ of the null model. This gives the log-likelihood ratio between the two models. We approximate the evidence by $\log P_{manhat}(\{E_{\vec{u}}\}) \approx \log P(\{\vec{E}_{\vec{u}}\}|\vec{\Psi}^*)P(\vec{\Psi}^*)$, where $\vec{\Psi}^* = \arg\max_{\vec{\Psi}} P(\{\vec{E}_{\vec{u}}\}|\vec{\Psi})P(\vec{\Psi})$. This approximation is a lower bound to the true evidence and we argue that it is a good approximation because of the sharpness of the peaks in the posterior $P(\{\vec{E}_{\vec{u}}\}|\vec{\Psi})P(\vec{\Psi})$, see figure (9) (this figure plots the log posterior, so the posterior is considerably sharper).

The experimental results, i.e. the plots of $\log P_{manhat}(\{E_{\vec{u}}\})/P_{null}(\{E_{\vec{u}}\})$ in figure (15), show that all the images reported in section (6.1, 6.2,6.3), satisfy the Manhattan world assumption (according to our model selection test). It is therefore not surprising that the algorithms estimates of the vanishing point were accurate for these images. The figure also shows a plausible trend: indoor images best satisfy the Manhattan world assumption followed by outdoor images and then by rural images. The figure also shows ten miscellaneous images which are not expected to satisfy the Manhattan world assumption. These images include an

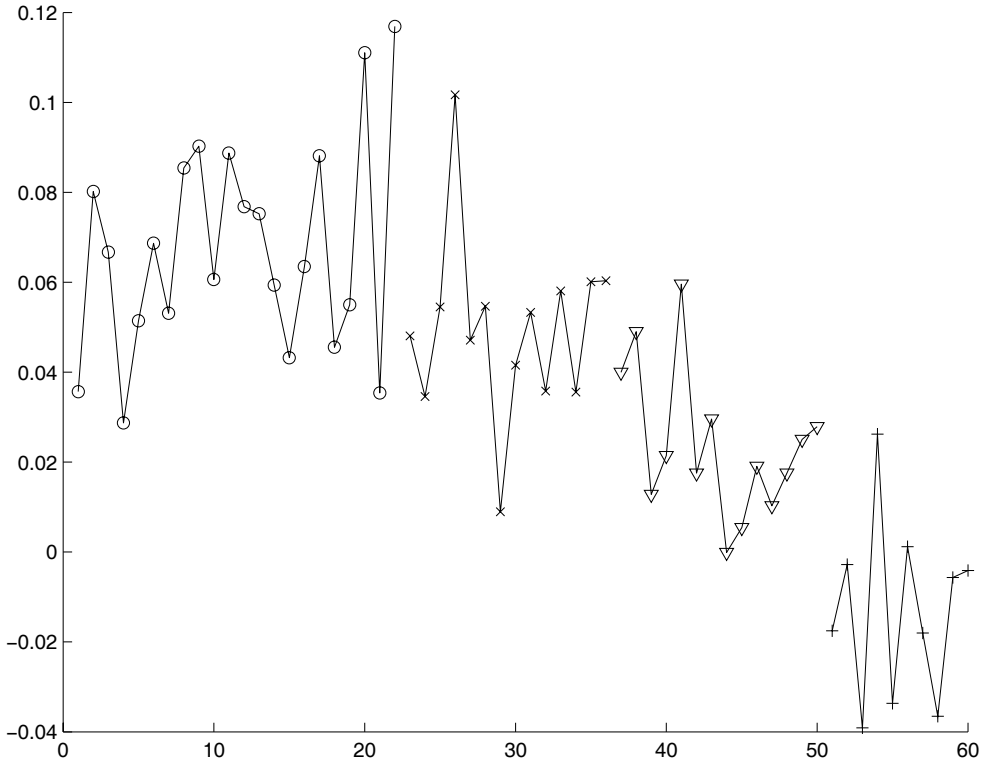

Figure 15: We plot the (normalized) loglikehood ratio $(1/N)\log(P_{manhat}(\{\vec{E}_{\vec{u}}\})/P_{null}(\{\vec{E}_{\vec{u}}\})$ of the (approximated) Manhattan model with respect to the null hypothesis (where N is the image size). The vertical axis is the log-likelihood ratio and the horizontal axis is the index label of the images. We indicate indoor, outdoor, Sowerby, and miscellaneous images by circles, crosses ($\times$), triangles, and pluses ($+$) respectively.

underwater image, see figure (16), and almost all have log-likelihoods than are less than zero and considerably lower than those for indoor, outdoor, and rural scences. The main exception is the fourth image which is a photograph of a helicopter, see figure (13), and where the hill silhouettes are mistakenly interpreted as horizontal x lines.

It is also interesting to plot the evidence, $\log(P_{manhat}(\{\vec{E}_{\vec{u}}\})$, for the Manhattan model alone, see figure (16). (As above, we approximate this sum by the dominant contribution given by $\vec{\Psi}^*$). The evidence is useful for indicating trends to determine what classes of images fit the Manhattan world assumption. To avoid biases caused by different images sizes, we normalize the evidence by the image size and plot L/N. (Our conditional independence assumption, if correct, implies that the evidence scales linearly with the number of pixels). Observe that the data is very high dimensional and so the probabilities of any image will be very small.

7 Outliers in Manhattan world

We now describe how the Manhattan world model may help for the task of detecting target objects in background clutter. To perform such a task effectively requires modelling the properties of the background clutter in addition to those of the target object. It has recently been appreciated (Ratches, Walters, Buser and Guenther 1997) that simple models of background clutter based on Gaussian probability distributions are often inadequate and that better performance can be obtained using alternative probability models (Zhu, Lanteman and Miller 1998).

The Manhattan world assumption gives an alternative, and complementary,

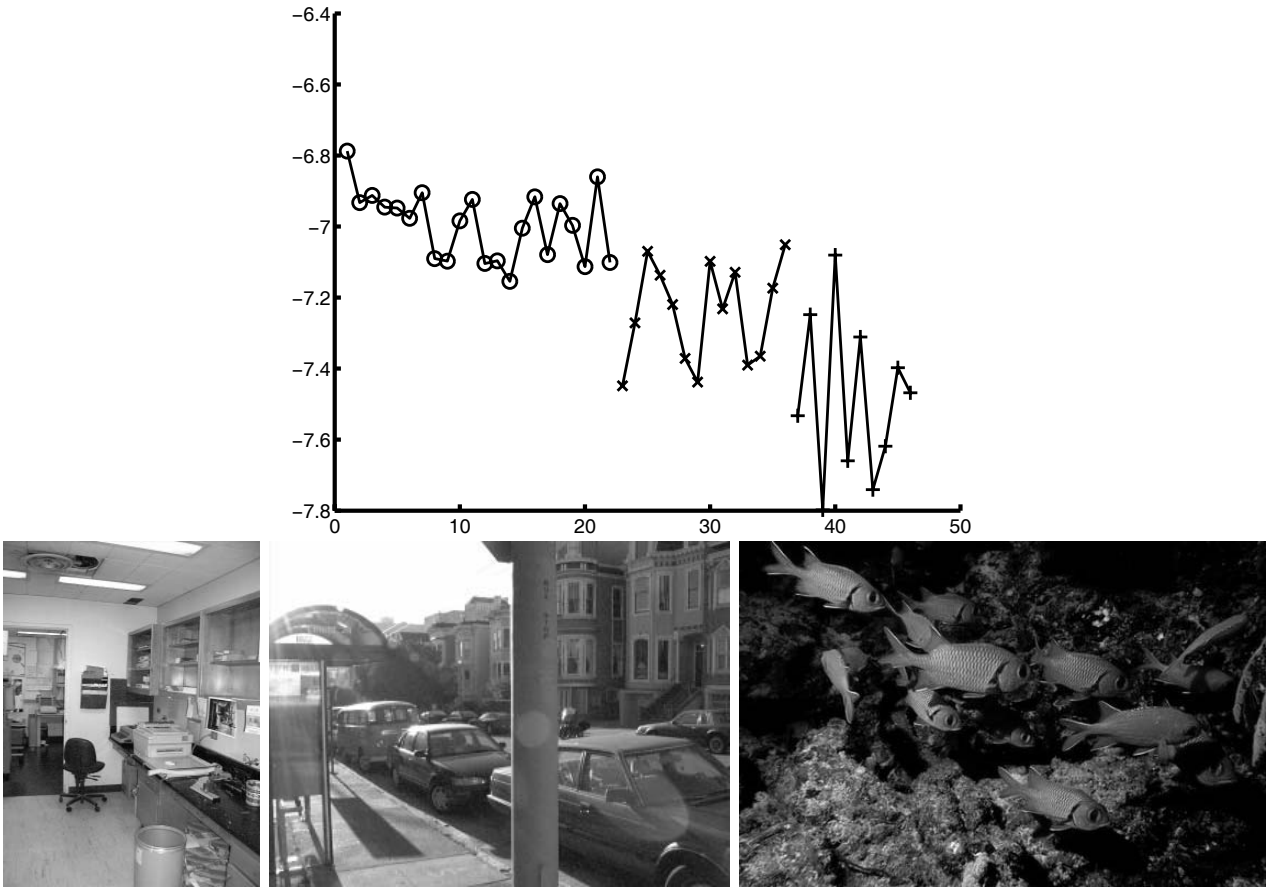

Figure 16: Evidence per pixel (approximated), $1/N \log(P_{manhat}(\{\vec{E}_{\vec{u}}\})$, evaluated in three domains: indoor images (labeled by o's), outdoor urban images (labeled by x's), and miscellaneous images (labeled by +'s). A representative image from each domain is shown below. The trend is that the evidence per pixel decreases, on average, as we go from images with strong Manhattan structure to images without Manhattan structure.

way of probabilistically modelling background clutter. The background clutter will correspond to the regular structure of buildings and roads and its edges will be aligned to the Manhattan grid. The target object, however, is assumed to be unaligned (at least, in part) to this grid. *Therefore many of the edges of the target object will be assigned to model 4 by the algorithm.* This enables us to significantly simplify the detection task by removing all edges in the images except those assigned to model 4. (Of course, further processing is required to group these outlier edges into coherent targets).

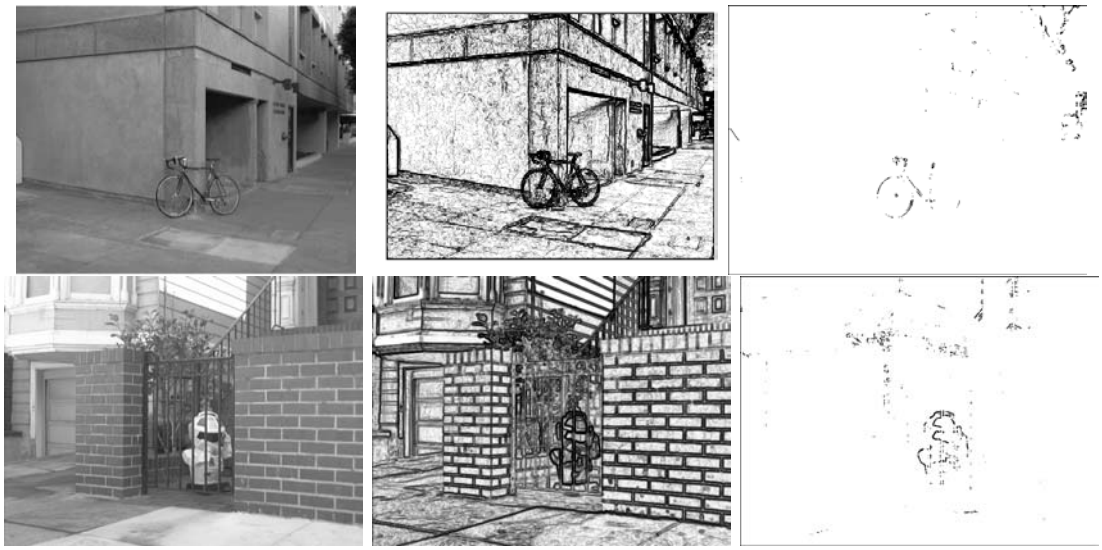

Figure 17: Bikes (top row) and robots (bottom row) as outliers in Manhattan world. The original image (left) and the edge maps (center) computed as $\log P_{on}(E_{\vec{u}})/P_{off}(E_{\vec{u}})$ – see Konishi *et al* 1999 – displayed as a grayscale image where black is high and white is low. In the right column we show the edges assigned to model 4 (i.e. the outliers) in black. Observe that the edges of the bike and robot are now highly salient (and would make detection simpler) because most of them are unaligned to the Manhattan grid.

This idea is demonstrated in Figures 17, 18, where the targets are a bike, a

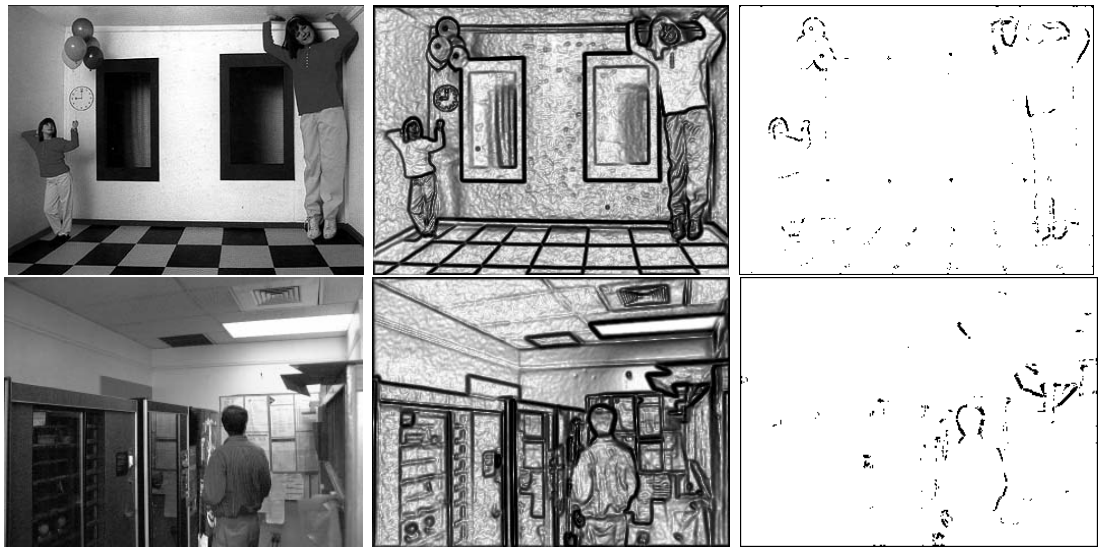

Figure 18: People in Manhattan world: the Ames room twins (top) and a co-author (bottom). Same conventions as in preceding figure. The Ames room actually *violates* the Manhattan assumption but human observers, and our algorithm, interpret it as if it satisfied the assumptions. In fact, despite appearances, the two people in the Ames room are really the same size.

robot and people. At each pixel, our algorithm decides whether the most likely interpretation is model 4 or any of the other models. More precisely, it decides "outlier" if $P(m_{\vec{u}} = 4|\vec{E}_{\vec{u}}, \vec{\Psi}^*)/P(m_{\vec{u}} \neq 4|\vec{E}_{\vec{u}}, \vec{\Psi}^*) > T$ and "non-outlier" otherwise, where $T \approx 0.4$ and $\vec{\Psi}^*$ is the MAP estimate of the camera orientation. Observe how most of the edges in the image are eliminated as target object candidates because of their alignment to the Manhattan grid. The bike, robot and people stand out as outliers to the grid.

The Ames room, see the top row of Figure (18), is a geometrically distorted room which is constructed so as to give the false impression that it is built on a Cartesian coordinate frame when viewed from a special viewpoint. Human observers assume that the room is indeed Cartesian despite all other visual cues to the contrary. This distorts the apparent size of objects so that, for example, humans in different parts of the room appear to have very different sizes. In fact, a human walking across the room will appear to change size dramatically. Our algorithm, like human observers, interprets the room as being Cartesian and helps identify the humans in the room as outlier edges which are unaligned to the Cartesian reference system.

This simple example illustrates a method of modelling background clutter which we refer to as *scene clutter* because it is effectively the same as defining a probability model for the entire scene. Observe that scene clutter models require external variables – in this case the $\vec{\Psi}$ camera orientation – to determine the orientation of the viewer relative to the scene axes. These variables must be estimated to help distinguish between target and clutter. This differs from standard models used for background clutter (Ratches, Walters, Buser and Guenther 1997, Zhu, Lanterman and Miller 1998) where no external variable is used.

8 Consistency Check of the Prior

Ideally, the parameters defining the model assignment prior $P(m_{\vec{u}})$ should be learned based on the ground-truth classification of pixels in Manhattan-type images into the five model categories. However, no such ground-truth information is available, so we determined the parameters on the basis of both measurements and guesswork. In the segmentations supplied with the Sowerby dataset images about 10% of all pixels are edges. It is plausible to assume that 40% of all edges are outliers (model 4) and that x, y and z edges occur in roughly equal proportions. Hence we arrived at the prior frequencies $0.02, 0.02, 0.02, 0.04, 0.9$ for $m_{\vec{u}} = 1, 2, \ldots, 5$.

In this section we describe a consistency check for the prior frequencies by estimating the frequencies of each model assignment *using* the above prior and conditioning on image data. Our results show, see table (1), that the estimated frequencies of the different edge types are similar to our prior frequencies. This is only a crude check because, clearly, our estimated frequencies will be biased towards our choice of prior frequencies.

Our check is based on the following observation: assume a Bayesian model $P(X|Y) = P(Y|X)P(X)/P(Y)$, where Y is the observables and X is the variable to be estimated. Let $\{Y_i\}$ be a set of samples from $P(Y)$. Then $\sum_i P(X|Y_i) = \sum_i P(Y_i|X)P(X)/P(Y_i) \sim \sum_Y P(Y)P(Y|X)P(X)/P(Y) = P(X)$ as the number of samples tends to infinity. Hence the posterior averaged over a set of representative samples should converge to the prior. Now because our images are large, and we are assuming conditional independence, it is plausible that we get sufficient number of samples from each image to ensure that the estimated edge frequencies in each image are roughly equal to the prior frequencies (this is an ergodic, or

self-averaging, assumption).

We can compute the empirical frequency h_k of each model assignment k (h_1 is the empirical frequency of x lines, h_2 is the empirical frequency of y lines, etc.) in an image. We define h_k as the mean frequency of model k in the image, conditioned on all the image data and the MAP estimate of the camera orientation:

$$h_k =< (1/N) \sum_{\vec{u}} \delta_{k,m_{\vec{u}}} >_{P(\{m_{\vec{u}}\}|\{\vec{E}_{\vec{u}}\},\vec{\Psi}^*)} = (1/N) \sum_{\vec{u}} P(m_{\vec{u}} = k|\vec{E}_{\vec{u}}, \vec{u}, \vec{\Psi}^*),$$

where $P(m_{\vec{u}} = k|\vec{E}_{\vec{u}}, \vec{u}, \vec{\Psi}^*) = P(m_{\vec{u}} = k)P(\vec{E}_{\vec{u}}|m_{\vec{u}} = k, \vec{u}, \vec{\Psi}^*)/Z$ and $Z = \sum_{k'=1}^{5} P(m_{\vec{u}} = k')P(\vec{E}_{\vec{u}}|m_{\vec{u}} = k', \vec{u}, \vec{\Psi}^*)$.

We show results on 23 indoor images in Table 1. The empirical frequencies are fairly consistent with the prior model frequencies, although in many of the images the frequency of z lines appears to be higher than the x and y frequencies. Averaging the empirical frequencies across an entire domain might be a useful way to adapt the Manhattan prior to new domains.

Another useful consistency check on the prior is to calculate the MAP estimate of the model assignment at each pixel *conditioned* on the entire image. To evaluate this we calculate the most likely model assignment at each pixel given the image data there and the global MAP estimate $\vec{\Psi}^*$ of the camera orientation. More precisely, at each pixel u we find the value of $m_{\vec{u}}$ that maximizes $P(m_{\vec{u}}|\vec{E}_{\vec{u}}, \vec{\Psi}^*)$. We show results on two images in Figure 19, which demonstrates that pixels are being classified appropriately.

Image	1	2	3	4	5	6	7	8	9	10	11	12
h_1	0.015	0.018	0.027	0.018	0.032	0.021	0.016	0.025	0.025	0.026	0.025	0.020
h_2	0.023	0.019	0.014	0.032	0.036	0.030	0.028	0.045	0.044	0.030	0.025	0.028
h_3	0.024	0.072	0.051	0.019	0.017	0.056	0.035	0.065	0.070	0.033	0.063	0.068
h_4	0.025	0.038	0.033	0.031	0.035	0.036	0.033	0.054	0.056	0.055	0.035	0.038
h_5	0.912	0.854	0.875	0.900	0.879	0.856	0.888	0.812	0.806	0.857	0.852	0.800
Image	13	14	15	16	17	18	19	20	21	22	23	
h_1	0.017	0.032	0.029	0.026	0.033	0.022	0.026	0.010	0.018	0.018	0.023	
h_2	0.028	0.055	0.026	0.027	0.044	0.020	0.022	0.028	0.056	0.024	0.020	
h_3	0.057	0.035	0.028	0.052	0.066	0.036	0.052	0.044	0.083	0.033	0.108	
h_4	0.035	0.065	0.041	0.035	0.065	0.043	0.041	0.026	0.060	0.033	0.049	
h_5	0.862	0.813	0.875	0.859	0.793	0.879	0.859	0.892	0.783	0.892	0.800	

Table 1: Empirical estimates of prior model frequencies for 23 indoor images.

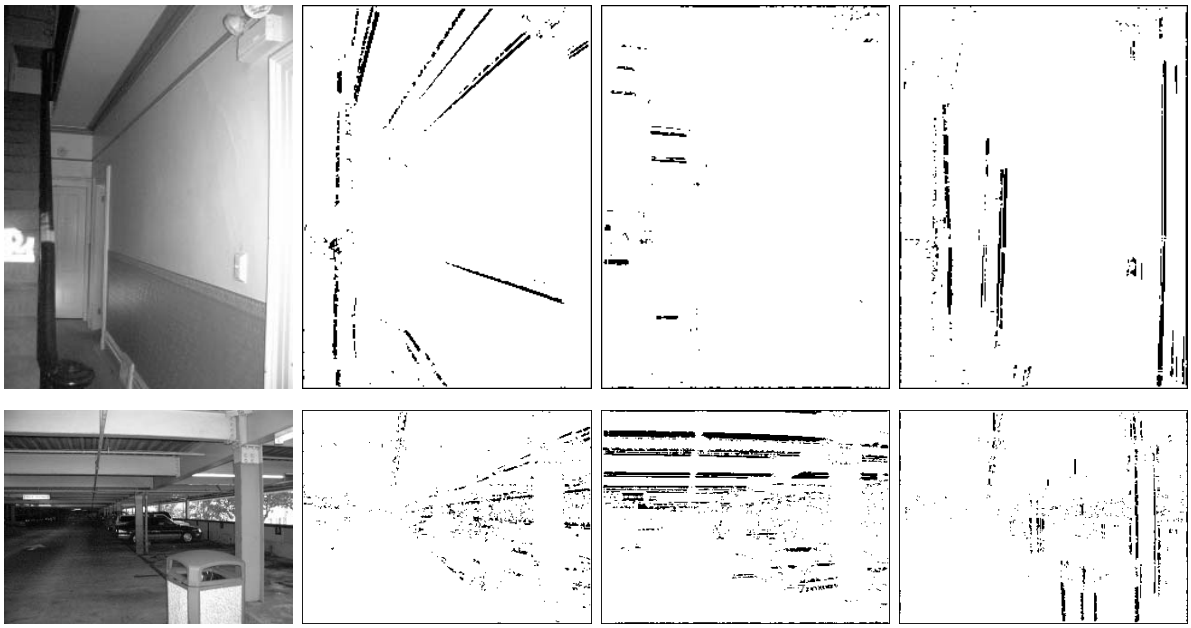

Figure 19: Detection of x, y, z lines. In each row, the original image on far left is followed by images showing the locations of pixels inferred to be on x, y and z lines, respectively, from center left to far right.

9 Summary and Conclusions

We developed a Bayesian model for exploiting the Manhattan world assumption and showed that it gave good estimates of vanishing points (e.g. viewpoint) as compared to manual estimation. In addition, the model is able to detect outlier edges, which are not aligned to the Manhattan world grid, and which may be useful for detecting objects. We formulated a null hypothesis model and used model comparison to test whether an image obeyed the Manhattan World assumption. This demonstrated that the Manhattan world assumption applies to a range of images, rural and otherwise, in addition to urban scenes. Moreover, the simplicity of the algorithm makes it suitable for implementation by an artificial retina (Burgi 2002).

Our work adds to the growing literature on the statistics of natural images and is, perhaps, the first to determine statistical regularities which depend explictly on the three-dimensional scene structure. We expect that there are many further image statistical regularities of this type which might be exploited by biological and artificial vision systems.

More recently, Deutscher, Isard and MacCormick (2002) have made use of the Manhattan world assumption as a component of a surveillance system (Isard and MacCormick 2001). Deutscher *et al* have implemented a stochastic search algorithm which is faster than the algorithms we use in this paper. In addition, they extended the Manhattan formulation to estimate focal length as well as camera orientation and demonstrated reasonably accuracy, although the approach was less accurate than standard methods requiring calibration patterns or motion estimation.

Acknowledgments

We want to acknowledge funding from NSF with award number IRI-9700446, support from the Smith-Kettlewell core grant, and from the Center for Imaging Sciences with Army grant ARO DAAH049510494. It is a pleasure to acknowledge email conversations with Song Chun Zhu about scene clutter. We gratefully acknowledge the use of the Sowerby image dataset at British Aerospace and thank Andy Wright for bringing it to our attention. We thank two referees for helpful feedback and, in particular, the suggestion that we should introduce a null hypothesis model to compare with the Manhattan model.

References

- Balboa, R. and Grzywacz, N.M. (2000). The Minimal Local-Asperity Hypothesis of Early Retinal Lateral Inhibition. *Neural Computation*. **12**, pp 1485-1517.

- Blake, A. and Yuille, A.L. (Eds). (1992). *Active Vision*. MIT Press. Cambridge, MA.

- Brillault-O'Mahony, B. (1991) New Method for Vanishing Point Detection. *Computer Vision, Graphics, and Image Processing*. 54(2). pp 289-300.

- Burgi, P-Y. (2002). Bayesian Algorithms for Autonomous Vision Systems. Technical Report 1040. Swiss Centre for Electronics and Microtechnology. Neuchatel, Switzerland.

- Chen, H., Belhumeur, P., Jacobs, D. (2000). In Search of Illumination Invariants. In *Proceedings IEEE Conference on Computer Vision and Pattern*

Recognition.

- Coughlan, J. and Yuille, A.L. (1999). Manhattan World: Compass Direction from a Single Image by Bayesian Inference. *Proceedings International Conference on Computer Vision* ICCV'99. Corfu, Greece.

- Coughlan, J. and Yuille, A.L. (2000). The Manhattan World Assumption: Regularities In Scene Statistics Which Enable Bayesian Inference. *Proceedings Neural Information Processing Systems* (NIPS '00). Denver, CO.

- Cover, T.M. and Thomas, J.A. (1991) *Elements of Information Theory.* Wiley Interscience Press. New York.

- Deutscher, J. Isard,M. and MacCormick, J. (2002). Automatic Camera Calibration from a Single Manhattan Image. *Proceedings of the European Conference on Computer Vision.* ECCV 2002. Springer-Verlag. pp 175-188.

- Faugeras, O.D, (1993) *Three-Dimensional Computer Vision.* MIT Press.

- Hartley, R. and Zisserman, A. (2000) *Multiple View Geometry in Computer Vision.* Cambridge University Press. Cambridge, England.

- Isard, M. and MacCormick, J. (2001). BraMBLe: A Bayesian multiple-blob tracker. *Proc. 8th Int. Conf. Computer Vision.* Vol. 2, pp 34-41.

- Konishi, S., Yuille, A.L., Coughlan, J.M. and Zhu, S.C. (1999). Fundamental Bounds on Edge Detection: An Information Theoretic Evaluation of Different Edge Cues. *Proc. Int'l conf. on Computer Vision and Pattern Recognition.*

- Konishi,S.M., Yuille, A.L., Coughlan, J.M. and Zhu, S.C. (2003). Statistical Edge Detection: Learning and Evaluating Edge Cues. *Pattern Analysis and Machine Intelligence*. In press. 2003.

- Lutton, E., Maître, H. and Lopez-Krahe, J. (1994) Contribution to the determination of vanishing points using Hough transform. *IEEE Trans. on Pattern Analysis and Machine Intelligence*. 16(4). pp 430-438.

- Mathews, J., Walker, R.L. (1970) *Mathematical Methods of Physics.* The Benjamin/Cummings Publishing Co.

- Huang, J. and Mumford, D. (1999). Statistics of Natural Images and Models. In *Proceedings Computer Vision and Pattern Recognition* CVPR'99. Fort Collins, Colorado.

- Mundy, J.L. and Zisserman, A. (Eds). (1992) *Geometric Invariants in Computer Vision*. MIT Press.

- Oliva, A. and Torralba, A. (2001). Modeling the Shape of the Scene: A Holistic Representation of the Spatial Envelope. *International Journal of Computer Vision*, 42(3), pp 145-175.

- Ratches, J.A., Walters, C.P., Buser, R.G. and Guenther, B.D. (1997). Aided and Automatic Target Recognition Based upon Sensory Inputs from Image Forming Systems. *IEEE Trans. on PAMI*, vol. 19, No. 9.

- Shufelt, J.A. (1999). Performance Evaluation and Analysis of Vanishing Point Detection Techniques. *IEEE Trans. on PAMI*, vol. 21, No. 3.

- Torr, P. and Zisserman, A. (1998). Robust Computation and Parameterization of Multiple View Relations. In *Proceedings of the International Conference on Computer Vision*. ICCV'98. Bombay, India. pp 727-732.

- Yuille, A.L. and Coughlan, J.M. (2000). Fundamental Limits of Bayesian Inference: Order Parameters and Phase Transitions for Road Tracking . *Pattern Analysis and Machine Intelligence* PAMI. Vol. 22. No. 2.

- Yuille, A.L., Coughlan, J.M., Wu, Y-N, and Zhu, S.C. (2001). Order Parameters for Minimax Entropy Distributions: When does high level knowledge help? *International Journal of Computer Vision*. 41(1/2), pp 9-33.

- Zhu, S.C., Lanterman, A. and Miller, M.I. (1998). Clutter Modeling and Performance Analysis in Automatic Target Recognition. In *Proceedings Workshop on Detection and Classification of Difficult Targets*. Redstone Arsenal, Alabama.