The horizon line is an important property for a wide variety of image understanding tasks. As such, many methods have been introduced to estimate the horizon line from a single image, primarily geometric methods which assume the presence of specific cues in the scene (e.g., vanishing points). These purely geometric methods are limited in their real-world capability, require extensive tuning, and are tested on benchmark datasets designed to showcase their ability. We introduce a large, realistic evaluation dataset, Horizon Lines in the Wild (HLW), containing natural images with labeled horizon lines.

1. Introduction

Estimating the location of the horizon line in an image is one of the most fundamental geometric problems in computer vision. Knowledge of the horizon line enables a wide variety of applications, including: image metrology [3], more accurate pedestrian and vehicle detection [5], and improving the look of consumer photographs [7].

Despite this demonstrated importance, progress in this research direction has stagnated and nearly all recent methods that focus on this problem make assumptions about the presence of particular geometric objects in the scene, such as vanishing points, repeated textures, and coplanar circles. Existing benchmark datasets for single image horizon line estimation are designed to highlight geometric methods, contributing to this stagnation. In an effort to address this, we introduce a new benchmark dataset containing real-world images with labeled horizon lines. Our dataset is significantly larger and more diverse when compared with existing benchmark datasets for horizon line detection.

Our contributions can be summarized as follows: 1) a novel approach for using structure from motion to automatically label images with horizon lines, 2) using this approach, introducing a large new evaluation dataset of natural images with labeled horizon lines, and 3) evaluating the most recent state-of-the-art method on this dataset.

1.1. Horizon Line: Geometric Definition

The image location of the horizon line is defined as the projection of the line at infinity for any plane which is orthogonal to the local gravity vector. The gravity vector often coincides with the local ground-plane surface normal, but not always. This is distinct from the problem of detecting the sky-line, which is the set of points where the sky and the ground meet.

A camera is defined by its extrinsic and intrinsic parameters. A point in the world, \( X_i \), is related to a pixel in \( p_{ci} \) in camera \( c \) as follows:

\[
[u_{ci}, v_{ci}, 1]^T = p_{ci} \propto K_c (R_c X_i + t_c)
\]

where \( R_c \) is the camera orientation, \( t_c \) is the camera translation, and \( K_c \) is the intrinsic calibration. For our camera coordinates we assume that the positive \( x \)-direction is to the right, the positive \( y \)-direction is up, and the viewing direction is down the negative \( z \)-axis. Using this parameterization, the world viewing direction of our camera is \( R_c^T [0, 0, -1]^T \). Assuming that the world-vector \([0, 1, 0]^T\) points in the zenith direction, the horizon line in our image is defined as the set of pixels, \( p \), where \( p^T K_c^{-T} R_c [0, 1, 0]^T = 0 \). If the intrinsic calibration, \( K_c \), of the camera is known then the horizon line provides a sufficient set of constraints to estimate the tilt and roll in world coordinates.
2. A New Dataset for Horizon Line Detection

We introduce Horizon Lines in the Wild (HLW), a large dataset of real-world images with labeled horizon lines, captured in a diverse set of environments. We begin by describing limitations in existing datasets for evaluating horizon line detection methods and then describe our approach for leveraging structure from motion to automatically label images with horizon lines.

2.1. Limitations of Existing Datasets

There are two main datasets that have been used in recent work on estimating horizon lines: the Eurasian Cities Dataset [1] (ECD) and the older York Urban Dataset [4] (YUD). We argue that these datasets have outlived their usefulness; they are too small and the images do not reflect the diversity of environments in which real-world horizon line detections methods must work.

ECD is the predominant benchmark dataset used for evaluating automatic vanishing point detection algorithms. It consists of 103 outdoor images captured in large urban areas, many of which do not support the Manhattan world assumption [2], i.e., that the structures in the image largely obey a Cartesian reference frame causing regularities in the scene statistics which can be exploited (e.g., that the contained line segments fit mostly into three mutually orthogonal directions). Of these images, the first 25 are used for model fitting, with the remainder used for testing. Of the 78 testing images, a majority are considered quite easy. Due to a combination of the few number of testing images and the small number of “challenging” images, the difference in performance between various methods often comes down to a single image. As Lezama et al. [8] note, there is even a duplicated testing image in ECD, with each instance having a different ground-truth horizon line.

The older YUD dataset is similarly small (102 images, first 25 for model fitting) and is seen as too easy: the images are captured in a confined area with a single camera, there are relatively fewer outlier line segments, the scenes obey the Manhattan world assumption, and there is zero camera roll. It is worth noting that during the construction of both ECD and YUD, each image was hand-labelled with a ground-truth horizon line through a manual process akin to the following: identify families of parallel line segments, estimate a vanishing point for each, and compute the horizon line from the horizontal vanishing points using a least squares fit. This process is slow, error prone, and severely limits the diversity of scenes.

It is our belief that the limitations of these datasets have caused useful progress in this research area to stagnate. The current state-of-the-art methods are quite slow, which is reasonable when you have a small testing dataset. For example, we find that the current state-of-the-art technique [8] requires approximately 30 seconds per image on YUD and 1 minute per image on ECD (results obtained using code made available by the authors). These methods have also focused on a particular processing pipeline: detect line segments, find vanishing points, then globally optimize to find a consistent scene interpretation. The reliance on vanishing points limits these methods to regions with many man-made structures. There is clearly a need for a larger and more diverse dataset for evaluating horizon line estimation methods.

2.2. Leveraging Structure from Motion

Manually labeling images with horizon lines is time-consuming and requires specific geometric cues to be present. Instead, we introduce a novel technique for automatically labeling images with horizon lines using structure from motion (SfM), which we then employ to generate a large dataset. A similar strategy is used in a recent work by Kendall et al. [6] to generate camera poses from a video of a scene to be used as ground truth for experiments related to camera relocalization.

The output of SfM is the extrinsic and intrinsic camera parameters for a subset of the input images. Typically these images are downloaded from photo-sharing websites, such as Flickr, around major landmarks. The extrinsic coordinates output by SfM algorithms typically have an unknown global orientation and translation. Since our focus is the horizon line, we just need to estimate the global up direction (the yaw of the reconstruction is irrelevant to our needs). The standard approach used to estimate global orientation is to average the image ‘up’ directions in world coordinates. The implicit assumption of this approach is that the expected tilt and roll of a camera is zero. While this works well in many cases, it fails in scenes with a single dominant landmark that is viewed from one direction (e.g., Notre Dame in Paris).

In practice, we found that we get more reliable world zenith direction estimates if we instead only assume that the expected roll of a camera is zero. For a given set of images, we solve for the world direction of the points at infinity in the left, [-1,0,0], and right, [1,0,0], directions. Given a set of these points, we use singular value decomposition to estimate a basis for the horizon plane, ignoring images that are clearly in the wrong orientation (Figure 1).

We combine our approach with images and camera models provided in the 1DSfM datasets [9]. Starting with 15 high-quality SfM models, we filtered out anomalous images, fit and manually validated a global horizon line for each model, and then projected the line back into each image. The resulting dataset, HLW, contains 19,809 images. From each model, we hold out 100 images at random, including holding out two models completely, resulting in 2018 images to be used for evaluation.
2.3. Comparisons with Existing Datasets

A montage of sample images from HLW are shown next to a montage of sample images from ECD in Figure 2. Even when considering this small set of images, there is clearly much greater diversity of scene types in HLW (e.g., zoomed in view of a statue, elevated view of a city). The scenes in ECD consist primarily of urban images with large buildings in the background.

In addition, we analyze the distributions of horizon lines for images in HLW versus ECD/YUD. We parameterize the horizon line using the parametric equation of a line, $x \cos \theta + y \sin \theta = \rho$, where $\rho$ is the perpendicular distance from the origin to the line and $\theta$ is the angle the line makes with the horizontal axis. Figure 3 shows the joint distribution over $\theta$ ($x$-axis) and $\rho$ ($y$-axis). As observed, images in HLW span a much larger range of possible horizon lines.

3. Evaluation

We evaluate the current state-of-the-art method by Lezama et al. [8] on HLW. The standard error metric used for horizon line detection is the maximum distance from the detection to the ground truth in image space, normalized by the height of the image, which we refer to as horizon detection error. This is often reported for a set of images as the area under the curve of the cumulative histogram of errors. Barinova et al [1] motivate the use of horizon detection error as the standard accuracy measure for automatic vanishing point detection algorithms. Figure 4 visualizes the result. The large relative performance difference compared to other benchmarks highlights the challenging nature of our dataset.

4. Conclusion

We introduced Horizon Lines in the Wild (HLW), a new dataset for single image horizon line estimation, to address the limitations of existing horizon line detection datasets. HLW is several orders of magnitude larger than any existing dataset for horizon line detection, has a much wider variety of scenes and camera perspectives, and wasn’t constructed to highlight the value of any particular geometric cue. While HLW is not perfect (limited to the types of im-
ages that survive SfM, propagated errors from poorly reg-
istered images), our hope is that it will continue to drive
advances on this important problem in the future. The
dataset is available for download at http://cs.uky.
edu/~jacobs/data/.

References

[1] O. Barinova, V. Lempitsky, E. Tretiak, and P. Kohli. Geomet-
ric image parsing in man-made environments. In European
Conference on Computer Vision, 2010. 2, 3

[2] J. M. Coughlan and A. L. Yuille. Manhattan world: Compass
direction from a single image by bayesian inference. In IEEE
International Conference on Computer Vision, 1999. 2

[3] A. Criminisi, I. Reid, and A. Zisserman. Single view metro-
logy. International Journal of Computer Vision, 40(2):123–
148, 2000. 1

[4] P. Denis, J. Elder, and F. Estrada. Efficient edge-based meth-
ods for estimating manhattan frames in urban imagery. In Eu-
ropean Conference on Computer Vision, 2008. 2

[5] D. Hoiem, A. A. Efros, and M. Hebert. Putting objects in per-
spective. International Journal of Computer Vision, 80(1):3–
15, 2008. 1

[6] A. Kendall, M. Grimes, and R. Cipolla. Convolutional net-
works for real-time 6-dof camera relocalization. In IEEE In-
ternational Conference on Computer Vision, 2015. 2

[7] H. Lee, E. Shechtman, J. Wang, and S. Lee. Automatic upright
adjustment of photographs. In IEEE Conference on Computer
Vision and Pattern Recognition, 2012. 1

[8] J. Lezama, R. Grompone von Gioi, G. Randall, and J.-M.
Morel. Finding vanishing points via point alignments in image
primal and dual domains. In IEEE Conference on Computer
Vision and Pattern Recognition, 2014. 2, 3

[9] K. Wilson and N. Snavely. Robust global translations with
1dsfm. In European Conference on Computer Vision, 2014. 2