A Simple and Effective Method to Detect Orthogonal Vanishing Points in Uncalibrated Images of Man-Made Environments

G. Simon, A. Fond & M.-O. Berger

Université de Lorraine, INRIA Nancy - Grand Est, LORIA - MAGRIT, France

Abstract

This paper presents an effective and easy-to-implement algorithm to compute orthogonal vanishing points in uncalibrated images of man-made scenes. The main contribution is to estimate the zenith and the horizon line before *detecting the vanishing points, using simple properties of the central projection and exploiting accumulations of oriented segments around the horizon. Our method is fast and yields an accuracy comparable, and even better in some cases, to that of state-of-the-art algorithms.*

Categories and Subject Descriptors (according to ACM CCS): I.4.1 [Image Processing & Computer Vision]: —Imaging geometry

1. Introduction

Finding orthogonal vanishing points (VPs) in a photography has many potential applications in computer graphics, including perspective correction, architecture reconstruction and texture extraction. Surprisingly, while this problem has been extensively studied in the literature, manual solutions are still used in most existing software. Figure 1 shows an example of using the "Photo Match" tool in Trimble SketchUp ©. The user has to adjust two green bars and two red bars, so that the extension lines of the bars intersect at two orthogonal VPs. Once this calibration step is performed, the scene geometry can be recovered up to a scaling factor, by drawing 3D shapes directly on the image.

Camera self-calibration from monocular images generally follows two steps [KZ02, Tar09, XOH13, LGRM14]. First, lines are grouped into pencils, whose centers are considered as potential VPs. Then, an orthogonality measure is evaluated for every triplet of VPs and the most plausible triplet is taken as the so-called Manhattan frame, and used to compute the focal length. A drawback of this approach is that complex and time-consuming techniques have to be used to solve the general problem of VP detection, while only three particular VPs are finally used. In this paper, we show that, under certain conditions that are satisfied in many practical situations, the horizon line (shown in yellow in Fig. 1), can be found directly using a well known property of the central projection. Once the horizon line is known, finding two or three orthogonal VPs, along with the focal length and all horizontal VPs (hVPs) is a fast and simple procedure. Experiments made on three public datasets show competitive results with regard to state-of-the-art methods.

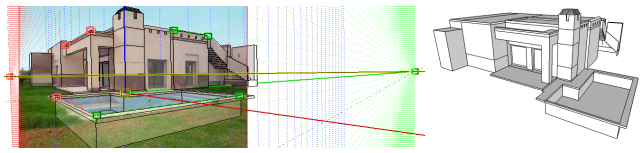


Figure 1: *Manual adjustment of orthogonal VPs and 3D modeling by photo matching in Trimble SketchUp ©.*

2. Related Work

There is a vast literature on the problem of VP detection in uncalibrated images. [KZ02] use the Expectation-Maximization (EM) algorithm, which iteratively estimates the coordinates of VPs as well as the probabilities of individual line segments belonging to a particular vanishing direction. Although EM is often sensitive to initialization, a very rough procedure is used for this step. Several attempts have been made to obtain a more accurate initialization. [Tar09] estimate VP hypotheses in the image plane using pairs of edges and compute consensus sets using the J-linkage algorithm. The same framework is used in [XOH13], though a probabilistic consistency measure is proposed, which shows better performance. In [LGRM14], the problem is solved in the dual domain where pencil of lines are collinear points. The use of a robust point alignment detector based on the *a contrario* framework leads to the potential VPs. Unfortunately, since any two parallel lines intersect in a VP, lines grouping remains a difficult problem which can yield a large number of VPs, including many false positive. To tackle this issue, a few works enforce some geometric constraints during VP detection. [WH12] present a RANSAC-based approach using a solution for estimating three orthogonal VPs and focal length from a set of four lines, aligned with either two or three orthogonal directions. [TBKL12] use different layers of geometric primitives, including the horizon line, and construct energy functions for each

DOI: 10.2312/egsh.20161008

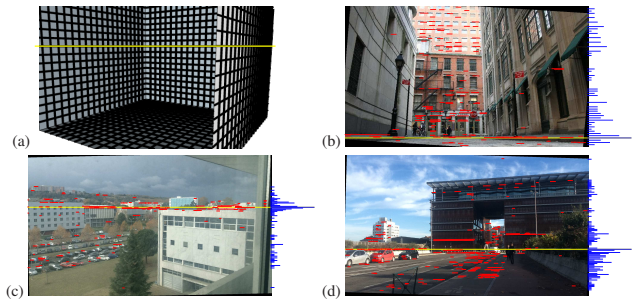


Figure 2: *Main trick of our method: any horizontal line at the height of the camera's optical center projects to the horizon line.*

layer with respect to hypothesized VPs. VPs are then estimated by minimizing the overall energy across layers.

All the aforementioned works, except [LGRM14], incorporate variations of the RANSAC algorithm and/or iterative optimization techniques. This may result in high computational cost and/or complex tuning of parameters. The method proposed in [LGRM14] is deterministic and non-iterative, but it still suffers from a high computational cost as shown in our experiments.

3. Principle of the Method

Our method is based on a simple observation that any horizontal line at the height of the camera's optical center projects to the horizon line *regardless of its direction* (Fig. 2(a)). Therefore, if an image is rotated so that the horizon line, denoted by $\mathcal{H}$ in the following, is horizontal in the image space, one can expect an accumulation of horizontal line segments at the height of $\mathcal{H}$. Figures 2(b)-(d) show three example images, rotated so that $\mathcal{H}$ is horizontal. Line segments (LSs) were detected in these images using the LSD algorithm [GJMR12], from which horizontal line segments (hLSs) were selected within a tolerance of ± 0.5 degrees (red lines). A histogram of the hLSs' y-coordinates, measured at the centroids of the segments, is displayed on the right side of each image. In Fig. 2(b), a lot of hLSs are located at different heights of the facades in the background, which are nearly parallel to the image plane. However, hLSs detected on the left and right facades, which are nearly perpendicular to the image plane, are mainly located at the height of $\mathcal{H}$, yielding a peak in the histogram at that height. The peak is even more noticeable in Fig. 2(c), where none of the predominant facades are parallel to the image plane.

Doors, windows, floor separation lines but also man-made objects such as cars (Fig. 2(d)), road signs, street furniture, and so on, often appear at eye level, so that the expected accumulation of hLSs around $\mathcal{H}$ is surprisingly observed in many images. In order to quantify more precisely the validity of this statement, we used three public datasets (DSs) : York Urban [DEE08] (102 images of resolution 640×480, 57 outdoors and 45 indoors), Toulouse Vanishing Points [AGC15] (114 images of resolution 1920×1080, 74 outdoors and 40 indoors), and Eurasian Cities [TBKL12] (103 images of various resolutions, all outdoors). Ground truth (GT) data are provided in these DSs, from which the horizon line can be obtained. Fig. 3(a) shows a histogram of 21499 signed distances (divided by the image height) between hLSs' centroids and GT horizon lines (in rotated images), summed over all images in the DSs. The Laplacian-like shape of this histogram is noticeable.

Our method has two prerequisites: (i) The horizon line must pass through the image. This condition is satisfied in most practical situations. For instance, $\mathcal{H}$ is outside the image for only 2% of images in the three datasets. (ii) Line segments must be dense on, or at least near, the horizon line. This condition is generally satisfied in outdoor environments, but it may be violated, especially in indoor scenes. Fig. 3(b),(c) show the histograms of signed distances obtained for the 234 outdoor and (resp.) 85 indoor images in the DSs. The peak is sharper in the first histogram, which suggests that our method is more suited to outdoor scenes (which is confirmed by some results shown in Sec. 5). To reduce failures in cases where the density of LSs around $\mathcal{H}$ is low, we consider the first N modes of the hLSs' y-coordinates, instead of the first one only.

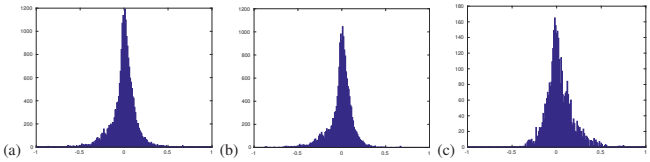


Figure 3: *Histogram of signed distances, summed over (a) all, (b) outdoor and (c) indoor images in the three datasets.*

4. Algorithm

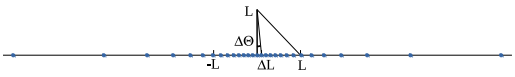
We assume the image size is $H \times W$, the principal point coincides with the image center, and the pixel aspect ratio is one. 2D points are expressed in pixel coordinates with the origin located at the image center, so that the intrinsic matrix $\mathbf{K}$ is a diagonal matrix with diagonal entries $f, f, 1$. Since the horizon line is orthogonal with the line connecting the vertical VP (known as the zenith) and the principal point, finding the rotation that transforms $\mathcal{H}$ to a horizontal line amounts to finding the zenith. Our algorithm consists in five main steps (all parameter values are given in Tab. 1):

1. Detect LSs in the image using [GJMR12].
2. Find the zenith (denoted by $\mathcal{Z}$) by brute-force search with decreasing density sampling, and rotate the LSs so that the line from $\mathcal{Z}$ to the image center is vertical (Sec. 4.1).
3. Select hLSs from the rotated LSs within a tolerance of $\pm\varepsilon$ and compute the first N modes of the hLSs' y-coords (Sec. 3).
4. For each mode, compute potential hVPs and a score (Sec. 4.2).
5. Solve for orthogonal VPs and focal length using $\mathcal{Z}$ and the hVPs corresponding to the highest score (Sec. 4.3).

4.1. Detection of the Zenith

As mentioned by several authors (e.g. in [XOH13]), the zenith usually locates vertically and far from the image center. Therefore, a simple brute-force algorithm can be used to find it reliably. Lines passing through the image center, uniformly spaced by an angle ε in a range $[-\Phi, \Phi]$ around the vertical axis, are successively evaluated as potential directions of $\mathcal{Z}$. In order to

limit the computation time, we use a gradually decreasing density sampling. A line $\mathcal{L}$, defined by a position vector $\mathbf{c}$ (here $\mathbf{c} = (0,0)^T$) and a unit direction vector $\mathbf{d}$, is sampled at positions $\{\mathbf{p} \circ s(k)\}_{-k_\infty \leq k \leq k_\infty}$, where $\mathbf{p}(\lambda) = \mathbf{c} + \lambda\mathbf{d}$, $s(k) = L\tan(k\Delta\Theta)$, with $\Delta\Theta = \arctan(\Delta L/L)$, and $k_\infty = \lfloor \pi/(2\Delta\Theta) \rfloor$. $\mathbf{p} \circ s(k)$ generates samples from $\mathbf{c}$ to a point far on the line, with decreasing density. ΔL and L are two positive constants that allow us to control the initial density ($1/\Delta L = 1/(s(1) - s(0))$) and the scattering of the sample points (half of the samples are between $\mathbf{p}(-L)$ and $\mathbf{p}(L)$).



The use of the tangent function is motivated by the fact that the angle between the optical axis and a vanishing direction in the 3D camera frame is arctangential in the distance between the projected direction and the principal point. Each sample point is scored by the number of LSs consistent with that point. We use as consistency measure the angle between the LS and the line from the centroid of the LS to the sample point, which must be less than ε in absolute value. Sample points whose distance to image center are less than half the image height are not evaluated ($\mathcal{Z}$ is assumed outside the image). Finally, the sample point that gets the highest score is taken as $\mathcal{Z}$. Note that even in the case $\mathcal{Z}$ is at infinity, its direction should be correctly obtained with the highest score reached at $\mathbf{p} \circ s(\pm k_\infty)$.

4.2. Detection of the Horizontal VPs

After step 3., several horizon line hypotheses are available. Let the line $\mathcal{L} : y = y_h$ be one such hypothesis. Then, a potential "dominant" hVP can be computed by sampling $\mathcal{L}$ according to $\mathbf{p} \circ s(k)$, with $\mathbf{c} = (0, y_h)^\top$ and $\mathbf{d} = (1,0)^\top$, and searching for the global maximum of the scores, as proceeded in Sec. 4.1. The dominant hVP may be reached at $\mathbf{p} \circ s(\pm k_\infty)$ in the case of an infinite VP. In order to detect other potential hVPs, we look for peaks in the curve of scores $c(k)$, computed at $\{\mathbf{p} \circ s(k)\}_{k \in [-k_\infty, k_\infty]}$ (an example curve is shown in Fig. 4). More precisely, a median filter of size M is applied to $c(k)$, and the obtained values are subtracted from $c(k)$, yielding a curve $d(k)$ where, in general, peaks corresponding to hVPs are very sharp. Then, all values of $d(k)$ below T times the median of $|d(k)|$ are set to 0, as well as all values in the interval $[k_0 - \lfloor M/2 \rfloor, k_0 + \lfloor M/2 \rfloor]$, where k_0 is the index of the dominant hVP. Finally, we iterate the following step until all values of $d(k)$ are 0: the index $k_i, i > 0$ of the maximum of $d(k)$ is determined and values in the interval $[k_i - \lfloor M/2 \rfloor, k_i + \lfloor M/2 \rfloor]$ are set to 0. All this procedure is repeated for each mode obtained at step 3., and the mode maximizing $c(k_0) + c(k_1)$ (with $c(k_1) = 0$ for modes where only the dominant hVP is detected) is selected as $\mathcal{H}$.

4.3. Camera Self-Calibration

Inputs of the self-calibration procedure are the coordinates $(0, y_z)$ of $\mathcal{Z}$, the height y_h of $\mathcal{H}$, and the positions $\left\{\mathbf{h}_i = (x_i, y_h)^\top = \mathbf{p} \circ s(k_i)\right\}_{0 \leq i < n}$ of $n \geq 1$ hVPs. $\mathcal{Z}$ is considered infinite when $|y_z| > L_\infty$, and a hVP when $|x_i| > L_\infty$. In the case when $\mathcal{Z}$ is finite, the focal length f is computed using the following procedure. If more than one finite hVPs have been found, each pair of them is examined as potential orthogonal VPs.

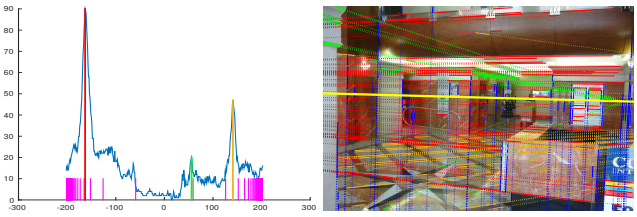


Figure 4: *Left: Curve of scores $c(k)$ obtained for the image on the right. The red, green, orange vertical lines show the positions of three hVPs. The magenta ticks are regularly spaced by W on $s(k)$. Right: LSs consistent with the hVPs (solid lines with the same resp. colors) with their centroids joined to the VPs (dashed lines). LSs consistent with the zenith are represented in blue, and $\mathcal{H}$ in yellow.*

For each pair $(\mathbf{h}_k, \mathbf{h}_l)$, a hypothetical focal length f_{kl} is computed from $\mathbf{v}_k^\top \mathbf{v}_l = 0$, where $\left\{\mathbf{v}_i = \mathbf{K}_{kl}^{-1}\widetilde{\mathbf{h}}_i\right\}_{i=k,l}$ (with $\widetilde{\ }$ denoting homogeneous coordinates), which leads to $f_{kl} = \sqrt{-\mathbf{h}_k \cdot \mathbf{h}_l}$ [WH12]. All pairs for which f_{kl} is not a real number, or is out of the interval [0.28W,3.8W] (similar to e.g. [XOH13]) are discarded. Hypothetical coordinates of the zenith $\mathbf{z}_{kl} = (0, y_{kl})^\top$ are inferred from the remaining pairs, using $\widetilde{\mathbf{z}}_{kl} = \mathbf{K}_{kl}(\mathbf{v}_k \times \mathbf{v}_l)$, and the pair for which $d = |s^{-1}(y_{kl}) - s^{-1}(y_z)|$ is smallest is selected as orthogonal hVPs, provided that $d < D$. If only one finite hVP has been found, the focal length can still be obtained from $\mathcal{Z}$ and $\mathcal{H}$ using the simple formula $f = \sqrt{-y_z y_h}$. However, this procedure is less reliable than the previous one, as the result can not be verified using another VP. Finally, if no finite hVP has been found, the height of $\mathcal{H}$ is not reliable. This situation occurs when a plane parallel to the horizontal axis of the camera is the only source of hLSs. In that case, f can not be obtained, but the (infinite) dominant hVP can still be used along with the zenith to orthorectify the observed plane, up to an aspect ratio. In the case when $\mathcal{Z}$ is infinite, f can not be obtained. Again, vertical planes in the direction of any (finite or infinite) hVP can be orthorectified.

5. Results and Conclusions

Fig. 5 reports example results. The same parameters were used for all experiments shown in this section (Tab. 1), with the exception that a higher value of Φ is used for the Eurasian DS, that contains several images with a high camera roll. K is an integer that determines the density of sampling ($K = 7$ in most experiments).

Section 3		Section 4.1				Section 4.2		Section 4.3	
$B^\dagger$	$H/4$	ε	0.5 deg	L	W	M	2^K	L_∞	$32W$
N	32	Φ	$\pi/32$	ΔL	$W/2^K$	T	4	D	4

†Number of bins in the histogram of hLSs' y-coordinates.

Table 1: *Parameters of the method ($\Phi = \pi/16$ for Eurasian).*

The horizon estimation is evaluated on the same DSs (York Urban and Eurasian Cities) and with the same protocol of [VZ12, XOH13, LGRM14], and in fact we report their results for all but the proposed methods. Fig. 6 shows the percentage of test images that achieve an error smaller than the values on the x-axis. Following [VZ12], we report a numerical value as the percentage of area under the curve (AUC) in the subset $[0, 0.25] \times [0, 1]$. Our method

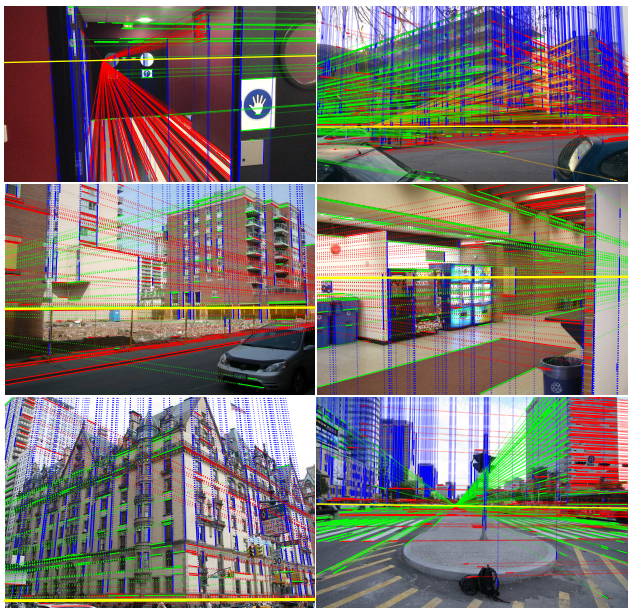
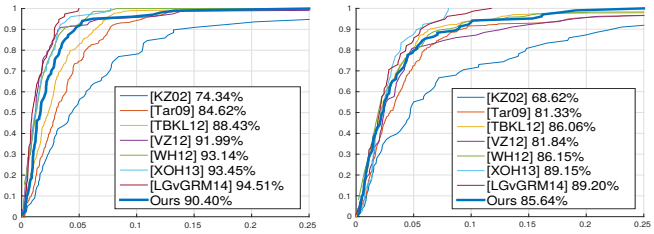


Figure 5: *Example results. Graphic conventions are the same as in Fig. 4(right). VPs of the Manhattan frame are shown in R, G, B.*

is competitive with state-of-the-art algorithms, and especially as we use the same (unlearned) parameters for the three DSs, while in e.g. [TBKL12, WH12, LGRM14], parameters are learned for each DS, using their 25 first images.

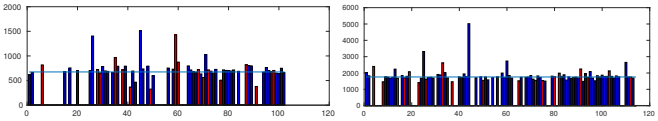


York Urban DS Eurasian Cities DS

Figure 6: *Quantitative evaluation of the horizon line.*

Computation of f is assessed with the York and Toulouse DSs (GT values of f are not included in the Eurasian DS). Fig. 7 shows the value of f in each image, with a blue or a red bar depending on whether f was obtained from three or (resp.) two orthogonal VPs (a blank is left where f could not be obtained). The relative error of the median value of f is 4.4% for the York DS, and -0.01% for the Toulouse DS, whose images have three times higher resolution.

Tab. 2 shows the computation times (in sec) and AUC (separated between outdoor and indoor images) obtained for the York and Toulouse DSs, for K varying from 4 to 8. The method was implemented in Matlab and run on I7-3520M CPU. While the computation time roughly doubles each time K is incremented by one (that is, increases linearly with $1/\Delta L$), the AUC does not vary noticeably after $K = 6$. In all experiments, our method shows a significantly higher AUC for outdoor than for indoor scenes. As an example,



York Urban DS Toulouse VP DS

Figure 7: *Focal lengths obtained from two (red bars) or three (blue bars) VPs. The ground truth value is shown as a horizontal line.*

these results are compared with those of one of the most recent state-of-the-art methods [LGRM14], using the default, unlearned parameters of [LGRM14] and the non-accelerated, high-precision, mode of execution (Tab. 2, last row). The mean execution time for the York DS is practically the same as with our algorithm when $K = 7$ (but $7.8\times$ higher when $K = 4$). However, computation times are very high for the Toulouse DS (up to 25 min for one frame), whereas the accuracy is poor (AUC of 79% against 87.3% with our method, which is also $50\times$ faster in average when $K = 4$). Thus, while detecting a lot of line segments helps our method to achieve high performance, this can penalize other methods such as [LGRM14], for which lines grouping becomes a harder task. These experiments show that our method is computationally fast and efficient in various environments, and particularly outdoors.

	York Urban DS (640 × 480)							**Toulouse DS** (1920 × 1080)						
K	mean	std	min	max	AUC	out.	in.	mean	std	min	max	AUC	out.	in.
4	1.6	0.7	0.3	3.8	85.7	87.5	84.5	5.2	2.8	0.4	11.7	87.3	91.0	81.4
5	3.0	1.3	0.4	8.0	86.6	89.4	84.1	9.4	5.1	0.5	21.7	87.3	91.4	80.5
6	6.1	2.8	0.8	15.9	89.1	90.8	88.0	21.0	11.2	1.0	46.6	89.5	92.4	84.9
7	12.1	5.5	1.6	35.0	90.4	91.6	89.5	38.9	22.1	1.7	90.6	88.7	92.8	82.0
8	24.2	11.1	3.7	64.0	88.7	92.1	85.4	72.8	41.0	2.8	167.4	89.0	92.4	83.6
[LGRM14]	12.5	17.6	1.1	119.9	88.7	89.2	88.7	257.8	338.8	1.2	1476.3	79.0	84.6	69.7

Table 2: *Computation times and AUC against the sampling density.*

References

[AGC15] ANGLADON V., GASPARINI S., CHARVILLAT V.: The Toulouse Vanishing Points Dataset. In *Mult. Syst. Conf.* (2015). 2

[DEE08] DENIS P., ELDER J. H., ESTRADA F. J.: Efficient edge-based methods for estimating manhattan frames in urban imagery. In *ECCV* (2008). 2

[GJMR12] GROMPONE VON GIOI R., JAKUBOWICZ J., MOREL J.-M., RANDALL G.: LSD: a Line Segment Detector. *IPOL 2* (2012). 2

[KZ02] KOSECKA J., ZHANG W.: Video compass. In *ECCV* (2002). 1

[LGRM14] LEZAMA J., GROMPONE VON GIOI R., RANDALL G., MOREL J.-M.: Finding vanishing points via point alignments in image primal and dual domains. In *CVPR* (2014). URL: http://dev.ipol.im/~jlezama/vanishing_points/. 1, 2, 3, 4

[Tar09] TARDIF J.-P.: Non-iterative approach for fast and accurate vanishing point detection. In *ICCV* (2009). 1

[TBKL12] TRETYAK E., BARINOVA O., KOHLI P., LEMPITSKY V.: Geometric image parsing in man-made environments. *IJCV 97*, 3 (2012). 1, 2, 4

[VZ12] VEDALDI A., ZISSERMAN A.: Self-similar sketch. In *ECCV* (2012). 3

[WH12] WILDENAUER H., HANBURY A.: Robust camera self-calibration from monocular images of Manhattan worlds. In *CVPR* (2012). 1, 3, 4

[XOH13] XU Y., OH S., HOOGS A.: A minimum error vanishing point detection approach for uncalibrated monocular images of man-made environments. In *CVPR* (2013). 1, 2, 3