Robust Camera Self-Calibration from Monocular Images of Manhattan Worlds

Horst Wildenauer, Allan Hanbury
Vienna University of Technlology
Institute of Software Technology and Interactive Systems
horst.wildenauer@gmail.com, hanbury@ifs.tuwien.ac.at

Abstract

We focus on the detection of orthogonal vanishing points using line segments extracted from a single view, and using these for camera self-calibration. Recent methods view this problem as a two-stage process. Vanishing points are extracted through line segment clustering and subsequently likely orthogonal candidates are selected for calibration. Unfortunately, such an approach is easily distracted by the presence of clutter. Furthermore, geometric constraints imposed by the camera and scene orthogonality are not enforced during detection, leading to inaccurate results which are often inadmissible for calibration. To overcome these limitations, we present a RANSAC-based approach using a minimal solution for estimating three orthogonal vanishing points and focal length from a set of four lines, aligned with either two or three orthogonal directions. In addition, we propose to refine the estimates using an efficient and robust Maximum Likelihood Estimator. Extensive experiments on standard datasets show that our contributions result in significant improvements over the state-of-the-art.

1. Introduction

In their seminal work, Coughlan & Yuille [2] pointed out that imagery of man-made environments can be often characterized by a predominance of orthogonal structures, coining the name *Manhattan world*. Such orthogonal structures provide invaluable cues about a camera's orientation w.r.t. the world coordinate frame and its internal parameters (mostly the focal length). This information is usually extracted from three finite, mutually orthogonal vanishing points [1]. The detection of vanishing points has been applied to scene understanding and single view reconstruction of indoor scenes [6, 10], architecture reconstruction [17], detection of rectangular structures [9, 12] and multi-view stereo [4].

1.1. Related Work & Contributions

We concern ourselves with the detection of orthogonal vanishing points using sets of line segments extracted from a single, uncalibrated view. In [18], Rother suggests an exhaustive search over vanishing point hypotheses obtained from all possible line intersections to find dominant orthogonal directions and the most plausible camera parameters. In practice Rother's method suffers from high computational cost, which other methods try to reduce by a two-step process. First, vanishing points are estimated from concurrent line segments either through iterative procedures [14, 15] or by simultaneous clustering [8, 20]. Then, from a plausible orthogonal vanishing point triplet the camera calibration is estimated [1, 7, 11].

Iterative methods perform RANSAC-based clustering of line segments, making use of line intersections to generate vanishing point hypotheses. After the vanishing point with maximal support is found, its consensus set is removed and the procedure is repeated in search of remaining vanishing points. This approach suffers from severe limitations: (a) Line segments are often compatible with more than one vanishing point. Depending on spatial tolerance, they can be either wrongly fused into one cluster, or multiple detections for one vanishing point occur. (b) The vanishing points are in general not compliant with constraints imposed by the camera and the scene orthogonality, causing inaccuracies or complete failure in calibration.

Recently, Tardif [20] attacked the first problem, using a robust clustering technique specifically designed for treatment of multiple models. Among others, Kosecka & Zhang [8] suggested simultaneous optimization of vanishing points using Expectation-Maximization. However, orthogonality and camera constraints are not enforced and it is not clear if their initialization based on line segment orientation finds all relevant vanishing points.

Our RANSAC-based approach addresses both limitations in a unified framework. We explicitly exploit orthogonality and camera constraints during hypotheses generation, and thereby make better use of the available data. After the RANSAC stage the quality of the results can be refined, for

978-1-4673-1228-8/12/$31.00 ©2012 IEEE

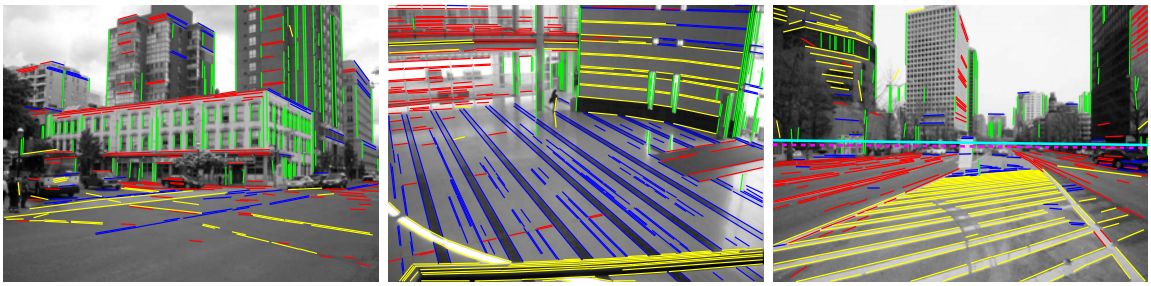

Figure 1. Exemplary results: Line segments are classified as belonging to three orthogonal directions (red, green, blue) or outliers (yellow). The rightmost image additionally shows the estimated horizon (dashed magenta line) together with the ground truth (solid cyan line).

which we utilize a robust mixture model-based Maximum Likelihood Estimator. Fig. 1 shows example results of the proposed approach. To summarize, the contributions of this paper are:

- We propose a minimal solution for the problem of estimating orthogonal vanishing points and focal length from a set of four lines, being aligned with either two, or three orthogonal directions. Based on this, we deploy a RANSAC-based approach that exploits orthogonality to efficiently classify line segments into parallel and mutually orthogonal groups.
- Following the idea of [14], we refine vanishing point positions and focal length in a continuous fashion. To this end, we propose a computationally efficient and robust Maximum Likelihood Estimator utilizing the heavy-tailed Cauchy distribution.
- In extensive experiments we show that our contributions result in a significant improvement of the state-of-the-art in camera calibration from vanishing points. Furthermore, we demonstrate that our approach gives excellent results when applied to the problem of horizon estimation for non-Manhattan scenery.

1.2. Notation

Vectors are typeset bold, e.g. $\mathbf{v}$ and matrices in capital, *sans-serif* letters, e.g. the rotation matrix R. Homogeneous 3-vectors are used to represent points and lines in the projective plane $\mathbb{P}^2$. E.g., for a line or a vanishing point we write $\mathbf{l}_i = (l_{i1}, l_{i2}, l_{i3})^\mathsf{T}$ and $\mathbf{v}_i = (v_{i1}, v_{i2}, v_{i3})^\mathsf{T}$. The intersection of two lines is given by $\mathbf{l}_i \times \mathbf{l}_j$, where $\times$ denotes the cross-product. Likewise a line connecting points $\mathbf{x}_i$, $\mathbf{x}_j$ is computed as $\mathbf{x}_i \times \mathbf{x}_j$. For the unsigned orthogonal distance of a point to a line, we write $d(\mathbf{x}_j, \mathbf{l}_i) = |\mathbf{l}_i^\mathsf{T}\mathbf{x}_j / (x_{j3}\sqrt{l_{i1}^2 + l_{i2}^2})|$.

Unless stated otherwise, vanishing points will be always denoted by $\mathbf{v}$. Image line segments and the associated implicit line representation will be designated by the symbols s and $\mathbf{l}$ respectively.

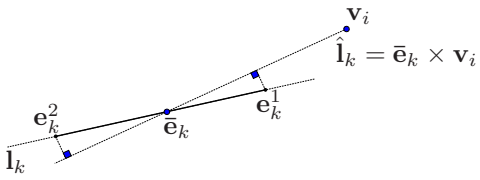

Figure 2. Illustration of the consistency measure in use. The ideal line passes through $\mathbf{v}_i$ and intersects the line segments centroid.

2. Line Segment Error

To assess the degree to which a line segment is consistent with a given vanishing point, we adopt the image-based consistency measure suggested in [22]. Here the error of a line segment s_k w.r.t. a vanishing point $\mathbf{v}_i$ is quantified by the orthogonal distance of one line segment endpoint (e.g., $\mathbf{e}_k^1$) to a line passing through both the vanishing point and the centroid $\bar{\mathbf{e}}_k$ of the line segment (see Fig. 2). Or, more formally: $\mathrm{dist}(\mathrm{s}_k, \mathbf{v}_i) = d(\mathbf{e}_k^1, \bar{\mathbf{e}}_k \times \mathbf{v}_i)$.

Besides treating finite and infinite vanishing points in the same way, it measures error in the image, which is usually preferred in computer vision [5, 18]. In [20], Tardif experimentally demonstrated the proposed measure to give superior results when comparing with approaches based on the Gaussian-sphere (see e.g., [8]). Furthermore, it was shown to be a computationally efficient approximation to Liebowitz's error-in-endpoints model [11],

3. Camera Calibration from Vanishing Points

For a camera with zero skew and unit aspect ratio, it is well known that focal length f and principal point $\mathbf{p}$ can be determined from three finite, mutually orthogonal vanishing points [1]. In the image, such a triad of vanishing points form the vertices of a self-polar triangle with orthocenter inside the triangle [11, 5]. Using the orthogonality constraints imposed by the vanishing points $\mathbf{v}_i, i = 1, 2, 3$:

$$\mathbf{v}_1^\mathsf{T}\omega\mathbf{v}_2 = \mathbf{v}_1^\mathsf{T}\omega\mathbf{v}_3 = \mathbf{v}_2^\mathsf{T}\omega\mathbf{v}_3 = 0, \qquad (1)$$

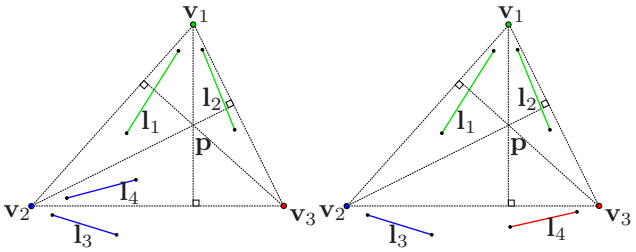

Figure 3. Examples of two admissible configurations of 4 lines. Left: Two line pairs meet in two vanishing points. Right: One pair meets in one vanishing point, the other lines pass through the remaining two vanishing points.

where $\omega = \mathsf{K}^{-\mathsf{T}}\mathsf{K}^{-1}$ denotes the *image of the absolute conic* (IAC) and K the camera calibration matrix, the principal point is computed as the triangle orthocenter. With $\mathbf{p}$ known, f can be recovered from any of the three constraints.

Unfortunately humans tend to take pictures with little camera inclination and often a vanishing point will be far from the image center. In such situations, computing the principal point is ill-conditioned [11] and it is common to set it to the image center [17]. This practice, as Kanatani & Sugayawa [7] demonstrated, stably provides better focal length estimates.

In our work, we will exploit the mild assumption of a known principal point directly in the vanishing point detection stage, computing the focal length and three orthogonal vanishing points from four image lines. The minimal solution to this problem is explained in the next section.

3.1. Minimal Solution

For the discussion of the minimal solution we assume the principle point to be known and, without loss of generality, write the camera calibration matrix in its simplified form as

$$\mathsf{K}_f = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2}$$

In this case, the IAC has the form of a diagonal matrix $\omega_f = \mathrm{diag}(1/f^2, 1/f^2, 1)$. Let us further suppose we are given four lines $\mathbf{l}_i, i = 1, 2, 3, 4$ obtained from line segments detected in the image. It turns out that one needs to consider 9 configurations of line directions, admissible for computing three orthogonal vanishing points and the focal length. Two cases can be distinguished: In the first case it is assumed that the lines are drawn in two pairs, each pair meeting in a vanishing point consistent with a different direction. This gives 3 configurations. In the second case one pair of lines is compatible with one direction while the remaining two lines are along the other directions, generating 6 additional configurations. See Fig. 3 for examples of these two cases. Note that the assignment of a line to a specific direction is not relevant for the enumeration of admissible configurations — it is only relevant which pair of lines meets in a common vanishing point.

In the first case we may suppose that lines are paired as $(\mathbf{l}_1, \mathbf{l}_2)$ and $(\mathbf{l}_3, \mathbf{l}_4)$ so that two vanishing points are given by

$$\mathbf{v}_1 = \mathbf{l}_1 \times \mathbf{l}_2 \;\text{ and }\; \mathbf{v}_2 = \mathbf{l}_3 \times \mathbf{l}_4. \tag{3}$$

Putting (3) into the orthogonality constraint (1) one can easily obtain the focal length as [1]

$$f = \sqrt{\frac{v_{11}v_{21} + v_{12}v_{22}}{-v_{13}v_{23}}}. \tag{4}$$

Since f (and hence K) is known, the third orthogonal vanishing point can be readily constructed:

$$\mathbf{v}_3 = \mathsf{K}_f \left(\left(\mathsf{K}_f^{-1}\mathbf{v}_1 \right) \times \left(\mathsf{K}_f^{-1}\mathbf{v}_2 \right) \right). \tag{5}$$

For the treatment of the second case let us assume that a pair of lines $(\mathbf{l}_1, \mathbf{l}_2)$ meets in a common vanishing point $\mathbf{v}_1 = \mathbf{l}_1 \times \mathbf{l}_2$. The remaining lines $\mathbf{l}_3$ and $\mathbf{l}_4$ are concurrent with $\mathbf{v}_2$ and $\mathbf{v}_3$ respectively. The vanishing line polar to $\mathbf{v}_1$ [5] can then be written as

$$\mathbf{h} = \omega_f \mathbf{v}_1 = \left(\frac{v_{11}}{f^2}, \frac{v_{12}}{f^2}, v_{13} \right)^{\mathsf{T}}, \tag{6}$$

and the two vanishing points orthogonal to $\mathbf{v}_1$ are expressed as the intersection of the remaining lines with $\mathbf{h}$:

$$\mathbf{v}_2 = \mathbf{h} \times \mathbf{l}_3 \;\text{ and }\; \mathbf{v}_3 = \mathbf{h} \times \mathbf{l}_4. \tag{7}$$

Now we have to find $f > 0$ such that the orthogonality between $\mathbf{v}_2$ and $\mathbf{v}_3$ is satisfied. Substituting (6) into (7) we may rewrite the constraint $\mathbf{v}_2^{\mathsf{T}}\omega_f\mathbf{v}_3 = 0$ as

$$\frac{f^4 a + f^2 b + c}{f^6} = 0, \tag{8}$$

$$\begin{aligned} \text{with } a =& v_{13}^2 l_{31} l_{41} + v_{13}^2 l_{32} l_{42} \\ b =& v_{12}^2 l_{31} l_{41} - v_{11} v_{12} l_{32} l_{41} - v_{11} v_{13} l_{33} l_{41} - \\ & v_{11} v_{12} l_{31} l_{42} - v_{12} v_{13} l_{33} l_{42} - v_{11} v_{13} l_{31} l_{43} - \\ & v_{12} v_{13} l_{32} l_{43} + v_{11}^2 l_{32} l_{42} \\ c =& v_{11} l_{33} l_{43} + v_{12}^2 l_{33} l_{43}. \end{aligned}$$

As one can see, the numerator in (8) is a fourth-degree polynomial with even terms only. This can be interpreted as quadratic in the unknown f^2 which can be easily solved for. In general, at most two real solutions $f > 0$ are obtained. Substituting f back into (6) the vanishing line and subsequently from (7) the remaining vanishing points are determined. Note that the vanishing point triangle can still be constructed for real $f^2 < 0$. However, in this case the principal point does not lie inside the triangle, as required.

4. RANSAC

Using the results from Sec. 3.1 we suggest a RANSAC-based approach that efficiently generates hypotheses for three orthogonal vanishing points in one sample step. In a standard implementation, one repeatedly draws a *minimal sample set* (MSS) of 4 line segments and evaluates the consensus sets for all hypotheses generated by the 9 configurations. Here, a line segment s_k is classified as inlier w. r. t. the hypothesis H, if it has an error $\mathrm{dist}(\mathrm{s}_k, \mathbf{v}_i^H)$ (see Sec. 2) smaller than a predefined threshold t for any of the three vanishing points $\mathbf{v}_i^H, i = 1, 2, 3$.

To avoid the wasteful evaluation of superfluous hypotheses, we filter out likely incorrect solutions (e.g., when four lines meet in only one vanishing point) during hypothesis generation. To this end, we sample $4 + 1$ line segments, rejecting hypotheses for which the 5th line segment is an outlier. In addition, hypotheses are discarded if a vanishing point lies on a line segment, or if a line segment is compatible with more than one vanishing point. For a valid vanishing point triplet, the latter only happens when a line segment lies in the vicinity of a vanishing line [18]. In practice this is negligible as it occurs only in a small number of cases. Overall, the rejection scheme reduces the number of hypotheses to be tested on the data by a factor of ≈ 7.

4.1. Number of Samples

The number of trials k_{trial} necessary to select an all inlier sample at least once with user-defined failure probability α is given by $k_{trial} = \frac{\log(\alpha)}{\log(1-P)}$, where P is the probability of taking an uncontaminated sample from the data. To determine P one has to consider that the MSS is drawn from three different sets of lines consistent with the orthogonal vanishing points and one outlier set. That is, we have a multi-type population.

Let N denote the number of lines, and $N_i, i = 1, 2, 3$ the individual inlier population sizes. The number of lines taken from the i-th population is given by m_{ij}, with $\sum_{i=1}^{3} m_{ij} = 4$. In our problem we have $m_{ij} \in \{0, 1, 2\}$ (see Sec. 3.1), resulting in six possibilities of drawing the MSS of four lines from three populations (for simplicity, we exclude the additionally sampled 5th line from the discussion). Using the multivariate hypergeometric distribution we get

$$P = \frac{1}{\binom{N}{4}} \sum_{j=1}^{6} \prod_{i=1}^{3} \binom{N_i}{m_{ij}}. \tag{9}$$

However, a-priori knowledge about expected population sizes is not available in our setting. Also, (9) cannot be used to adaptively estimate the necessary number of iterations during RANSAC execution. The reason is that the encountered consensus set will be contaminated by outliers, causing erroneous estimates for individual population sizes and hence P. In practice one can get around this problem by choosing conservative values for k_{trial}. As we have detailed earlier in the section, our algorithm effectively rejects a large number of geometrically implausible solutions already in the hypothesis generation stage. In experiments, we found that the simple strategy of stopping after $k_{data} = 500$ valid hypotheses have been evaluated provided a good balance between accuracy and run-time.

5. Mixture Model

Once the dominant orthogonal vanishing points $\mathbf{v}_i$ as well as f (and thus K_f) have been found using RANSAC, a camera rotation matrix R can be obtained straightforwardly [18]:

$$\mathsf{R} = \left(\pm \frac{\mathsf{K}_f^{-1}\mathbf{v}_1}{\|\mathsf{K}_f^{-1}\mathbf{v}_1\|}, \pm \frac{\mathsf{K}_f^{-1}\mathbf{v}_2}{\|\mathsf{K}_f^{-1}\mathbf{v}_2\|}, \pm \frac{\mathsf{K}_f^{-1}\mathbf{v}_3}{\|\mathsf{K}_f^{-1}\mathbf{v}_3\|} \right) \quad \text{s.t. } \det(\mathsf{R}) = 1. \tag{10}$$

To refine these estimates we employ a mixture model similar to the one presented in [19]. Each line segment is assumed to be created by four causes: (1)–(3) It aligns with one of three Manhattan directions, (4) It is an outlier. More formally, let us assume we are given n line segments $\mathrm{s}_k, k = 1 \ldots n$. Let $\Psi = \{f, \mathsf{R}\}$ be the set of parameters determining the image position of the vanishing points $\mathbf{v}_i(\Psi) = \mathsf{K}_f \mathbf{r}_i$, with $\mathbf{r}_i$ being the i-th column of the rotation matrix. We model the likelihood of a line segment being caused by the Manhattan frame as a mixture model:

$$P(\mathrm{s}_k|\Psi) = \sum_{i=1}^{3} \theta_i P(\mathrm{s}_k|\mathbf{v}_i(\Psi)) + \theta_4 P(\mathrm{s}_k|O), \tag{11}$$

where $P(\mathrm{s}_k|\mathbf{v}_i(\Psi))$ is the likelihood of a line segment belonging to a particular vanishing point and $P(\mathrm{s}_k|O)$ models the contribution of the outlier process. The so-called mixing coefficients $\Theta = \{\theta_{i=1\ldots4} | \theta_i \geq 0 \wedge \sum_{i=1}^{4} \theta_i = 1\}$ can be thought of as priors on the proportion of line segments being generated by each of the four causes.

Assuming conditional independence between line segments (c.f. [2]), Ψ can be estimated using a maximum likelihood estimator of the form

$$\Psi^\star = \arg\max_{\Psi,\Theta} \sum_k \log P(\mathrm{s}_k|\Psi, \Theta). \tag{12}$$

In [14] the errors in line segment endpoints are assumed to be i.i.d. zero-mean Gaussians and the maximization of (12) was conducted using Expectation Maximization (EM). It is not clear however, if the Gaussian is able to model frequently encountered errors caused by radial distortion, or imperfections in the line segment extraction process. Therefore we extracted approximately 2000 line segments from 5 annotated images and assigned them to known ground truth

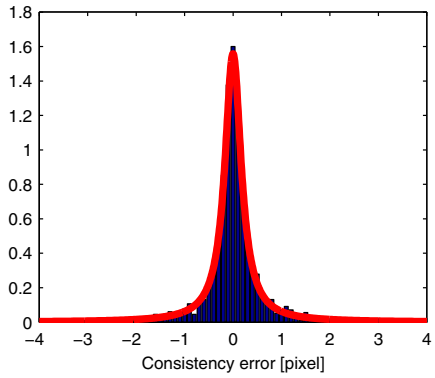

Figure 4. Empirical density of the image-based consistency measure for line segments with fitted Cauchy-distribution

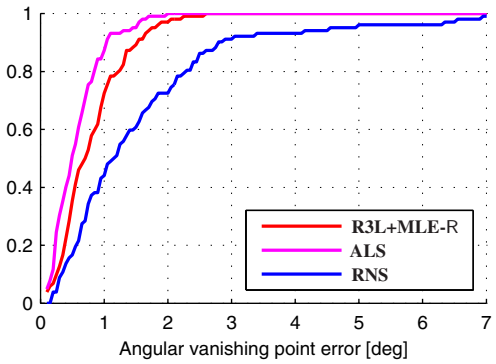

Figure 5. Cumulative histogram of mean angular vanishing point error (YUD). The ordinate shows the fraction of trials with angular errors less than the respective value on the the abscissa.

vanishing points using the error measure from Section 2. The empirical error distribution obtained has heavier tails than a Gaussian and resembles a Cauchy distribution (see Fig. 4). Thus, we model the line segment likelihood as

$$P(\mathrm{s}_k|\mathbf{v}_i) = \frac{1}{\pi}\left(\frac{\gamma}{\mathrm{dist}(\mathrm{s}_k, \mathbf{v_i})^2 + \gamma^2}\right), \qquad (13)$$

where γ is the scale parameter specifying the width of the of the likelihood function. Due to the heavy tails of the Cauchy distribution, the modelling of the outlier process is of lesser importance and we found it to be sufficient to add a small constant component (the Cauchy distribution evaluated at 5 pixel distance) to ensure numerical stability.

For the maximisation of the log-likelihood, we use the BFGS algorithm [16], which is efficient due to the simple form of the line segment likelihood. The rotation matrix R is parametrized by the *exponential map* as described in [5]. The priors Θ are initialized using population size estimates from RANSAC's inlier set and are kept fixed during optimization. In preliminary experiments not reported here, we compared our approach with the Gaussian error model and found it to give more stable results for a wider range of scale parameter values. Also, the BFGS optimization performed around twice as rapidly as the EM algorithm.

6. Experimental Results

Our methods were implemented in MATLAB, with RANSAC (Sec. 4) and the optimization of the mixture model (Sec. 5) being done in C++. For line segment detection we employ the Canny detector, followed by edge linking and polygonalization into straight edge segments. Resulting candidates are refined by a Total Least Squares fit and segments shorter than 20 pixels are rejected.

6.1. Manhattan Frame Detection

We evaluated the proposed algorithms on the task of finding three orthogonal vanishing points using the York Urban Database (YUD). It consists of 102 images of urban environments taken by a calibrated camera. Ground truth vanishing points were individually computed from manually annotated image lines, using a Gaussian sphere-based method followed by orthogonalization [3]. The following variants of the methods described in our paper were tested:
R3L+MLE-R: We assume a calibrated camera and estimate its rotation with our BFGS-based algorithm (Sec. 5). For initialization of R we implemented the RANSAC approach described in [13] adopting the consistency measure from Sec. 2 to determine the consensus set.
R4L: Estimation of focal length f and camera rotation R using the 4-line RANSAC introduced in Sec. 4.
R4L+MLE-fR: Refinement of f and R with our BFGS-based method, initialized by **R4L**.

For all tests on the YUD the RANSAC inlier threshold was set to $t = 0.5$ pixel and the scale parameters of the Cauchy-based likelihood functions were set to $\gamma = 0.3$. Tests with **R4L+MLE-**fR were conducted with the principal point set to the image center. Each test was repeated 100 times and we report the cumulative results.

In the case of rotation estimation for a calibrated camera (**R3L+MLE-**R) we compare with the approaches **ALS** and **RNS** from Mirzae & Roumeliotis [13] (results were provided to us by the authors). Fig. 5 shows the cumulative histogram for the mean angular deviation from a ground truth vanishing point triplet. **RNS** denotes their robust approach using automatically detected line segments. **ALS** denotes the best possible result, operating on labelled ground truth lines. As one can see, our approach is almost as good as **ALS** and detects all orthogonal vanishing points with a mean error of less than 3 degrees.

Additionally, we analysed the consistency of detected vanishing points w.r.t. three groups of ground truth line segments (corresponding to the orthogonal directions). In Fig. 6 the cumulative results for **ALS**, **RNS**, and

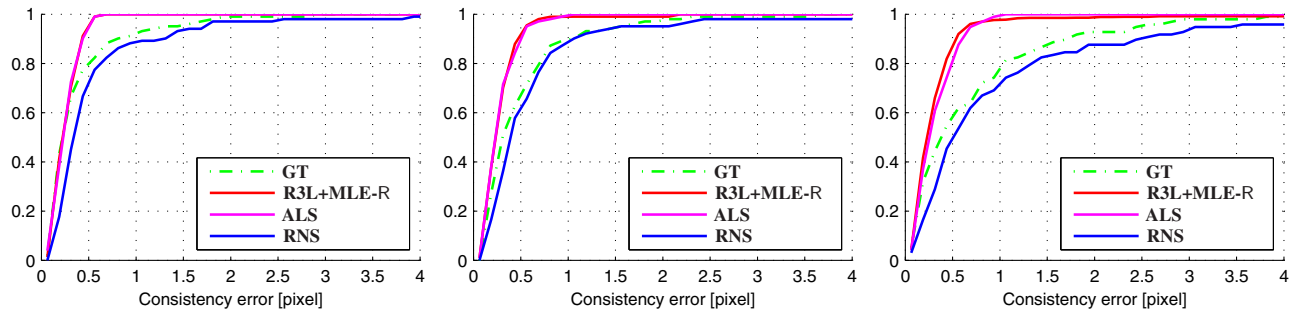

Figure 6. Cumulative histograms of vanishing point consistency error w.r.t. three groups of ground truth line segments. The ordinate shows the fraction of trials with errors less than the respective consistency (see Sec. 2) on the abscissa. Left: Vertical vanishing point, middle and right: Horizontal vanishing points.

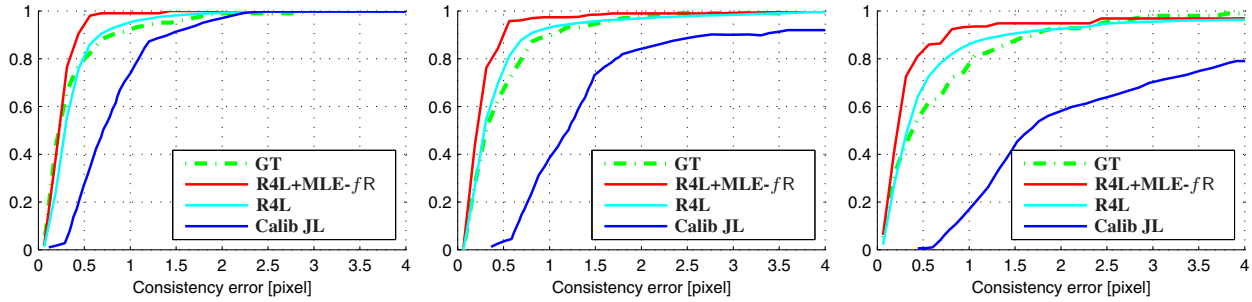

Figure 7. Cumulative histograms of vanishing point consistency error w.r.t. ground truth line segments (uncalibrated case)

R3L+MLE-R are shown. **GT** denotes the consistency achieved using ground truth vanishing points. One can see that **ALS** gives lower error than ground truth. This can be attributed to the sub-optimal two-stage process in which ground truth vanishing points have been obtained [13]. Surprisingly **R3L+MLE-**R works almost as well as **ALS** and consistently outperforms **RNS** and even ground truth. To be fair, it has to be pointed out that **R3L+MLE-**R and **RNS** have different implementations for line segment detection. However, the difference can more likely be explained by our use of an image-based line segment error in the RANSAC-stage and the robust likelihood model during refinement. Also, as the same consistency measure is used for vanishing point detection, and evaluation, a slight bias towards our method can be expected. But, as Fig. 5 clearly shows, our approach also fares much better than **RNS** when angular deviation from vanishing points is compared.

For an evaluation of **R4L** and **R4L+MLE-**fR (uncalibrated camera), we compare with the results reported by Tardif [20] for his **Calib JL** method. As in the previous test, we plot the cumulative results for the consistency of detected vanishing points in Fig. 7. One can see that our estimates are always better than the results achieved by **Calib JL**. Both algorithms work well for the vertical vanishing point and tend to give poorer results for horizontal ones. In fact, **R4L** and **R4L+MLE-**fR completely fail to detect all orthogonal directions only in a small number of cases, both yielding better than ground truth results for the majority of images. Examples where the algorithm succeeded are shown in Fig. 1. Line segments are assigned to the vanishing points with smallest consistency error. Segments with errors greater than 2 pixels are marked as outliers. Fig. 9 depicts four images causing failure. In these there were not enough line segments supporting all vanishing points sufficiently, or another structure obscured the orthogonal frame.

In Fig. 8 the cumulative statistics of the error in focal length estimation is depicted. It is worth noting that **Calib JL** utilizes the known camera calibration to select the most orthogonal triplet from a larger set of vanishing points. The focal length estimation is merely a way of assessing the accuracy of the detected vanishing points. One can see that our algorithm considerably outperforms its competitor in quality of the focal length estimates. In addition, we compare with the calibration method proposed by Nebehay & Pflugfelder [14]. For the YUD they reported 50%, and 64% successfully calibrated images at a relative focal length error of less than 5% and 10% respectively. For calibration the principal point was fixed at the image center and EM-refinement similar to ours has been deployed. Using **R4L+MLE-**fR we obtain 74% and 88% successful cali-

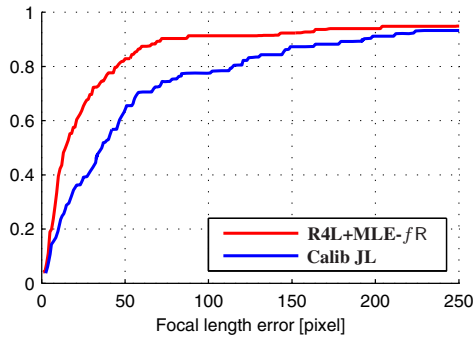

Figure 8. Cumulative histogram for focal length error (YUD).

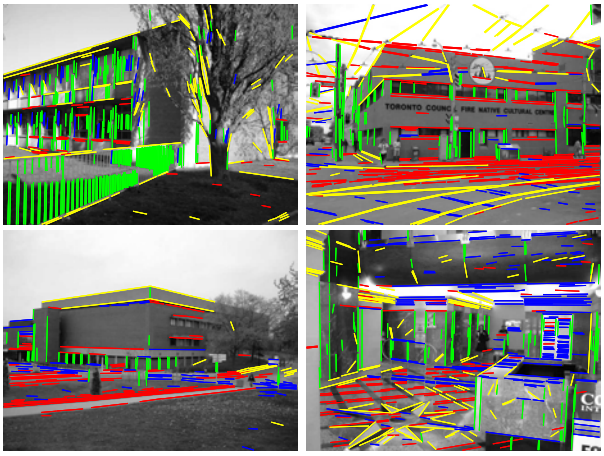

Figure 9. Examples of failures of the **R4L+MLE-fR** method

brations at the same error rates. With 65% and 82% **R4L** is still significantly better than its competition.

6.2. Horizon Estimation

In a second experimental run, we tested how well our **R4L+MLE-fR** method performs on the task of horizon estimation, comparing against the **GeometricParsing**-method recently proposed method by Tretyak *et al*. [21]. To this end, we evaluated the algorithm on the challenging Eurasian Cities Dataset (EAD) [21]. The dataset consists of 103 images of urban scenes, containing a significant proportion of images violating the Manhattan assumption.

Following the protocol of Tretyak *et al*. we retained the first 25 images for parameter selection and run **R4L+MLE-fR** on the remaining 78 images. For this experiment, images were scaled to a maximum side length of 640 pixels and we set the RANSAC threshold to $t = 1.0$ and the scale of the likelihood functions as $\gamma = 1.0$. The number of data evaluations was fixed at $k_{data} = 1000$. As in previous evaluations, the principal point was fixed on the image center. Here, the horizon is computed as the vanishing line polar to

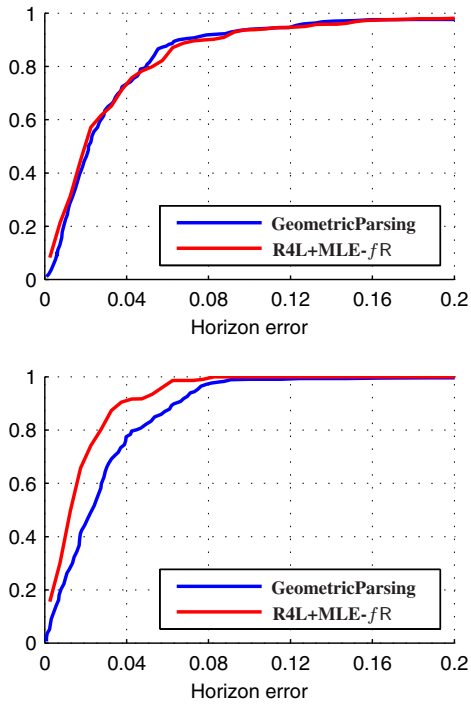

Figure 10. Cumulative histograms for horizon estimation error on EAD (top) and YUD (bottom).

the vanishing point with the largest distance from the image center in the y-direction. The horizon error is determined by the maximum Euclidean distance from the ground truth horizon in the image x-direction, divided by the image height. Examples are given in Fig. 11. Fig. 10 compares the outcome of our experimental run with the results reported in [21] — for completeness, results for the YUD (obtained with the settings described in Sec. 6.1) are also depicted.

As expected, we outperform **GeometricParsing** on the YUD, since the images depict mostly dominant orthogonal structures. On the much harder EAD our method performed as well as **GeometricParsing** which is interesting as this algorithm has been specifically designed for imagery not compliant with the Manhattan assumption. Our algorithm, despite the difficulties, was able to either pick-out orthogonal structures (see Fig. 11, lower-left image) or at least took a triplet of non-orthogonal but spatially well separated and dominant vanishing points (Fig. 11, upper-left image). Of course, in the latter case focal length estimates are incorrect.

6.3. Timings

On a Core i7 CPU, for an average of 500 line segments per image of the YUD, **R4L** and the BFGS optimizer took a mean time of 5ms and 19ms respectively. On the task of horizon detection on the EAD, **R4L+MLE-fR** took a mean time of 35ms. This is on average more than 800 times less

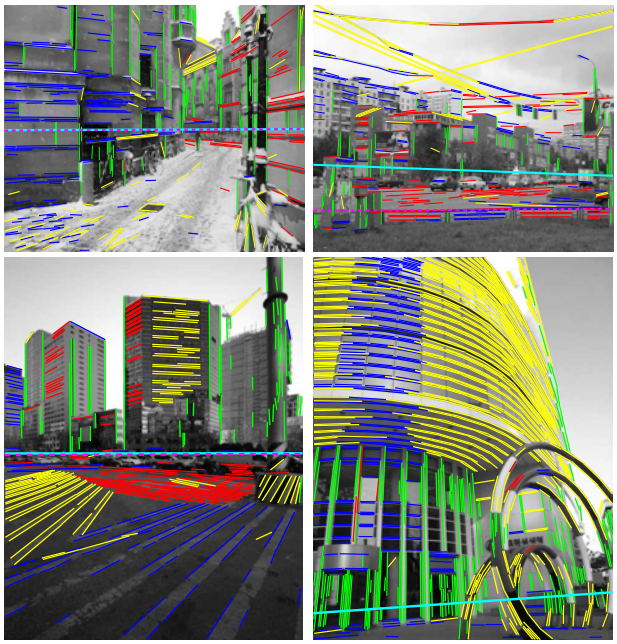

Figure 11. EAD horizon estimation examples. Ground truth horizon lines are depicted as solid cyan, estimates are shown as dashed magenta lines. Left: Successful detections. Right: Failures.

than the timings reported by Tretyak *et al*. [21].

7. Conclusion

This paper presents a novel approach for robust and accurate detection of orthogonal vanishing points in single images of a *Manhattan world*. We introduce an efficient RANSAC-based line segment classifier that uses quadruples of lines to generate hypotheses for focal length and three orthogonal vanishing points at once. Experiments show that the proposed approach results in a significant improvement over the state-of-the-art in detection of orthogonal vanishing points and camera self-calibration.

References

[1] B. Caprile and V. Torre. Using vanishing points for camera calibration. *IJCV*, 2, 1990.

[2] J. M. Coughlan and A. L. Yuille. The manhattan world assumption: Regularities in scene statistics which enable bayesian inference. In *NIPS*, 2000.

[3] P. Denis, J. H. Elder, and F. J. Estrada. Efficient edge-based methods for estimating Manhattan frames in urban imagery. In *ECCV*, 2008.

[4] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski. Manhattan-world stereo. In *CVPR*, 2009.

[5] R. I. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Second edition, 2004.

[6] V. Hedau, D. Hoiem, and D. Forsyth. Recovering the spatial layout of cluttered rooms. In *ICCV*, 2009.

[7] K. Kanatani and Y. Sugaya. Statistical optimization for 3-d reconstruction from a single view. *IEICE Trans.Inf. & Syst.*, (10), 2005.

[8] J. Kosecká and W. Zhang. Video compass. In *ECCV*, 2002.

[9] J. Kosecká and W. Zhang. Extraction, matching, and pose recovery based on dominant rectangular structures. *CVIU*, 100(3), 2005.

[10] D. C. Lee, M. Hebert, and T. Kanade. Geometric reasoning for single image structure recovery. In *CVPR*, 2009.

[11] D. Liebowitz. *Camera calibration and reconstruction of geometry from images*. PhD thesis, University of Oxford, Dept. Engineering Science, 2001.

[12] B. Micusík, H. Wildenauer, and J. Kosecka. Detection and matching of rectilinear structures. In *CVPR*, 2008.

[13] F. M. Mirzaei and S. I. Roumeliotis. Optimal estimation of vanishing points in a manhattan world. In *ICCV*, 2011.

[14] G. Nebehay and R. Pflugfelder. A self-calibration method for smart video cameras. In *EmbedCV09*, 2009.

[15] M. Nieto and L. Salgado. Non-linear optimization for robust estimation of vanishing points. In *ICIP*, 2010.

[16] J. Nocedal and S. Wright. *Numerical Optimization*. Springer, second edition, 2006.

[17] D. Robertson and R. Cipolla. *Practical Image Processing and Computer Vision*, chapter Architectural modelling. Wiley, 2011.

[18] C. Rother. *Multi-view reconstruction and camera recovery using a real or virtual reference plane*. PhD thesis, KTH, 2003.

[19] G. Schindler and F. Dellaert. Atlanta world: An expectation maximization framework for simultaneous low-level edge grouping and camera calibration in complex man-made environments. In *CVPR*, 2004.

[20] J.-P. Tardif. Non-iterative approach for fast and accurate vanishing point detection. In *ICCV*, 2009.

[21] E. Tretyak, O. Barinova, P. Kohli, and V. Lempitsky. Geometric image parsing in man-made environments. *IJCV*, 2011.

[22] H. Wildenauer and M. Vincze. Vanishing point detection in complex man-made worlds. In *ICIAP*, 2007.