OPTIMAL VANISHING POINT DETECTION AND ROTATION ESTIMATION OF SINGLE IMAGES FROM A LEGOLAND SCENE

Wolfgang Förstner

University Bonn, Institute for Geodesy and Geoinformation
Department of Photogrammetry, Nussallee 15, 53121 Bonn, Germany
wf@ipb.uni-bonn.de, http://ipb.uni-bonn.de

Commission III/1

The paper presents a method for automatically and optimally determining the vanishing points of a single image, and in case the interior orientation is given, the rotation of an image with respect to the intrinsic coordinate system of a lego land scene. We perform rigorous testing and estimation in order to be as independent on control parameters as possible. This refers to (1) estimating vanishing points from line segments and the rotation matrix, (2) to testing during RANSAC and during boosting lines and (3) to classifying the line segments w. r. t. their vanishing point. Spherically normalized homogeneous coordinates are used for line segments and especially for vanishing points to allow for points at infinity. We propose a minimal representation for the uncertainty of homogeneous coordinates of 2D points and 2D lines and rotations to avoid the use of singular covariance matrices of observed line segments. This at the same time allows to estimate the parameters with a minimal representation. The vanishing point detection method is experimentally validated on a set of 292 images.

1 INTRODUCTION

The orientation of a camera with respect to a world coordinate system is a classical task of photogrammetry. It can only be determined in case some information about the world is available and can be identified in the image. Determining the camera pose using reference points or lines visible in the image are classical and use either the spatial resection or the direct linear transformation, depending on whether the camera is calibrated or not. Reducing the amount of prior knowledge leads to the area of inverse perspective, where partial information about the object and the image can be derived using generic geometric properties of the scene, especially the existence of parallel or orthogonal straight lines. Going beyond, recent attempts to exploit illumination effects visible in an image allow to derive rough estimates of the orientation of the camera or the direction of the light source.

This paper presents a method for detecting vanishing points and possibly the rotation matrix of the camera w. r. t. the intrinsic coordinate system of a Lego land or Manhattan world with two or three mutually orthogonal sets of straight lines visible in the image. The paper is motivated by two observations of previous approaches: (1) The difficulty to represent the space of solutions, the unit sphere in two or three dimensions, within a search or estimation process and (2) lack of a standard algorithmic sequence for statistically rigorously inferring the desired parameters.

The main goal of the paper therefore is to apply rigorous testing, classification and estimation in all steps of a stochastic algorithm in order to reduce the number of control parameters and to obtain optimal results in terms of reliability and accuracy. To exploit the framework of projective geometry, where uncertain elements regularly are characterized with singular covariance matrices and during estimation require additional constraints, we propose a minimal representation of the uncertainty and the parameters to be estimated. These two means at the same time solve the two above mentioned problems of representation and rigorous estimation.

The paper is organized as follows. We first give a short review on previous approaches for handling the two mentioned problems within vanishing point detection and rotation estimation. We then introduce a minimal representation overcoming singularities, which is used for a method for vanishing point detection and rotation estimation. We experimentally validate the method on a set of 292 images.

Notation. We distinguish between the name, say χ, of an entity and its representation, say $\mathbf{x}$. We distinguish homogeneous vectors or matrices, say the 3-vector $\mathbf{x}$ or the 3×3-matrix K, from Euclidean vectors or matrices, say the 2-vector $\boldsymbol{x}$ and the 3×3-matrix R. Stochastical variables, say $\underline{\mathrm{x}}$, are underscored. The unit matrix is denoted with I_n, the skew symmetric 3×3-matrix $\mathsf{S}(\mathbf{x})$ of a 3-vector induces the cross product, thus $\mathbf{x} \times \mathbf{y} = \mathsf{S}(\mathbf{x})\mathbf{y}$. In case an equation contains homogeneous entities on both sides, the equality sign $=$ means equality up to a factor $\neq 0$. Concatenation of scalars, vectors and matrices follows MATLAB style: horizontal concatenation reads $[a, b]$, vertical reads $[a; b]$.

2 RELATED WORK

Vanishing point detection has been attacked at least since Barnard's influential paper 1983, addressing the unit sphere representation of the space of all vanishing points. When referring to one vanishing point all lines not going through this point are outliers highly robust methods are required, e. g. clustering [Straforini et al., 1992], the Hough transform [Tuytelaars et al., 1998] and random sample consensus [Wildenauer and Vincze, 2007]. Heuvel [1998] included an estimation of the rotation matrix. A good review is given by Rother [2000]. However, quite a number of recent publications address special aspects, such as including lens distortion [Grammatikopoulos et al., 2007], focussing on road scenes [Kong et al., 2009], efficient clustering processes [Schmitt and Priese, 2009], or a new method based on the socalled J-linkage algorithm for efficient clustering [Tardif, 2009].

Statistical modelling in the context of vanishing point detection has been adressed by Collins and Weiss [1990] who used models for the uncertainty of lines and vanishing point for optimal estimation, [Heuvel, 1998] who tracked the uncertainty through the sequence of decision steps or Coughlan and Yuille [2003], who include generic knowledge of the distribution of the lines, which was the basis for Deutscher et al. [2002], where also the principle distance was estimated, and Schindler and Dellaert [2004], who use an expectation-maximization approach to find more that one triplet of vanishing points.

Our approach, similar to the one of Heuvel [1998], tracks the uncertainty from the automatic line detection, via the vanishing

point detection to the estimation of the rotation. We however, our approach (1) exploits the uncertainty estimate from the line segment finder, (2) starts with finding vanishing points, and only in case of known interior orientation imposes the rectangular constraints, and (3) is tested on a large set of images.

3 MINIMAL REPRESENTATION FOR UNCERTAIN POINTS ON THE UNIT SPHERE

3.1 Motivation

Uncertain points $\mathcal{X}$ in the Euclidean plane are usually represented by a pair $\{\boldsymbol{\mu}_x, \Sigma_{xx}\}$ where $\boldsymbol{\mu}_x$ is a 2-vector of the mean and Σ_{xx} the 2×2-covariance matrix. It may be visualized by a standard ellipse, $(\boldsymbol{x} - \boldsymbol{\mu}_x)^\mathsf{T}\Sigma_{xx}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x) = 1$, with bounding box being of size $2\sigma_x \times 2\sigma_y$. In case we have an observation $\boldsymbol{x}$ for the mean vector it can also be represented by the a pair $\{\boldsymbol{x}, \Sigma_{xx}\}$, a situation which we will assume in the following to simplify notation. This representation directly can be transferred to homogeneous coordinates with 3-vector $\mathbf{x}$ and the corresponding 3×3-covariance matrix Σ_{xx}. In order to avoid the free scaling and be able to represent points at infinity it is of advantage to work with *spherically normalized* homogeneous coordinates $\mathbf{x}^s$:

$$\mathbf{x}^s = \mathsf{N}(\mathbf{x}) := \mathbf{x}/|\mathbf{x}|\,. \tag{1}$$

Given the covariance matrix Σ_{xx} for the unnormalized homogeneous vector, the covariance matrix of the spherically normalized vector is

$$\Sigma_{\mathrm{x}^s\mathrm{x}^s} = \boldsymbol{J}_{xx^s}\Sigma_{\mathrm{xx}}\boldsymbol{J}_{xx^s}^\mathsf{T}\,, \quad \text{with} \quad \boldsymbol{J}_{xx^s} = (\boldsymbol{I}_3 - \mathbf{x}^s\mathbf{x}^{s\mathsf{T}})\,/\,|\mathbf{x}| \tag{2}$$

where the Jacobian, except for the factor is a projection matrix with rank 2. Therefore the covariance matrix $\Sigma_{x^sx^s}$ has rank 2 and null space $\mathbf{x}^s$. It can be visualized by a standard ellipsoid: Its centre sits on the unit sphere S^2 at $\mathbf{x}^s$, it is flat in the direction $\mathbf{x}^s$ and indicates the directional uncertainty of the vector $\mathbf{x}^s$. Actually the distribution of the unit vectors is no Gaussian distribution any more, as all uncertain points sit on the unit sphere. We assume the directional uncertainty to be small enough, such that an approximation of the spherical distribution of $\underline{\mathbf{x}}^s$ by a Gaussian distribution is acceptable. This Taylor expansion, neglecting second order terms, decreases with the square of the angular uncertainty, which is acceptable, namely below 1 ‰, in case the directional uncertainty is below 1°.

The singularity of the covariance matrix has a number of consequences: (1) the standard ellipse can not be given explicitly, (2) we cannot express the Mahalanobis distance of a point close to a given point μ_x, as we also would need the inverse covariance matrix of the difference vector, which is singular, (3) we cannot express a weighted least squares estimation using the inverse covariance matrix as weight matrix, (4) in case we estimate some point on the unit sphere, we need an additional constraint, which increases the number of parameters per unknown 2D point to four, which in case of large systems may be inefficient. These disadvantages is the motivation to change the representation.

The argumentation can be transferred to all other geometric entities in 2D and 3D space, especially to 2D lines, which we will use for representing straight line segments.

3.2 A Minimal Representation

A minimal representation would do with two parameters per point and a 2×2 covariance matrix. Without loosing the possibility to handle points at infinity we can use the following representation.

1. We use the spherically normalized vector $\mathbf{x}^s$ for representing a point, either working in the projective plane, where $\mathbf{x}^s$ and $-\mathbf{x}^s$ represent the same point or working in the oriented projective plane, distinguishing point vectors with different sign.

In the following we always assume homogeneous vectors to be spherically normalized, and omit the superscript for simplifying the notation.

2. We only use the uncertainty of the directional vector in the tangent plane. This either can be done by locally spanning a two-dimensional coordinate system (s, t) in the tangent plane or by rotating the uncertain point into the north (or the south-) pole, there working in the (x, y) system of the Euclidean plane and omitting the third coordinate. This leaves us with a regular 2×2-covariance matrix. The situation is sketched in Fig. 1 The minimal rotation from the north pole $\boldsymbol{e}_3$ to the point $\mathbf{x}$ is given by [McGlone et al., 2004, p. 51]

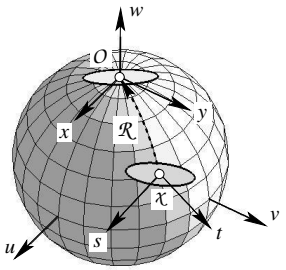

Figure 1: Transfer of the uncertainty of a point on the unit sphere S^2 from the point $\mathcal{X}$ to the north pole $O(\boldsymbol{e}_3)$ by a minimal rotation $\mathcal{R}$, thereby transporting the coordinate frame (st), tangent to $\mathcal{X}$ to (xy), tangent to O: working at O with two parameters is equivalent to working in the two-dimensional tangent space at $\mathcal{X}$ embedded in the three-dimensional space (uvw).

$$\begin{aligned} \boldsymbol{R}(\boldsymbol{e}_3, \mathbf{x}) &= \left[\begin{array}{cc|c} \frac{1+x_3-x_1^2}{1+x_3} & -\frac{x_1x_2}{1+x_3} & x_1 \\ -\frac{x_1x_2}{1+x_3} & \frac{1+x_3-x_2^2}{1+x_3} & x_2 \\ -x_1 & -x_2 & x_3 \end{array}\right] \\ &= [\boldsymbol{J}_x(\mathbf{x}) \mid \mathbf{x}] \end{aligned} \tag{3}$$

The rotation of the uncertain point $\underline{\mathbf{x}}$ to the north pole yields the *reduced covariance matrix*

$$\Sigma_{x_rx_r} = \boldsymbol{J}_x(\mathbf{x})^\mathsf{T}\,\Sigma_{\mathrm{xx}}\,\boldsymbol{J}_x(\mathbf{x}) \tag{4}$$

with the Jacobian $\boldsymbol{J}_x$ being the left two columns of $\boldsymbol{R}(\mathbf{x}, \boldsymbol{e}_3)$. The index $_x$ indicates that the Jacobian refers to a 3-vector. The inverse relation, the covariance matrix Σ_{xx} of the homogeneous point for a given reduced covariance matrix $\Sigma_{x_rx_r}$:

$$\Sigma_{\mathrm{xx}} = \boldsymbol{J}_x(\mathbf{x})\,\Sigma_{x_rx_r}\,\boldsymbol{J}_x^\mathsf{T}(\mathbf{x})\,. \tag{5}$$

3. We measure the difference of two neighbouring points either in the tangent plane or after having rotated both points with the same transformation into the north (or the south-) pole. As we can transfer the uncertainty by variance propagation a comparison then can be based on the Mahalanobis distance.

Thus, in case two neighbouring points $\mathcal{X}$ and y are given, the Mahalanobis distance is determined by first reducing the two points, thus by rotating the wit the same rotation close to the north pole. For this we need two steps. First we need to make the covariance matrices consistent by centering them at some good approximation $\mathbf{x}^a$ for the common mean. This is achieved by rotating the

uncertainties from $\mathbf{x}$ and $\mathbf{y}$ to $\mathbf{x}^a$ by a minimal rotation, e. g.

$$\Sigma^a_{\mathrm{xx}} = \mathsf{R}(\mathbf{x}, \mathbf{x}^a)\, \Sigma_{\mathrm{xx}}\, \mathsf{R}^{\mathsf{T}}(\mathbf{x}, \mathbf{x}^a) \tag{6}$$

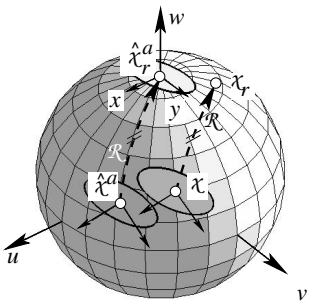

Figure 2: Estimation is performed at the north-pole in order to exploit the regularity of the reduced covariance matrix of the observations. This first needs to be transformed into the approximate point $\mathbf{x}^a$, then - together with $\mathbf{x}$ - rotated ($\mathcal{R}$) into the north-pole, allowing to skip the third coordinate - and yielding the reduced observation $\boldsymbol{x}_r$. The reduced approximate value is $\mathbf{0}$,

Second we need to derive the *reduced homogeneous coordinates* by rotating the complete situation to the north pole and omitting the third coordinate, thus applying $\boldsymbol{J}^{\mathsf{T}}_x(\mathbf{x}^a)$. This leads to the reduced coordinates

$$\boldsymbol{x}_r = \boldsymbol{J}^{\mathsf{T}}_x(\mathbf{x}^a)\mathbf{x}\,, \qquad \boldsymbol{y}_r = \boldsymbol{J}^{\mathsf{T}}_x(\mathbf{x}^a)\mathbf{y}\,, \tag{7}$$

and the corresponding covariances

$$\Sigma^a_{x_r x_r} = \boldsymbol{J}^{\mathsf{T}}_x(\mathbf{x}^a)\, \Sigma^a_{\mathrm{xx}}\, \boldsymbol{J}_x(\mathbf{x}^a) \tag{8}$$

$$\Sigma^a_{y_r y_r} = \boldsymbol{J}^{\mathsf{T}}_x(\mathbf{x}^a)\, \Sigma^a_{\mathrm{yy}}\, \boldsymbol{J}_x(\mathbf{x}^a) \tag{9}$$

which finally allows us to determine the Mahalanobis distance in the classical manner:

$$d^2 = (\boldsymbol{x}_r - \boldsymbol{y}_r)^{\mathsf{T}} \left(\Sigma_{x_r x_r} + \Sigma_{y_r y_r}\right)^{-1} (\boldsymbol{x}_r - \boldsymbol{y}_r) \tag{10}$$

The inversion always is possible in case the rank of the covariance matrix of the homogeneous point is two.

Therefore in the following we will represent a 2D point $\mathcal{x}$ using its homogeneous coordinates and its reduced covariance matrix:

$$\boxed{\mathcal{x}: \quad \{\mathbf{x}, \Sigma_{x_r x_r}\}\,.} \tag{11}$$

where the approximate value is chosen to be identical to $\mathbf{x}$, thus $\mathbf{x}^a = \mathbf{x}$. As also the full covariance is used, it may be of advantage to also store the full 3×3 covariance matrix in case it is to be used frequently and the number of elements to be handled is not too large.

The same representation is used for 2D lines $\mathcal{l} : \{\mathbf{l}, \Sigma_{l_r l_r}\}$, where $\mathbf{l}$ are the spherically normalized homogeneous coordinates, reflecting the interpretation of 2D lines on the unit sphere as already done by Barnard [1983].

4 DETECTING VANISHING POINTS

We assume a set of N straight line segments $\mathcal{l}_n, n = 1, ..., N$ be derived with some image processing tool. We represent them as infinitely long straight lines $\{\mathbf{l}, \Sigma_{l_r l_r}\}_n$ in some camera coordinate system:

1. In all cases we assume the camera to be straight line preserving.

2. We assume the image centre to be identical to the principle point, the skew between the axes to be 0 and the scales in x and y direction to be equal.

3. The principle distance is either given or assumed to be identical to the longer side of the image.

Thus the calibration matrix is assumed to be

$$\mathsf{K} = \begin{bmatrix} c & 0 & x_h \\ 0 & c & y_h \\ 0 & 0 & 1 \end{bmatrix} \tag{12}$$

Given the line segments ${}^p\mathbf{l}$ in the pixel-coordinate system we obtain their coordinates in the camera system from $\mathbf{l} = \mathsf{K}^{\mathsf{T}}\, {}^p\mathbf{l}$. In case the above mentioned assumptions are close to reality, we after this transformation will be able to apply the orthogonality constraints between the derived vanishing point vectors, otherwise, the transformation from pixel to camera coordinates just serves for conditioning. In the experimental section we give more details on the derivation of the line segments and their accuracy.

We again assume the homogeneous coordinates to be spherically normalized and the uncertainty is represented by the reduced covariance matrices derived from the image processing tool. The line segments belong to one of four classes, either to one of the three vanishing points of the Lego land scene or to none of these three points. Let the n-th line segments $\mathcal{l}_n$ belong to class $c_k, k \in \{0, 1, 2, 3\}$, $k = 0$ indicating the line being an outlier.

The optimization problem is to find classes c_n and three vanishing points $\mathbf{x}_j$ such that the robustified function

$$\Omega^2(c_n, \mathbf{x}_j) = \sum_{j=0}^{3} \sum_{n=1}^{N} \delta(c_n - j)\rho(d^2(\mathbf{x}_j, \mathbf{l}_n)) \tag{13}$$

is minimal. Here we use the Kronecker delta, the function $\rho(t^2) = \min(T^2, t^2)$ with the threshold $T = 3$ and the Mahalanobis distance of the vanishing point $\mathcal{x}_j$ from the line segment $\mathcal{l}_n$

$$d^2(\mathbf{x}_j, \mathbf{l}_n) = \frac{e^2(\mathbf{x}_j, \mathbf{l}_n)}{\sigma^2_{e_{jn}}} = \frac{(\mathbf{x}_j^{\mathsf{T}}\mathbf{l}_n)^2}{\mathbf{x}_j^{\mathsf{T}} \Sigma_{\mathbf{l}_n \mathbf{l}_n} \mathbf{x}_j}\,. \tag{14}$$

We assume $\rho(d^2(\mathbf{x}_0, \mathbf{l}_n)) = T^2$ to include line segments going to no vanishing points. In case the classes c_n are not known, the optimization function in general is not convex and shows many local optima, requiring specific optimization procedures such as used in Coughlan and Yuille [2003]; Deutscher et al. [2002]; Schindler and Dellaert [2004]. We apply a stochastic algorithm for finding the classes. A RANSAC approach appears to be most suitable, due to the simplicity of the geometric situation and the lack of approximate values. Therefore we need to perform the following steps:

1. For each vanishing point, (1) use a RANSAC approach for identifying preliminary classes, (2) determine a robust Maximum likelihood type estimation for the vanishing points, given the classes, (3) perfom a boosting step to find more lines belonging to that vanishing point and (4) finally again perform a robust ML type estimation. This actually is a procedure for robustly finding the pole $\mathbf{x}_j$ of the best fitting circle through a subset of the points $\mathbf{l}_n$ on the unit sphere.

2. Given all vanishing points test whether two of them are identical.

3. determine the best class for each line segment and repeat the ML type estimation for all vanishing points.

4. Enforce the orthogonality constraint onto the vanishing point directions in the camera system, which results from the orthogonality of the main directions of the Lego land scene.

We describe the individual steps in more detail.

4.1 Detecting vanishing points using RANSAC

Random sample consensus (RANSAC, Fischler and Bolles [1981]) has undergone many modifications since its invention. We use the version MLESAC, where, instead of choosing the solution with the maximum number if inliers, we choose the randomly found pair of lines minimizing $\Omega^2(\mathbf{x}) = \sum_n \rho(d^2(\mathbf{l}_n, \mathbf{x}))$. At the same time we treat all line segments with $d^2 < T^2$ as inliers.

Though the putative vanishing point $\mathcal{X} = \mathcal{l}_m \cap \mathcal{l}_{m'}$, depending on the randomly chosen line segments $\mathcal{l}_m$ and $\mathcal{l}_{m'}$ is uncertain, we treat it as a trial point within a random search procedure, and therefore do not need to take its uncertainty into account - in contrast to a procedure, where the other $N-2$ line segments are tested to determine the size of the consensus set of lines for the hypothesis (m, m').

4.2 ML-type estimation of a vanishing point

Given a set $\{\mathcal{l}_n, n \in \mathcal{N}_j\}$ lines segments which are supposed to support the vanishing point $\mathcal{x}_j$, we perform an estimation of the point. Based on approximate values $\mathbf{l}_n^a, n \in \mathcal{N}_j$ for the line segments and $\mathbf{x}_j^a$ for the j-th vanishing point, we search for estimates (omitting the index j in this section for simplicity)

$$\widehat{\mathbf{l}}_n = \mathbf{l}_n + \widehat{\mathbf{v}}_n = \widehat{\mathbf{l}}_n^a + \widehat{\Delta\mathbf{l}}_n\,, \quad \widehat{\mathbf{x}} = \widehat{\mathbf{x}}^a + \widehat{\Delta\mathbf{x}} \tag{15}$$

fulfilling the nonlinear constraints

$$g_n(\widehat{\mathbf{l}}_n, \widehat{\mathbf{x}}) = \widehat{\mathbf{l}}_n^\mathsf{T}\,\widehat{\mathbf{x}} = 0\,, \quad n \in \mathcal{N}_j \tag{16}$$

As the covariance matrix of the observations is singular, we transform the model into the equivalent model using the reduced observations $\boldsymbol{l}_{rn}$ and the reduced approximated point $\boldsymbol{x}_r^a$:

$$\boldsymbol{l}_{rn} = \boldsymbol{J}_x^\mathsf{T}(\widehat{\mathbf{l}}_n^a)\,\mathbf{l}_n \qquad \boldsymbol{x}_r^a = \boldsymbol{J}_x^\mathsf{T}(\widehat{\mathbf{x}}^a)\,\mathbf{x}^a \equiv \mathbf{0} \tag{17}$$

leading to the new constraints for the reduced values

$$g_n(\widehat{\boldsymbol{l}}_{rn}, \widehat{\boldsymbol{x}}_r) = 0\,, \quad n \in \mathcal{N}_j \tag{18}$$

We now iteratively improve the approximate values for the fitted observations and the estimated point using a linear substitute model, derived by Taylor expansion of the constraints. This reads as

$$g_n(\widehat{\boldsymbol{l}}_{rn}^a, \widehat{\boldsymbol{x}}_r^a) + \boldsymbol{a}_n^\mathsf{T}\,\widehat{\Delta\boldsymbol{x}}_r + \boldsymbol{b}_n^\mathsf{T}\,\widehat{\Delta\boldsymbol{l}}_{rn} = 0 \quad D(\boldsymbol{l}_{rn}) = \Sigma^a_{\mathsf{l}_{rn}\mathsf{l}_{rn}}\,, \tag{19}$$

where the reduced covariance matrix is determined using (6) and (8). With the approximate values for the residuals $\widehat{\boldsymbol{v}}_{rn}^a = \widehat{\boldsymbol{l}}_{rn}^a - \boldsymbol{l}_{rn}$ and the estimated residuals $\widehat{\boldsymbol{v}}_{l_{rn}} = \widehat{\boldsymbol{v}}_{rn}^a + \widehat{\Delta\boldsymbol{l}}_{rn}$ we want to find best corrections $\widehat{\Delta\boldsymbol{l}}_{rn}$ and $\widehat{\Delta\boldsymbol{x}}_r$ minimizing

$$\Omega^2(\widehat{\Delta l}_{rn}, \widehat{\Delta\boldsymbol{x}}_r) = \sum_{n\in\mathcal{N}_j} \widehat{\boldsymbol{v}}_{rn}^\mathsf{T}\,(\Sigma^a_{\mathsf{l}_{rn}\mathsf{l}_{rn}})^{-1}\,\widehat{\boldsymbol{v}}_{rn} \tag{20}$$

under the given constraints. Observe the covariance matrix $\Sigma_{\mathsf{l}_{rn}\mathsf{l}_{rn}}$ depends on the unknown parameters. This setup only is possible, as the covariance matrix of the reduced observations is regular. The classical solution is $\widehat{\Delta\boldsymbol{x}}_r = \Sigma_{\widehat{\mathbf{x}}\widehat{\mathbf{x}}}\,\boldsymbol{A}_r^\mathsf{T}(\boldsymbol{B}_r^\mathsf{T}\Sigma^a_{l_rl_r}\boldsymbol{B}_r)^{-1}\boldsymbol{c}_g$ with

$$\underset{N\times 2}{\boldsymbol{A}_r} = [\boldsymbol{a}_{rn}^\mathsf{T}]\,, \quad \underset{N\times 2N}{\boldsymbol{B}_r^\mathsf{T}} = \mathrm{Diag}(\boldsymbol{b}_{rn}^\mathsf{T}) \;\; c_{g_n} = -\mathbf{x}^{a\mathsf{T}}\,\mathbf{l}_n^a + \boldsymbol{b}_{rn}^\mathsf{T}\,\widehat{\boldsymbol{v}}_{rn}^a$$

with the Jacobians, both being 1×2-vectors $\boldsymbol{a}_{rn}^\mathsf{T} = \widehat{\mathbf{l}}_n^{a\mathsf{T}}\,\boldsymbol{J}_x(\widehat{\mathbf{x}}^a)$, $\boldsymbol{b}_{rn}^\mathsf{T} = \mathbf{x}_r^{a\mathsf{T}}\,\boldsymbol{J}_x(\widehat{\mathbf{l}}_n^a)$ and the covariance matrix of the reduced coordinates of the estimated point

$$\Sigma_{\widehat{x}_r\widehat{x}_r} = (\boldsymbol{A}_r^\mathsf{T}(\boldsymbol{B}_r^\mathsf{T}\Sigma^a_{l_rl_r}\boldsymbol{B}_r)\boldsymbol{A}_r)^{-1} \tag{21}$$

The corrections for the fitted observations are

$$\widehat{\Delta\boldsymbol{l}}_r = \Sigma^a_{\mathsf{l}_r\mathsf{l}_r}\boldsymbol{B}_r\Sigma_{\widehat{x}_r\widehat{x}_r}(\boldsymbol{c}_g - \boldsymbol{A}_r\widehat{\Delta\boldsymbol{x}}_r) - \widehat{\boldsymbol{v}}^a \tag{22}$$

We now need to determine the updated fitted observations $\widehat{\mathbf{l}}_n$ and the updated point coordinates $\widehat{\mathbf{x}}$. These we obtain by rotating the vectors $\mathsf{N}([\widehat{\Delta\boldsymbol{l}}_{rn}; 1])$ and $\mathsf{N}([\widehat{\Delta\boldsymbol{x}}_r; 1])$ back to the vicinity of the approximate values:

$$\widehat{\mathbf{l}}_n^{(\nu+1)} = \boldsymbol{R}^\mathsf{T}(\widehat{\mathbf{l}}_n^{(\nu)}, \boldsymbol{e}_3)\,\mathsf{N}\left(\begin{bmatrix}\widehat{\Delta\boldsymbol{l}}_{rn}\\ 1\end{bmatrix}\right) \tag{23}$$

$$\widehat{\mathbf{x}}^{(\nu+1)} = \boldsymbol{R}^\mathsf{T}(\widehat{\mathbf{x}}^{(\nu)}, \boldsymbol{e}_3)\,\mathsf{N}\left(\begin{bmatrix}\widehat{\Delta\boldsymbol{x}}_r\\ 1\end{bmatrix}\right) \tag{24}$$

Finally, we determine the estimated variance factor from the weighted sum of the residuals: $\widehat{\sigma}_0^2 = \Omega/(N_j - 2)$.

As some of the constraints may not be fulfilled, we apply a ML-type estimation, by eliminating those constraints where the Mahalanobis distance d_n of the constraint c_{g_n} from 0 is significant, using $\underline{d}_n^2 = \underline{c}_{g_n}^2/\sigma_{c_{g_n}}^2 \sim \chi_1^2$ with $\sigma_{c_{g_n}}^2 = \boldsymbol{b}_{rn}^\mathsf{T}\,\Sigma^a_{\mathsf{l}_{rn}\mathsf{l}_{rn}}\,\boldsymbol{b}_{rn}$.

4.3 Boosting for finding lines belonging to a vanishing point

Now, as we have an initial good estimate for the vanishing point, we search for more lines supporting it. This boosting step, statistically checks the inner products $f_n = \widehat{\mathbf{x}}^\mathsf{T}\mathbf{l}_n$, using $\underline{d}_n^2 = \underline{f}_n^2/\sigma_{f_n}^2 \sim \chi_1^2$. It takes the uncertainty $\Sigma_{\widehat{\mathbf{x}}\widehat{\mathbf{x}}}$ from (21) of the estimated vanishing point into account when determining the variance $\sigma_{f_n}^2$ of f_n. Thus we use $\sigma_{f_n}^2 = \widehat{\mathbf{x}}^\mathsf{T}\Sigma_{\mathsf{l}_n\mathsf{l}_n}\widehat{\mathbf{x}} + \mathbf{l}_n^\mathsf{T}\Sigma_{\widehat{\mathbf{x}}\widehat{\mathbf{x}}}\mathbf{l}_n$. This procedure is repeated three times, in each step only using those line segments which have not been previously identified.

4.4 Classifying the line segments

We now assume we have up to three vanishing points $\mathcal{x}_j(\mathbf{x}_j, \Sigma_{\mathbf{x}_j\mathbf{x}_j})$, omitting the hat $\widehat{\ }$, both for convenience, and as these vanishing points will be treated as input to the final estimate. As the sequence of finding the vanishing points may favour the first and possibly the second vanishing point, we now perform a classification of all line segments. We perform the same test with the up to three test statistics $d_{nj}^2 = f_{nj}^2/\sigma_{f_{nj}}^2$ per line segment $\mathcal{l}_n$. In case the log likelihood ratio $r_{n,jj'} = d_{nj}^2/d_{nj'}^2$ is smaller than a factor, say 1/25, for all $j' \neq j$, then the line segment is classified as belonging to class j, thus $c_n = j$. Otherwise it is classified as not decidable, indicating it is either an outlier or a line segment close to a line joining two vanishing points. After the lines have been classified as belonging to one of the three vanishing points, the coordinates of the vanishing point again are determined using a ML-type estimation.

4.5 Enforcing the orthogonality constraint

We finally enforce the orthogonality constraint in case we know the interior orientation of the camera. As the coordinate vectors

of the vanishing points are mutually statistically independent, we have the full statistical information from all relevant line segments.

We now treat the coordinates $(\mathbf{x}_j, \Sigma_{\mathbf{x}_j\mathbf{x}_j})$ of the vanishing points achieved from the individual detection, boosting and estimation as observations, which need to be corrected. Based on approximate values $\mathbf{x}_j^a$, which in the first iteration are identical to $\mathbf{x}_j$, we again obtain $\widehat{\mathbf{x}}_j = \mathbf{x}_j + \widehat{\mathbf{v}}_j = \widehat{\mathbf{x}}_j^a + \widehat{\Delta\mathbf{x}}_j\,, \quad j = 1,2,3$ to fulfill the orthogonality constraint. The model for enforcing the three orthogonality constraints is $\boldsymbol{g}(\widehat{\mathbf{x}}_i) = \mathbf{0}$ with $g_1 = \widehat{\mathbf{x}}_2^\mathsf{T}\widehat{\mathbf{x}}_3$, $g_2 = \widehat{\mathbf{x}}_3^\mathsf{T}\widehat{\mathbf{x}}_1$, $g_3 = \widehat{\mathbf{x}}_1^\mathsf{T}\widehat{\mathbf{x}}_2$. After reducing the observations $\widehat{\boldsymbol{x}}_{rj} = \boldsymbol{J}_x^\mathsf{T}(\widehat{\mathbf{x}}_j)\,\mathbf{x}_j, j = 1,2,3$ in order to be able to handle the singularity of the covariance matrices $\Sigma_{\mathbf{x}_j\mathbf{x}_j}$ we obtain the reduced model $\boldsymbol{g}(\widehat{\boldsymbol{x}}_{ri}) = \mathbf{0}$ with $g_1 = \widehat{\boldsymbol{x}}_{r2}^\mathsf{T}\widehat{\boldsymbol{x}}_{r3} = 0$, $g_2 = \widehat{\boldsymbol{x}}_{r3}^\mathsf{T}\widehat{\boldsymbol{x}}_{r1} = 0$, $g_3 = \widehat{\boldsymbol{x}}_{r1}^\mathsf{T}\widehat{\boldsymbol{x}}_{r2} = 0$. The linearized model therefore is $\boldsymbol{c}_g(\widehat{\boldsymbol{x}}^a) + \boldsymbol{B}_r^\mathsf{T}\widehat{\Delta\boldsymbol{x}}_r = \mathbf{0}$ or explicitely

$$\begin{bmatrix} \widehat{\mathbf{x}}_2^{a\mathsf{T}}\widehat{\mathbf{x}}_3^a \\ \widehat{\mathbf{x}}_3^{a\mathsf{T}}\widehat{\mathbf{x}}_1^a \\ \widehat{\mathbf{x}}_1^{a\mathsf{T}}\widehat{\mathbf{x}}_2^a \end{bmatrix} + \begin{bmatrix} \mathbf{0}^\mathsf{T} & \widehat{\boldsymbol{x}}_{r3}^{a\mathsf{T}} & \widehat{\boldsymbol{x}}_{r2}^{a\mathsf{T}} \\ \widehat{\boldsymbol{x}}_{r3}^{a\mathsf{T}} & \mathbf{0}^\mathsf{T} & \widehat{\boldsymbol{x}}_{r1}^{a\mathsf{T}} \\ \widehat{\boldsymbol{x}}_{r2}^{a\mathsf{T}} & \widehat{\boldsymbol{x}}_{r1}^{a\mathsf{T}} & \mathbf{0}^\mathsf{T} \end{bmatrix} \begin{bmatrix} \widehat{\Delta\boldsymbol{x}}_{r1} \\ \widehat{\Delta\boldsymbol{x}}_{r2} \\ \widehat{\Delta\boldsymbol{x}}_{r3} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

The reduced covariance matrices $\Sigma^a_{x_{rj}x_{rj}}$ of the observations are derived following (6) and (8), first transforming the rotation into the approximate point and second transforming them to the north (or south) pole and omitting the third, the zero component.

Minimizing $\Omega = \sum_{j=1}^3 \widehat{\boldsymbol{v}}_{rj}^\mathsf{T}\,(\Sigma^a_{x_{rj}x_{rj}})^{-1}\,\widehat{\boldsymbol{v}}_{rj}$ under the three constraints yields the classical solution for the update for the fitted observations $\widehat{\Delta\boldsymbol{x}}_r = \Sigma^a_{x_{rj}x_{rj}}\boldsymbol{B}_r(\boldsymbol{B}_r^\mathsf{T}\Sigma_{x_{rj}x_{rj}}\boldsymbol{B}_r)^{-1}(\boldsymbol{c}_g + \boldsymbol{B}_r^\mathsf{T}\boldsymbol{x}_r) + \boldsymbol{x}_r$. They are used to obtain improved approximate values for the fitted values of the vanishing point coordinates, as in (24). In spite of the low redundancy of $R = 3$ it useful to determine and report the estimated variance factor $\widehat{\sigma}_0^2 = \Omega/3$.

5 EXPERIMENTS

5.1 Used data

We perform two tests, one using uncalibrated images for investigating reliabity of the vanishing point detections and a second using partially calibrated images, where the principal distance is known.

In both cases we automatically derive straight line segments. They are represented by their centroid $\boldsymbol{x}_0$ [pel], their length l [pel], their direction ϕ. This allows to give the stochastical properties by the standard deviation σ_q of the centroid across the line and the standard deviation σ_ϕ of the direction. They are derived from a ML-estimation using the edge elements and are approximately

$$\sigma_q = \frac{1}{\sqrt{l}}\,\sigma_e\,, \quad \sigma_\phi = \sqrt{\frac{12}{l^3 - l}}\,\sigma_e \tag{25}$$

where σ_e is standard deviation of an edge element, which depends on the manner of subpixel positioning and always is smaller than the rounding error $1/\sqrt{12}$ [pel]. In our context mainly the angular accuracy is relevant. Using the techniques described by Meidow et al. [2009] the spherically normalized coordinates $\mathbf{l} := \mathbf{l}^s$ of all line segments together with their singular covariance matrix $\Sigma_{\mathbf{ll}}$ are determined. For testing we always take a high significance level of $S = 0.9999$. We employ the adaptive determination of the number of trials in the RANSAC procedure as described by Hartley and Zisserman [2000].

The processing time for each image is in the order of a few seconds, including the edge detection program, written in C, and the vanishing point detection, using non optimized MATLAB code.

5.2 Detecting vanishing points

The first group of experiments addresses the quality of the vanishing point detection. We evaluate the detection procedure on three levels of accuracy.

Visual evaluation. First we check the reliability of the vanishing point detection. We downloaded 140 Google images named 'building' or 'batiment' (cf. Fig. 3) with a minimum side length of 768 pixels. No interior orientation is known, some images are images sections, some are graphics, some show significant lens distortion. For each image we visually identified the number of vanishing point, a human could find, and - by inspecting the color coded line segments and the directions to the vanishing points - the number of correctly found vanishing points. The result is shown in the table 1. From the 102 images, where 3 or more vanishing points could be detected, in only 7 images the system found only one vanishing points, whereas in 28 images two vanishing points were detected. From the 95 images, mostly with facades, where only two vanishing points could be detected, in 90 % the system could find both vanishing points. In three images no vanishing points could be detected even by a human. This is coherent with the experiment on the ETRIMS-data base [Korč and Förstner, 2009], where in all 60 images of facades both vanishing points could be detected.

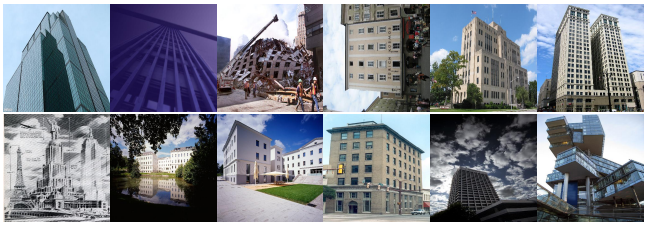

Figure 3: 12 of 140 building images, taken from Google ('building', 'batiment'). Such images are used for evaluating the vanishing point detection.

	0	1	2	3	4	5	6
0	3	0	0	0	0	0	0
1	0	0	3	4	0	0	0
2	0	0	92	25	2	0	1
3	0	0	0	65	4	0	1

Table 1: Horizontal: number of vanishing points a human could detect. Vertical: the number of vanishing points correctly detected by the algorithm. 140 Google images and 60 eTRIMS images.

The software gives an internal estimate for the accuracy of the vanishing points. We compared this with the number of line segments supporting a vanishing point. In the Google data set of 40 images on an average 90 lines supported a vanishing point, the mean standard deviation of the direction is appr. 0.3°. On an average we obtain an internal estimate for the accuracy $\sigma_d \approx 2.5°/\sqrt{n}$, which is a lower bound for the real accuracy. Of course, no check on the orthogonality of the vanishing points could be performed as the intrinsic parameters are not available for these images.

Comparison with manual measurements. In 54 images we measured all vanishing points manually, by interactively identifying two straight lines pointing towards a vanishing point. We compared these directions with the one determined by the system.

The histogram of the 143 angular differences between the manual and the automatic measurements is shown in fig. 4. About 1/3 of the manually measured vanishing points show a difference larger than 6 degrees, a threshold taken from the histogram. The other 95 differences indicate an average angular difference of 1.1°. As a spatial direction has two degrees of freedom, this corresponds to a standard deviation in each angular direction of appr. 0.7°, which is slighly larger than twice the internally estimated accuracy of 0.3°.

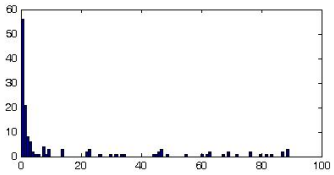

Figure 4: Histogram of angular differences between manually measured and automatically computed vanishing points

Finally, we evaluated the the method on the 102 images of the YORK URBAN data base [Denis et al., 2008], with images of size 580× 640 pixels. There the vanishing points were manually determined using multiple edges per vanishing point. From the 306 vanishing points 286 (93 %) were automatically detected. They showed an angular difference of 1.7°, corresponding to a directional uncertainty of 1.2°. The lower accuracy probably results from the lower resolution, thus the shorter line segments used. An example, where only two vanishing points were found is shown in fig. 5.

Figure 5: Image with only two vanishing points found correctly. The black star is the estimated vanishing point, the white diamond the reference point. The green line segments in the image are identified as outliers before boosting. After boosting the distance to the reference point is diminuished by appr. 30 %. The blue vanishing point can be automatically be identified as outlier, as $\widehat{\sigma}_0^2 = 126.7^2$ is significanly larger than 1.

6 CONCLUSIONS AND FUTURE WORK

We presented a new method for vanishing point detection, which tracks the uncertainty from the image data to the final result and empirically evaluated the result on 292 images. The estimation was consistently performed with homogeneous coordinates, for which we proposed to use a minimal representation for the uncertainty.

The system easily can be extended to determine parameters for lens distortion, as proposed by Grammatikopoulos et al. [2007]. It up to now does not perform enough self diagnosis, in order to reliably determine the number of correct vanishing points. For a practical use a priori knowledge should be integratable in a flexible manner, e. g. constraints on the orientation or the prior knowledge about the relevance of line segments.

References

Barnard, S. T., 1983. Interpreting perspective images. AI 21(4), pp. 435–462.

Collins, R. T. and Weiss, R. S., 1990. Vanishing Point Calculation as a Statistical Inference on the Unit Sphere. In: ICCV'90, IEEE Computer Society Intl. Conf. on Computer Vision, IEEE Computer Society, Los Alamitos, CA, pp. 400–403.

Coughlan, J. M. and Yuille, A. L., 2003. Manhattan world: orientation and outlier detection by bayesian inference. Neural Comput. 15(5), pp. 1063–1088.

Denis, P., Elder, J. H. and Estrada, F. J., 2008. Efficient edge-based methods for estimating manhattan frames in urban imagery. In: ECCV '08: Proceedings of the 10th European Conference on Computer Vision, Springer-Verlag, Berlin, Heidelberg, pp. 197–210.

Deutscher, J., Isard, M. and MacCormick, J., 2002. Automatic camera calibration from a single manhattan image. In: ECCV '02: Proceedings of the 7th European Conference on Computer Vision-Part IV, Springer-Verlag, London, UK, pp. 175–205.

Fischler, M. A. and Bolles, R. C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM 24(6), pp. 381–395.

Grammatikopoulos, L., Karras, G. and Pets, E., 2007. An automatic approach for camera calibration from vanishing points. 62(1), pp. 64–76.

Hartley, R. I. and Zisserman, A., 2000. Multiple View Geometry in Computer Vision. Cambridge University Press.

Heuvel, F. A. V. D., 1998. Vanishing point detection for architectural photogrammetry. In: Proc. ECCV '02, Springer, pp. 652–659.

Kong, H., Audibert, J.-Y. and Ponce, J., 2009. Vanishing point detection for road detection. Computer Vision and Pattern Recognition, IEEE Computer Society Conference on 0, pp. 96–103.

Korč, F. and Förstner, W., 2009. eTRIMS Image Database for interpreting images of man-made scenes. Technical Report TR-IGG-P-2009-01.

McGlone, C. J., Mikhail, E. M. and Bethel, J. S., 2004. Manual of Photogrammetry. Am. Soc. of Photogrammetry and Remote Sensing.

Meidow, J., Beder, C. and Förstner, W., 2009. Reasoning with uncertain points, straight lines, and straight line segments in 2D. International Journal of Photogrammetry and Remote Sensing 64, pp. 125–139.

Rother, C., 2000. A new approach for vanishing point detection in architectural environments. In: In Proc. 11th British Machine Vision Conference, pp. 382–391.

Schindler, G. and Dellaert, F., 2004. Atlanta world: an expectation maximization framework for simultaneous low-level edge grouping and camera calibration in complex man-made environments. pp. I: 203–209.

Schmitt, F. and Priese, L., 2009. Vanishing point detection with an intersection point neighborhood. In: Discrete Geometry for Computer Imagery, Vol. 5810, Springer Berlin / Heidelberg, pp. 132–143.

Straforini, M., Coelho, C., Campani, M. and Torre, V., 1992. The recovery and understanding of a line drawing from indoor scenes. 14(2), pp. 298–303.

Tardif, J.-P., 2009. Non-Iterative Approach for Fast and Accurate Vanishing Point Detection. In: Proceedings 12th International Conference on Computer Vision.

Tuytelaars, T., Gool, L. V., Proesmans, M., Moons, T. and Mi, E., 1998. The cascaded hough transform as an aid in aerial image interpretation. In: In Proc. ICCV, b, pp. 67–72.

Wildenauer, H. and Vincze, M., 2007. Vanishing point detection in complex man-made worlds. In: ICIAP '07: Proceedings of the 14th International Conference on Image Analysis and Processing, IEEE Computer Society, Washington, DC, USA, pp. 615–622.