Scene based camera pose estimation in Manhattan worlds

Darko Vehar[a,c], Rico Nestler[b,c], and Karl-Heinz Franke[c]

[a]Ilmenau University of Technology, Computer Graphics Group
[b]Ilmenau University of Technology, Group of Quality Assurance and Industrial Image
Processing, P.O. Box 100565, 98684 Ilmenau, Germany
[c]Zentrum für Bild- und Signalverarbeitung (ZBS) e.V., Werner-von-Siemens-Straße 10, 98693
Ilmenau, Germany

ABSTRACT

This paper presents a principle for scene-related camera calibration in Manhattan worlds. The proposed estimation of extrinsic camera parameters from vanishing points represents a useful alternative to the traditional target-based calibration methods, especially in large urban or industrial environments. We analyse the effects of errors in the calculation of camera poses and derive general restrictions for the use of our approach. In addition, we present methods for calculating the position and orientation of several cameras to a world coordinate system and discuss the effect of imprecise or incorrectly calculated vanishing points. Our approach was evaluated with real images of a prototype for human-robot collaboration installed at ZBS e.V.. The results were compared with a perspective n-Point (PnP) method.

Keywords: 3D reconstruction, Manhattan world, extrinsic wide-baseline calibration, image based calibration, line segmentation, line classification, vanishing points

1. INTRODUCTION AND MOTIVATION

A prerequisite step and an important challenge for a 3D reconstruction with multiple cameras according to the geometric principle of triangulation is the precise estimation of the camera's imaging properties and their spatial relation to a common world coordinate system. The later is also referred to as the camera pose.

Traditional techniques determine the *intrinsic* imaging properties of a camera from a picture of a calibration object with known geometry. Using known world-to-pixel correspondences, the internal camera parameters can be computed precisely with calibration methods.

Calibration techniques can also be used to determine the *extrinsic* camera parameters—position and orientation of the camera in world coordinates. However, when wide-baseline camera systems for observation of large workspaces have to be calibrated extrinsically, this form of calibration becomes impractical. An easy-to-handle calibration object should ideally cover the entire measuring space but is often imaged too small or is only partially visible in the camera frustums. In such cases, the estimation of the target point positions is uncertain or inaccurate due to the image resolution. Furthermore, scenes objects, such as machines and equipment, interfere with the positioning and visibility of the calibration objects significantly. Thus, in practice, often only a part of the measuring volume can be included in the system calibration.

In urban and industrial environments a variety of 3D object edges usually appear parallel to each other. Furthermore, these parallel lines can be assigned to three orthogonal directions, as they occur on 3D objects with mainly right angles. A scene with these stated constraints is referred to as *Manhattan world*[1] and is represented with a *Manhattan world model*. In the perspective image, the characteristics of this world model form the basis for the proposed approach of scene-related camera calibration. We exemplify how a camera pose with respect to the Manhattan world and subsequently, the relative orientation and position between two cameras

---

Further author information: (Send correspondence to D.V.)
D.V.: E-mail: darko.vehar@tu-ilmenau.de, Telephone: +49 3677 6897686
R.N.: E-mail: rico.nestler@tu-ilmenau.de, Telephone: +49 3677 6897685
K.-H.F.: E-mail: karl-heinz.franke@zbs-ilmenau.de, Telephone: +49 3677 6897681
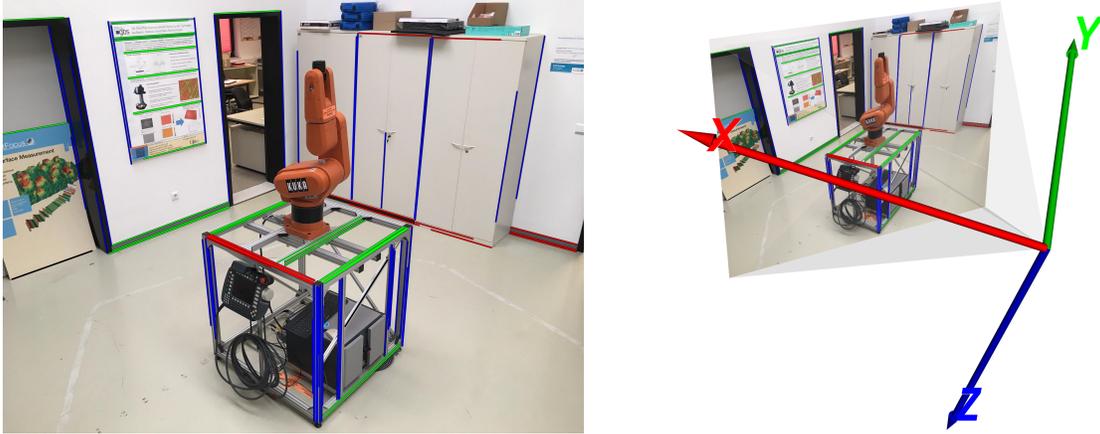
Figure 1. Prototype for a human-robot collaboration installed at the ZBS e.V.. The line segments in the left image are shown in colours according to the assigned axes of a Manhattan world. The image on the right shows a world coordinate system that was derived from vanishing points.

are calculated, given images of a Manhattan world with intrinsically precalibrated, distortion-free cameras. The method described here is part of a calibration toolkit *3D-EasyCalib*$^{\mathrm{TM}}$ developed by the ZBS e.V., which solves several geometrical calibration problems and can be used with minimal expert knowledge.

2. THEORY

2.1 Pinhole camera model and camera calibration

The foundation for the proposed approach is the perspective projection of a 3D Manhattan world scene modelled by an ideal pinhole camera. The pinhole camera model defines the transformation of the three-dimensional Euclidean space onto the two-dimensional image plane according to the principle of central projection. If the 3D coordinate system of the world is set equal to that of the camera with the origin at point C, then by using a notation of Refs. 2, the projection can be described compactly with the equation

$$\boldsymbol{x} = K[I|\boldsymbol{0}]\boldsymbol{X_C} \,. \tag{1}$$

Image point $\boldsymbol{x}$ and world point $\boldsymbol{X_C}$ are represented by homogeneous vectors. Both are uniquely determined up to a scale factor. The index C shows that the world point $\boldsymbol{X_C}$ is defined in the camera coordinate system. The 3×3 matrix

$$K = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

is called the camera matrix or the calibration matrix of the pinhole camera model. It consists in its basic form, with zero skew and unit aspect ratio, of the focal length f, the distance of the image plane from the camera centre, and of the principal point (p_x, p_y). Both are expressed here in terms of pixel dimensions.

If the point $\boldsymbol{X}$ is defined in a 3D world, for example in a Manhattan world, and the camera pose (orientation and position) in this world is defined with the rotation matrix R and the translation $\boldsymbol{t}$, then X must first be rigidly transformed into the camera coordinate frame with $\boldsymbol{X_C} = \begin{bmatrix} R & \boldsymbol{t} \\ \boldsymbol{0}^T & 1 \end{bmatrix} \boldsymbol{X}$ before it is projected (1) onto the image plane. In summary, the well-known algebraic relationship of the projective camera is derived as (see Fig. 2, left)

$$\boldsymbol{x} = K[R|\boldsymbol{t}]\boldsymbol{X} \,. \tag{3}$$

Traditional calibration approaches, such as in Refs. 3, calculate the unknown camera parameters K, R, $\boldsymbol{t}$ with the help of a geometric calibration target whose known world points $\{\boldsymbol{X_i}\}$ and corresponding image points $\{\boldsymbol{x_i}\}$ are used to obtain the solution of (3). Also for the method presented in this article, the cameras are precalibrated intrinsically (K) according to this principle.
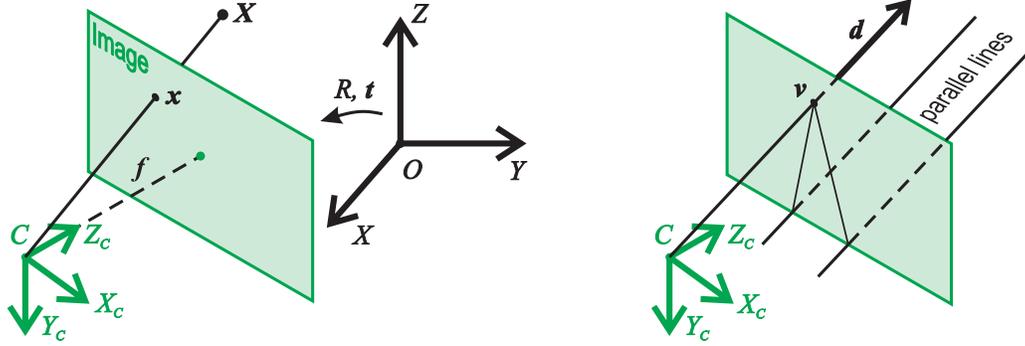
Figure 2. Pinhole geometry of a camera placed at the coordinate origin C. Camera and Manhattan world coordinate frames are related via rotation and translation (left). Formation of a vanishing point as intersection of two imaged parallel world lines (right).

2.2 Vanishing points in a Manhattan world

A set of parallel lines with direction $\boldsymbol{d}$ defined in the 3D camera coordinate system intersect in the ideal point, the point at infinity, $\boldsymbol{X_\infty} = (\boldsymbol{d}^T, 0)^T$ of the associated projective space. The perspective projection of this ideal point on the image plane is called *vanishing point*. It can be calculated by inserting the ideal point into the equation (1) as $\boldsymbol{v} = K\boldsymbol{d}$. Analogously, the inverse projection of the vanishing point is defined as a ray through the camera centre and the point $\boldsymbol{v}$ with the direction $\boldsymbol{d} = K^{-1}\boldsymbol{v}$ as shown in Fig. 2, right. Consequently, for all vanishing points found in the image of a Manhattan world, which belong to three mutually perpendicular directions, the vectors $\boldsymbol{d_1} = K^{-1}\boldsymbol{v_1}$, $\boldsymbol{d_2} = K^{-1}\boldsymbol{v_2}$ and $\boldsymbol{d_3} = K^{-1}\boldsymbol{v_3}$ are parallel to the axes of a Manhattan world which are orthogonal to each other (see Fig. 3, left).

2.3 Derivation of the camera pose's rotation matrix R

The normalized vectors $\boldsymbol{d_1}$, $\boldsymbol{d_2}$, $\boldsymbol{d_3}$ derived according to the above considerations form an orthonormal basis of a world coordinate system

$$\boldsymbol{r_1} = \frac{\boldsymbol{d_1}}{\|\boldsymbol{d_1}\|}, \boldsymbol{r_2} = \frac{\boldsymbol{d_2}}{\|\boldsymbol{d_2}\|}, \boldsymbol{r_3} = \frac{\boldsymbol{d_3}}{\|\boldsymbol{d_3}\|} . \tag{4}$$

The sought rotation matrix $R = \begin{bmatrix}\boldsymbol{r_1}, \boldsymbol{r_2}, \boldsymbol{r_3}\end{bmatrix}$ is composed of these basis vectors. To construct a valid rotation in a right-handed coordinate system, the order of the $\boldsymbol{r_1}$, $\boldsymbol{r_2}$ and $\boldsymbol{r_3}$ has to be chosen accordingly.

2.4 Derivation of the camera pose's translation vector $\mathbf{t}$

The translation vector $\mathbf{t}$ cannot be computed unambiguously from an image without additional knowledge of the scene. A practical way to estimate the translation is to use known distances in the scene. For example, two known point pairs $\{\boldsymbol{x_i}, \boldsymbol{X_i}\}$ can be used to solve (3) with the previously determined rotation matrix R. Consequently, errors from the calculation of the rotation could propagate in the determination of the translation.

An alternative was presented in Refs. 4. The required parameters consist of a known direction $\boldsymbol{d}$ that is parallel to one of the axes of the Manhattan world, the real length of a line segment on the direction $\boldsymbol{d}$, and the image coordinates of the start and end points of the depicted line segment. The principle of this approach is shown in Fig. 3, right.

From the similarity of the triangles $CO'I$ and COP, the distance of the point O from the camera centre is $\|\overrightarrow{CO}\| = \|\overrightarrow{CO'}\| \|\overrightarrow{OP}\| / \|\overrightarrow{O'I}\|$ and thus the translation vector $\boldsymbol{t}$ can be determined as

$$\boldsymbol{t} = \overrightarrow{CO} = \|\overrightarrow{CO}\| \frac{\overrightarrow{CO'}}{\|\overrightarrow{CO'}\|} . \tag{5}$$
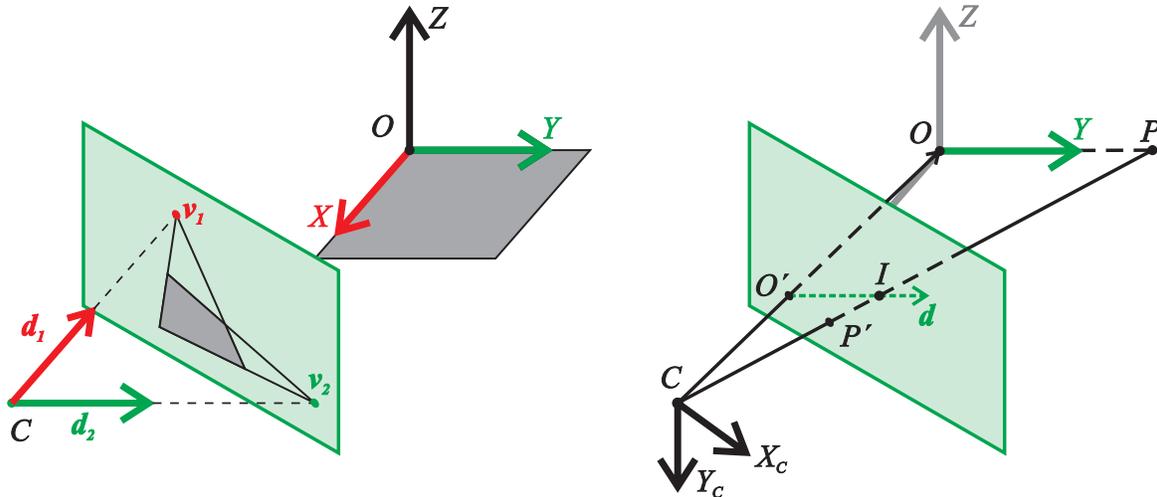
Figure 3. The derivation of the rotation (left) and the translation (right) from vanishing points. The 3D points O' and P' lie on the image plane and are defined in the camera coordinate system with its origin in C.

3. VANISHING POINT DETECTION

It was shown above how vanishing points can be used to derive extrinsic camera parameters. A vanishing point results as the intersection of all parallel lines defined in 3-space which are projected perspectively onto an image plane, as shown in the Fig. 2, right. As the accuracy of the vanishing point positions is crucial for the precision of the parameters derived therefrom, a short revision of the methods for the vanishing point detection is in place. This is followed by a short review of approaches for line segmentation in the section below, which is a common prerequisite for the vanishing point detection algorithms.

Since the first successful attempt by Barnard[7] in 1983, many methods have been proposed for the detection of vanishing points in images. Most approaches consist of three steps: line extraction, lines to vanishing point assignment and vanishing point position estimation. In particular, the clustering of the lines into groups sharing the same vanishing point distinguishes the methods, since this can be done either in the image space or in a transformed space, for example in Hough space or on the Gaussian sphere. Manhattan world assumption can either be used in the clustering-step or in the vanishing point computation.

According to our own research,[8] the agglomerative clustering method, called J-Linkage[9, 10] produced the best results because of its robustness against outliers, that occur in images of non-ideal Manhattan worlds. Similar to a RANSAC scheme, J-Linkage evaluates vanishing point hypotheses using consensus sets of lines. The separation of the clusters is based on the Jaccardi distance.

3.1 Procedures for line segment extraction

An established and widespread approach uses the robust Canny edge detector and a parametric line search, e.g. in Hough space, to extract lines from an image. Its susceptibility to faults (i.e false-positive lines), especially in regions with fine textures, leads to erroneous results of the subsequent vanishing point detection.

We evaluated two alternative methods, the *Fast Line Detector*[5] (FLD) and the *Line Segment Detector*[6] (LSD). Both output line segments with sub-pixel accuracy and produce subjectively correct results. The FLD processes the results of the Canny detector by using a co-linearity constraint to track straight line segments along contours. The LSD splits the image into so-called line support regions, i.e. regions of similar local gradient orientation. These regions are merged into the line segments by a region growing approach.

Figure 4. Input image of a test scene (left), line segments extracted with FLD (center) and LSD (right).

4. PROPOSED METHOD AND EVALUATION

The theory for the camera pose estimation from an image of a Manhattan world and the J-Linkage algorithm for the vanishing point detection from pre-detected line segments were merged into a processing pipeline of our proposed method. It consists of four consecutive steps: line segment extraction, line clustering, vanishing point calculation and camera pose (rotation and translation) computation.

The process steps were carried out exemplary with an image of the sample scene shown in Fig. 4, left. Colour images were converted to grayscale before processing. The intrinsic parameters of the camera are: $f = 1721.11$ pixels, principal point $(p_x, p_y) = (1001.15, 753.91)$ at an image resolution of 2016×1512 pixels. The intrinsic calibration including a lens distortion correction was done beforehand.

4.1 Line segment extraction

Utilizing an intensity image as an input, both algorithms (LSD and FLD) parameterize line segments as start and end points with subpixel-precision. Some parameters, e.g. minimum line length, may be adjusted depending on the resolution of the input image and scene properties to obtain the best results.

For the example image (Fig. 4, left) FLD outputs 563 and LSD 611 line segments. According to the subjective evaluation, the consistency of the results is similar. Statistical evaluation of the line segments length shows, that the LSD favours shorter line segments with a smaller length variance compared to the FLD. Since no preference on this basis can be made at this point, the final method selection takes place in the context of the subsequent processing steps.

4.2 Edge clustering and vanishing point calculation

Estimated line segments were input to the J-Linkage algorithm and labelled to corresponding vanishing points as output. Since a Manhattan world was assumed, the three clusters with the largest number of assigned line segments correspond to the wanted Manhattan world axes (vanishing points).

The intersection of lines in a cluster was computed using a least squares method. In addition to the calculated vanishing point, the root-mean-square (RMS) over the smallest distances of the lines to the intersection (vanishing point) was computed. This represents the accuracy of the line detectors and gives an indication of the validity of the estimated vanishing points in case of degenerate camera orientation (Sec. 4.3). Some results of the here favoured FLD and the LSD methods are shown in the Tab.1.

The RMS of the LSD line segments was slightly smaller than that of the FLD in every case. The J-linkage output nearly twice as many clusters for lines extracted with FLD (28) as for LSD (11). Both results depend on the precision of the extracted line segments. The evaluation of additional images resulted in similar differences in favour of the LSD, so it is the recommended method for the line segmentation step.

Table 1. Comparison of two favoured line extraction methods with coordinates of the estimated vanishing points and their accuracy for the scene shown in Fig. 4. The value of #v is the number of distinctive vanishing points found by J-Linkage algorithm.

Method (#v)	v_1	v_2	v_3
FLD (28)	$(-1054.88, 43.10)$	$(2672.13, -20.03)$	$(1087.19, 4668.74)$
RMS	40.57	21.86	59.54
LSD (11)	$(-1087.41, 30.96)$	$(2653.04, -5.65)$	$(1086.31, 4711.57)$
RMS	27.47	21.44	36.4
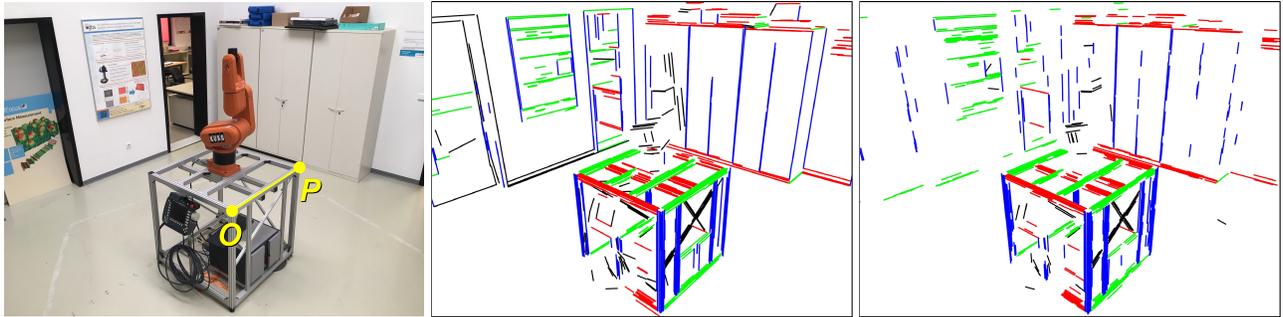
Figure 5. Input image (left), line clustering based on the lines extracted with FLD (middle) and LSD (right). The edges are shown in colour according to the assigned vanishing points. Black lines could not be assigned to vanishing points of the imaged Manhattan world.

4.3 Estimation of the camera pose's rotation

For the calculation of the rotation matrix according to (4), the vanishing points were arranged clockwise with respect to the principle point. For the direction vectors computed from the left image in Fig. 4, the pairwise angles between the basis vectors were: $\angle(\boldsymbol{r_1}, \boldsymbol{r_2}) = 89.49°$, $\angle(\boldsymbol{r_2}, \boldsymbol{r_3}) = 89.5°$ and $\angle(\boldsymbol{r_1}, \boldsymbol{r_3}) = 90.36°$. The resulting rotation matrix was orthogonalized with the help of Rodrigues' rotation formula.

Depending on the orientation of the camera in the Manhattan world, a rotation matrix from (4) can not always be estimated. These cases occur when one or two, hence all three, axes of the camera coordinate system are parallel to axes of the world coordinate system. Thus, the projections of parallel lines in the world are also parallel on the image plane and the associated vanishing point is at the point of infinity.

If only one infinite vanishing point is present, the corresponding direction vector can be calculated as the cross product of the other two. The second case should be obviated when possible, for example by changing the camera pose, as further analysis is otherwise impossible.

4.4 Estimation of the camera pose's translation

The computation of the equation (5) of camera pose's translation is trivial. Only point I can possibly be misinterpreted. It results as an intersection of the rays $C + \lambda_1 \overrightarrow{CP'}$ and $O + \lambda_2 \boldsymbol{d}$ in 3-space as shown in Fig. 3, right.

We selected points O and P on the upper side of the robot frame (see Fig. 5, left), determined their image coordinates and computed the translation vector $\boldsymbol{t}$ according to the equation (5). The length of the side of the frame is $\|\overrightarrow{OP}\| = 880$ mm. Thereby the distance from the camera centre to the origin of the Manhattan world is $\|\boldsymbol{t}\| = 2411.37$ mm and the orientation, for easier interpretation presented as Euler angles (z, y', x" convention), is: $yaw = 138.98°$, $pitch = -0.68°$, $roll = 66.72°$.

4.5 Estimation of the relative pose between two cameras

The motivation for the proposed scene-based calibration was not only to estimate the exterior orientation of multiple cameras in a common world frame but to determine their relative orientation and position to each

Table 2. Vanishing points and RMS of the shortest distances of clustered lines to the corresponding vanishing points.

	v_1	v_2	v_3
Fig. 6, top	$(818.41, 4805.54)$	$(-1551.14, -137.14)$	$(2435.64, 53.72)$
RMS	43.81	40.54	30.03
Fig. 6, bottom	$(639.56, 2826.98)$	$(-9053.44, -2453.98)$	$(1696.53, -559.24)$
RMS	33.06	113.97	23.95

Table 3. Comparison of camera poses with respect to Manhattan world. Translation and rotation of the camera in Fig. 6, bottom with respect to camera in Fig. 6, top.

	$\|\boldsymbol{t_{ref}}\|$ [mm]	$\|\boldsymbol{t}\|$ [mm]	$\|\boldsymbol{t_{ref}} - \boldsymbol{t}\|$ [mm]	θ [°]
Fig. 6, top	2712.89	2710.76	2.13	0.60
Fig. 6, bottom	2092.83	2091.54	1.34	0.28
	$\|\boldsymbol{t_{21,ref}}\|$ [mm]	$\|\boldsymbol{t_{21}}\|$ [mm]	$\|\boldsymbol{t_{21,ref}} - \boldsymbol{t_{21}}\|$ [mm]	θ_{21} [°]
Fig. 6, stereo	2750.18	2746.08	7.57	0.33

other. Given a camera pair with known extrinsics, a selected point in the world is transformed to the respective camera frames with equations $\boldsymbol{X_1} = \begin{bmatrix} R_1 & \boldsymbol{t_1} \\ \boldsymbol{0}^T & 1 \end{bmatrix} \boldsymbol{X}$ and $\boldsymbol{X_2} = \begin{bmatrix} R_2 & \boldsymbol{t_2} \\ \boldsymbol{0}^T & 1 \end{bmatrix} \boldsymbol{X}$. Substitution of the $\boldsymbol{X}$ from the second equation with the $\boldsymbol{X}$ from the first one gives us the orientation and translation of the camera C_2 with respect to C_1 as

$$R_{21} = R_2 R_1^T \quad \text{and} \quad \boldsymbol{t_{21}} = \boldsymbol{t_2} - R\boldsymbol{t_1} \,. \tag{6}$$

It should be noted, that the R_1, R_2, $\boldsymbol{t_1}$ and $\boldsymbol{t_2}$ have to be defined with respect to the same coordinate system. In the procedure described in Sec. 4.3 the rotation matrices were computed by manually assigning the world axes. This step will be automated with image-based matching methods in the future.

5. COMPARISON OF THE PROPOSED APPROACH WITH A PNP METHOD

To compare the presented scene-based calibration approach with a traditional target-based method, a calibration target was placed in a test scene and aligned with the Manhattan world. The coordinate system of the Manhattan world is in this case equal to the coordinate system of the calibration target. We determined the camera pose $(R_{ref}, \boldsymbol{t_{ref}})$ using an iterative PnP algorithm[11] and then using our proposed method $(R, \boldsymbol{t})$. The control points of the target were calculated with sub-pixel accuracy. The measured length of the line segment required for the calculation of the translation according to (5) is $\|\overrightarrow{OP}\| = 720$ mm. The intrinsic parameters of the cameras are equal to the intrinsics in the previous section.

We chose the following metrics to compare the results. The error in the translation is represented by the Euclidean norm of the vector difference $\|\boldsymbol{t_{ref}} - \boldsymbol{t}\|$. The difference of the orientations is determined by the angle error $\theta = \cos^{-1}\left(\frac{1}{2}(tr(R_{ref}R^T) - 1)\right)$ which defines a metric for 3D rotations.[12]

The vanishing points with associated RMS distances are shown in the Tab. 2. In the Fig. 6, second row, right, the green line segments are almost parallel to each other and the calculated vanishing point $\boldsymbol{v_2}$ lies distinctly outside of the image. The associated RMS error is as expected the largest. Nonetheless, the basis vectors $\angle(\boldsymbol{r_1}, \boldsymbol{r_2}) = 90.1°$, $\angle(\boldsymbol{r_2}, \boldsymbol{r_3}) = 90.1°$ and $\angle(\boldsymbol{r_1}, \boldsymbol{r_3}) = 89.57°$ are almost perpendicular to each other, thereby no additional processing steps were necessary.

The results of the camera poses with respect to the world and the resulting relative orientation and position of the cameras to each other are shown in the Tab. 3. The observed differences in the translations were only a few millimetres large. The error in the rotation was smaller than one angular degree. The overall accuracy of the proposed method was only slightly worse than that of the PnP method. The evaluation of additional images confirmed these results. In the case of the stereo setup, Tab. 3, bottom, the computation errors of the relative pose between two cameras accumulate as expected.
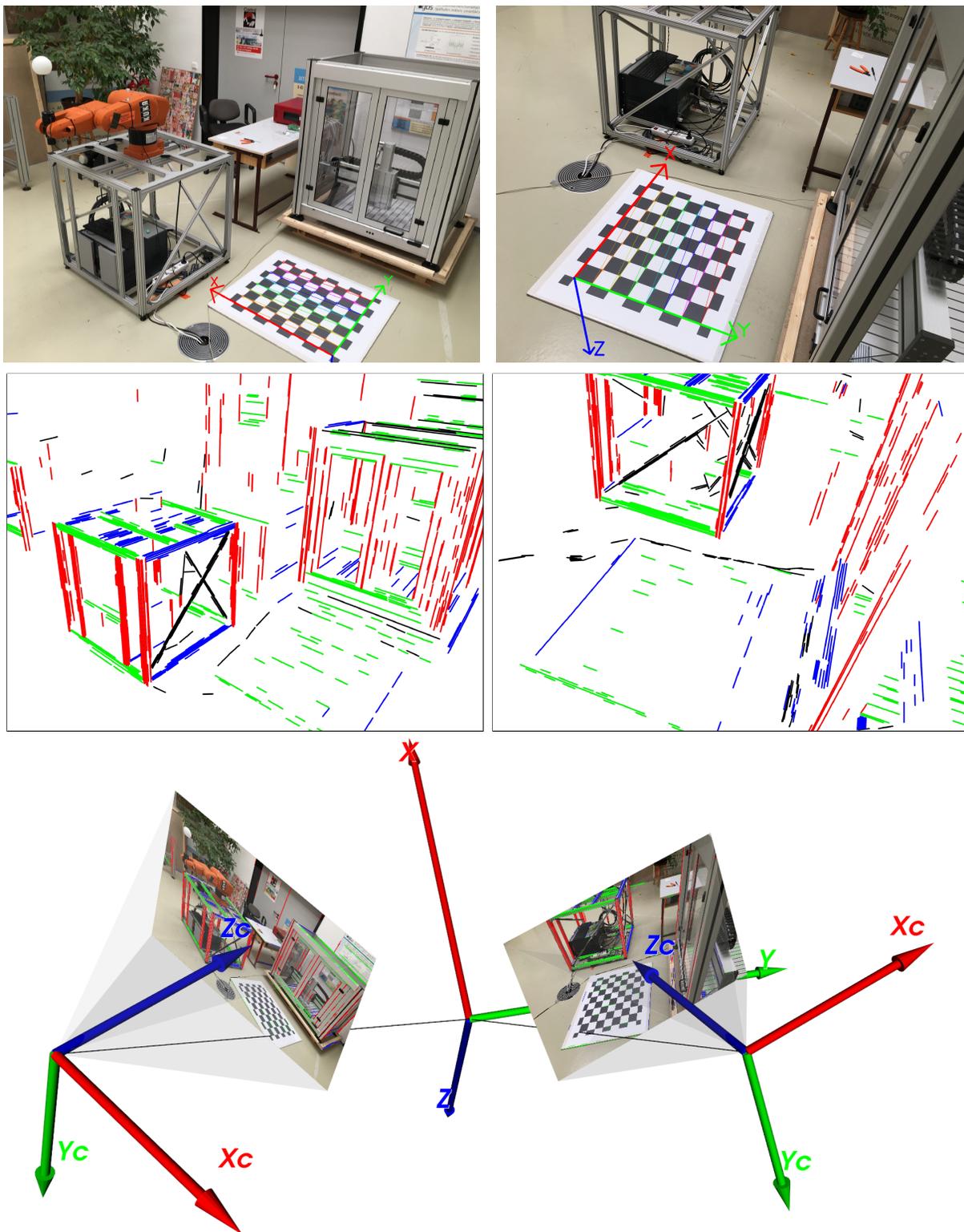
Figure 6. Input images of two different camera views of the same scene with a coordinate system of the target found by our method (top row). The classified line segments from LSD and J-linkage algorithm are shown in the second row. 3D view at the bottom shows both cameras and the Manhattan world coordinate system.

6. SUMMARY AND FUTURE WORK

In environments that do not allow target-based extrinsic calibration and are represented by a Manhattan world, the proposed method of scene-related calibration is a useful alternative. The comparison with a traditional approach shows that our procedure achieves an accuracy that is suitable for many applications.

Manhattan world assumption is a prerequisite for the techniques presented in this paper. When a scene only partially fulfils the constraints of this model, for example, when only two orthogonal directions can be extracted from the imaged scene, the Manhattan world can still be identified. We can supplement the missing information by using suitable heuristics or improve the erroneous results. With the error metrics presented in the article, such situations can be reliably identified and automatically corrected. These challenges are the subject of our further research.

ACKNOWLEDGMENTS

REFERENCES

[1] Coughlan, J. M. and Yuille, A. L., "Manhattan world: Compass direction from a single image by bayesian inference," in [*Proceedings of the International Conference on Computer Vision*], **2**, 941–947, IEEE Computer Society, Washington, DC, USA (1999).

[2] Hartley, R. I. and Zisserman, A., [*Multiple View Geometry in Computer Vision*], Cambridge University Press, New York, second ed. (2004).

[3] Zhang, Z., "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(11), 1330–1334 (2000).

[4] Guillou, E., Meneveaux, D., Maisel, E., and Bouatouch, K., "Using vanishing points for camera calibration and coarse 3d reconstruction from a single image," *The Visual Computer* **16**(7), 396–410 (2000).

[5] Lee, J. H., Lee, S., Zhang, G., Lim, J., Chung, W. K., and Suh, I. H., "Outdoor place recognition in urban environments using straight lines," in [*IEEE International Conference on Robotics and Automation*], 5550–5557 (2014).

[6] Grompone von Gioi, R., Jakubowicz, J., Morel, J.-M., and Randall, G., "Lsd: a line segment detector," *Image Processing On Line* **2**, 35–55 (2012).

[7] Barnard, S. T., "Interpreting perspective images," *Artificial Intelligence* **21**, 435–462 (Nov. 1983).

[8] Rehawi, L., *Geometrical Estimation of Multiple Cameras in a Manhattan-World*, bachelor's thesis, TU Ilmenau - Department of Computer Science and Automation / SECSY, ZBS e.V. (2018).

[9] Toldo, R. and Fusiello, A., "Robust multiple structures estimation with j-linkage," in [*Computer Vision – ECCV 2008*], 537–547, Springer Berlin Heidelberg, Berlin, Heidelberg (2008).

[10] Tardif, J., "Non-iterative approach for fast and accurate vanishing point detection," in [*2009 IEEE 12th International Conference on Computer Vision (ICCV)*], 1250–1257, IEEE Computer Society, Los Alamitos, CA, USA (2009).

[11] Bradski, G., "The opencv library," *Dr. Dobb's Journal of Software Tools* (2000).

[12] Huynh, D. Q., "Metrics for 3d rotations: Comparison and analysis," *Journal of Mathematical Imaging and Vision* **35**, 155–164 (Oct. 2009).