The toulouse vanishing points dataset

Vincent Angladon, Simone Gasparini, Vincent Charvillat

The Toulouse Vanishing Points Dataset

Vincent Angladon[1,2], Simone Gasparini[1], and Vincent Charvillat[1]

[1]Université de Toulouse; INPT – IRIT; 118 Route de Narbonne, F-31062 Toulouse, France,
{vincent.angladon, simone.gasparini, vincent.charvillat}@irit.fr

[2]Telequid; Toulouse, France

ABSTRACT

In this paper we present the Toulouse Vanishing Points Dataset, a public photographs database of Manhattan scenes taken with an *iPad Air 1*. The purpose of this dataset is the evaluation of vanishing points estimation algorithms. Its originality is the addition of Inertial Measurement Unit (IMU) data synchronized with the camera under the form of rotation matrices. Moreover, contrary to existing works which provide vanishing points of reference in the form of single points, we computed uncertainty regions. The Toulouse Vanishing Points Dataset is publicly available at http://ubee.enseeiht.fr/tvpd

Categories and Subject Descriptors

I.4.8.h [**Artificial Intelligence**]: Image Processing and Computer Vision—*Sensor fusion*; I.2.10.b [**Artificial Intelligence**]: Vision and Scene Understanding—*3D/stereo scene analysis*; I.4.1.b [**Image Processing and Computer Vision**]: Digitization and Image Capture—*Imaging geometry*

General Terms

Performance

Keywords

Vanishing points, RGB images, IMU data, Dataset

1. INTRODUCTION

Image understanding requires the analysis of the geometric properties of the image: since the perspective projection is a non-invertible mapping between the 3D dimensional scene and the 2D image plane, the depth information is lost, thus making image interpretation a challenging task [5]. Studying and analysing the geometric properties of an image is thus crucial to recover the spatial layout of the scene.

A well known geometric entity that can be used as strong cue for image understanding is the *vanishing point*. Under the perspective projection, parallel lines in the scene are mapped to a pencil of lines that intersect in a so-called vanishing point (VP), an image point that is the projection of the intersection of the parallel lines at infinity. In a calibrated camera, a vanishing point gives the 3D direction of the pencil of lines. Detecting a VP can thus provide a strong constraint on the scene geometry. For example, most man-made scenes consist of three orthogonal dominant directions, *i.e.* there are three main sets of parallel lines; this is often referred to as the "Manhattan World" [13]. By detecting these three orthogonal VPs associated to the sets of parallel lines, some information about the camera and the scene can be inferred: *e.g.* the camera can be calibrated [9, 39] and its rotation w.r.t. the scene can be estimated [2, 22, 24]. Vanishing points can be used as priors to constrain the 3D reconstruction of such scenes [18]. Recently, vanishing points have received a lot of interest in many works dealing with the indoor and outdoor scene understanding and reconstruction from a single image [21, 25, 28] as a fundamental cue for recovering the spatial layout of the scene [30, 33].

Another important source of information that can help the interpretation of a scene is the inertial data. In the last years we witnessed the development and the large diffusion of mobile devices equipped with inertial measurement unit (IMU), such as accelerometers, magnetometers, and gyroscopes. Thanks to such sensors, the absolute orientation and the gravity vector of the camera can be estimated for each taken picture. Inertial data has been widely used in robotics in combination with the visual data in order to estimate the movement and the pose of the robots [12]. The recent diffusion of mobile devices has fostered their adoption in many computer vision and multimedia applications, such as 3D reconstruction [36], in order to provide better estimation of the camera movement, especially when the visual data is affected by, *e.g.*, occlusions and motion blur.

The fusion between inertial and visual data is thus becoming an interesting topic because of their complementarity. Inertial data is indeed computationally cheap but suffers from drift and measurement noise; visual data can provide more precise and stable measurements but it is computationally more expensive. In the case of VP detection, the orientation and gravity vector provided by the IMU sensors can be used as priors for driving and easing the process of VP detection. In this paper, we propose a new dataset collecting images of indoor and outdoor "Manhattan" scenes taken with a modern mobile device equipped with IMU sen-

sors. Each image stores the inertial data (rotation matrix) of the moment at which it has been captured. The dataset provides a ground truth for the VPs for each image: instead of providing a single point for each of the 3 orthogonal directions, we rather supply uncertainty regions in which the triplet of orthogonal VPs shall lie, according to the manually extracted segments.

The main contributions of this work are the creation of a new dataset of photographs associated with the camera orientation provided by IMU data and a new method to compute regions of uncertainty for the location of the VPs using line segments. To the best of our knowledge, there is no public dataset containing photographs associated with IMU data. In this paper we present our efforts to create one such dataset.

The paper is organized as follows: Section 3 describes related datasets used for the evaluation of VPs algorithms and their limitations. Section 2 gives some background on VPs computation from image processing and how to use IMU data as a prior for estimating the VPs. Section 4 explains how we computed our reference VPs with the ground truth segments while Section 5 describes the methodology used to collect the data. Finally, Section 6 concludes the paper.

2. BACKGROUND

In the following subsections, we provide a short background on the problem of estimating vanishing points with a general, non-exaustive overview of the main techniques. We also present an overview of the IMU sensors that are typically found on modern mobile devices, the data they provide and how it can be used to ease the VP detection and estimation.

2.1 Vanishing point detection

The detection of vanishing points requires the extraction of geometric features in the image, such as image gradients, lines or line segments, which can be clustered to estimate the VPs. Each cluster contains a pencil of lines corresponding to parallel 3D lines of the scene. This task can be considered a "chicken-and-egg" problem: if the feature clustering is known, then the VPs can be easily estimated as the point that minimizes a certain distance measure w.r.t. the features of each cluster. Conversely, if the VPs are given, the feature clustering is easily solved by assigning each feature to the "closest" VP (w.r.t. a certain distance measure). For this reason, a prior knowledge about the scene or the camera orientation given by the IMU sensor might ease the problem, as described in the next section.

In this work we focus on the VP detection in "Manhattan scenes", in which there exist 3 dominant, mutually orthogonal directions. The above problem is thus constrained to the retrieval of a triplet of mutually orthogonal VPs. As stated in Section 1, this problem is crucial for many computer vision applications dealing with indoor and outdoor scene in man-made environments.

Generally, most of the proposed methods in the literature for estimating orthogonal VPs employ line segments as image features, which can be extracted with advanced image processing techniques [19] based on a-contrario approach. Other approaches consider image gradients [10], low level features that provide local orientation information, and that mostly used to detect a single, dominant VP.

The estimation procedure then follows two main steps: the *clusterization* of the line segments and the *VP estimation* for each cluster. Various techniques have been suggested in the literature for the clusterization of the lines: Hough based methods [5], RANSAC frameworks [1, 31, 40] and J-linkage algorithm [37]. The second step relies on the estimation of the VP for each cluster as the point that minimizes an error function W for each segment line of the cluster. Several formulations have been proposed for W, such as point-line distance error functions [3, 31, 37], orientation error functions [15, 29, 32], or probabilistic error functions [11, 41]. The final triplet is then chosen among the possible solutions or, as in [31] the orthogonality constraint can be enforced during the VP estimation process.

Some recent methods do not follow a tightly two-step process but they rather try to solve the problem globally: [6] tries to find the rotation (*i.e.* a triplet of VP) that maximizes the number of clustered segments. [3] follows a global approach in which the clusterization and the VP estimation are solved simultaneously as an *Uncapacitated Facility Location* problem. Another global approach proposed by [24] casts the VP detection in an Expectation Maximization framework.

2.2 Inertial data

In the last decades we witnessed a notable breakthrough in microelectronics which brought low-cost miniaturized silicon sensors to common mobile devices such as smart-phones and tablets. In particular, Inertial Measurement Unit (IMU) sensors usually consist of accelerometers measuring the acceleration of the device, and gyroscopes measuring the rate of change of the device's orientation. The IMU measurements can provide good accuracy information on the position, velocity, and attitude over a short period of time. On the other hand, they are usually corrupted by different types of error sources such as sensor noises, scale factor and temperature dependent bias, which are nonlinear and difficult to characterize [17]. Moreover, they all provide derived measures (acceleration, angular velocity), which need to be integrated to compute the current position and attitude, thus causing error accumulation and a significant drift in the position and the attitude over the course of time. These problems can be mitigated by employing optimal estimation and filtering techniques such as the Kalman filter [23].

In this work we consider the attitude data provided by the IMU sensor, which is normally given w.r.t. the direction of the Earth magnetic North pole. This can be considered as an estimation of the camera orientation, under the realistic assumption that the two reference systems are aligned [12]. Usually the device operating system allows the developers to retrieve the attitude information in the form of a rotation matrix $\mathtt{R}$, encoding the yaw, pitch and roll angles of the device. From this matrix, it is easy to recover the vertical VP (*i.e.* the zenith) of a Manhattan scene: $\mathbf{v}_{\text{zenith}} = \mathtt{K}\,\mathtt{R}\,[\mathbf{0}\;\mathbf{0}\;\mathbf{1}]^{\mathsf{T}}$, where $\mathtt{K}$ is the calibration matrix of the camera.

The vertical VP is dual to the horizon line, *i.e.* the projection of the plane containing the camera center, whose normal is parallel to the direction of the vertical VP. Since the two remaining VPs of the Manhattan scene are mutually orthogonal w.r.t. the detected vertical VP, the detection problem can be thus reduced to the search of one VP along the horizon line.

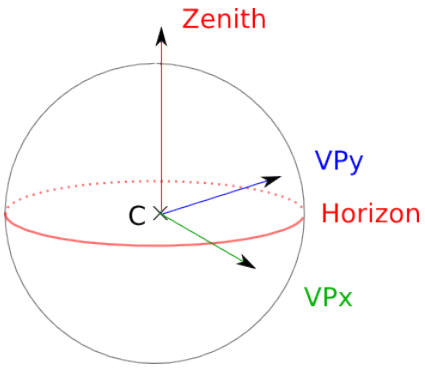

Figure 1: The three VPs form an orthogonal frame. C is the optical center of the camera. The knowledge of the zenith enables to reduce the search of the two VPs orthogonal to the zenith, VP_x and VP_y on the horizon.

Figure 2: The green square is the estimated VP by [15] (associated horizon in red). The light green point (associated horizon dashed) is the VP after orthogonalization of the Manhattan directions: it lies far from the common intersection zone of the associated line segments.

3. EXISTING DATASETS

There are two main well known datasets that have been used to compare VP detection methods. The *York Urban Database* [15, 16], published in 2008, was the first extensive dataset for VPs estimation algorithms evaluation in Manhattan scenes. It is the most popular dataset used by most of the works to assess the effectiveness of the proposed method. This dataset consists of 102 indoor and outdoor images of Manhattan scenes. Each image is provided with the hand-made ground truth line segments: an interactive program is used to select and identify line segments with sub-pixel precision and assign to each of them the corresponding Manhattan direction, so that 3 clusters of lines are obtained. The camera is calibrated using a subset of the images: assuming a natural camera (*i.e.* square pixels), the focal length and the principal point are estimated in a non linear optimization process by enforcing the mutual orthogonality of the estimated VP triplets. The VPs are estimated using the algorithm proposed by [11]. It must be noted that this algorithm computes VP using a statistical framework from a given set of line segment clusters, and each VP is thus estimated separately and no orthogonality constraint is enforced. Then an orthogonal frame is fitted to each triplet to enforce the constraint. This yields to an orthogonal solution which is not necessary optimal given the statistical distribution of the line intersections used for the estimation. The resulting Manhattan directions, indeed, can be quite far from the line segments intersections as it can be seen in Figure 2. The obtained orthogonal solution might be a biased solution that may not be suitable to be used as reference to evaluate and compare VPs estimation algorithms.

Recently, the *PKU Campus Database* [26, 27] has been proposed as a VP dataset consisting of 200 indoor and outdoor photographs of Manhattan scenes. The dataset is inspired by the York Urban Database and indeed it has been built in a similar way, except for camera calibration that has been done off-line using the method proposed by [35] and [42]. The line segments are detected automatically but the algorithm is not described. Finally, the VP are computed using the same method as the York Dataset, thus being affected by the same issues described above.

For the sake of completeness, we also mention the *Eurasian Cities Dataset* [4, 38], which collects 103 outdoor urban images. However, the dataset was built with a different scope in mind, focusing on the more general scenes that do not necessarily fit the Manhattan hypothesis. An interesting characteristic is that the authors provide, among other ground truth data, the horizon for each image, computed using a least square minimization of the horizontal VPs. Our dataset provides the same information obtained directly from the IMU data.

Vanishing points algorithms are generally evaluated by comparing the position of the estimated VP with the reference VP provided by the dataset. This scheme assumes that the reference VPs are not biased, which is not the case of the previously mentioned datasets.

Creating a ground truth for VP detection is a hard and challenging task, even if the images are manually annotated: as pointed out in [41], many deviations from a perfect imaging system such as camera noise, camera calibration errors, line segment extraction error, *etc.* affect the estimation of the ground truth orthogonal VPs, and only optimal or sub-optimal solution can be found for them. The only way to have *real* ground truth data for the VP would be the use of synthetic images, in which all the parameters are known by design, or using real images and highly accurate and costly instruments (*e.g.* electronic theodolites) to measure the actual attitude of the camera w.r.t. the Manhattan scene.

A more meaningful approach that we are proposing in this dataset is to provide an uncertainty region for the locations of the VPs, as opposed to single points. This information can be used to reject or accept the solution of an algorithm (the solution is respectively outside or inside the region).

4. GROUND TRUTH CREATION

The construction of a vanishing points dataset requires two elements: photos and reference vanishing points. We decided to compute the reference VPs with hand-labeled line segments which must be accurately drawn. The uncertainty of a ground truth line segment comes from the selection of the two extrema, which can be modeled with circular regions of uncertainty around the extrema (see Figure 3). Shufelt [34] was the first to introduce the error modeling for line segment endpoints in a VP detection algorithm. The true position of a line segment endpoint is assumed to lie *among all the possible locations within its pixel.* The lines

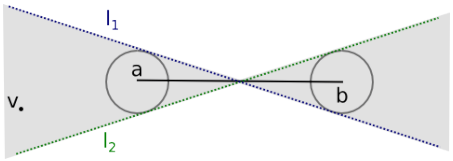

Figure 3: The uncertainty of a ground truth line segment is modeled with circular regions of uncertainty around the two extrema a and b. The lines connecting all these possible endpoints sweep an area bounded by two lines, called *double wedge* (the area in grey), in which the associated VP v should lie.

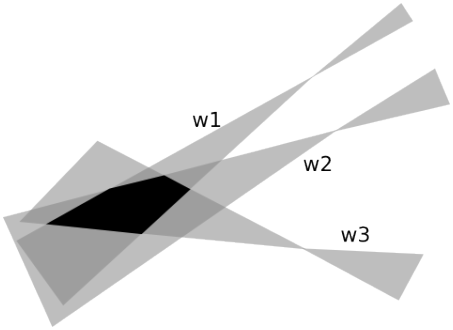

Figure 4: The intersection of the different double wedges w1, w2, w3 associated to the image of parallel lines of the scene is a convex polygon in which the VP should lie.

connecting all these possible endpoints sweep an area which is bounded by two lines, $\mathbf{l}_1$ and $\mathbf{l}_2$ as in Figure 3. This area is called a *double wedge* [7] (the grey area in Figure 3). In his proposed method the Gaussian sphere [5] is divided into accumulators, each wedge region is projected on the sphere and the corresponding accumulators are incremented. The maxima on the sphere then represent the directions of the VPs.

More recently, Xu [41] introduced a probabilistic consistency measure, which models the uncertainty of endpoint locations with a 1D Gaussian which is then used in an EM framework to estimate the VPs. Contrary to Xu, we followed a geometrical approach because our objective is to compute a confidence region for the solution, rather than finding one VP solution. In the confidence regions, all the possible VPs are equiprobable since we do not assume it is less unlikely to commit a two pixels error on an endpoint rather than one pixel. In this sense, our approach to find the regions is closer to [34]: we work in the image plane, and instead of using accumulators, we compute the exact geometric intersection of the double wedges.

Assuming the real line associated with the annotated line segment is contained in its double wedge, the intersection of all the double wedges of a given line segment cluster forms a region in which the VP should lie (see Figure 4).

Using double wedges to model the uncertainty of the line segment is interesting as they naturally take into account the length of the segment. In general, long line segments should be more robust as they mitigate the annotation error of the two extrema. A long line segment, indeed, has a thinner double wedge, and thus it will contribute to narrow down the uncertainty region of the associated VP. Conversely, short segments have wider wedges which do not help to reduce the uncertainty region.

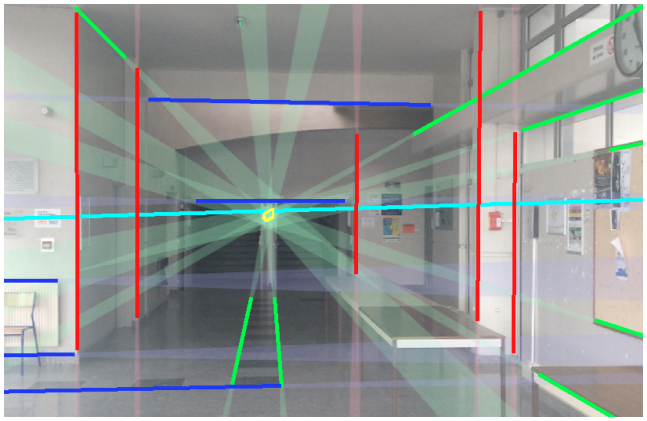

Figure 5: The yellow polygon is the intersection of the green double wedges. The cyan line is the horizon computed from the IMU data.

We reformulate the double wedge intersection problem in term of Boolean operations on half-planes. Let $\mathbf{l}_1$ and $\mathbf{l}_2$ be the bounding lines (see Figure 3) of a segment [**a b**]. Without loss of generality, consider the half-planes h_1 and h_2 bounded by $\mathbf{l}_1$ and $\mathbf{l}_2$ respectively, and both containing **a**. The double wedge w associated to [**a b**] is defined as

$$w = (h_1 \cap h2) \cup \left(\overline{h_1} \cap \overline{h_2}\right),$$

where $\overline{h_i}$ denotes the complementary of h_i, *i.e.* the other half of the plane.

The intersection of the double wedges of all the line segments thus requires the computation of the intersections and unions of the half-planes h_i of each line segment, which is a well-known computational geometry problem treated in [7]. The computation of the intersection is performed in the projective plane, which is equivalent to performing the computation on the Gaussian sphere: this allows us to compute the intersection of parallel lines and to handle the case of VPs at infinity.

5. DATA COLLECTION METHODOLOGY

At the time of writing, the dataset contains 114 photographs (40 indoor and 74 outdoor). The photos were taken at different moments of the day and therefore have various exposures. A majority of the indoor scenes contain low levels of clutter (chairs, sofas, ...). In contrast, a majority of outdoor scenes contain a lot of occluding objects such as trees and vehicles, making the estimation of VP more challenging. The photos were taken holding the camera in different attitudes in order to have a sufficient variety of poses: post-hoc analysis revealed a mean and maximal absolute angular value between the camera principal axis and the horizon of 6.7° and 26° respectively. Figure 6 shows some selected photos from the dataset.

We collected the photos using an *iPad Air 1* running *iOS 8* in landscape mode with a 1920x1080 resolution and using the following *iOS* capture presets: automatic white balance, auto exposition and fixed focus. The auto-focus was disabled because it can add a significant random lag between the

Figure 6: Some photos of the dataset with their ground truth line segments (red, green, blue), the horizon line computed from the IMU data (cyan line) and the polygon of the uncertainty region computed on the red line segments. On the top row, the horizon lines computed with the IMU data do not intersect the yellow polygons because of the bias of the IMU data.

moment the shutter button is pressed and the effective shot of the photo.

Instead of using the raw data values of the accelerometers, gyroscopes and the magnetometer, we used the *CMDeviceMotion* class of the *iOS SDK* which provides high level data such as the gravity and the attitude of the device through sensor fusion algorithms not detailed in the official documentation. A 30Hz sampling rate was set to collect the IMU data. We developed a specific application for recording the device orientation provided by the *CMDeviceMotion* class along the taken photos. The source code of the application is available for download at the dataset website.

Camera and IMU calibration.

The camera was calibrated offline using Bouguet camera calibration toolbox [8] to estimate the intrinsic parameters. Experiments on the IMU sensors holding the *iPad* on a try square shown that in the worst case, we could obtain 2 degrees of error on the roll and pitch values. This bias is visible in the Figure 6, where the horizon lines computed with the IMU data do not intersect the polygons of the uncertainty regions associated to the VPs orthogonal to the zenith. In addition, no calibration of the IMU sensors is performed since our observations revealed that the flatness of the ground is less reliable and repeatable than the orientation values returned without setting a reference attitude, *e.g.* the ground.

A known issue affecting mobile devices is the synchronization between the data provided by the IMU sensors and the image provided by the camera [20]. Due to the low-cost design, most devices do not synchronize the IMU and image data using, *e.g.*, a global common timestamp. In [20] it is shown that this may be critical when designing methods that fuse IMU and image data. Our preliminary experiments demonstrated that for our device the IMU data could be not

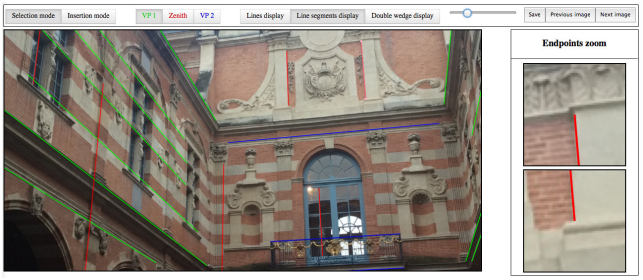

Figure 7: The web application used to annotate the images with line segments.

synchronized w.r.t. the orientation computed using the image. The mean lag between the IMU data and the camera frames was found to be 16ms with a standard deviation of 140ms. To take into consideration this uncertainty and provide smoother data, we computed the attitude matrix as the average rotation [14] over a time window covering the mean lag measured during the preliminary experiments.

Line segments creation.

In order to generate the ground truth, a web application has been developed (see Figure 7) to let the users accurately draw the line segments and to associate them with one of the three Manhattan directions (as in [15]). A post-hoc analysis revealed a mean of 16.7 segments per photo. We assumed a 4 pixels accuracy around the endpoints clicked by the users. This value has been determined experimentally as the average value that ensured that the intersection of the double wedges was not empty and contained the VP solution provided by [3]. We also made comparisons on the York Urban Database, see Figure 8. As expected, the VP computed with [11] lie in our uncertainty regions. Since the orthogonalization process is independent of the line segments, the orthogonalized VP do not always lie in the uncertainty regions (see Section 3).

The source code of the application is available for download at the dataset web-page.

Data organisation.

All the pictures are in the same folder. The IMU data are stored in the EXIF *UserComment* field of the photos. The data are given in JSON format and they contain the attitude of the mobile device at the time of the shot, represented as a change of basis matrix from the world reference frame to the camera frame.

The line segments ground truth are stored in JSON format in a separate file in the format `imagename.txt`. Finally, a `imagename.mat` Matlab binary file is also provided to easily access to the line segments and the mobile device attitudes.

Dataset license.

The source code is distributed under the terms of the *BSD licence* and the dataset content under the terms of the *Creative Common by-nc-sa Licence.*

6. CONCLUSION

In this paper we presented a new photographs dataset of indoor and outdoor Manhattan scenes for the evaluation and comparison of vanishing points estimation algorithms.

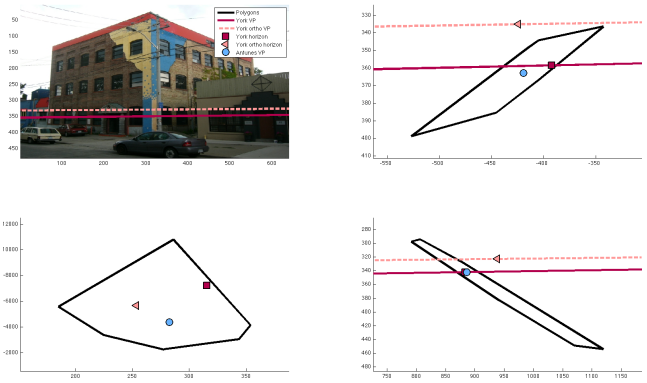

Figure 8: Comparison on the York Urban Database. The red squares (VP provided by the dataset using [11], associated horizon in red) and the blue circles (VP computed using [3]) lie in our uncertainty regions. The orthogonalized VP are represented with pink triangles (associated horizon dashed) are not in our polygons.

This dataset is the only one to our knowledge to include IMU data. Instead of providing *real* ground truth data for the VPs, we opted for a more meaningful approach consisting in computing uncertainty regions for the location of the vanishing point. These regions are provided in the form of polygons and are computed by intersecting the double edges of the ground truth line segments. We believe that the use of IMU data, despite their bias, can ease the computation of vanishing points and provide more robust results with images containing a majority of outlier line segments. We hope our works will stimulate the design and the comparison of algorithms using IMU data which are widely used in robotics and in mobile device applications. The Toulouse Vanishing Points Dataset is available for download at http://ubee.enseeiht.fr/tvpd

7. REFERENCES

[1] D. G. Aguilera, J. G. Lahoz, and J. Finat Codes. A new method for vanishing points detection in 3D reconstruction from a single view. In *Proceedings of the ISPRS Working Group V/4 Workshop*, 2005.

[2] M. Antone and S. Teller. Automatic recovery of relative camera rotations for urban scenes. In *Proceedings of the 200 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2000)*, volume 2, pages 282–289. IEEE Comput. Soc, 2000.

[3] M. Antunes and J. P. Barreto. A Global Approach for the Detection of Vanishing Points and Mutually Orthogonal Vanishing Directions. *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR2013)*, pages 1336–1343, June 2013.

[4] O. Barinova, V. Lempitsky, E. Tretyak, and P. Kohli. Eurasian Cities Dataset. http://graphics.cs.msu.ru/en/research/projects/msr/geometry, 2010.

[5] S. Barnard. Interpreting perspective images. *Artificial Intelligence*, 21(4):435–462, Nov. 1983.

[6] J. C. Bazin, C. Demonceaux, P. Vasseur, K. Ikeuchi, and M. Pollefeys. Globally optimal line clustering and vanishing point estimation in Manhattan world. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR2012)*, number 1, pages 638–645. IEEE, June 2012.

[7] M. d. Berg, O. Cheong, M. v. Kreveld, and M. Overmars. *Computational Geometry: Algorithms and Applications*. Springer-Verlag TELOS, Santa Clara, CA, USA, 3rd ed. edition, 2008.

[8] J. Y. Bouguet. Camera calibration toolbox for Matlab, 2008.

[9] B. Caprile and V. Torre. Using vanishing points for camera calibration. *International Journal of Computer Vision*, 4(2):127–139, Mar. 1990.

[10] J. Choi, W. Kim, H. Kong, and C. Kim. Real-time vanishing point detection using the Local Dominant Orientation Signature. In *Proceedings of the 2011 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON)*, pages 1–4. IEEE, 2011.

[11] R. Collins and R. Weiss. Vanishing point calculation as a statistical inference on the unit sphere. In *Proceedings of the 1990 IEEE International Conference on Computer Vision (ICCV 1990)*, pages 400–403. IEEE Comput. Soc. Press, 1990.

[12] P. Corke, J. Lobo, and J. Dias. An Introduction to Inertial and Visual Sensing. *The International Journal of Robotics Research*, 26(6):519–535, June 2007.

[13] J. M. Coughlan and A. L. Yuille. Manhattan world: Orientation and outlier detection by bayesian inference. *Neural Comput.*, 15(5):1063–1088, May 2003.

[14] W. Curtis, A. Janin, and K. Zikan. A note on averaging rotations. In *Proceedings of the IEEE Virtual Reality Annual International Symposium*, pages 377–385. IEEE, 1993.

[15] P. Denis, J. H. Elder, and F. J. Estrada. Efficient Edge-Based Methods for Estimating Manhattan Frames in Urban Imagery. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Proceedings of the 2008 European Conference on Computer Vision (ECCV 2008)*, pages 197–210. Springer Berlin Heidelberg, 2008.

[16] P. Denis, J. H. Elder, and F. J. Estrada. York Urban Line Segment Database. http://www.elderlab.yorku.ca/YorkUrbanDB/, 2008.

[17] N. El-Sheimy, H. Hou, and X. Niu. Analysis and Modeling of Inertial Sensors Using Allan Variance. *IEEE Transactions on Instrumentation and Measurement*, 57(1):140–149, Jan. 2008.

[18] Y. Furukawa, B. Curless, S. M. Seitz, and R. Szeliski. Manhattan-world stereo. In *Proceedings of the 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, pages 1422–1429, 2009.

[19] R. Grompone von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall. LSD: a Line Segment Detector. *Image Processing On Line*, 2:35–55, Mar. 2012.

[20] C. Ham, S. Lucey, and S. Singh. Hand Waving Away Scale. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, editors, *Proceedings of the 2014 European Conference on Computer Vision (ECCV 2014)*, volume 8692 of *Lecture Notes in Computer Science*, pages 279–293, Cham, 2014. Springer International Publishing.

[21] V. Hedau, D. Hoiem, and D. Forsyth. Recovering the spatial layout of cluttered rooms. *2009 IEEE 12th International Conference on Computer Vision*, pages 1849–1856, Sept. 2009.

[22] A. T. Joseph J. Lim, Hamed Pirsiavash. Parsing ikea objects: Fine pose estimation. In *Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV 2013)*, 2010.

[23] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35, 1960.

[24] J. Košecká and W. Zhang. Video Compass. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Proceedings of the 2002 European Conference on Computer Vision (ECCV2002)*, pages 476–490. Springer Berlin Heidelberg, 2002.

[25] J. Košecká and W. Zhang. Extraction, matching, and pose recovery based on dominant rectangular structures. *Computer Vision and Image Understanding*, (February 2005), 2005.

[26] B. Li, K. Peng, X. Ying, and H. Zha. PKU Campus Database. http://www.cis.pku.edu.cn/vision/vpdetection/, 2012.

[27] B. Li, K. Peng, X. Ying, and H. Zha. Vanishing point detection using cascaded 1D Hough Transform from single images. *Pattern Recognition Letters*, 33(1):1–8, Jan. 2012.

[28] B. Micusik, H. Wildenauer, and J. Kosecka. Detection and matching of rectilinear structures. In *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR2008)*, pages 1–7. IEEE, June 2008.

[29] M. Nieto and L. Salgado. Real-time robust estimation of vanishing points through nonlinear optimization. In N. Kehtarnavaz and M. F. Carlsohn, editors, *Proceedings of SPIE 7724, Real-Time Image and Video Processing 2010*, pages 772402–772402–14, Apr. 2010.

[30] S. Ramalingam, J. K. Pillai, A. Jain, and Y. Taguchi. Manhattan Junction Catalogue for Spatial Reasoning of Indoor Scenes. In *Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV2013)*, pages 3065–3072. IEEE, June 2013.

[31] C. Rother. A new approach to vanishing point detection in architectural environments. *Image and Vision Computing*, 20(9-10):647–655, Aug. 2002.

[32] 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 *Proceedings of the 2004 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2004)*, volume 1, pages 203–209. IEEE, 2004.

[33] A. G. Schwing, S. Fidler, M. Pollefeys, and R. Urtasun. Box in the Box: Joint 3D Layout and Object Reasoning from Single Images. In *Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV2013)*, pages 353–360. IEEE, Dec. 2013.

[34] J. Shufelt. Performance evaluation and analysis of vanishing point detection techniques. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 21(3):0–6, 1999.

[35] P. Sturm and S. Maybank. On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications. In *Proceedings of the 1999 IEEE Conference on Computer Vision and Pattern Recognition (CVPR1999)*, volume 1, pages 1432–1437, 1999.

[36] P. Tanskanen, K. Kolev, L. Meier, F. Camposeco, O. Saurer, and M. Pollefeys. Live metric 3d reconstruction on mobile phones. In *Proceedings of the 2013 IEEE International Conference on Computer Vision*, Washington, DC, USA, 2013. IEEE Computer Society.

[37] J.-P. Tardif. Non-iterative approach for fast and accurate vanishing point detection. In *Proceedings of the 2009 IEEE International Conference on Computer Vision (ICCV2009)*, pages 1250–1257. IEEE, Sept. 2009.

[38] E. Tretyak, O. Barinova, P. Kohli, and V. Lempitsky. Geometric Image Parsing in Man-Made Environments. *International Journal of Computer Vision*, 97(3):305–321, Sept. 2011.

[39] H. Wildenauer and A. Hanbury. Robust camera self-calibration from monocular images of Manhattan worlds. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR2012)*, pages 2831–2838. IEEE, June 2012.

[40] H. Wildenauer and M. Vincze. Vanishing Point Detection in Complex Man-made Worlds. In *Proceedings of the 2007 International Conference on Image Analysis and Processing (ICIAP 2007)*, pages 615–622. IEEE, Sept. 2007.

[41] Y. Xu, S. Oh, and A. Hoogs. A Minimum Error Vanishing Point Detection Approach for Uncalibrated Monocular Images of Man-Made Environments. In *Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR2013)*, pages 1376–1383. IEEE, June 2013.

[42] Z. Zhang. A flexible new technique for camera calibration. *International Journal of Computer Vision*, 22:1330–1334, 2000.