Assessment of camera orientation in Manhattan scenes using information from optical and inertial sensors

Evgeny Myasnikov
Geoinformatics and Information Security department
Samara National Research University;
Image Processing Systems Institute of RAS - Branch of the FSRC "Crystallography and Photonics" RAS
Samara, Russia
mevg@geosamara.ru

***Abstract*—In the present paper, the solution to the problem of assessing the orientation of a camera is performed under the condition of two main limitations. The first limitation is the analysis of Manhattan scenes only. The second one is the presence of an accelerometer in a mobile device. To assess the characteristics of the proposed solution, a data set was prepared containing both photos and accelerometer readings, as well as information about the true orientation of the device. Experimental studies were carried out using the prepared data set.**

Keywords*—*camera orientation, vanishing point, Manhattan scenes, accelerometer, inertial sensor

I. Introduction

Assessing the camera orientation is one of the most important tasks in three-dimensional computer vision. Typically, camera orientation is estimated using calibration patterns, and it requires human interaction. For this reason, automatic methods for assessing the orientation are of particular interest.

Despite the presence of various sensors in modern mobile devices, such as an accelerometer, compass, etc., their use for orientation estimation is limited due to the low accuracy and the influence of noise [1]. For this reason, both optical information and information from the sensors of mobile devices are used to determine the orientation of the camera.

In this paper, we consider a method for assessing the orientation of a camera, based on the analysis of the position of vanishing points [2], i.e. the points in the plane of a perspective image, in which projections of mutually parallel lines of three-dimensional space converge. In this case, the problem is solved under the condition of two main limitations. The first limitation is the restriction of the class of analyzed scenes only to Manhattan scenes [3], in which the lines are aligned along three main mutually orthogonal directions. Vivid examples of such scenes are photographs of city buildings (the lines of building facades may possess these characteristics), road scenes (border of the roadway, markings, poles), indoor scenes (borders of rooms, furniture lines, decoration elements - panels, tiles, etc.). The second limitation is the presence of an accelerometer on a mobile device.

The orientation of the camera in this paper is determined sequentially in several stages. At the first stage, using the inertial sensor readings, the direction to the first vanishing point corresponding to the direction of gravity is determined. After that, the position of the first vanishing point is refined along vertical lines in the optical image. At the second stage, the vanishing points of the horizontal lines of the main and side facades are determined. So the found vanishing points, taking into account the data of the inertial sensor, determine the orientation of the camera. The method for determining vanishing points described in this paper is based on the idea described in [4], according to which the search for horizontal vanishing points can be performed along the horizon line defined by a plane orthogonal to the direction of the vertical vanishing point.

Unfortunately, common data sets for the evaluation of vanishing point assessment methods (see, for example, [5]) do not contain information from inertial sensors. For this reason, their use for evaluating methods similar to those described in this paper is possible only in the mode of sensors emulation, as it was done, for example, in [6].

For the above reason, to evaluate the characteristics of the proposed solution, we prepared our own data set containing both photos and accelerometer readings as well as information about the true camera orientation. Experimental studies were carried out using the prepared data set.

It should be noted that the initial implementation of the algorithm for determining vanishing points was previously described in [6]. Thus, in the present work, the previously proposed approach is further developed and studied, using the data set prepared as part of the work.

The work is organized as follows. Section 2 describes the developed method for assessing camera orientation. Section 3 describes the modeling technique and conducts experimental studies. The work ends with a conclusion and a list of used literature.

II. Method

As it was mentioned in the introduction, the described method consists of sequentially determining three vanishing points, followed by finding the orientation of the camera. The general scheme of the method is presented in Fig. 1.

First, preliminary processing of the image received from the camera is performed. In particular, it is scaled and rotated with an accuracy of 90 degrees in accordance with the information received from the inertial sensor. If necessary, the vector received from the sensor is transformed so as to correspond to the direction of gravity for the rotated image.

After preliminary processing by one of the known methods, for example, by the Canny method [7], contours are extracted from the image. The extracted contours are traced and the segments of straight lines are searched. The found segments form the set L, which will be used subsequently to find vanishing points.

Further, the information obtained from the inertial sensor is used for a preliminary assessment of the first vanishing point VP_1. It is assumed that the direction to the first vanishing point corresponds, up to a sign, to the gravity vector.

In the direction obtained, a set L_1 of segments is selected such that the lines corresponding to this set deviate from the direction to VP_1 no more than the predefined angle.

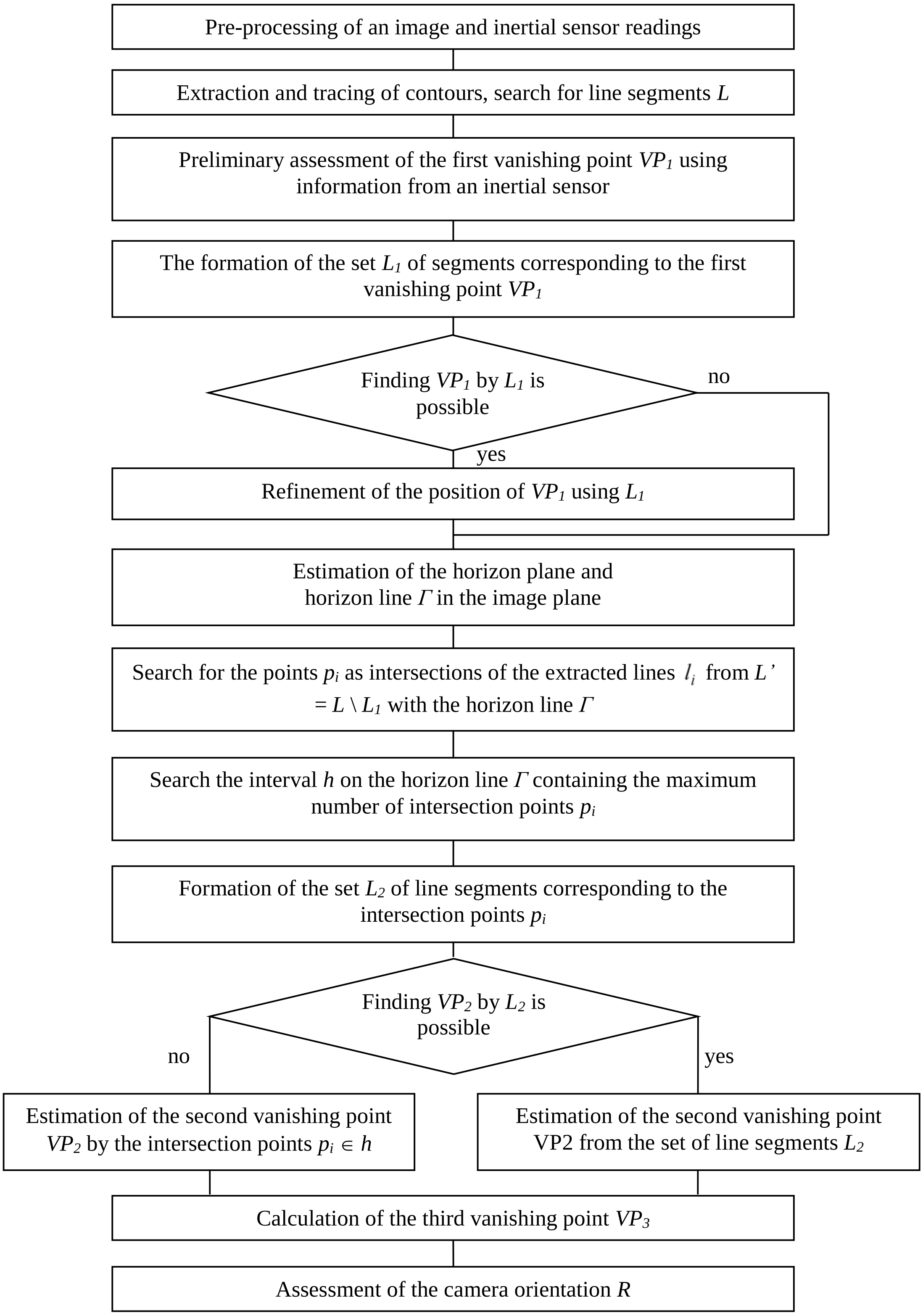


Fig. 1. General scheme of the method.

If there are enough selected segments, the first vanishing point is refined by a weighted summation of the points determined by all possible segments from L_1. Moreover, segments of greater length have more weight. If the assessment of VP_1 by L_1 is not possible, the initial estimation of VP_1 is used for further processing.

At the next stage, the direction to VP_1 is used to determine the horizon line plane as the plane passing through the origin of the modeled optical system and orthogonal to the direction to VP_1 (refined gravity vector). In addition to the plane, the horizon line Γ is also determined as the projection of a line, which is in the horizon plane and non-orthogonal to the image, on the image plane.

Further, for all the lines $l_i \in L'$ extracted in the image, with the exception of the lines used earlier to find the vanishing point VP_1 ($L' = L \setminus L_1$), the intersection points with the horizon line Γ are determined. A search is made for such a segment h (with a predetermined angular size) on the horizon line, at which the maximum number of intersection points pi falls.

After this, we form the set L_2 of line segments, for which the intersections p_i with the horizon line Γ fall in the indicated interval h. If there are enough selected segments, the second vanishing point is estimated by weighted summation of the intersection points determined by all possible segments from L_2. If the estimation of VP_2 by L_2 is not possible, the weighted sum of the points of intersection of the corresponding lines with the horizon line Γ is taken for the position of VP_2. In both cases, the segments of longer length have more weight when determining VP_2.

After determining two vanishing points, the third is found as a vector orthogonal to the vectors corresponding to the first and second points: $V_3 = V_1 \times V_2$.

After finding vanishing points, the camera orientation can be found as follows:

$$R=[r_1\ r_2\ r_3],$$

where R is the rotation matrix, and vectors r_1 r_2 r_3 are calculated as

$$r_1 = mK^{-1}VP_1,\ r_2 = mK^{-1}VP_2,\ r_3 = r_1 \times r_2,$$

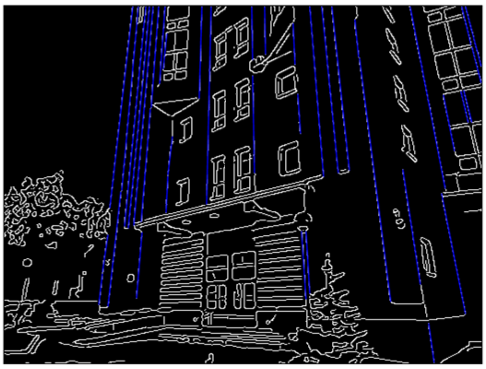
where m is the scale factor, K is the matrix of internal parameters of the camera [8] containing information on the focal length, pixel size, tilt, the shift of the image center relative to the optical axis.

In general, the proposed method is the development of the previously described method [6], the main idea of which [4] is to search for horizontal vanishing points along the horizon line defined by a plane orthogonal to the direction of the vertical vanishing point. Compared with the previous implementation, both the individual steps of the method underwent changes (the search for segments on the contours is now carried out according to the criterion of maximum deviation, the weighted summation takes into account the lengths of the segments of lines, which are separated from each other by a sufficient distance, the second vanishing point is refined without using histograms), as well as the general scheme of the method (now contains branches that increase the reliability of determining vanishing points, as well as the actual orientation estimation stage).

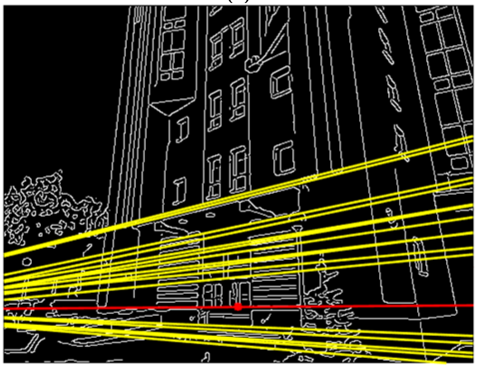
III. Experiments

To study the method described above, we used our own specially prepared data set. This set was collected using the Huawei Honor 9 lite smartphone [9]. Its camera has a CMOS BSI sensor with an f / 2.2 aperture, a focal length of 3.46 mm, and produces a color image of 12.98 MP. To collect images and inertial sensor data, we developed an Android application that stores both the captured images and a custom number of accelerometer readings recorded prior to the shot.

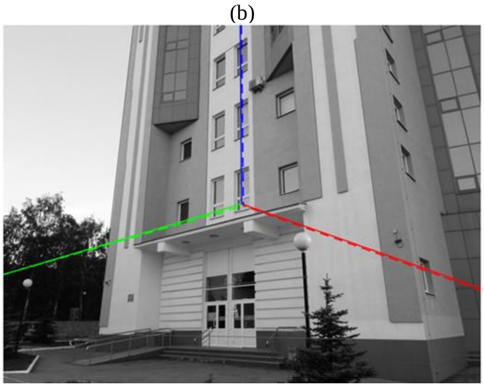
To obtain information about the true position of the camera, several (from 3 to 7 for each vanishing point) lines were manually selected that reliably determine the directions to the true vanishing points. This procedure was performed at 2x magnification, and normalized vanishing points obtained using selected lines were considered as true vanishing points. At the moment, the described data set consists of 40 images of buildings with the corresponding inertial sensor data and true orientation data.



(a)



(b)



(c)

Fig. 2. An example of the method: a) extracted contours (white) and the set of lines segments corresponding to the first vanishing point (blue); b) the horizon line (red) and the set of lines defining the second vanishing point (red); c) directions to true (dashed lines) and estimated (solid lines) vanishing points.

An example demonstrating the various stages of the proposed method is shown in Fig. 2.

To assess the quality of the developed method, modeling was performed according to the following scheme:

- for each image from the prepared data set, three vanishing points and camera orientations relative to the building depicted in the photograph were determined;

- using information about the true position, for each vanishing point, the error was calculated as the angular deviation of the direction to the estimated vanishing point from the true direction;

- based on the data obtained for each vanishing point, a histogram of the angular deviation of the found points from their true values was constructed, and the average value of such a deviation was also calculated.

The experimental results are shown in the following Figure 3.

Each of the histograms shown in the figure shows the angular deviation of the estimated vanishing point from its true position. In the ideal case, such a histogram should have a single column on the left side (first), which means the minimum deviation of the vanishing point from the true values for all test images. As can be seen from the above figures, in most cases the position of the three vanishing points was made with a deviation of up to 2º, while the deviation exceeded 4º was observed for only 3 of 40 images. The average error values were: 1.69º, 1.54º, and 1.88º for the first, second, and third points, respectively.

It should be noted that using only information from the inertial sensor (see the histogram in Fig. 1 (a)) provided a greater level of errors in determining the direction to the first vanishing point. The average error value was 3.7º when using only an inertial sensor versus 1.69º when refined with an optical image. Thus, the accuracy of the algorithm can be improved in conditions of noisy readings of the gravity vector by selecting parameters. Another way to increase accuracy may be to use previously obtained estimates in processing a video stream, which is the subject of future research.

IV. Conclusion

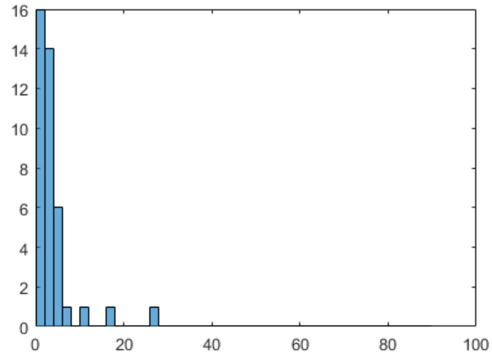
The method for automatic assessment of the orientation of a camera in Manhattan scenes using information from optical and inertial sensors is proposed and investigated. To study the developed technique, the data set was created containing digital images of buildings, readings of inertial sensors, as well as information about the true position of vanishing points obtained by careful manual marking of the source images.

The described method is simple to implement, and undemanding to computing resources. Its use allows reducing the average level of errors in determining the orientation in more than 2 times compared with the inertial sensor.

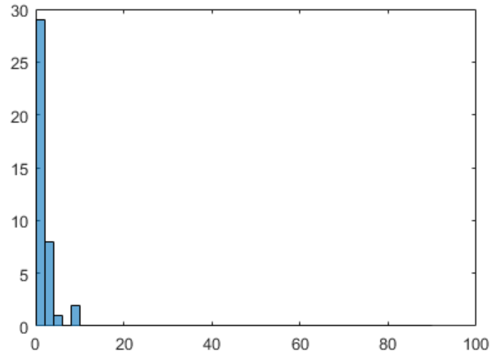
As a direction for further work, it is planned to expand the method for assessing the orientation and position of the camera when working with a video stream.

Acknowledgment

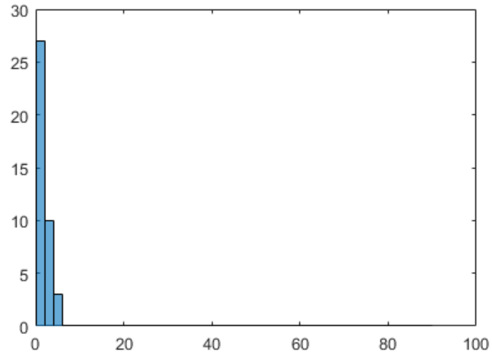
The work was partly funded by RFBR according to the research project 17-29- 03190-ofi-m in parts of «2. Method» - «3. Experiments» and by the Russian Federation Ministry of Science and Higher Education within a state contract with the «Crystallography and Photonics» Research Center of the RAS in parts «1. Introduction» and «4. Conclusion».



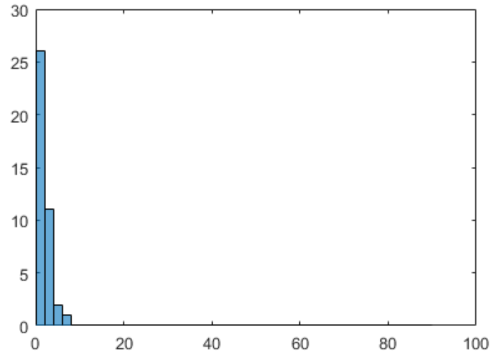
(a)



(b)



(c)



(d)

Fig. 3. Estimation of the method quality. Histograms of the angles deviations of directions to the vanishing points from their true values (in degrees): a) the first vanishing point, estimated by inertial sensor readings; b) the first vanishing point, refined by an optical image; c) the second vanishing point; d) the third vanishing point.

REFERENCES

[1] V.V. Myasnikov and E.A. Dmitriev, "The accuracy dependency investigation of simultaneous localization and mapping on the errors from mobile device sensors," Computer Optics, vol. 43, no. 3, pp. 492-503, 2019. DOI: 10.18287/2412-6179-2019-43-3-492-503.

[2] B. Caprile and V. Torre, "Using vanishing points for camera calibration," International Journal of Computer Vision, vol. 4, no. 2, pp. 127-139, 1990.

[3] J.M. Coughlan and A.L. Yuille, "Manhattan World: compass direction from a single image by Bayesian inference," Proceedings of the Seventh IEEE International Conference on Computer Vision, vol.2, pp. 941-947, 1999.

[4] V. Angladon, S. Gasparini and V. Charvillat, "The Toulouse vanishing points dataset," Proceedings of the 6th ACM Multimedia Systems Conference (MMSys '15), Portland, United States, 2015.

[5] P. Denis, J.H. Elder and F. Estrada, "Efficient Edge-Based Methods for Estimating Manhattan Frames in Urban Imagery," Proc. European Conference on Computer Vision, vol. 5303, pp. 197-211, 2008.

[6] E. Myasnikov, "Automatic search for vanishing points on mobile devices," CEUR Workshop Proceedings, vol. 2391, pp. 216-221, 2019.

[7] J. Canny, "A Computational Approach To Edge Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, no. 6, pp. 679-698, 1986.

[8] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision," Cambridge, Cambridge University Press, 2004.

[9] Huawei.com, "HONOR 9 Lite," 2020. [Online]. URL: https://consumer.huawei.com/ru/support/phones/honor-9-lite.