Camera Pose Determination and 3-D Measurement From Monocular Oblique Images With Horizontal Right Angle Constraints

Xiaodong Xiong, Yongjun Zhang, Junfeng Zhu, and Maoteng Zheng

***Abstract*—This letter introduces a novel method for camera pose determination from monocular urban oblique images. Horizontal right angles that widely exist in urban scenes are used as geometric constraints in the camera pose determination, and the proposed 3-D measurement method using a monocular image is presented and then used to check the accuracy of the recovered image's exterior orientation parameters. Compared to the available vertical-line-based camera pose determination method, our new method is more accurate.**

***Index Terms*—Horizontal right angle, monocular image, oblique image, pose determination, 3-D measurement.**

I. Introduction

THREE-dimensional object measurement and reconstruction from monocular image is currently a hot topic both in computer vision and photogrammetry. The prerequisite of single-image-based 3-D reconstruction is the camera pose determination, which is usually realized by space resection using points or lines with known coordinates in the object space [1], [2]. However, such coordinate measurements of these features may not always be available, and scene constraints, such as vanishing points, parallel lines, circles, rectangles, etc., thus are used to recover the image's exterior orientation parameters (EOPs). Gonzalez-Aguilera *et al.* proposed effective methods for the intrinsic and extrinsic camera parameter recovery and 3-D reconstruction using vanishing points and other geometric constraints [3]–[6]. Tommaselli and Reiss presented a photogrammetric method for determining the dimensions of flat surfaces using vertical and horizontal lines as well as lasermeter measurements [7]. Wang *et al.* conducted camera calibration and 3-D reconstruction using the vanishing points of three mutually orthogonal directions, mutually orthogonal line segments of equal length or with a known length ratio, circle, and any two pairs of parallel lines in space that may not be coplanar or orthogonal [8], [9]. Under the assumption that, in man-made environments, the majority of the lines align with three principal directions of the world coordinate frame, Kosecka and Zhang proposed a dominant rectangular structure extraction method and used it in the recovery of the camera pose and the matching across widely separated views [10].

This letter deals with camera pose determination and 3-D measurement using a monocular urban oblique image without ground measurements (e.g., ground control points). The aerial oblique images are of large format, and each contains hundreds of buildings. Thus, scene constraints such as horizontal lines, vertical lines, right angles, and vanishing points widely exist in the oblique images. Gerke already presented an effective method to integrate linear horizontal, vertical, and right-angled scene structures into the bundle adjustment of oblique image sequences [11], but this method needs the scene structures to be seen in at least two images, which is not suitable for monocular image orientation. As the buildings are of different orientations and any one building on the image is small, there are no dominant vanishing points or dominant rectangular structures on these images, which make the vanishing-point-based or rectangular-structure-based method unsuitable. Considering that many horizontal right angles can be obtained in the urban scene because most buildings have horizontal roofs and orthogonal corners, horizontal right angles are used as scene constraints to recover the image EOPs (also called "image orientation") in this letter.

The remainder of this letter is organized as follows. The principle of horizontal right angle constraints is first introduced. Then, the horizontal right-angle-based monocular image orientation method is discussed in detail, which is followed by its application for building measurement. Finally, we present the experimental results in Section V and our conclusion in Section VI.

Manuscript received February 14, 2014; revised March 17, 2014; accepted April 1, 2014. This work was supported in part by the National Natural Science Foundation of China under Grants 41171292 and 41322010, by the National Basic Research Program of China (973 Program; No. 2012CB719904), by the Fundamental Research Funds for the Central Universities under Grant 2012213020207, and by the Academic Award for Excellent Ph.D. Candidates Funded by the Ministry of Education of China under Grant 5052012213001.

The authors are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: zhangyj@whu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2014.2315918

II. Principle of Horizontal Right Angle Constraints

A building outline on an oblique image is shown in Fig. 1. Obviously, in object space, the building outline is a rectangle; however, on the oblique image, the projected building outline is no longer a rectangle. We hereafter call the orthogonal building corners on a horizontal rooftop abc, bcd, cda, and dab, as shown in Fig. 1, the horizontal right angles, and the monocular image orientation is based on these horizontal right angles.

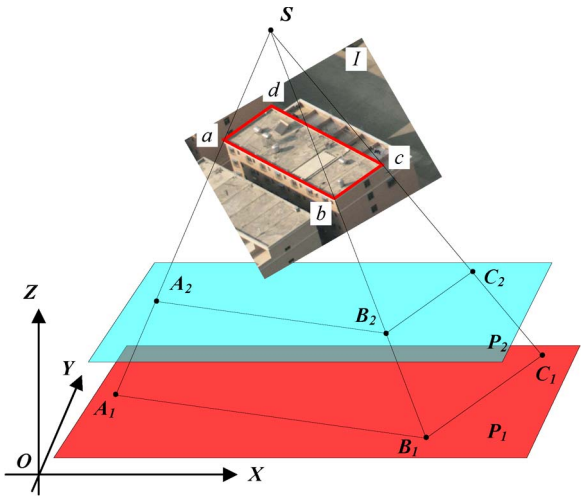

Fig. 1. Character of horizontal right angle projective transformation geometry.

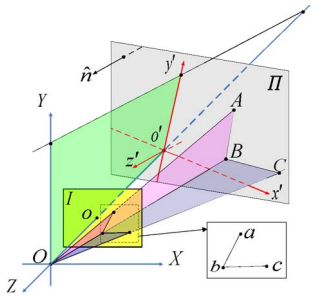

Fig. 2. Definition of frames and the projection of horizontal right angle.

An analysis of the perspective transformation attribute of the horizontal right angle is demonstrated in Fig. 1. Planes P_1 and P_2 represent two horizontal planes with different elevation values in the object space, and plane I represents the oblique aerial image, while S represents the perspective center. It is easy to determine that, if the EOPs of the oblique image are recovered, when projecting the building corner that is on the oblique image (e.g., abc in Fig. 1) to an arbitrary object space horizontal plane (e.g., planes P_1 or P_2 in Fig. 1), the projected corner (e.g., corners $A_1B_1C_1$ or $A_2B_2C_2$ in Fig. 1) will always be a right angle. This is the horizontal right angle constraint adopted in this letter.

III. Monocular Image Orientation

Since no ground measurements are available, the image orientation only can be done in a specifically defined frame, and the calculated EOPs thus are relative values with respect to a specific horizontal plane.

A. Definition of Frames

Fig. 2 shows the frames defined in this letter. $O-XYZ$ is the image space frame, where coordinate origin O is the perspective center of the aerial image, and the OZ axis passes through image principal point o and is perpendicular to image plane I. Plane Π represents a defined horizontal plane in the object space which passes through the point (0, 0, Z_0) in the $O-XYZ$ frame. Object space frame $o'-x'y'z'$ is defined according both to the $O-XYZ$ frame and object space plane Π, where coordinate origin o' is the intersection point of axis OZ and plane Π, axis $o'y'$ is the intersection line of plane $O-YZ$ and plane Π, and axis $o'z'$ is parallel to the normal vector of plane Π.

Based on the aforementioned analysis, the monocular image orientation in this letter can be divided into two steps: first, calculate the equation of plane Π in image space frame $O-XYZ$, and second, build object space frame $o'-x'y'z'$, and calculate the image's EOPs in this frame. The horizontal right angle constraints are used to recover the object plane equation.

In image space frame $O-XYZ$, suppose that the equation of object space plane Π is as follows:

$$n_x(X-0)+n_y(Y-0)+Z-Z_0=0 \tag{1}$$

where $(n_x, n_y, 1)$ is the normal vector of the plane, and the plane passes through the point (0, 0, Z_0). Z_0 can be set as -500 m.

As shown in Fig. 2, suppose that the coordinates of image point b on the image space frame are $(x_b, y_b, -f)$; then, the equation of ray Ob from perspective center O to image point b can be represented as

$$\begin{cases} X-0=x_b^* t \\ Y-0=y_b^* t \\ Z-0=-f^* t \end{cases} \tag{2}$$

where f is the focus length of the image, t is an arbitrary real number, (X, Y, Z) is an arbitrary point which lies on the ray Ob.

Get the value of t by substituting (2) into (1). Then, the coordinates of intersection point B between ray Ob and plane Π can be calculated as follows by substituting the value t into (2):

$$\begin{cases} X_B=x_b Z_0/(n_x x_b+n_y y_b-f) \\ Y_B=y_b Z_0/(n_x x_b+n_y y_b-f) \\ Z_B=-f Z_0/(n_x x_b+n_y y_b-f). \end{cases} \tag{3}$$

B. Monocular Image Orientation With Horizontal Right Angles

Suppose that, as shown in Fig. 2, polyline abc is a building corner on the image. If polyline abc is projected onto plane Π, object space polyline ABC is obtained. Therefore, angle $\angle ABC$ must be a right angle, as follows:

$$\overrightarrow{AB}\cdot\overrightarrow{BC}=0. \tag{4}$$

Suppose that the object space coordinates of points A, B, and C are (X_A, Y_A, Z_A), (X_B, Y_B, Z_B), and (X_C, Y_C, Z_C), respectively. Then, (4) can be rewritten as

$$(X_A-X_B)(X_C-X_B)+(Y_A-Y_B)(Y_C-Y_B)+(Z_A-Z_B)(Z_C-Z_B)=0. \tag{5}$$

Calculate the object space coordinates of points A, B, and C with (3), and substitute them into (5), and after some

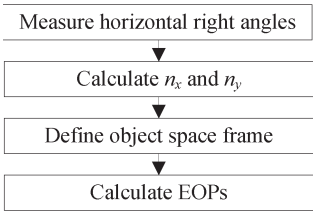

Fig. 3. Workflow of the proposed monocular image orientation method.

mathematical derivation, the horizontal right angle constraint in monocular image orientation is as follows:

$$G(n_x, n_y) = F_1 n_x^2 + F_2 n_y^2 + F_3 n_x n_y + F_4 n_x + F_5 n_y + F_6 = 0 \tag{6}$$

where $F_i (i = 1, \ldots, 6)$ represents the coefficients, and their values can be calculated from the image space coordinates of points a, b, and c.

The error equation of the horizontal right angle constraint-based monocular image orientation can be written as follows by mathematical derivation:

$$V_G = \frac{\partial G}{\partial n_x} dn_x + \frac{\partial G}{\partial n_y} dn_y + G(n_{x0}, n_{y0}) \tag{7}$$

where $\partial G/\partial n_x = 2F_1 n_x + F_3 n_y + F_4$ and $\partial G/\partial n_y = 2F_2 n_y + F_3 n_x + F_5$.

Moreover, the value of $G(n_{x0}, n_{y0})$ is calculated from the initial values of n_x and n_y with (6). The initial values of both n_x and n_y can be set as zero.

As can be seen from (7), only the image coordinates of the building corners in the image space frame are included in the equation, so no ground measurements are needed in the monocular image orientation. In total, there are two unknown values (n_x and n_y) in the error equation, and one horizontal right angle introduces one equation; therefore, theoretically, at least two horizontal right angles are needed. Practically reliable solutions are obtained through least square adjustment using more than two horizontal right angles.

Once the normal vector of plane Π is calculated, object space frame $o' - x'y'z'$ can be defined as shown in Fig. 2. Then, the image's EOPs in this object space frame can be calculated using rigid-body transformation (including three translations and three orientation angles and is similar to the absolute orientation model used in [12]). Specifically, three points lying on the three axis of frame $o' - x'y'z'$ are first defined with given coordinates, respectively. Second, the coordinates of these three points in $O - XYZ$ frame are calculated according to the parameters of plane Π. Finally, the rigid-body transformation parameters between the two frames are calculated using these three points, and the image's EOPs are transformed to object space. The workflow of the proposed method is given in Fig. 3 as a summary.

In this letter, we suppose that the camera is calibrated. However, the proposed method is easy to be broadened to cope with noncalibrated cameras by introducing self-calibration parameters in (6).

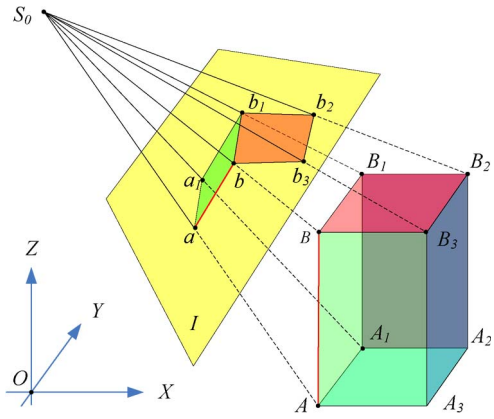

Fig. 4. Demonstration of building reconstruction using monocular oblique image.

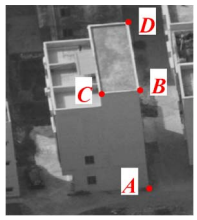

Fig. 5. Points measured on the building for 3-D measurement.

IV. Building Reconstruction Using Monocular Oblique Image

As shown in Fig. 4, plane I represents the aerial image, and point S_0 represents the perspective center. In the object space, a box model $AA_1A_2A_3 - BB_1B_2B_3$ represents a building, and on the image, the points $(a, a_1, b, b_1, b_2, b_3)$ are the projected image points of the building vertexes $(A, A_1, B, B_1, B_2, B_3)$, respectively.

Suppose that the image's EOPs are recovered, specifically that the coordinates of the perspective center are (X_S, Y_S, Z_S) and the rotation matrix is R. The image plane coordinates of points a and b are (x_a, y_a) and (x_b, y_b), respectively. Then, the object space coordinates (X_B, Y_B) of vertex B can be calculated as follows by rearranging the traditional collinearity equation if elevation value Z_B is given (can be set as $Z_S - 500$):

$$\begin{cases} X_B = X_S + (Z_B - Z_S)\frac{a_1 x_b + a_2 y_b - a_3 f}{c_1 x_b + c_2 y_b - c_3 f} \\ Y_B = Y_S + (Z_B - Z_S)\frac{b_1 x_b + b_2 y_b - b_3 f}{c_1 x_b + c_2 y_b - c_3 f} \end{cases} \tag{8}$$

where a_i, b_i, c_i $(i = 1, 2, 3)$ are the elements of the rotation matrix R.

The coordinates of the other points (B_1, B_2, B_3), which have the same elevation values with point B, can also be calculated using (8). Since point A and point B lie on an identical vertical line, they have the same horizontal coordinates. Then, the elevation coordinates of point A can be calculated as follows:

$$\begin{cases} Z_A = Z_S + (X_B - X_S)\frac{c_1 x_a + c_2 y_a - c_3 f}{a_1 x_a + a_2 y_a - a_3 f} & if (X_B \neq X_S) \\ Z_A = Z_S + (Y_B - Y_S)\frac{c_1 x_a + c_2 y_a - c_3 f}{b_1 x_a + b_2 y_a - b_3 f} & if (Y_B \neq Y_S). \end{cases} \tag{9}$$

TABLE I
RECOVERED IMAGE'S EOPS AND 3-D MEASUREMENT RESULTS

Image No.	Methods and Number of Constraints		EO (Xs (m), Ys (m), Zs (m), φ (degree), ω(degree), κ(degree))	Amount of Buildings	Building Reconstruction Errors (m)					
					Mean		Max		RMSE	
					L	H	L	H	L	H
1	**A**	**9**	-0.86 , -331.16, 374.61 0.131350, 41.477521, -0.000000	**5**	**-0.03**	**-0.08**	**-0.43**	**-1.00**	**0.21**	**0.48**
	V	9	1.90, -339.78, 366.82 -0.297228, 42.808033, -0.000000	5	0.12	-0.79	0.37	-1.35	0.29	0.85
2	**A**	**9**	3.21, -329.11, 376.40 -0.487794, 41.164676, -0.000000	**5**	**0.14**	**-0.22**	**0.17**	**-0.74**	**0.14**	**0.38**
	V	9	5.53, -324.24, 380.58 -0.832265, 40.426345, 0.000000	5	0.14	-0.32	0.18	-1.10	0.15	0.72
3	**A**	**9**	6.67, -334.31, 371.74 -1.027592, 41.961108, 0.000000	**5**	**0.00**	**-0.38**	**0.37**	**-0.62**	**0.21**	**0.43**
	V	9	6.44, -338.48, 367.95 -1.003349, 42.607135, 0.000000	5	0.03	-0.65	0.40	-0.97	0.22	0.69
4	**A**	**9**	-6.43, -357.54, 349.46 1.054683, 45.650346, -0.000000	**5**	**0.20**	**-0.04**	**0.85**	**-0.55**	**0.55**	**0.41**
	V	9	3.08, -357.81, 349.23 -0.504566, 45.694615, 0.000000	5	0.26	0.03	0.67	1.59	0.47	1.09
5	**A**	**9**	-3.55, -355.08, 352.01 0.577887, 45.247042, -0.000000	**5**	**0.33**	**0.06**	**0.77**	**0.88**	**0.46**	**0.53**
	V	9	-5.37, -367.04, 339.49 0.905476, 47.230023, 0.000000	5	0.48	-0.26	1.21	-1.34	0.71	0.94
6	**A**	**9**	-6.21, -364.39, 342.32 1.039791, 46.783551, 0.000000	**5**	**0.40**	**0.09**	**0.86**	**-0.93**	**0.55**	**0.61**
	V	9	-3.57, -364.15, 342.62 0.596189, 46.743457, -0.000000	5	0.12	-0.48	0.92	-0.97	0.56	0.66
Total	**A**	-	-	**30**	**0.17**	**-0.10**	**0.86**	**-1.00**	**0.39**	**0.48**
	V	-	-	30	0.19	-0.41	1.21	1.59	0.45	0.84

If $X_B \neq X_S$ and $Y_B \neq Y_S$, Z_A will take the average value given by the aforementioned two equations in (9).

V. EXPERIMENTS

A. Dataset

In order to test the feasibility of the proposed method, oblique images and LiDAR data in the same area were used in our experiments. The LiDAR system was Trimble Harrier 56 with POS/AV 510, the horizontal/vertical accuracy of the LiDAR system was 0.25/0.15 m, and the laser beam divergence angle was 0.5 mrad. The LiDAR flight height was 680 m, and the average point density and point space were 3.4 points/m and 0.55 m, respectively. The oblique camera was SWDC-5, the flight height was 800 m, and the average ground sampling distance was 0.07 m. The data sets were acquired at Yangjiang, Guangdong, China, on July 2013. SWDC-5 is an oblique photography system (composed of five digital cameras) developed by the Chinese Academy of Survey and Mapping. In this letter, the images from the forward view and the backward view were used. Their camera parameters (focus length, pixel size, and format size) are (9.9847E + 001 mm, 0.006 mm, and 8184*6114) and (9.9880E + 001 mm, 0.006 mm, and 8200*5960), respectively. Images used here were already corrected for lens distortion and principal point offset.

B. Camera Pose Determination and 3-D Measurement

Six oblique aerial images were used to test the proposed method, three of which were forward view and the others were backward view. The vertical-line-based monocular image orientation method introduced by Zhang *et al.* [13] was also conducted as a comparison experiment. The experiments were conducted in the following three steps. 1) A certain number of horizontal right angles (or vertical lines) were manually measured on each oblique image. 2) The EOPs of each oblique image were calculated using the proposed method (or vertical-line-based method). 3) Several evenly distributed buildings were measured on each oblique image, their side length values were calculated using the method mentioned previously, and their side length values measured from the LiDAR data were used as reference data to check the accuracy of the 3-D measurement from the monocular image.

For each cubic-shaped building, four vertexes (such as points A, B, C, and D shown in Fig. 5) were measured on the aerial image. Meanwhile, the object space coordinates of these vertexes were manually measured in the LiDAR data. Since the accuracy of the LiDAR points was relatively high, the measured values in the LiDAR data were deemed as ground truth. Considering that the scale of the 3-D measurement could not be determined without ground measurements, we deemed that the width of the building (i.e., the length of BC for the building in Fig. 5) was the same as the width measured in the LiDAR data; thus, the scale was fixed. For the purpose of evaluating the accuracy of the recovered image's EOPs, we compared the building height (the length of BA) and the building length (the length of BD) calculated from single-image-based 3-D measuring with the LiDAR data measurements.

Table I lists the recovered EOPs for six oblique images and the 3-D measurement results for each image. The symbol "A" denotes the horizontal right-angle-based method, and the symbol "V" denotes the vertical-line-based method. For each

image, the "A" and "V" methods are tested, respectively. For each method, nine constraints are used (i.e., nine horizontal right angles for "A" and nine vertical lines for "V"). The symbol "L" denotes the errors of the building length, and the symbol "H" denotes the errors of the building height. When the horizontal right-angle-based method was used, the total root-mean-square errors (rmse) of the building length and the building height were 0.39 and 0.48 m, respectively. Considering that the average LiDAR point distance is 0.55 m, the accuracy of the 3-D measurement was deemed acceptable. When the vertical-line-based method was used, the total rmse values of the building length and the building height were 0.45 and 0.84 m, respectively. Obviously, the horizontal right-angle-based method was superior to the vertical-line-based method. The reason is as follows. 1) Measuring vertical lines may be difficult because they are usually occluded by other buildings. 2) Vanishing points of vertical lines are infinite image points; if the measured vertical lines are not well distributed, slight inaccuracy of line measurements will result in large errors in the positions of calculated vanishing points [14].

During the iteration procedure of "A" method, the convergence condition is the absolute difference values between the previous and current steps for both the two parameters (of n_x and n_y), are smaller than $1.0e-012$ (i.e., $|n_x(i)-n_x(i-1)|<1.0e-012$ and $|n_y(i)-n_y(i-1)|<1.0e-012$, i is the current iteration number $n_x(i)$ is the value of n_x in i_{th} iteration). When performing the "A" method with the six images, the average iteration number is 9. For the calculation of any of the six images, the time consumed for both of the "A" and "V" methods is negligible (less than 1 ms).

VI. Conclusion

In this letter, the principle of horizontal right angle constraints first has been presented, and then, the proposed method of camera pose determination from a monocular oblique image using the horizontal right angle constraints has been described. With the purpose of verifying the feasibility of the camera pose determination method, 3-D measurement experiments were conducted with several urban oblique aerial images, and the vertical-line-based method was also tested for the purpose of comparison. The experimental results showed that our method was more accurate than the vertical-line-based method. However, when only gabled-roof buildings are available, our method will fail. Since the automatic registration of oblique images remains a challenge due to image deformation [15], future work could address adopting horizontal right angle constraints in the oblique image matching procedure. Specifically, this would include automatically detecting the horizontal right angles and recovering the oblique image's EOPs, and then eliminating the image deformation by rectifying the oblique image to a horizontal plane.

Acknowledgment

The authors would like to thank the anonymous reviewers and members of the editorial team for the comments and contributions and M. Wang from the University of South Florida, Tampa, FL, USA, for refining the use of English in this letter.

References

[1] J.-Y. Han, J. Guo, and J.-Y. Chou, "A direct determination of the orientation parameters in the collinearity equations," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 313–316, Mar. 2011.

[2] A. F. Habib, H. T. Lin, and M. F. Morgan, "Line-based modified iterated Hough transform for autonomous single-photo resection," *Photogramm. Eng. Remote Sens.*, vol. 69, no. 12, pp. 1351–1357, Dec. 2003.

[3] D. Gonzalez-Aguilera and J. Gomez-Lahoz, "From 2D TO 3D through modelling based on a single image," *Photogramm. Rec.*, vol. 23, no. 122, pp. 208–227, Jun. 2008.

[4] D. Gonzalez-Aguilera, J. Gomez-Lahoz, and P. Rodriguez-Gonzalvez, "An automatic approach for radial lens distortion correction from a single image," *IEEE Sensors J.*, vol. 11, no. 4, pp. 956–965, Apr. 2011.

[5] D. Gonzalez-Aguilera and J. Gomez-Lahoz, "Forensic terrestrial photogrammetry from a single image," *J. Forensic Sci.*, vol. 54, no. 6, pp. 1376–1387, Nov. 2009.

[6] J. Garcia-Gago, J. Gomez-Lahoz, J. Rodríguez-Méndez, and D. González-Aguilera, "Historical single image-based modeling: The case of Gobierna Tower, Zamora (Spain)," *Remote Sens.*, vol. 6, no. 2, pp. 1085–1101, 2014.

[7] A. Tommaselli and M. Reiss, "A photogrammetric method for single image orientation and measurement," *Photogramm. Eng. Remote Sens.*, vol. 71, no. 6, pp. 727–732, Jun. 2005.

[8] G. Wang, H.-T. Tsui, Z. Hu, and F. Wu, "Camera calibration and 3D reconstruction from a single view based on scene constraints," *Image Vis. Comput.*, vol. 23, no. 3, pp. 311–323, Mar. 2005.

[9] G. Wang, J. Wu, and Z. JI, "Single view based pose estimation from circle or parallel lines," *Pattern Recog. Lett.*, vol. 29, no. 7, pp. 977–985, May 2008.

[10] J. Kosecka and W. Zhang, "Extraction, matching, and pose recovery based on dominant rectangular structures," *Comput. Vis. Image Understanding*, vol. 100, no. 3, pp. 274–293, Dec. 2005.

[11] M. Gerke, "Using horizontal and vertical building structure to constrain indirect sensor orientation," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 3, pp. 307–316, May 2011.

[12] T.-A. Teo and S.-H. Huang, "Automatic co-registration of optical satellite images and airborne LiDAR data using relative and absolute orientations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 5, pp. 2229–2237, Oct. 2013.

[13] J. Zhang, Y. Zhang, and Z. Zhang, "The feasibility study on the aerial triangulation over urban area constrained by vertical lines," in *Proc. 4th SPIE Int. Symp. Multispectral Image Process. Pattern Recog.*, 2005, pp. 473–479.

[14] R. Wang and F. P. Ferrie, "Self-calibration and metric reconstruction from single images," *Int. Archives Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 37, no. B5, pp. 639–644, 2008.

[15] H. Yang, S. Zhang, and Y. Wang, "Robust and precise registration of oblique images based on scale-invariant feature transformation algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 783–787, Jul. 2012.

Camera Pose Determination and 3-D Measurement From Monocular Oblique Images With Horizontal Right Angle Constraints

Xiaodong Xiong, Yongjun Zhang, Junfeng Zhu, and Maoteng Zheng

***Abstract*—This letter introduces a novel method for camera pose determination from monocular urban oblique images. Horizontal right angles that widely exist in urban scenes are used as geometric constraints in the camera pose determination, and the proposed 3-D measurement method using a monocular image is presented and then used to check the accuracy of the recovered image's exterior orientation parameters. Compared to the available vertical-line-based camera pose determination method, our new method is more accurate.**

***Index Terms*—Horizontal right angle, monocular image, oblique image, pose determination, 3-D measurement.**

I. Introduction

THREE-dimensional object measurement and reconstruction from monocular image is currently a hot topic both in computer vision and photogrammetry. The prerequisite of single-image-based 3-D reconstruction is the camera pose determination, which is usually realized by space resection using points or lines with known coordinates in the object space [1], [2]. However, such coordinate measurements of these features may not always be available, and scene constraints, such as vanishing points, parallel lines, circles, rectangles, etc., thus are used to recover the image's exterior orientation parameters (EOPs). Gonzalez-Aguilera *et al.* proposed effective methods for the intrinsic and extrinsic camera parameter recovery and 3-D reconstruction using vanishing points and other geometric constraints [3]–[6]. Tommaselli and Reiss presented a photogrammetric method for determining the dimensions of flat surfaces using vertical and horizontal lines as well as lasermeter measurements [7]. Wang *et al.* conducted camera calibration and 3-D reconstruction using the vanishing points of three mutually orthogonal directions, mutually orthogonal line segments of equal length or with a known length ratio, circle, and any two pairs of parallel lines in space that may not be coplanar or orthogonal [8], [9]. Under the assumption that, in man-made environments, the majority of the lines align with three principal directions of the world coordinate frame, Kosecka and Zhang proposed a dominant rectangular structure extraction method and used it in the recovery of the camera pose and the matching across widely separated views [10].

This letter deals with camera pose determination and 3-D measurement using a monocular urban oblique image without ground measurements (e.g., ground control points). The aerial oblique images are of large format, and each contains hundreds of buildings. Thus, scene constraints such as horizontal lines, vertical lines, right angles, and vanishing points widely exist in the oblique images. Gerke already presented an effective method to integrate linear horizontal, vertical, and right-angled scene structures into the bundle adjustment of oblique image sequences [11], but this method needs the scene structures to be seen in at least two images, which is not suitable for monocular image orientation. As the buildings are of different orientations and any one building on the image is small, there are no dominant vanishing points or dominant rectangular structures on these images, which make the vanishing-point-based or rectangular-structure-based method unsuitable. Considering that many horizontal right angles can be obtained in the urban scene because most buildings have horizontal roofs and orthogonal corners, horizontal right angles are used as scene constraints to recover the image EOPs (also called "image orientation") in this letter.

The remainder of this letter is organized as follows. The principle of horizontal right angle constraints is first introduced. Then, the horizontal right-angle-based monocular image orientation method is discussed in detail, which is followed by its application for building measurement. Finally, we present the experimental results in Section V and our conclusion in Section VI.

Manuscript received February 14, 2014; revised March 17, 2014; accepted April 1, 2014. This work was supported in part by the National Natural Science Foundation of China under Grants 41171292 and 41322010, by the National Basic Research Program of China (973 Program; No. 2012CB719904), by the Fundamental Research Funds for the Central Universities under Grant 2012213020207, and by the Academic Award for Excellent Ph.D. Candidates Funded by the Ministry of Education of China under Grant 5052012213001.

The authors are with the School of Remote Sensing and Information Engineering, Wuhan University, Wuhan 430079, China (e-mail: zhangyj@whu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2014.2315918

II. Principle of Horizontal Right Angle Constraints

A building outline on an oblique image is shown in Fig. 1. Obviously, in object space, the building outline is a rectangle; however, on the oblique image, the projected building outline is no longer a rectangle. We hereafter call the orthogonal building corners on a horizontal rooftop abc, bcd, cda, and dab, as shown in Fig. 1, the horizontal right angles, and the monocular image orientation is based on these horizontal right angles.

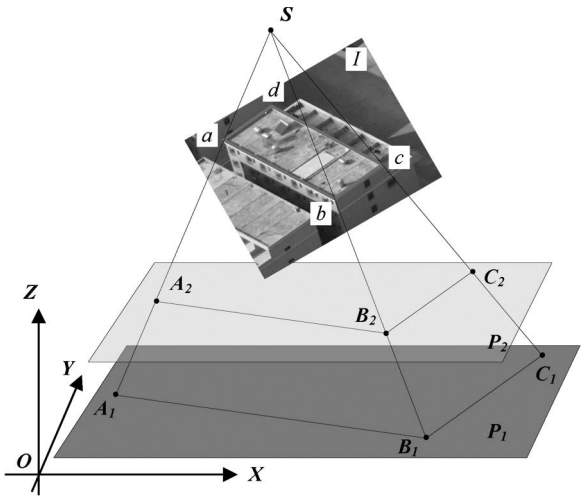

Fig. 1. Character of horizontal right angle projective transformation geometry.

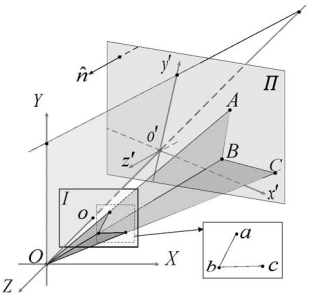

Fig. 2. Definition of frames and the projection of horizontal right angle.

An analysis of the perspective transformation attribute of the horizontal right angle is demonstrated in Fig. 1. Planes P_1 and P_2 represent two horizontal planes with different elevation values in the object space, and plane I represents the oblique aerial image, while S represents the perspective center. It is easy to determine that, if the EOPs of the oblique image are recovered, when projecting the building corner that is on the oblique image (e.g., abc in Fig. 1) to an arbitrary object space horizontal plane (e.g., planes P_1 or P_2 in Fig. 1), the projected corner (e.g., corners $A_1B_1C_1$ or $A_2B_2C_2$ in Fig. 1) will always be a right angle. This is the horizontal right angle constraint adopted in this letter.

III. Monocular Image Orientation

Since no ground measurements are available, the image orientation only can be done in a specifically defined frame, and the calculated EOPs thus are relative values with respect to a specific horizontal plane.

A. Definition of Frames

Fig. 2 shows the frames defined in this letter. $O-XYZ$ is the image space frame, where coordinate origin O is the perspective center of the aerial image, and the OZ axis passes through image principal point o and is perpendicular to image plane I. Plane Π represents a defined horizontal plane in the object space which passes through the point (0, 0, Z_0) in the $O-XYZ$ frame. Object space frame $o'-x'y'z'$ is defined according both to the $O-XYZ$ frame and object space plane Π, where coordinate origin o' is the intersection point of axis OZ and plane Π, axis $o'y'$ is the intersection line of plane $O-YZ$ and plane Π, and axis $o'z'$ is parallel to the normal vector of plane Π.

Based on the aforementioned analysis, the monocular image orientation in this letter can be divided into two steps: first, calculate the equation of plane Π in image space frame $O-XYZ$, and second, build object space frame $o'-x'y'z'$, and calculate the image's EOPs in this frame. The horizontal right angle constraints are used to recover the object plane equation.

In image space frame $O-XYZ$, suppose that the equation of object space plane Π is as follows:

$$n_x(X-0)+n_y(Y-0)+Z-Z_0=0 \tag{1}$$

where $(n_x, n_y, 1)$ is the normal vector of the plane, and the plane passes through the point (0, 0, Z_0). Z_0 can be set as -500 m.

As shown in Fig. 2, suppose that the coordinates of image point b on the image space frame are $(x_b, y_b, -f)$; then, the equation of ray Ob from perspective center O to image point b can be represented as

$$\begin{cases} X-0=x_b^* t \\ Y-0=y_b^* t \\ Z-0=-f^* t \end{cases} \tag{2}$$

where f is the focus length of the image, t is an arbitrary real number, (X, Y, Z) is an arbitrary point which lies on the ray Ob.

Get the value of t by substituting (2) into (1). Then, the coordinates of intersection point B between ray Ob and plane Π can be calculated as follows by substituting the value t into (2):

$$\begin{cases} X_B=x_b Z_0/(n_x x_b+n_y y_b-f) \\ Y_B=y_b Z_0/(n_x x_b+n_y y_b-f) \\ Z_B=-f Z_0/(n_x x_b+n_y y_b-f). \end{cases} \tag{3}$$

B. Monocular Image Orientation With Horizontal Right Angles

Suppose that, as shown in Fig. 2, polyline abc is a building corner on the image. If polyline abc is projected onto plane Π, object space polyline ABC is obtained. Therefore, angle $\angle ABC$ must be a right angle, as follows:

$$\overrightarrow{AB}\cdot\overrightarrow{BC}=0. \tag{4}$$

Suppose that the object space coordinates of points A, B, and C are (X_A, Y_A, Z_A), (X_B, Y_B, Z_B), and (X_C, Y_C, Z_C), respectively. Then, (4) can be rewritten as

$$(X_A-X_B)(X_C-X_B)+(Y_A-Y_B)(Y_C-Y_B)+(Z_A-Z_B)(Z_C-Z_B)=0. \tag{5}$$

Calculate the object space coordinates of points A, B, and C with (3), and substitute them into (5), and after some

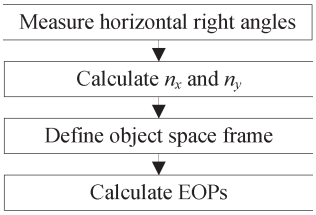

Fig. 3. Workflow of the proposed monocular image orientation method.

mathematical derivation, the horizontal right angle constraint in monocular image orientation is as follows:

$$G(n_x, n_y) = F_1 n_x^2 + F_2 n_y^2 + F_3 n_x n_y + F_4 n_x + F_5 n_y + F_6 = 0 \tag{6}$$

where $F_i (i = 1, \ldots, 6)$ represents the coefficients, and their values can be calculated from the image space coordinates of points a, b, and c.

The error equation of the horizontal right angle constraint-based monocular image orientation can be written as follows by mathematical derivation:

$$V_G = \frac{\partial G}{\partial n_x} dn_x + \frac{\partial G}{\partial n_y} dn_y + G(n_{x0}, n_{y0}) \tag{7}$$

where $\partial G/\partial n_x = 2F_1 n_x + F_3 n_y + F_4$ and $\partial G/\partial n_y = 2F_2 n_y + F_3 n_x + F_5$.

Moreover, the value of $G(n_{x0}, n_{y0})$ is calculated from the initial values of n_x and n_y with (6). The initial values of both n_x and n_y can be set as zero.

As can be seen from (7), only the image coordinates of the building corners in the image space frame are included in the equation, so no ground measurements are needed in the monocular image orientation. In total, there are two unknown values (n_x and n_y) in the error equation, and one horizontal right angle introduces one equation; therefore, theoretically, at least two horizontal right angles are needed. Practically reliable solutions are obtained through least square adjustment using more than two horizontal right angles.

Once the normal vector of plane Π is calculated, object space frame $o' - x'y'z'$ can be defined as shown in Fig. 2. Then, the image's EOPs in this object space frame can be calculated using rigid-body transformation (including three translations and three orientation angles and is similar to the absolute orientation model used in [12]). Specifically, three points lying on the three axis of frame $o' - x'y'z'$ are first defined with given coordinates, respectively. Second, the coordinates of these three points in $O - XYZ$ frame are calculated according to the parameters of plane Π. Finally, the rigid-body transformation parameters between the two frames are calculated using these three points, and the image's EOPs are transformed to object space. The workflow of the proposed method is given in Fig. 3 as a summary.

In this letter, we suppose that the camera is calibrated. However, the proposed method is easy to be broadened to cope with noncalibrated cameras by introducing self-calibration parameters in (6).

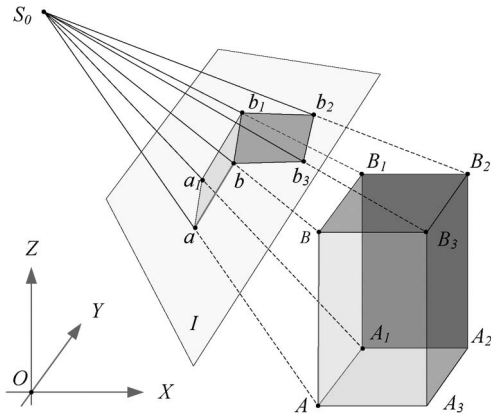

Fig. 4. Demonstration of building reconstruction using monocular oblique image.

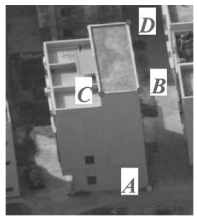

Fig. 5. Points measured on the building for 3-D measurement.

IV. Building Reconstruction Using Monocular Oblique Image

As shown in Fig. 4, plane I represents the aerial image, and point S_0 represents the perspective center. In the object space, a box model $AA_1A_2A_3 - BB_1B_2B_3$ represents a building, and on the image, the points $(a, a_1, b, b_1, b_2, b_3)$ are the projected image points of the building vertexes $(A, A_1, B, B_1, B_2, B_3)$, respectively.

Suppose that the image's EOPs are recovered, specifically that the coordinates of the perspective center are (X_S, Y_S, Z_S) and the rotation matrix is R. The image plane coordinates of points a and b are (x_a, y_a) and (x_b, y_b), respectively. Then, the object space coordinates (X_B, Y_B) of vertex B can be calculated as follows by rearranging the traditional collinearity equation if elevation value Z_B is given (can be set as $Z_S - 500$):

$$\begin{cases} X_B = X_S + (Z_B - Z_S)\frac{a_1 x_b + a_2 y_b - a_3 f}{c_1 x_b + c_2 y_b - c_3 f} \\ Y_B = Y_S + (Z_B - Z_S)\frac{b_1 x_b + b_2 y_b - b_3 f}{c_1 x_b + c_2 y_b - c_3 f} \end{cases} \tag{8}$$

where a_i, b_i, c_i $(i = 1, 2, 3)$ are the elements of the rotation matrix R.

The coordinates of the other points (B_1, B_2, B_3), which have the same elevation values with point B, can also be calculated using (8). Since point A and point B lie on an identical vertical line, they have the same horizontal coordinates. Then, the elevation coordinates of point A can be calculated as follows:

$$\begin{cases} Z_A = Z_S + (X_B - X_S)\frac{c_1 x_a + c_2 y_a - c_3 f}{a_1 x_a + a_2 y_a - a_3 f} & if (X_B \neq X_S) \\ Z_A = Z_S + (Y_B - Y_S)\frac{c_1 x_a + c_2 y_a - c_3 f}{b_1 x_a + b_2 y_a - b_3 f} & if (Y_B \neq Y_S). \end{cases} \tag{9}$$

TABLE I
RECOVERED IMAGE'S EOPS AND 3-D MEASUREMENT RESULTS

Image No.	Methods and Number of Constraints		EO (Xs (m), Ys (m), Zs (m), φ (degree), ω(degree), κ(degree))	Amount of Buildings	Building Reconstruction Errors (m)					
					Mean		Max		RMSE	
					L	H	L	H	L	H
1	**A**	**9**	-0.86 , -331.16, 374.61 0.131350, 41.477521, -0.000000	**5**	**-0.03**	**-0.08**	**-0.43**	**-1.00**	**0.21**	**0.48**
	V	9	1.90, -339.78, 366.82 -0.297228, 42.808033, -0.000000	5	0.12	-0.79	0.37	-1.35	0.29	0.85
2	**A**	**9**	3.21, -329.11, 376.40 -0.487794, 41.164676, -0.000000	**5**	**0.14**	**-0.22**	**0.17**	**-0.74**	**0.14**	**0.38**
	V	9	5.53, -324.24, 380.58 -0.832265, 40.426345, 0.000000	5	0.14	-0.32	0.18	-1.10	0.15	0.72
3	**A**	**9**	6.67, -334.31, 371.74 -1.027592, 41.961108, 0.000000	**5**	**0.00**	**-0.38**	**0.37**	**-0.62**	**0.21**	**0.43**
	V	9	6.44, -338.48, 367.95 -1.003349, 42.607135, 0.000000	5	0.03	-0.65	0.40	-0.97	0.22	0.69
4	**A**	**9**	-6.43, -357.54, 349.46 1.054683, 45.650346, -0.000000	**5**	**0.20**	**-0.04**	**0.85**	**-0.55**	**0.55**	**0.41**
	V	9	3.08, -357.81, 349.23 -0.504566, 45.694615, 0.000000	5	0.26	0.03	0.67	1.59	0.47	1.09
5	**A**	**9**	-3.55, -355.08, 352.01 0.577887, 45.247042, -0.000000	**5**	**0.33**	**0.06**	**0.77**	**0.88**	**0.46**	**0.53**
	V	9	-5.37, -367.04, 339.49 0.905476, 47.230023, 0.000000	5	0.48	-0.26	1.21	-1.34	0.71	0.94
6	**A**	**9**	-6.21, -364.39, 342.32 1.039791, 46.783551, 0.000000	**5**	**0.40**	**0.09**	**0.86**	**-0.93**	**0.55**	**0.61**
	V	9	-3.57, -364.15, 342.62 0.596189, 46.743457, -0.000000	5	0.12	-0.48	0.92	-0.97	0.56	0.66
Total	**A**	-	-	**30**	**0.17**	**-0.10**	**0.86**	**-1.00**	**0.39**	**0.48**
	V	-	-	30	0.19	-0.41	1.21	1.59	0.45	0.84

If $X_B \neq X_S$ and $Y_B \neq Y_S$, Z_A will take the average value given by the aforementioned two equations in (9).

V. EXPERIMENTS

A. Dataset

In order to test the feasibility of the proposed method, oblique images and LiDAR data in the same area were used in our experiments. The LiDAR system was Trimble Harrier 56 with POS/AV 510, the horizontal/vertical accuracy of the LiDAR system was 0.25/0.15 m, and the laser beam divergence angle was 0.5 mrad. The LiDAR flight height was 680 m, and the average point density and point space were 3.4 points/m and 0.55 m, respectively. The oblique camera was SWDC-5, the flight height was 800 m, and the average ground sampling distance was 0.07 m. The data sets were acquired at Yangjiang, Guangdong, China, on July 2013. SWDC-5 is an oblique photography system (composed of five digital cameras) developed by the Chinese Academy of Survey and Mapping. In this letter, the images from the forward view and the backward view were used. Their camera parameters (focus length, pixel size, and format size) are (9.9847E + 001 mm, 0.006 mm, and 8184*6114) and (9.9880E + 001 mm, 0.006 mm, and 8200*5960), respectively. Images used here were already corrected for lens distortion and principal point offset.

B. Camera Pose Determination and 3-D Measurement

Six oblique aerial images were used to test the proposed method, three of which were forward view and the others were backward view. The vertical-line-based monocular image orientation method introduced by Zhang *et al.* [13] was also conducted as a comparison experiment. The experiments were conducted in the following three steps. 1) A certain number of horizontal right angles (or vertical lines) were manually measured on each oblique image. 2) The EOPs of each oblique image were calculated using the proposed method (or vertical-line-based method). 3) Several evenly distributed buildings were measured on each oblique image, their side length values were calculated using the method mentioned previously, and their side length values measured from the LiDAR data were used as reference data to check the accuracy of the 3-D measurement from the monocular image.

For each cubic-shaped building, four vertexes (such as points A, B, C, and D shown in Fig. 5) were measured on the aerial image. Meanwhile, the object space coordinates of these vertexes were manually measured in the LiDAR data. Since the accuracy of the LiDAR points was relatively high, the measured values in the LiDAR data were deemed as ground truth. Considering that the scale of the 3-D measurement could not be determined without ground measurements, we deemed that the width of the building (i.e., the length of BC for the building in Fig. 5) was the same as the width measured in the LiDAR data; thus, the scale was fixed. For the purpose of evaluating the accuracy of the recovered image's EOPs, we compared the building height (the length of BA) and the building length (the length of BD) calculated from single-image-based 3-D measuring with the LiDAR data measurements.

Table I lists the recovered EOPs for six oblique images and the 3-D measurement results for each image. The symbol "A" denotes the horizontal right-angle-based method, and the symbol "V" denotes the vertical-line-based method. For each

image, the "A" and "V" methods are tested, respectively. For each method, nine constraints are used (i.e., nine horizontal right angles for "A" and nine vertical lines for "V"). The symbol "L" denotes the errors of the building length, and the symbol "H" denotes the errors of the building height. When the horizontal right-angle-based method was used, the total root-mean-square errors (rmse) of the building length and the building height were 0.39 and 0.48 m, respectively. Considering that the average LiDAR point distance is 0.55 m, the accuracy of the 3-D measurement was deemed acceptable. When the vertical-line-based method was used, the total rmse values of the building length and the building height were 0.45 and 0.84 m, respectively. Obviously, the horizontal right-angle-based method was superior to the vertical-line-based method. The reason is as follows. 1) Measuring vertical lines may be difficult because they are usually occluded by other buildings. 2) Vanishing points of vertical lines are infinite image points; if the measured vertical lines are not well distributed, slight inaccuracy of line measurements will result in large errors in the positions of calculated vanishing points [14].

During the iteration procedure of "A" method, the convergence condition is the absolute difference values between the previous and current steps for both the two parameters (of n_x and n_y), are smaller than $1.0e-012$ (i.e., $|n_x(i)-n_x(i-1)|<1.0e-012$ and $|n_y(i)-n_y(i-1)|<1.0e-012$, i is the current iteration number $n_x(i)$ is the value of n_x in i_{th} iteration). When performing the "A" method with the six images, the average iteration number is 9. For the calculation of any of the six images, the time consumed for both of the "A" and "V" methods is negligible (less than 1 ms).

VI. Conclusion

In this letter, the principle of horizontal right angle constraints first has been presented, and then, the proposed method of camera pose determination from a monocular oblique image using the horizontal right angle constraints has been described. With the purpose of verifying the feasibility of the camera pose determination method, 3-D measurement experiments were conducted with several urban oblique aerial images, and the vertical-line-based method was also tested for the purpose of comparison. The experimental results showed that our method was more accurate than the vertical-line-based method. However, when only gabled-roof buildings are available, our method will fail. Since the automatic registration of oblique images remains a challenge due to image deformation [15], future work could address adopting horizontal right angle constraints in the oblique image matching procedure. Specifically, this would include automatically detecting the horizontal right angles and recovering the oblique image's EOPs, and then eliminating the image deformation by rectifying the oblique image to a horizontal plane.

Acknowledgment

The authors would like to thank the anonymous reviewers and members of the editorial team for the comments and contributions and M. Wang from the University of South Florida, Tampa, FL, USA, for refining the use of English in this letter.

References

[1] J.-Y. Han, J. Guo, and J.-Y. Chou, "A direct determination of the orientation parameters in the collinearity equations," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 313–316, Mar. 2011.

[2] A. F. Habib, H. T. Lin, and M. F. Morgan, "Line-based modified iterated Hough transform for autonomous single-photo resection," *Photogramm. Eng. Remote Sens.*, vol. 69, no. 12, pp. 1351–1357, Dec. 2003.

[3] D. Gonzalez-Aguilera and J. Gomez-Lahoz, "From 2D TO 3D through modelling based on a single image," *Photogramm. Rec.*, vol. 23, no. 122, pp. 208–227, Jun. 2008.

[4] D. Gonzalez-Aguilera, J. Gomez-Lahoz, and P. Rodriguez-Gonzalvez, "An automatic approach for radial lens distortion correction from a single image," *IEEE Sensors J.*, vol. 11, no. 4, pp. 956–965, Apr. 2011.

[5] D. Gonzalez-Aguilera and J. Gomez-Lahoz, "Forensic terrestrial photogrammetry from a single image," *J. Forensic Sci.*, vol. 54, no. 6, pp. 1376–1387, Nov. 2009.

[6] J. Garcia-Gago, J. Gomez-Lahoz, J. Rodríguez-Méndez, and D. González-Aguilera, "Historical single image-based modeling: The case of Gobierna Tower, Zamora (Spain)," *Remote Sens.*, vol. 6, no. 2, pp. 1085–1101, 2014.

[7] A. Tommaselli and M. Reiss, "A photogrammetric method for single image orientation and measurement," *Photogramm. Eng. Remote Sens.*, vol. 71, no. 6, pp. 727–732, Jun. 2005.

[8] G. Wang, H.-T. Tsui, Z. Hu, and F. Wu, "Camera calibration and 3D reconstruction from a single view based on scene constraints," *Image Vis. Comput.*, vol. 23, no. 3, pp. 311–323, Mar. 2005.

[9] G. Wang, J. Wu, and Z. JI, "Single view based pose estimation from circle or parallel lines," *Pattern Recog. Lett.*, vol. 29, no. 7, pp. 977–985, May 2008.

[10] J. Kosecka and W. Zhang, "Extraction, matching, and pose recovery based on dominant rectangular structures," *Comput. Vis. Image Understanding*, vol. 100, no. 3, pp. 274–293, Dec. 2005.

[11] M. Gerke, "Using horizontal and vertical building structure to constrain indirect sensor orientation," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 3, pp. 307–316, May 2011.

[12] T.-A. Teo and S.-H. Huang, "Automatic co-registration of optical satellite images and airborne LiDAR data using relative and absolute orientations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 5, pp. 2229–2237, Oct. 2013.

[13] J. Zhang, Y. Zhang, and Z. Zhang, "The feasibility study on the aerial triangulation over urban area constrained by vertical lines," in *Proc. 4th SPIE Int. Symp. Multispectral Image Process. Pattern Recog.*, 2005, pp. 473–479.

[14] R. Wang and F. P. Ferrie, "Self-calibration and metric reconstruction from single images," *Int. Archives Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. 37, no. B5, pp. 639–644, 2008.

[15] H. Yang, S. Zhang, and Y. Wang, "Robust and precise registration of oblique images based on scale-invariant feature transformation algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 4, pp. 783–787, Jul. 2012.