Automatic Georeferencing of Images Acquired by UAV′s

Elder M. Hemerly

Systems and Control Department, Technological Institute of Aeronautics, São José dos Campos-São Paulo, Brazil

Abstract: This paper implements and evaluates experimentally a procedure for automatically georeferencing images acquired by unmanned aerial vehicles (UAV′s) in the sense that ground control points (GCP) are not necessary. Since the camera model is necessary for georeferencing, this paper also proposes a completely automatic procedure for collecting corner pixels in the model plane image to solve the camera calibration problem, i.e., to estimate the camera and the lens distortion parameters. The performance of the complete georeferencing system is evaluated with real flight data obtained by a typical UAV.

Keywords: Automatic georeferencing, unmanned aerial vehicle (UAV), camera calibration, image processing, lens distortion.

1   Introduction

The determination of ground control points (GCP) is in general a costly and slow process, hence a considerable bottleneck in georeferencing. For instance, in agricultural applications, it would demand surveying of points in remote crop areas. In some other applications, such as in rescuing missions or operation in hostile environments, the determination of GCP can be impossible.

By automatic georeferencing in this paper is meant the capability of mapping image points to the reference frame, also called the mapping frame or the navigation frame, without the use of GCP′s. However, additional data is required: the attitude and position of the camera in the navigation frame. This requirement ties naturally with the increasing demand for georeferencing of images obtained by UAV′s, since here the required data is available anyway, for use by the autopilot (attitude) and the guidance system (position). This demand comes from several areas, including agriculture, security and rescuing applications, to name a few, where the use of unmanned aerial vehicle (UAV′s) has been shown to be of utmost importance.

A critical step in image georeferencing concerns camera calibration, since the camera model is needed in the process, for mapping image pixel to position in the reference frame. The literature on camera calibration is large, but main references can be singled out, namely Tsai[1], Zhang[2] and Bouguet[3]. Different models can be used, but the basic procedure is to estimate the camera parameters by some optimization algorithm, whose performance index is the error between the position of pixels in different images obtained from a standard pattern, called model plane, where the pixel positions are known beforehand, since it is constructed by the user.

Typically in the literature concerning camera calibration, including the aforementioned ones, is the necessity of interaction with the user, usually via mouse, for marking regions in the images from which pixels should be extracted. This hampers the implementation of a truly automated camera calibration procedure.

Several approaches for UAV′s based georeferencing have been proposed recently. In [4], global positioning system

(GPS) information was used to provide the coordinates of the aerial photo centre, but variations in pitch and roll were not catered for, thereby restricting the UAV attitude. An alternative approach was presented by Zhou et al.[5] in the sense that GPS is not used directly. Instead, at least four well-distributed GCP are required. This hypothesis can restrict the applications to areas where good features in the images are available. In [6], GPS information was used, but sequential triangulation was also needed for updating the camera parameters. A robust image matching procedure had then to be applied in real time for finding tie points. Very recently, Xiang and Tian[7] proposed a method for georeferencing, but with two limitations: the orientation and position of the camera with respect to the UAV was not generic, and the camera calibration procedure requires interaction with the user, i.e., it is not performed automatically.

This paper implements a system for performing automatic georeferencing of images obtained by UAV′s. It differs from the results available in the literature in two accounts: 1) the camera position and orientation with respect to the UAV are not restricted, hence no simplification is possible in the pixel mapping, and 2) the pixel positions necessary for camera calibration are obtained entirely by image processing, i.e., in an automatic way.

The main contributions of this paper are then the following: 1) proposition and implementation of an automatic procedure for obtaining pixel positions in the model plane image, for enabling camera calibration; 2) proposition and implementation of an automatic technique for georeferencing of images acquired by UAV′s, which encapsulates all the required changes in the coordinate frames, and 3) performance evaluation with flight data, obtained by a typical UAV.

2   Automatic corner localization for camera calibration

The camera model is defined by (1)–(4), where K is the camera intrinsic parameter matrix and k_i, $i = 1, \cdots, 5$ are the lens distortion parameters. For details about these parameters, see [2]. Camera calibration boils down to the estimation of these parameters, thereby enabling the mapping of image pixel position (x_i, y_i), in (2), to the navigation

frame position (x_n, y_n) in (1).

$$s \begin{bmatrix} \hat{x}_c \\ \hat{y}_c \\ 1 \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \\ 1 \end{bmatrix} \qquad (1)$$

$$s \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = K \begin{bmatrix} \left(1 + k_1 r^2 + k_2 r^4 + k_5 r^6\right) \hat{x}_c + d_1 \\ \left(1 + k_1 r^2 + k_2 r^4 + k_5 r^6\right) \hat{y}_c + d_2 \\ 1 \end{bmatrix} \qquad (2)$$

where $r = \sqrt{\hat{x}_c^2 + \hat{y}_c^2}$

$$\begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = \begin{bmatrix} 2k_3 \dot{x}_c \dot{y}_c + k_4(r^2 + 2\hat{x}_c) \\ k_3(r^2 + 2\hat{y}_c) + 2k_4 \hat{x}_c \hat{y}_c \end{bmatrix} \qquad (3)$$

$$K = \begin{bmatrix} k_x f & \gamma & x_{i0} \\ 0 & k_y f & y_{i0} \\ 0 & 0 & 1 \end{bmatrix}. \qquad (4)$$

Vectors r_1, r_2 and r_3 in (1) depend on the camera attitude and its translation with respect to the model plane, and these parameters are called extrinsic camera parameters. They must also be estimated in the calibration process. This is a nonlinear estimation problem, and the Levenberg-Marquardt method can be employed to solve it. Besides the 10 intrinsic camera parameters, for each image it is necessary to estimate 6 extrinsic parameters. Therefore, a total of $(10 + 6n)$ parameters must be estimated, where n is the number of images, whose minimum value is 3. Hence, at least 28 parameters must be estimated during camera calibration. The cost to be minimized by the Levenberg-Marquardt algorithm is

$$J(\theta) = \sum_{j=1}^{n} \sum_{k=1}^{m} \left\| \begin{bmatrix} x_{i\_jk} \\ y_{i\_jk} \end{bmatrix} - \begin{bmatrix} \hat{x}_{i\_jk}(\theta) \\ \hat{y}_{i\_jk}(\theta) \end{bmatrix} \right\|^2 \qquad (5)$$

where n is the number of images, m is the number of pixel positions extracted from each image, (x_{i_jk}, y_{i_jk}) is the pixel position extracted from the images, $(\hat{x}_{i\_jk}(\theta), \hat{y}_{i\_jk}(\theta))$ is the estimated pixel position, by using the present parameter estimates, and $\theta \in \mathbf{R}^{10+6n}$ is the unknown parameter vector whose estimation solves the camera calibration problem. It is well known in the literature see for instance Ravindra[8] that the Levenberg-Marquardt algorithm constitutes an important output error method for identification, i.e., it is effective for solving the parameter estimation problem when there is only measurement noise, such as the case here: noise comes from the imaging system and the pixel localization procedure.

A given pattern is necessary for camera calibration, and the one used here is a 7×7 arrangement of black squares, with size equal to $l = 15.7$ mm. The spacing between adjacent squares is also equal to l. The upper left square in this arrangement is hollow and is used only for providing the pattern orientation. Since the black squares corners are the pixels to be used for calibration, a total $48 \times 4 = 192$ pixel positions are required for each image. Bearing in mind that at least 3 images are necessary for calibration, there are a total of at least $3 \times 192 = 576$ pixel positions to be determined automatically.

It should be stressed that the camera calibration problem can be simplified by exploiting the structure of a particular

application. For instance, in [9] an application in intelligent transportation system (ITS) was considered, where the automatic and dynamic camera calibration was based on dark channel prior in homogenous fog weather condition.

Most procedures for camera calibration require the marking, by the user, of some regions within the image, typically using the mouse, so that the reference points can be obtained. Here an automatic procedure is proposed for getting these points, which here are the corners of black squares. The procedure for finding the corner positions is illustrated in Fig. 1 and is based on three complementary steps, till a sharp position estimate is obtained. The first step a bounding box procedure, which obtains the blunt estimate, see Fig. 1. The second one is based on Harris corner detector, which does not have enough accuracy for the camera calibration, and produces the refined estimate in Fig. 1. The third and last step, which produces the sharp estimate, is based on a gradient search for the corner, having the refined estimate as starting point. Hence, the convergence is fast.
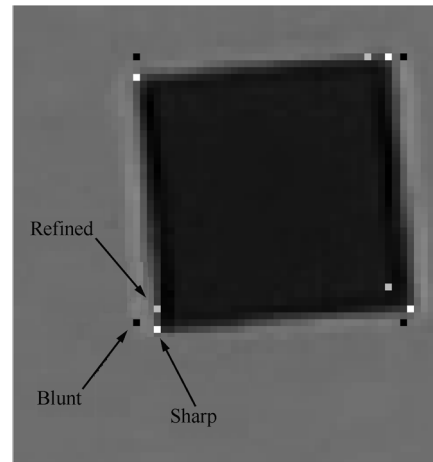

Fig. 1   Example of automatic corner localization procedure, based on 3 steps estimation

As can be seen in Fig. 1, the sharp position estimate is closer to the actual corners than the other two estimates. The sharp corner positions (x_{i_jk}, y_{i_jk}) are then used in the cost function defined by (5).

2.1   Example of camera calibration

Three images were obtained from the model plane, as shown in Fig. 2, by moving the camera, the Sony DSC-H70 still camera, with respect to the model plane. The exact camera orientation and position are immaterial, since this information is connected to the camera extrinsic parameters, which are also estimated in the process as mentioned before.

Fig. 2   Three images for the camera calibration example

Given the images in Fig. 2, the corner points were automatically located and fed to the Levenberg-Marquardt algorithm for estimating the camera parameters. The code

was written in Ansi C, for portability, with interface for Matlab to simplify the graphics handling. The estimated camera intrinsic parameter matrix was

$$\hat{K} = \begin{bmatrix} 743.7937 & -0.7729 & 566.3670 \\ 0 & 748.9947 & 345.5019 \\ 0 & 0 & 1 \end{bmatrix} \quad (6)$$

and the estimated lens distortion parameters were

$$\begin{bmatrix} \hat{k}_1 & \hat{k}_2 & \hat{k}_3 & \hat{k}_4 & \hat{k}_5 \end{bmatrix} =$$
$$[-0.01990.16290.00760.0039 - 0.3018]. \quad (7)$$

In order to qualitatively verify if the estimates given by (6) and (7) are adequate, the corner points in the reference image, whose positions are known by construction, can be mapped and superimposed onto the images in Fig. 2. By considering the third image and by marking the mapped points with white rectangles centered in the mapped pixel, the obtained result is shown in Fig. 3. It indicates qualitatively that the calibration was well performed, since the mapped pixels are located close to the rectangle corners.
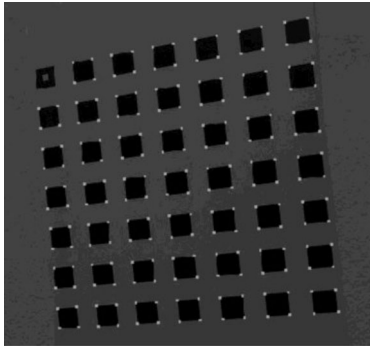

Fig. 3   Mapping of the points from the reference image superimposed onto the third image in Fig. 2, as small white rectangles

A quantitative measure for the camera calibration performance can be obtained from the reconstruction error, i.e., the difference between the corner positions in the images and the corresponding mapped positions from the corner points in the reference image. By considering all the 3 images in Fig. 2, with a total of 576 such points, the mean quadratic cost was 0.0943, therefore much smaller than 1 pixel. This means the camera calibration has been performed successfully. The execution times, for a PC with Pentium Dual-Core CPU 2.60 GHz and 1 152 × 648 image size, were: 1) automatic corner localization: 0.78 s per image, and 2) Levenberg-Marquardt algorithm: 5.38 s. Hence, the total execution time was 3(images) × 0.78 + 5.38 = 7.72 s. It must be stressed that camera calibration is not in general a real time problem. Hence the execution time is not at all a hard practical limitation.

3   Automatic georeferencing

The critical step in automatic georeferencing is to relate properly all the reference frames. These frames are shown in Fig. 4, and the basic problem is to map an image pixel position from the image (IMAG) frame to the navigation

(NAV) frame. As in [7], only north-east coordinates are of interest, and the height comes from the UAV navigation system. The earth-centered earth-fixed (ECEF) frame is necessary because typically the global navigation satellite system (GNSS) provides position information in this coordinate system.
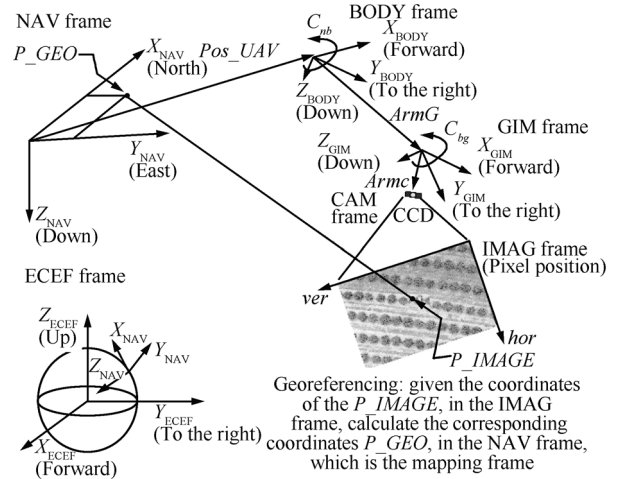

Fig. 4   Reference frames for automatic georeferencing

Let P_c be a point in the camera (CAM) frame and Pos_UAV the UAV position in the NAV frame. Then, the representation of P_c in the NAV frame is given by

$$P_{c\_n} = Pos\_UAV + C_{nb}\left(ArmG + C_{bg}\left(ArmC + C_{gc}P_c\right)\right) \quad (8)$$

which provides the translation between the CAM and NAV frames, where C_{nb} is the DCM between the UAV and the navigation frame, $ArmG$ is the camera gimbal lever with respect to the UAV CG, C_{bg} is the DCM between the UAV and the gimbal, $ArmC$ is the CCD lever with respect to the gimbal and C_{gc} is the direction cosine matrix (DCM) between the gimbal and the camera.

The relationship between the P_c coordinates and P_GEO is obtained from the camera model, namely,

$$\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix} = \begin{bmatrix} C_{cn} & T_c \end{bmatrix} \cdot \begin{bmatrix} x_n \\ y_n \\ z_n \\ 1 \end{bmatrix} \quad (9)$$

where $\begin{bmatrix} C_{cn} & T_c \end{bmatrix}$ are the camera extrinsic parameters, i.e., attitude and translation vectors. Additionally, the relation between the IMAG and CAM frames is given by

$$s\begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = K\begin{bmatrix} x_c \\ y_c \\ z_c \end{bmatrix}, \quad K = \begin{bmatrix} k_x f & \gamma & x_{i0} \\ 0 & k_y f & y_{i0} \\ 0 & 0 & 1 \end{bmatrix} \quad (10)$$

where s is a scale factor and K is the intrinsic camera parameter matrix obtained during the camera calibration process.

Now, by using: 1) UAV attitude and position; 2) the gimbal angles and 3) equation (8), C_{cn} and T_c are obtained. Hence, also by using the lens distortion parameters,

(9) and (10) can be solved for the coordinates x_n (North) y_n (East) in meters, thereby completing the georeferencing procedure. Numerically, georeferencing is a simple problem, involving only matrix algebra including matrix inversion. Hence, the computation time is basically function of the image size. For instance, in a PC with Pentium Dual-Core CPU 2.60 GHz, a $1\,152 \times 648$ image is georeferenced in approximately 0.58 s.

4 Experimental results

Two complementary experiments are now reported for evaluating the proposed automatic georeferencing procedure. The first one concerns indoor data, in a controlled environment, and the second one refers to a real UAV flight, were the length and width of a runway provide the ground truth.

4.1 Indoor data

A Sony DSC-H70 camera with resolution of $1\,152 \times 648$ pixels was used in this experiment and mounted on a tripod, which enable variation in camera attitude and position. Automatic corner localization for camera calibration was carried out as explained in Fig. 1. For ground truth, an image was taken with 4 circles on the floor, where the distances between their centers of gravity (CG's) form a square with known size of 1 m. This image is shown in Fig. 5, where there is a circle quite close to the image top border, in order to increase the necessity of adequate lens distortion calibration.

Fig. 5   Image for the indoor test

By using the electronic theodolite GTS-102N, from Topcon, a 3-dimensinal (3D) frame of reference was defined and each circle true coordinates were obtained. The accuracy of the true measurements were addressed to the GTS-102N accuracy, i.e., ± 2 mm. The electronic theodolite also provided the camera position and attitude. Therefore, by using the camera attitude $(pan; tilt) = (-14.08; -32.08)$ degrees, its position, $(X, Y, Z) = (-0.021; -0.037; -1.154)$ m, and the estimated camera and lens distortion parameters, georeferencing of the image in Fig. 5 was carried out and the result is shown in Fig. 6.

The true (measured) and estimated (via georeferencing) CG positions for the 4 circles in Fig. 6 are shown in Table 1.
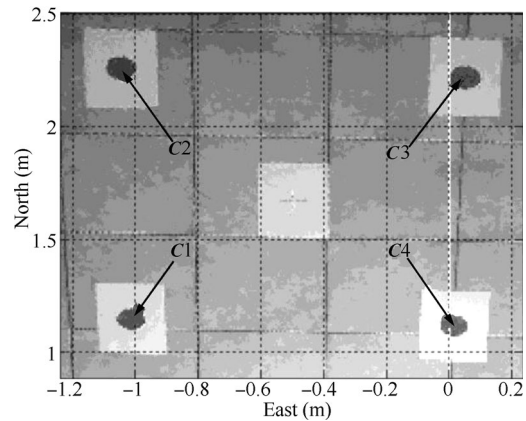

Fig. 6   Georeferenced image obtained from the original image shown in Fig. 5. True distance between circle CG's is 1 m

Table 1   True and estimated CG positions for circles in Fig. 6

Cricle	North (m)		East (m)	
	True	Georeferenced	True	Georeferenced
$C1$	1.232	1.161	-1.000	-1.018
$C2$	2.236	2.252	-1.000	-1.045
$C3$	1.236	2.212	0.0	0.048
$C4$	1.232	1.154	0.0	0.015

Since the number of estimated points is small, the error variance is not representative in this case. Hence, the maximum error is a better measure of performance. In this particular experiment, it amounts to 7.8 cm, which can be considered adequate. As a matter of fact, in any application this error is mainly a function of the camera attitude error. In order to illustrate this, Fig. 7 shows the equipotentials for position errors in the north direction, in meters, when the camera tilt angle is varied by 1 degree. The errors in the east direction are not presented, since they are much smaller, as expected.
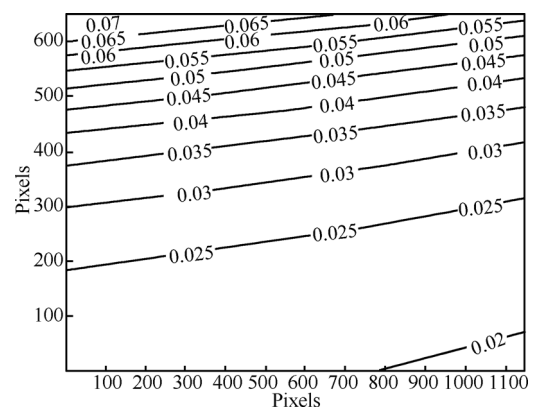

Fig. 7   Equipotentials in meters, for errors in north direction, when the camera tilt angle is varied by 1 degree. Axes in pixels (image size: $1\,152 \times 648$)

The correct way of interpreting Fig. 7 is as follows: consider, for instance, the circle $C3$ in Fig. 6, whose CG is located at pixel position $(hor, ver) = (947, 515)$ in Fig. 5, by considering the origin at the image lower right corner.

Then, from Fig. 7 it follows that an error of 1 degree in the camera tilt angle would produce an error of approximately 0.045 m in its CG localization, since this is the approximate equipotential for pixel position (947, 515) in Fig. 7.

The importance of Fig. 7 must be further highlighted: typically a georeferencing mission starts with a given accuracy requirement. This impacts upon the selection of an adequate UAV navigation system, since in a real time application the camera attitude error arises mainly due to the UAV attitude error. Indeed, the gimbal angles are known in advance, since they can be determined accurately from the gimbal controller. Therefore, the error equipotentials analysis aids in selecting a navigation system which meets the accuracy requirement.

An additional application could be in object tracking. For instance, Wang et al.[10] addressed the robust object tracking problem, in which the camera is supposed to be fixed. The extension of this method for the case in which there is camera movement, such as in UAV mounted camera, demands a prior phase of image registration, which can be performed by the procedure proposed here.

4.2 Flight data

The georeferencing is now evaluated by using real images obtained in a flight test. In this experiment, the Sony DSC-H70, already used in the indoor test, was installed on-board an UAV platform. The camera orientation was chosen to be nadir relative to the airframe.

The UAV is hand launched, hence suitable for georeferencing applications where no runways are available. The UAV platform and the camera installation are shown in Fig. 8.
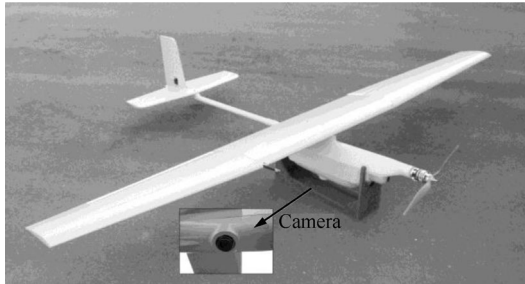

Fig. 8   UAV platform used in the flight test

The UAV in Fig. 8 is equipped with an air data attitude and heading reference system (ADAHRS) that uses microelectromechanical system (MEMS) sensors, including accelerometers, rate gyros, magnetometers, air data pressure sensors, along with a differential wide area augmentation systems (WAAS)-enabled GPS receiver. This device provides pitch and roll accuracy of 0.2 degrees, heading accuracy of 0.6 degrees, and position with an accuracy of 3 m circular error probable (CEP). The altitude is calibrated at sea level pressure of 1013 mbar. The UAV has an autopilot and a guidance system, running in the on board computer, for controlling and navigating the aircraft over the landmarks. The on board computer is programmed to shoot the camera at every 7 s. During image acquisition, the attitude, altitude and aircraft positions are grabbed from the ADAHRS and filed for further processing. The images are

taken directly from the camera secure digital (SD) card.

The flight test accuracy requirement would be defined in terms of the equipment specification above very straightforwardly, in the center of the image, as

$$CEP_{GEOREF} = CEP_{ADAHRSnb} + \varepsilon = (3 + \varepsilon)m \quad (11)$$

where ε is the error due the attitude and height accuracy, obtained by a procedure similar to that shown in Fig. 7.

4.3 Experimental setup: Ground truth

An image was obtained from an airclub runway, as shown in Fig. 9, whose length and width are known accurately. The values are 225 m and 10 m, respectively, and constitute the ground truth. The problem addressed in this section is to use the camera parameters, the UAV attitude and height, and the gimbal angle, so as to recover these two runway parameters. It would also be possible to establish a datum (point with precise GPS position) in the runway and try to recover it. However, in this case the main source of error would be the UAV navigation system itself and the delay in the image acquisition system, which are not the central points of this work.
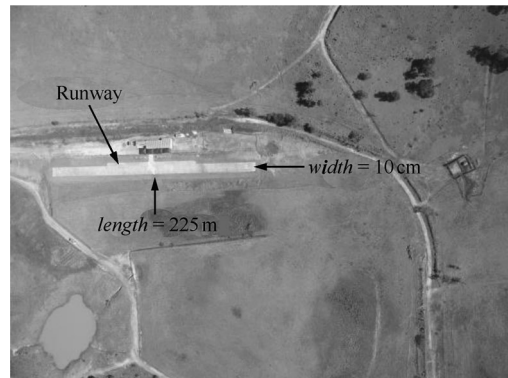

Fig. 9   Runway image obtained by the UAV camera, with $1\,152 \times 648$ pixels. Runway ground truth: $length = 225$ m, $width = 10$ m

By using the available parameters: 1) UAV attitude $(roll, pitch, yaw) = (0.0356;\ 0.057;\ 3.166)$ rad; 2) UAV height equal to 357.46 m; 3) gimbal angles $(pan, tilt) = (-\pi/2,\ \pi/2)$ rad and 4) camera parameters as in equations (6) and (7), the obtained georeferenced image is shown in Fig. 10. Only the relevant part of the image is shown, for better visualization.

By comparing the runway parameters in Figs. 9 and 10, it is concluded that the georeferencing procedure produced adequate results.

It is worth mentioning that in this experiment, the ADAHRS and GPS systems employed were those available. Therefore, no a priori accuracy could be imposed to the results. However, in a typical application, the user would start by taking into account the accuracy requirement. Based on that, these 2 systems would then be selected, as already explained after Fig. 7.
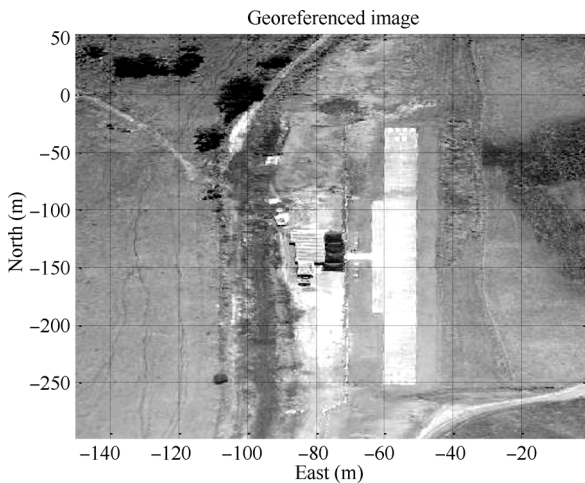
Fig. 10   Georeferenced image. Runway estimated parameters: $length = 224 \, \mathrm{m}$, $width = 9.5 \, \mathrm{m}$

5   Conclusions

An automatic georeferencing procedure has been modeled and implemented in this work. Since adequate camera calibration is necessary for good georeferencing accuracy, a technique for automatic localization of corner pixels in the model plane image has also been proposed. The performances of both camera calibration and automatic georeferencing techniques were evaluated experimentally, with adequate results.

It must be stressed that the automatic georeferencing accuracy depends on four main factors: 1) the UAV attitude and position systems accuracies; 2) the constraints in the flight height; 3) the camera resolution and 4) proper camera calibration. To some extent, 1), 3) and 4) can be properly dealt with, in order to comply with the requirements. However, 1) also impacts upon accuracy, to the extent that tight demands for accuracy can have as consequence the necessity of replacing the UAV attitude and position system altogether. Therefore, the procedure proposed in this paper for automatic georeferencing can also be useful to assist in selecting the UAV attitude and position hardwares.

A relevant byproduct of image georeferencing is the simplicity with which image mosaicking can be performed: simply align the georeferenced images according to their north and east coordinates, and apply some smoothing procedure in case of overlaps. In agriculture applications, for instance, mosaicking is necessary to provide covering of a large crop area. This application will be explored elsewhere.

6   Acknowledgments

References

[1] R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Transactions on Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 1987.

[2] Z. Y. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[3] J. Y. Bouguet. Camera Calibration Toolbox for Matlab, 2005. http://www.vision.caltech.edu/bouguetj/calib_doc/htmls/parameters.html.

[4] J. Mark, P. J. Hardin. Applications of inexpensive remotely piloted vehicles (RPVs) for collection of high-resolution digital imagery. In *Proceedings of the 20th Biennial Workshop on Aerial Photography, Videography, and High-resolution Digital Imagery for Resource Assessment*, Weslaco, TX, 2005.

[5] G. Q. Zhou, C. K. Li, P. G. Cheng. Unmanned aerial vehicle (UAV) real-time video registration for forest fire monitoring. In *Proceedings of the 2005 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, Seoul, Korea, pp. 1803–1806, 2005.

[6] K. Choi, I. Lee. Real-time georeferencing of image sequence acquired by a UAV multi-sensor system. In *Proceedings of the 2011 International Workshop on Multi-platform/Multi-sensor Remote Sensing and Mapping*, IEEE, Xiamen, China, pp. 1–6, 2011.

[7] H. T. Xiang, L. Tian. Method for automatic georeferencing aerial remote sensing (RS) images from an unmanned aerial vehicle (UAV) platform. *Biosystems Engineering*, vol. 108, no. 2, pp. 104–113, 2011.

[8] R. V. Jategaonkar. *Flight Vehicle System Identification: A Time Domain Methodology*, AIAA Press, Reston, Virginia-USA, 2006.

[9] H. J. Song, Y. Z. Chen, Y. Y. Gao. Velocity calculation by automatic camera calibration based on homogenous fog weather condition. *International Journal of Automation and Computing*, vol. 10, no. 2, pp. 143–156, 2013.

[10] Q. C. Wang, Y. H. Gong, C. H. Yang, C. H. Li. Robust object tracking under appearance change conditions. *International Journal of Automation and Computing*, vol. 7, no. 1, pp. 31–38, 2010.

Elder M. Hemerly graduated from Universidade Federal do Espírito Santo (UFES), Brazil in 1981. He received the M. Sc. degree from Technological Institute of Aeronautics (ITA), Brazil in 1985, and the Ph. D. degree from Imperial College-London in 1989. He is currently a professor at the Electronics Engineering Division of ITA, in the area of control systems.

    His research interests include system identification, adaptive control and sensor fusion.

    E-mail: hemerly@ita.br