EXPERIENCES IN GPS/IMU CALIBRATION. RIGOROUS AND INDEPENDENT CROSS-VALIDATION OF RESULTS

V. Casella, M. Franzini

DIET – University of Pavia - Italy – (vittorio.casella, marica.franzini)@unipv.it

KEY WORDS: Photogrammetry, GPS/INS, Direct Georeferencing, Calibration, Accuracy.

ABSTRACT:

The paper concerns selected results from a vast work regarding calibration of GPS/IMU systems and quality of direct georeferencing in photogrammetry. Thanks to Pavia's Test Site (PTS) and to a complex structure of flights which were acquired above it, it is possible to perform rigorous and independent validation of results. This means: the possibility of calibrating on one flight and validating on another, totally independent; the usage of disjoint sets of points for calibration and for assessing the results of direct georeferencing. Issues are also investigated such as short term time stability of calibration, systematic errors caused by time recording delays of the GPS/IMU integrated system and reestimation of camera focal length.

1. INTRODUCTION

Direct georeferencing (DG) in photogrammetry is nowadays widely used and rather well known from the conceptual point of view. Nevertheless, there are aspects of this technology which are still research topics. Some belong to the applicative level and deal with the definition of best practices for flight execution and GPS/IMU system calibration. Several issues belong to this first group: the frequency of IMU calibrations, the way to manage for mis-calibrated values of focal length, the significance of the re-estimated focal length, the mathematical model to adopt for performing IMU calibration (one or two steps, three or six estimated parameters), whether it is better to use local Cartesian coordinates instead of mapping ones. They have all previously been faced but a broader case study is necessary, in our opinion, to derive general, widely accepted, conclusions.

There are other issues belonging to an intermediate level: the identification of systematic errors and the assessment of their size; the effects of time registration delays and the way to compensate for them. To obtain reliable conclusions for the above-cited items it is necessary to have large case studies with a very good ground truth: several independent flights characterized by different flight heights, provided with repeated and cross strips; many CKPs with a dense and uniform distribution, very well measured.

Finally, there are pure research aspects such as the stochastic model used in the calibration process. In the greater part of the currently running procedures, IMU calibration is performed with the so-called two step methodology and each image is given the same weight. Even if Kalman filtering (KF) produces each time variances for the directly measured exterior orientation parameters, and they vary significantly, they are not taken into account. Even though aerial triangulation programs calculate variance for the adjusted exterior orientations, such values are not usually considered.

The present paper illustrates selected results of a vast work about calibration of GPS/IMU systems and quality of DG in photogrammetry.

One of the aims of the paper is *independent validation* of results, meaning usage of CKPs totally unused in previous stages of the workflow and availability of three different blocks flown at as many different heights, for each of the two considered flights. Another significant issue is *cross validation*, as there are two identical but different and independent flights. *Time stability* will be evaluated also, even if only in the short term. Finally, *time recording delays* will be investigated.

Time stability of GPS/IMU calibration was systematically investigated by Michael Cramer (Cramer, 2002). He also deals with camera self calibration and accuracy of DG; he hasn't performed independent checks, as the same 21 points were used as both GCPs and CKPs; concerning the stochastic model, he has assigned each exterior orientation value the same variance, taken from literature. Finally, not having different flight heights, he couldn't investigate the problem of focal length re-estimation.

The OEEPE (now EuroSDR) test is a well known and widely used term of reference for studies concerning DG in photogrammetry. Within its frame, CKPs and GCPs were kept separate, while the same flight was used for calibration and testing, thus preventing studies on calibration stability; there were two different flight heights, allowing focal length re-estimation: a third one could also allow for an independent check of the obtained values.

Attempts at considering a better stochastic model were performed by (Skaloud and Schaer, 2003) and (Pinto et alii, 2004). The former tried to model the correlation between successive solutions of KF; the latter took into account variances of the exterior orientation determined by aerial triangulation, but not of those directly coming from KF; they also tackled the correlation between different states of KF.

2. THE TEST SITE AND THE DATASET

During the last five years, Pavia's Test Site (PTS) has been established. It has many relevant features which have been developed according to the needs of the ongoing research projects.

There is a high-quality GPS network, constituted of 13 vertices. It includes a GPS permanent reference station operated by the *Laboratorio di Geomatica* of the DIET Department of the University of Pavia. There are many different cartographies concerning Pavia. Their scales range from 1:500 to 1:100000. Concerning laser scanning, there are several datasets acquired with different sensors: Optech 1210, Toposys I, Optech 3033. There are also several check areas

constituted of GPS and classical ground surveying measurements of flat areas, such as tennis courts and car parks, ramps and also sections of terrain.
Finally there are many ground control and check points. There are around 180 artificial ones (AGCP) which are white squares of 35 cm. They homogeneously cover the whole test site, which is 6 x 4.5 km wide.

Figure 1. Distribution of the AGCPs over the test site

There are also 50 large artificial points of 50 cm, recently added, and 62 natural ground control points.

Four different flights were performed over the test site, by the Italian company CGR, whose planes are equipped with Applanix POS/AV 510 sensors. Two of them were performed with a camera whose focal length is 300 mm, while the others were taken with a 150 mm camera.

The flights are composed of a certain number of blocks, flown at different heights and characterized by the scales 1:5000, 1:8000 and 1:18000. These image scales are usually used in Italy to produce maps respectively at the scales 1:1000, 1:2000 and 1:10000.

Data for **flights 1 and 2**, which are used within this paper, were acquired with a Wild RC30 camera, equipped with a 150 mm lens. They are composed of three blocks whose structure is shown in Figure 2, Figure 3 and Figure 4. The **1:5000 block** has three ordinary parallel strips covering a part of the test site, flown in an East-West direction. The first strip, once completed, is immediately re-flown in reverse. There are two cross strips, at the head and tail of the block; each of them is re-flown in reverse at the end. The along-track overlapping is 60%, while the across-track one is 30%. The number of images taken is around 140.
The **1:8000 block** has seven ordinary parallel strips covering the whole test site, flown in the East-West direction. The first one is flown back and forth. There are two cross strips, at the head and tail of the block; each of them is flown back (at the end) and forth (at the beginning). The along-track overlapping is 60%, as is the across-track. The number of images is around 130.
The **1:18000 block** has a very simple structure and is constituted of two strips flown in the East-West direction, with the 60/60 overlapping. The number of images taken is around 20.

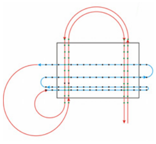

Figure 2.
Structure of the
1:5000 block

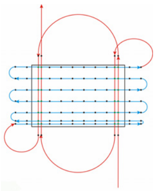

Figure 3.
Structure of the
1:8000 block

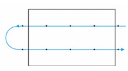

Figure 4.
Structure of the
1:18000 block

Flight	Date	Scale	Date	Focal length	Relative flight height	Overlapping	Strip number	Image number
1	14/5/2003	1:5000	14/05/03	150 mm	750 m	60/30	8	139
		1:8000	14/05/03	150 mm	1200 m	60/60	11	131
		1:18000	14/05/03	150 mm	2700 m	60/60	2	19
2	16/05/03	1:5000	16/05/03	150 mm	750 m	60/30	8	135
		1:8000	16/05/03	150 mm	1200 m	60/60	11	128
		1:18000	16/05/03	150 mm	2700 m	60/60	2	15
3	06/04/03	1:5000	06/04/03	300 mm	1500 m	60/30	8	146
		1:8000	06/04/03	300 mm	2400 m	60/60	11	145
4	17/03/03	1:8000	17/03/03	300 mm	2400 m	60/60	11	135

Table 1. Summary of the performed flights

3. QUALITY EVALUATION OF THE AERIAL TRIANGULATION

Aerial triangulation (AT) was performed for all the six blocks with the BLUH program of the University of Hannover. The AGCP set was split into two disjoint sets: the proper GCPs, which were used within the adjustment, and the CKPs, only used for independent quality assessment.

AT was performed with respect to a local cartesian reference system, where the subscripts *e*, *n* and *u* come from.
In order to evaluate the quality of the measurements which were performed on the images and to assess the attainable accuracy limit for the used images, validation was performed on the EO values determined by AT: object coordinates of CKPs were determined by photogrammetry and compared with the true ones. Table 2 summarizes the results: the second column indicates the number of the CKPs used; the third column reports

the number of observations, that is, the number of the photogrammetric measurements which were performed and checked; the procedures used implement single-model stereoplotting, which is normally used in map compilation, and a certain point can be measured more than once. Columns 4 to 6 report the averages of the differences between the stereoplotted coordinates and the true ones. Columns 7 to 9 report the standard deviations of the same differences.

Flight	Pts	Obs	$\mu_{\Delta e}$ [m]	$\mu_{\Delta n}$ [m]	$\mu_{\Delta u}$ [m]	$\sigma_{\Delta e}$ [m]	$\sigma_{\Delta n}$ [m]	$\sigma_{\Delta u}$ [m]
pv1-5000	167	613	-0.007	0.005	0.012	0.038	0.037	0.055
pv1-8000	193	1139	0.006	-0.006	0.010	0.046	0.055	0.087
pv1-18000	132	341	0.016	-0.014	-0.021	0.094	0.104	0.163
pv2-5000	147	570	-0.005	-0.006	0.012	0.053	0.054	0.076
pv2-8000	185	1068	0.007	0.004	-0.011	0.050	0.057	0.101
pv2-18000	159	382	0.009	0.029	-0.048	0.086	0.111	0.209

Table 2. Accuracy attainable with the AT-determined EO values

4. IMU CALIBRATION

IMU system calibration was performed six times, using all the considered blocks. The reference ATs were calculated with BLUH, while the following steps, calibration, direct orientation of the images and validation, were executed with Matlab procedures specifically written by the authors. Table 3 summarizes the results obtained : the second column shows the number of the images used; the next three columns contain the lever arm values, indicated with *D* (Delta), as they measure an offset; the following three columns contain the boresight misalignments, indicated with an *M*; the last six columns indicate the standard deviations of the above listed parameters.

Calibrations were performed at this stage with the usual two step procedure and the *D* and *M* vectors were determined by taking the simple arithmetic average of the differences between the AT-determined EOs and the directly measured ones, once they were converted to the same reference system.

Calibrations performed on the 1:18000 blocks are less reliable than the others, due to the reduced number of the images used . Generally speaking the misalignment estimation is very good, as figures are very often below the threshold indicated by the manufacturer, and rather stable from one calibration to another.

Lever arm estimation is less stable. In principle, if there weren't any systematic components, all the six lines of columns 3-5 should contain the same values. But at least one systematic error is present, due to the miscalibrated focal length; this bias is absorbed in the D_z estimation during calibration therefore, being the effects of the uncalibrated focal length depending on flight height, we must expect the D_z values to vary between the flights at different heights.

Nevertheless, variations of the planimetric components of D and variations of the altimetric one between different flights having the same height are not, in principle, to be expected. The observed differences could be due to random errors, which are present of course, and to other biases.

One further bias is probably due to time recording delays, as shown in Section 5. Concerning random errors, further papers will investigate the rigorous statistical discussion of the differences between diverse calibrations.

Nevertheless, Sections 6 to 8 will show that the differences in calibrations produce minimal effects on the accuracy of the stereoplotted coordinates, and this is the important thing.

Flight	Photo	D_x [m]	D_y [m]	D_z [m]	M_x [grad]	M_y [grad]	M_z [grad]	σ_{Dx} [m]	σ_{Dy} [m]	σ_{Dz} [m]	σ_{Mx} [grad]	σ_{My} [grad]	σ_{Mz} [grad]
pv1-5000	80	-0.174	-0.048	-0.146	-0.7227	0.1521	-0.0614	0.053	0.064	0.034	0.0075	0.0042	0.0049
pv1-8000	76	-0.134	-0.113	-0.220	-0.7204	0.1553	-0.0609	0.075	0.081	0.031	0.0044	0.0039	0.0041
pv1-18000	8	0.027	-0.076	-0.400	-0.7253	0.1586	-0.0633	0.101	0.092	0.050	0.0024	0.0039	0.0029
pv2-5000	84	-0.166	-0.015	-0.154	-0.7253	0.1522	-0.0605	0.086	0.098	0.040	0.0082	0.0053	0.0059
pv2-8000	77	-0.173	-0.041	-0.156	-0.7240	0.1546	-0.0617	0.094	0.105	0.034	0.0047	0.0044	0.0040
pv2-18000	8	-0.215	0.028	-0.292	-0.7238	0.1570	-0.0672	0.203	0.182	0.110	0.0024	0.0071	0.0033

Table 3. Summary of results obtained in the six calibrations performed

5. TIME RECORDING ISSUES

Time recording delays were initially investigated and results are shown in Table 4 for the flight *pv1-8000*. It summarizes the differences of the camera centre between the directly measured position and that determined by AT. The average value and the standard deviation are shown. It must be noted that the direct positions used in this section were obtained before IMU calibration, that is, using null values for lever arms and boresight misalignments.

The table shows results for the whole block and for four selected strips, respectively flown with the nose of the plane pointing North, South, East and West, as the final letter of the names in the first column highlights.

It is worth noticing that a time recording delay (the fact that a certain image is given the position that the camera had a moment before or after the shot) causes the presence of an offset in the camera position parallel to the flight direction. This is what happens: flight *pv1-8000-1-N* has a $\mu_{\Delta y}$ value of -9 cm and flight *pv1-8000-2-S* has a $\mu_{\Delta y}$ value of 12 cm. The same phenomenon happens with the East-West strips: the *pv1-8000-3-E* set has a $\mu_{\Delta x}$ value of -13 cm and the *pv1-8000-4-W* has 16 cm. When the whole block is considered, these systematic errors translate into enlarged random errors, as the results of set *pv1-8000* show.

It is clear that time recording delays are present, but certainly the considered data also contains other systematic errors whose origin must be investigated.

A final remark regards the fact that the above described significant systematic errors apparently don't affect accuracy. This is probably due to the capability that calibration has of somehow absorbing the greater part of this bias. Further work is necessary to separate the effects of time delays from other error sources: in subsequent papers an exhaustive study of this phenomenon will be presented and an attempt will be made to remove such errors.

Flight	Photo	$\mu_{\Delta x}$ [m]	$\mu_{\Delta y}$ [m]	$\mu_{\Delta z}$ [m]	$\sigma_{\Delta x}$ [m]	$\sigma_{\Delta y}$ [m]	$\sigma_{\Delta z}$ [m]
pv1-8000	76	0.001	-0.013	0.220	0.162	0.130	0.031
pv1-8000-1-N	4	0.034	-0.092	0.239	0.081	0.033	0.030
pv1-8000-2-S	6	-0.004	0.125	0.226	0.103	0.059	0.045
pv1-8000-3-E	8	-0.129	-0.162	0.253	0.071	0.023	0.018
pv1-8000-4-W	7	0.168	0.020	0.212	0.066	0.051	0.029

Table 4. Effects of time recording delays

6. ACCURACY OF DG WITH HOMOGENEOUS CALIBRATIONS

Homogeneous validations were performed for the six considered flights. Validation was performed in the same way described in Section 3, that is, stereoplotting approximately 160 CKPs. It is *homogeneous* because calibration and validation are calculated on the same flight, as columns 1 and 2 of Table 5 show: this is not representative of the daily working procedures, but allows us to define the best accuracy which is attainable with DG. Comparing Table 5 and Table 2 it can be noticed that DG accuracies are not too far from those of AT, in this homogeneous situation

Calibration	Validation	Pts	Obs	$\mu_{\Delta e}$ [m]	$\mu_{\Delta n}$ [m]	$\mu_{\Delta u}$ [m]	$\sigma_{\Delta e}$ [m]	$\sigma_{\Delta n}$ [m]	$\sigma_{\Delta u}$ [m]
pv1-5000	pv1-5000	167	613	0.054	-0.027	0.013	0.056	0.064	0.112
pv1-8000	pv1-8000	193	1139	0.055	-0.051	0.002	0.060	0.072	0.109
pv1-18000	pv1-18000	132	341	0.100	-0.068	0.065	0.107	0.138	0.219
pv2-5000	pv2-5000	147	570	0.085	-0.006	0.001	0.073	0.075	0.097
pv2-8000	pv2-8000	185	1068	0.097	-0.032	-0.007	0.057	0.073	0.109
pv2-18000	pv2-18000	159	382	0.182	-0.102	-0.142	0.099	0.144	0.288

Table 5. Accuracy of DG with homogenous calibration

7. CROSS VALIDATION WITHIN THE SAME FLIGHT

Cross validation was also performed, first within the same flight. Results will only be shown for the flight pv1, due to space reasons. Also, only calibrations calculated with the flights pv1-5000 and pv1-8000 will be used.

Results are summarized in Table 6, whose lines 2 and 6 coincide with lines 1 and 2 of Table 5. Passing from homogeneous calibration to a heterogeneous one, random errors maintain approximately the same size in general but sometimes planimetric components are increased by up to 50%. Concerning systematic errors, an increase of the *up* component is clearly visible: this highlights that the focal length value used, taken from the camera calibration report, is significantly different from the true one.

This systematic error is absorbed into the estimation of D_z and, if validation and calibration flights have approximately the same height, miscalibrated focal length is not too disturbing. But this is no longer valid if calibration and validation happen with two different flight heights: a systematic height error is visible, whose size is a function of the difference in height between calibration and validation flights.

Calibration	Validation	Pts	Obs	$\mu_{\Delta e}$ [m]	$\mu_{\Delta n}$ [m]	$\mu_{\Delta u}$ [m]	$\sigma_{\Delta e}$ [m]	$\sigma_{\Delta n}$ [m]	$\sigma_{\Delta u}$ [m]
pv1-5000	pv1-5000	167	613	0.054	-0.027	0.013	0.056	0.064	0.112
pv1-5000	pv1-8000	193	1139	0.052	-0.051	0.076	0.065	0.076	0.114
pv1-5000	pv1-18000	132	339	0.100	-0.088	0.283	0.165	0.190	0.220
pv1-8000	pv1-5000	167	613	0.052	-0.027	-0.061	0.061	0.073	0.114
pv1-8000	pv1-8000	193	1139	0.055	-0.051	0.002	0.060	0.072	0.109
pv1-8000	pv1-18000	132	340	0.111	-0.086	0.199	0.109	0.221	0.202

Table 6. Accuracy of DG with cross calibrations, within the same flight

8. CROSS VALIDATION WITH DIFFERENT FLIGHTS: SHORT TERM TIME STABILITY

Cross validation was also performed between flights *pv1* and *pv2* and results are summarized in Table 7. Results are surprisingly good, but it must be considered that the analyzed flights are separated by 48 hours. In the 8000 flights minor systematic errors appear in the height component. Their origin is that the re-estimated values of the focal length for the two flights show a difference of around 10 microns. In other words, the D_z components of lever arms, estimated within blocks of the same height, but on two different days, show a significant difference (Table 3).

Calibration	Validation	Pts	Obs	$\mu_{\Delta e}$ [m]	$\mu_{\Delta n}$ [m]	$\mu_{\Delta u}$ [m]	$\sigma_{\Delta e}$ [m]	$\sigma_{\Delta n}$ [m]	$\sigma_{\Delta u}$ [m]
pv2-5000	pv1-5000	167	613	0.054	-0.028	0.007	0.057	0.064	0.112
pv2-8000	pv1-8000	193	1139	0.057	-0.052	0.068	0.064	0.073	0.113
pv1-5000	pv2-5000	147	570	0.086	-0.006	0.008	0.073	0.075	0.097
pv1-8000	pv2-8000	185	1068	0.096	-0.033	-0.071	0.063	0.076	0.110

Table 7. Accuracy of DG with cross calibrations between different flights

9. RE-ESTIMATION OF FOCAL LENGTH

Re-estimation of focal length was performed for both flights *pv1* and *pv2*. Focal length corrections were calculated for each block, as AT was performed by jointly adjusting the usual photogrammetric observations as well as the measurements of the camera centre, performed by GPS; this required a careful weighting strategy, of course.

Noticeably, re-estimated focal length changes with the height therefore it is not strictly correct to assign a unique value to a flight, but this is done in the present section, for the sake of simplicity. Averaging the different results for flight *pv1*, it was established that the correction to apply to the nominal value of focal length was 30 microns.

GPS/IMU calibration was performed again, with the new focal length, for block *pv1-8000* only, and the results are shown in Table 8. Comparing the new calibration parameters with the previously determined ones, contained in the third line of Table 3, the only significant difference is the value D_z, as expected.

Validation was performed for the three blocks of flight *pv1* and the results are in Table 9. For homogeneous combination of calibration and validation, nothing changes, as line 3 of Table 9 and line 3 of Table 5 show. What is very interesting is that cross validations show very reduced systematic height effects: they are due to the residual differences of focal length between the various flight heights.

Flight	Photo	D_x [m]	D_y [m]	D_z [m]	M_x [grad]	M_y [grad]	M_z [grad]	σ_{Dx} [m]	σ_{Dy} [m]	σ_{Dz} [m]	σ_{Mx} [grad]	σ_{My} [grad]	σ_{Mz} [grad]
pv1-8000	76	-0.131	-0.111	0.019	-0.7204	0.1554	-0.0609	0.075	0.08	0.03	0.0043	0.0040	0.0041

Table 8. GPS/IMU calibration for flight *pv1-8000*, after focal length re-estimation

Calibration	Validation	Pts	Obs	$\mu_{\Delta e}$ [m]	$\mu_{\Delta n}$ [m]	$\mu_{\Delta u}$ [m]	$\sigma_{\Delta e}$ [m]	$\sigma_{\Delta n}$ [m]	$\sigma_{\Delta u}$ [m]
pv1-8000	pv1-5000	167	613	0.051	-0.027	0.026	0.061	0.073	0.115
pv1-8000	pv1-8000	193	1139	0.055	-0.051	0.002	0.060	0.072	0.108
pv1-8000	pv1-18000	132	340	0.110	-0.082	-0.088	0.110	0.225	0.209

Table 9. Accuracy of DG after focal length re-estimation

10. CONCLUSIONS

Various issues of GPS/IMU calibration and DG quality assessment were treated. Several combinations of the flights used for calibration and for validation were considered.

Accuracy was estimated using the nominal focal length and the re-estimated one. Short term calibration stability was studied. An initial attempt to recognize effects of time recording delays was made.

Further activities will deal with: systematic application of all the presented methodologies to all the available data; study of the time stability of the re-estimation of camera focal length; study of the variations of camera focal length across the various flight heights; rigorous statistical discussion of the variations between different calibrations and different focal length re-estimations; assessment of the errors induced by time recording delays.

11. BIBLIOGRAPHY

Cramer, M., 2002. Investigations on long term stability of system calibration for direct georeferencing. Final project study, Institute for Photogrammetry and Remote Sensing (ifp), Stuttgart, Germany. http://www.ifp.uni-stuttgart.de/publications/2002-/IGI_final_report.pdf (accessed 4 Apr. 2005).

Galetto, R., Casella, V., Spalla, A., Franzini, M., 2004. An Italian research project on direct photogrammetry. Proceedings of XXth ISPRS Congress, Istanbul, Turkey, published on DVD.

Heipke, C., Jacobsen, K., Wegmann, H., 2002. Analysis of the results of the OEEPE test “Integrated Sensor Orientation”. Test Report and Workshop Proceedings, Official Publication n°43, Werbedruck Schereckhase, Spangenberg, pp.31-45.

Honkavaara, E., Ilves, R., Jaakkola, J., 2003. Practical results of GPS/IMU/camera system calibration. Proceedings of International Workshop (WG I/5) "Theory, Technology and Realities of Inertial/GPS Sensor Orientation, Castelldefels, Spain, published on CD.

Pinto, L., Forlani, G., Passoni, D., 2004. Experimental tests on the benefits of a more rigorous model in IMU/GPS system calibration. Proceedings of XXth ISPRS Congress, Istanbul, Turkey, published on DVD.

Skaloud, J., Schaer, P., 2003. Toward a more rigorous boresight determination. Proceedings of International Workshop (WG I/5) "Theory, Technology and Realities of Inertial/GPS Sensor Orientation, Castelldefels, Spain, published on CD.

12. ACKNOWLEDGEMENTS

The research presented in this paper used data partially acquired within the frame of the National Research Project entitled *Integrated inertial positioning systems in aerial Photogrammetry*, co-funded by the Italian Ministry of the University for the year 2002, and chaired by prof. Galetto of the University of Pavia. The Italian company CGR is acknowledged for supplying additional required data.

The authors would like to thank the IPI Institute of the University of Hannover, namely prof. Christian Heipke and prof. Karsten Jacobsen for hosting Marica Franzini for a period of research and for allowing the free usage of the BLUH program.

The authors are also pleased to thank the technicians of the *Laboratorio di Geomatica* of the University of Pavia, Giuseppe Girone and Paolo Marchese, who performed most of the measurements used for the described research.