# INTEGRATING PHOTOGRAMMETRY AND INERTIAL SENSORS FOR ROBOTICS NAVIGATION AND MAPPING

*Fadi Bayoud, Jan Skaloud, Bertrand Merminod*
*Ecole Polytechnique Fédérale de Lausanne (EPFL)*
*Geodetic Engineering Laboratory*
*CH-1015, Lausanne*
*fadiatef.bayoud@epfl.ch, jan.skaloud@epfl.ch, bertrand.merminod@epfl.ch*

**Abstract**: Integrating visual and inertial sensors has become a common practice in navigation due to the increase in computer power, in algorithms advancement and in sensor improvements. One of the problems yet to be solved is the Simultaneous Localisation And Mapping (SLAM). SLAM is a term used in the robotics community to describe the problem of mapping the environment and at the same time using this map to determine (or to help in determining) the location of the mapping device. Classically, terrestrial robotics SLAM is approached using LASER scanners to locate the robot relative to the structured environment and at the same time to map this environment; however, outdoors robotics SLAM is not feasible with LASER. Recently, the use of visual methods, integrated with inertial sensors, has gained an interest. The current solutions use a single Kalman Filter with a state vector containing the map and the robot coordinates, which introduces non-linearity and complications to the filter, which then needs to run at high rates (20 Hz) with simplified navigation models. In this study, SLAM is developed using the Geomatics Engineering approach. Two filters are used in parallel: the Least-Squares Adjustment (LSA) for mapping and the Kalman Filter (KF) for navigation. Conceptually, the outputs of the LSA photogrammetric resection (position and orientation) are used as the KF external measurements. The filtered position and orientation are then employed in the LSA Photogrammetric intersection to map the surrounding features that are used as control features for the resection in the next epoch. In this manner, the KF takes the form of a navigation only filter using complete modelling, with a state vector containing the corrections to the navigation parameters and updated at low rates (1 to 2 Hz). Results show that this method is feasible with limitation induced from the quality of the images used. Although simulation showed that a position with accuracy of 5-10 cm for objects at distances of up to 10 meters is possible, in practice this is not achieved due to the low quality of the CCDs used. The environment of testing and/or of potential application is very crucial because of the importance of feature extraction's accuracy. In this paper, the methodology is presented, differences with other SLAM solution are pointed out, and numerical tests are provided and discussed.

## 1. Introduction

Robotics Simultaneous Localisation And Mapping (SLAM) is the problem of mapping the environment surrounding the robot and at the same time using this map to determine the location of the robot (Csorba, 1997; Newman 1999). Navigation and mapping systems are the core elements of SLAM, without which an exploring robot cannot do its job. The applications of self-navigating – exploring – robots are abundant, but one of the most important is: going to and exploring places where no man is safe to do. A map of the surrounding environment of the robot and a navigation system are essential for the robot to perform manoeuvres and in turn to complete its mission.

Traditionally, terrestrial robotics SLAM is approached using LASER scanners to locate the robot relative to the structured environment and to map this environment at the same time. LASER scanners have shown to be a very good tool where the accuracy of localisation is within the centimetre level. However, outdoors robotics SLAM is not feasible with LASER alone due to the absence of stereotypic features and environment's roughness. Therefore, different tools have to be used. Recently, the use of visual methods, integrated with inertial sensors, has gained an interest. These visual methods rely on exploitation of one or more cameras (or video). Yet, no clear indication about the mapping methods or integration algorithms is explicitly illustrated; one can consult the Journal of Robotics Systems (2004). Jung (2003) has used vision motion estimation to perform SLAM. His method relies on linking the pixel's motion rate on the images with the displacement of the robot. However, this filter has to run at very high rates and in case of vision loss, no other means to re-locate the robot is available.

These particular solutions use a single Kalman Filter with a state vector containing the map and the robot coordinates. This introduces high non-linearities, large state vector as well as other complications to the filter, which needs to run at high rates (20 Hz) with simplified navigation models. A classical tool for mapping is Photogrammetry, where sequence of stereo-images are captured and self-oriented. As for the navigation systems, the coupling of Inertial Navigation System (INS) and Global Navigation Satellite Systems (e.g., the GPS) is indispensable. In addition, photogrammetry can also be used for localisation – positioning.

In this paper, SLAM is solved by employing two CCD cameras and one IMU. Two filters are used in parallel: the Least-Squares Adjustment (LSA) for mapping and the Kalman Filter (KF) for navigation. Conceptually, the outputs of the LSA photogrammetric resection (position and orientation) are used as the KF external measurements. The filtered position and orientation are then employed in the LSA Photogrammetric intersection to map the surrounding that is used (as a set of control features) for the resection in the next epoch. In this manner, the KF takes the form of a navigation filter only, with a state vector containing the corrections to the navigation parameters. This way, the mapping and localisation can be updated at low rates (1 to 0.5 Hz) and more complete modelling of sensor errors is applied.

While photogrammetry alone can solve SLAM, the presence of an IMU is significant for automated feature detection. Moreover, the IMU is indispensable in cases where images cannot be used for reasons of visibility and when certain manoeuvres do not guarantee knowledge of the environment; e.g. when the robot or vehicle captures images of an unknown environment. To use photogrammetry to solve SLAM, full automation is required, which is still not fully achieved. Many attempts are directed towards the automation of photogrammetry; this goal has been still falling short due to the need of high level of artificial intelligence.

In the next section, the methodology to solve the SLAM problem by photogrammetry and IMU is introduced. The third section discusses briefly the photogrammetric mathematical model, the angular transformations and lever arm corrections that link the cameras' outputs with those of the

IMU. A numerical investigation is described and analysed in section Four. Finally, conclusions are drawn in the last section.

## 2. A Methodology for Combining Photogrammetry and IMU Outputs

The methodology followed in this research is different from the traditional SLAM solution pursued by the Robotics community. In the following, there will be two separate filters; the first is the Least-squares Adjustment (LSA) filter that will do the mapping and localisation by photogrammetry; the second is the Kalman Filter that performs the localisation by optimally integrating the localisation provided by the LSA with the IMU outputs.

The employed Kalman Filter (KF) is very similar to that used in the INS/GPS navigation. But instead of using the GPS positioning and velocity updates, the LSA outputs from photogrammetric resection, the Exterior Orientation parameters (EOP), position and orientation, will be used as updates. In this way, the KF is a navigation-only filter that: i) operates at the frequency of the update (e.g., 1 or 0.5 Hz), and ii) its state vector size is kept relatively small (e.g., 15 states) with homogeneous states that guarantee rapid convergence.

This procedure requires certain points to consider:

- Recursive LSA: the LSA solution of the epoch $k-1$ is used as observations for epoch $k$,
- Correlations between measurements and unknowns are carried from one epoch to the other.

Before starting, we need to initialise the system by defining its position and orientation with respect to a well-defined mathematical reference frame that has a physical meaning. This is important when the SLAM is solved on a global scale. This is called Georeferencing. The Georeferencing can be done in two ways:

1. Initialisation with GPS/INS, which demands open skies for the GPS signal, or
2. Initialisation with resection, which demands the existence of sufficient Ground Control Points (GCP) at the beginning of the survey.
3. Standing on a known point, and stable IMU alignment.

In the first case, open sky for the GPS is vital. The GPS/INS gives us the position and attitude of the IMU, which – after applying the lever arm and angles transformation – yield the EOP of the two cameras. As for the second case, theoretically at least three GCPs are required for the determination of the position and attitude of the two cameras by resection; however this is far from straightforward because of the sensitivity of the IMU to misalignments. Experience show that this is a very critical aspect of SLAM and thus a sufficient number of GCPs is required to be distributed all of the images, otherwise weak initialisation force the IMU to drift at the beginning of the mission.

Having the initialisation properly done, the vehicle moves and starts to map and localise itself as portrayed in chart of Figure 1 that shows the algorithm concept.

The data flow can be depicted as follows:

***Initialisation:***

1. Position and attitude of the two cameras considered as known
2. Intersection is employed to map features

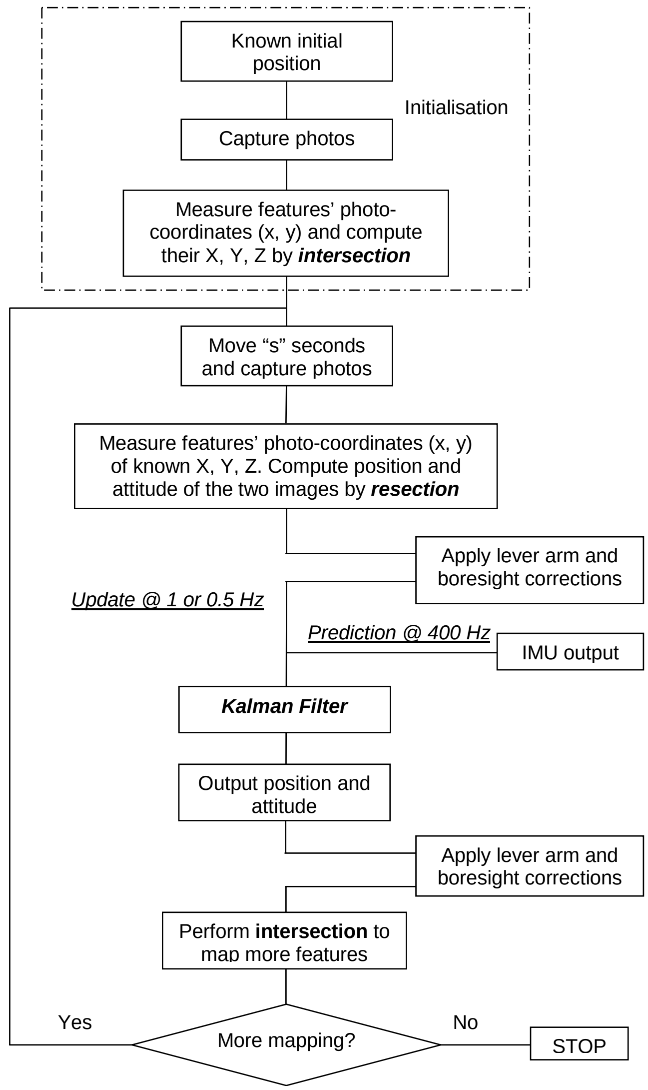


Figure 1: Flowchart of the Photogrammetric and IMU integration

### *After mapping enough features:*

1. Vehicle moves.
2. Resection computes the cameras' EOP by LSA using the features mapped from the previous cameras' location. (IMU predicted EOP can be used for feature extraction as well.)
3. Lever-arm and angles transformation (boresight) are applied to the EOPs to determine the IMU's position and attitude.
4. IMU outputs and IMU position and orientation derived from resection are integrated in a KF to compute filtered position and attitude of the current system location.
5. Lever-arm and angles transformation (and boresight) are applied to the filtered position and attitude to determine the EOP of the cameras.
6. Intersection is used to map more objects by LSA from the current location.
7. Vehicle moves and algorithm repeats.

Stages 3 and 5 are of great importance in the process of navigation and mapping, for the following facts (Figure 2):

- The cameras and the IMU are separate in space
- The outputs of photogrammetry and IMU belong to different reference systems; photogrammetry functions either in the camera space or object space and the IMU's outputs are in the body frame.

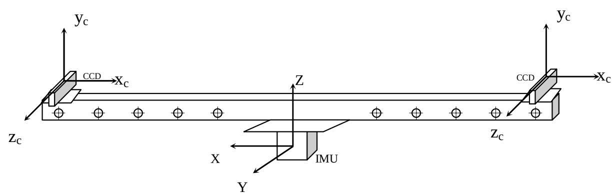


Figure 2: Mounting of the cameras and IMU

## 3. Photogrammetric Mathematical Model, Angle Transformation and Lever Arm

Mathematical properties govern the relationship between the image and the objects. The perspective centre, the object and its image are collinear, yielding a functional model called the co-linearity equation:

$$x = x_0 - c\frac{R_{11}(X - X_0) + R_{12}(Y - Y_0) + R_{13}(Z - Z_0)}{R_{31}(X - X_0) + R_{32}(Y - Y_0) + R_{33}(Z - Z_0)}$$
$$y = y_0 - c\frac{R_{21}(X - X_0) + R_{22}(Y - Y_0) + R_{23}(Z - Z_0)}{R_{31}(X - X_0) + R_{32}(Y - Y_0) + R_{33}(Z - Z_0)} \quad (1)$$

where $x, y$ are the photo-coordinates in the image frame, $X, Y, Z$ are the 3-D Coordinates in the object frame, $c$ is the focal length of the camera, $X_0, Y_0, Z_0$ are the 3-D Coordinates of the camera's perspective centre in the object frame, $x_0, y_0$ are the photo-coordinates of the projection of the perspective centre to the image plane, (theoretically, this projection point has to coincide with the principle point, which is the centre of the image frame, but in reality, it does not) and $R_{ij}$'s are the elements of the rotation matrix between the image and object frames, $\mathbf{R}_m^{\mathbf{c}}$.

Equations 1 describe the fundamental mathematical model for photogrammetric mapping, where it reveals the relationship between the image and the object coordinate systems. With this model, one can solve the basic problems of photogrammetric mapping, namely: resection and intersection:

- ***Resection***: In resection, the position and attitude (EOP) of an image are determined by having a set of at least three points with known coordinates in the object frame as well as in the image frame; these are the GCPs.
- ***Intersection***: In intersection, two images, with known EOPs, are used to determine the coordinates in the object frame of features found on the two images simultaneously, employing the principle of stereovision.

In principle, photogrammetry alone can be used to solve SLAM by employing recursively resection and intersection. This was analysed by a previous publication, Bayoud et al (2004); however, there are important shortcoming of such approach as mentioned in the Introduction.

### 3.1. Angles transformation and lever arm

The angle transformation applied in Stage 3 (going from resection to KF), is used to transform the orientation output of the resection from the mapping/camera frame to the earth/body frame to be consistent with the inertial output and KF states

$$\mathbf{R}_{\mathbf{b}}^{\mathbf{e}} = \mathbf{R}_{\mathbf{m}}^{\mathbf{e}} \left(\mathbf{R}_{\mathbf{m}}^{\mathbf{c}}\right)^{\mathbf{T}} \mathbf{T}_{\mathbf{b}^*}^{\mathbf{c}} \mathbf{R}_{\mathbf{b}}^{\mathbf{b}^*} \quad (2)$$

Where $\mathbf{T}_{\mathbf{b}^*}^{\mathbf{c}}$ is the rotation matrix between IMU and camera frames and depends on the definition of the axes (Figure 2), $\mathbf{R}_{\mathbf{b}}^{\mathbf{b}^*}$ is the boresight that contributes for the mounting imperfections, $\mathbf{R}_{\mathbf{b}}^{\mathbf{e}}$ is the rotation matrix between IMU body frame and Earth-Centred-Earth-Fixed (ECEF) frame, and $\mathbf{R}_{\mathbf{m}}^{\mathbf{e}}$ is the rotation matrix between ECEF and mapping frames.

The angle transformation applied in Stage 5 (going from KF to intersection) is used to transform the output of the inertial and KF to the camera's reference frame to perform the mapping. This is well documented in the relevant literature because it is the classical way of image direct-Georeferencing by GPS/INS (Skaloud and Schaer, 2003). The transformation is:

$$\mathbf{R}_{\mathbf{m}}^{\mathbf{c}} = \mathbf{T}_{\mathbf{b}^*}^{\mathbf{c}} \mathbf{R}_{\mathbf{b}}^{\mathbf{b}^*} \left(\mathbf{R}_{\mathbf{b}}^{\mathbf{e}}\right)^{\mathbf{T}} \mathbf{R}_{\mathbf{m}}^{\mathbf{e}} \quad (3)$$

The lever arm is divided as well into two parts depending on whether the process goes from resection to KF or from KF to intersection. In the first case, resection gives the coordinates of the cameras in the mapping frame. To these, the leverarm is added to obtain the coordinates of the IMU in the mapping frame, and then to transformed to the ECEF frame. Having the coordinates of the IMU, computed from resection, in the ECEF frame, they update the KF to determine the filtered position of the IMU.

### 3.2. Leverarm and Boresight calibration

The concept of the boresight $\mathbf{R}_{\mathbf{b}}^{\mathbf{b}^*}$ determination is well documented in the relevant literature (Bäumker et al., 2001; Skaloud and Schaer, 2003). In this work, an indirect procedure was followed to determine the two boresight matrices and two leverarm vectors of the left (L) and right (R) cameras. In the frame of the work carried out at the Geodetic Engineering Laboratory, a mapping system with a high-definition digital camera is well calibrated with respect to the IMU with known boresight and leverarm; to this system, the two CCDs were added (Figure 3). The boresight and leverarm of the two CCDs were first calibrated with respect to the high-definition digital camera by determining the EOP (by resection from GCP) of the three cameras in three different locations. Then, once these were computed, the link between the two CCDs and the IMU were directly made through the already known boresight and leverarm between the digital camera and the IMU.

## 4. Numerical Application

For testing the methodology and the set-up, two CCD cameras (progressive scan SONY XC-55, 640*480 square pixels of resolution 7.4 μm, and with a 6 mm lens) and an LN-200 IMU were used as instruments. Along, there were a synchronisation pulse, a Matrox Meteor-II/Multi-Channel frame grabber and a screen, IMU data acquisition box developed at the EPFL-TOPO (Skaloud and Viret, 2004), a laptop, and the power supply. The image grabbing was carried out at every second and was properly synchronised with the IMU via a synchronisation pulse. After a couple of minutes of inertial initialisation, the vehicle moved and started taking images. See Figure 3.

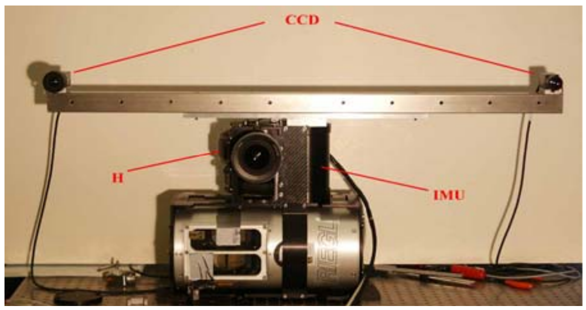


Figure 3: The system containing Sony CCD cameras, the LN-200 IMU, the High-resolution Hasseblad camera used in calibration and a Laser scanner not used in this study

### 4.1. Controlled Environment

Two tests were performed indoors in a controlled environment with the aim to evaluate the methodology and validate the software. In a controlled environment, accuracies are very similar to the simulations. Two sets of images, each containing 12 images in a 6 stereo-pairs with 3 to five seconds apart, were chosen. This means that the Kalman Filter update occurs every three to five seconds. Figure 4 portrays sample images of the second set. The initialisation was done using resection, after which, SLAM took over to map more features and locate the cameras.

According to the photogrammetric theory and simulations, the depth (here a X & Y-components) is geometrically weak because of the small stereo base of 1 meter. Also, it is known that a large number of GCPs is needed to determine an accurate EOP (8 to 10 GCP were used at each epoch). If more GCPs were used, more accurate determination of the EOP would be guaranteed, and this in turn assures accurate intersection. This was also confirmed by these two tests. Four Check Points (CHP) were used as control points at the end of the survey. The positioning accuracy is indirectly controlled by the accuracy of the CHPs; for a direct control on localisation, an outdoors test with GPS is needed. Tables 1 and 2 show the differences between the CHP of the two tests, respectively, determined by a theodolite and their SLAM estimated positions. An interesting remark is the harmony in the signs of the differences in the X and Y components; they are either positive or negative, while the Z component is more random (also the differences in the Z-component can be considered as statistically insignificant); this needs to be analysed in further tests to see if there is a trend or it is random behaviour.

| CHP | X | Y | Z |
|---|---|---|---|
| 1 | 1.9 | 3.9 | -0.5 |

| CHP | X | Y | Z |
|---|---|---|---|
| 1 | -4.9 | 4.5 | 0.1 |

| | | | |
|---|---|---|---|
| 2 | 6.7 | 11.9 | 0 |
| 3 | 6.2 | 10.1 | -0.8 |
| 4 | 4.4 | 7.0 | -0.4 |

Table 1: Validation of the first test, error on check points after 20 seconds & 6 photos -cm

| | | | |
|---|---|---|---|
| 2 | -1.1 | 1.1 | -1.7 |
| 3 | -1.8 | 0.5 | -1.4 |
| 4 | -5.5 | 7.5 | 3.3 |

Table 2: Validation of the second test, error on check points after 13 seconds & 6 photos - cm

Figure 4: Image examples of the second set

### 4.2. Outdoor test

The cameras used in the current set-up do not fit well the accuracies needed; this is mainly due to their short focal length (6 mm) without enough pixels. In addition, the stereo-base of 1 meter constrains the intersection geometry and limits the accuracy of the objects at distance larger than 10 meters. This set-up is very sensitive to poor lightening and far features, which need a close to an ideal field.

The initialisation was performed by determining the EOP by resection. This proved to be a critical phase due to the demand of high accuracy. Although the features were well distributed and measured accurately by a totalstation, they were poorly seen on the images (Figure 5, first image); furthermore some of them were further than 15 meters from the cameras. This caused large misalignments of the IMU and in turn caused the prediction to diverge as seen in Figure 7. However, due to the Kalman Filter updates, the INS solution is rapidly fixed in the third epoch and its predictions are consistent with the resection solution as can be seen in Figures 6 and 7.

Figure 5: Outdoor images

It is worthwhile to note the stability in the z-channel as seen in Figure 7, where after the third epoch the innovation does not exceed few centimetres. This is due to the strong geometry of this channel by photogrammetry as discussed above. It is also claimed that the differences in X & Y (the innovation) are due to the poor positioning from the resection; this analysis comes from the deduction that the Z-component of the IMU (its weak component) is close to the Z-component of resection (the strong photo component), so it is left that the differences in Figure 6 are due to the inaccurate resection and not due to the IMU drift. Due to the short stereo-base and poor feature recognition on the images the mapped features are determined with an accuracy of around 50 cm; this in turn degrades the accuracy of the resection that is dependent on the poor feature recognition as well. However, this means that the weighting in the Kalman Filter stochastic modelling needs more investigation.

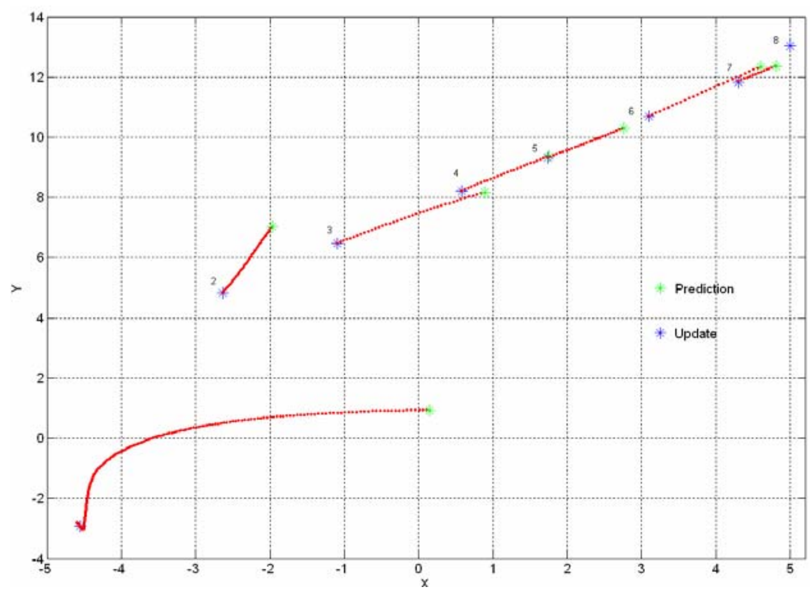


Figure 6: Differences between the prediction and update in X & Y (innovation)

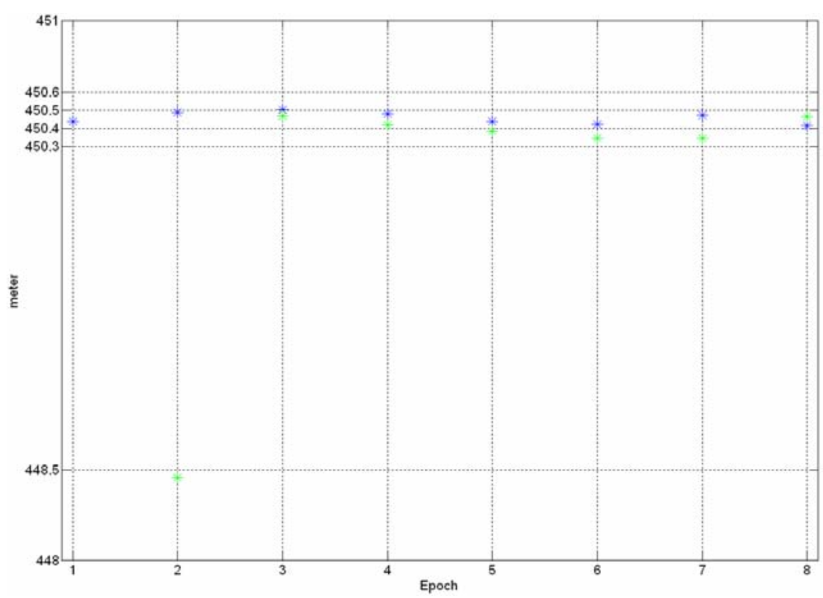


Figure 7: Differences between the prediction and update in Z (innovation)

## 5. Conclusions and Future Work

Robotics SLAM can be solved by Geomatics modus operandi. An integration procedure between an IMU and photogrammetry/cameras was developed and tested in this paper. The outputs of a pair of cameras are used first to localise the vehicle; this position is then used as an external measurements in a Kalman Filter whose prediction are the filtered outputs of an IMU. The Kalman filtered position is used, then, with the outputs of the two cameras to perform the mapping in a Least-Squares adjustment filter.

Conceptually, this integration is done in a loosely-coupled Kalman Filter. The algorithmic part of the filter, on the other hand, is far from simple and needs an understanding of system control, data handling, and priority managing. This was done in the preliminary software that was written for the purpose of his work.

The controlled numerical results showed that the methodology is promising and show the practical feasibility of our procedure. However, the image component is crucial and its choice has to be taken in care. Short focal length with low resolution highly limits the accuracy of the mapping that in turn degrades the quality of the vehicle's localisation. The short stereo-base has as well a similar disadvantage.

The conceptual advantages of using IMU/photo compared to only photos SLAM are considerable:

- Image sequences with limited or no overlaps do not halt the mapping
- IMU predicted EOP allows automated feature extractions
- Increased robustness and accuracy

An important remark from this work is the need of synergy between different specialisation for the advancement of specific problem. In this work two distinct approaches were combined, from which an apparent benefit for both become clear.

**References:**


[1] Bäumker M., F. J. Heimes (2001): New Calibration and Computing Method for Direct Georeferencing of Image and Scanner Data Using the Position and Angular Data of an Hybrid Inertial Navigation System. OEEPE Workshop, Integrated Sensor Orientation, Sep. 17-18, 2001; Hanover.

[2] Bayoud F. A., J. Skaloud, and B. Merminod (2004): Photogrammetry-Derived Navigation Parameters for INS Kalman Filter Updates. $XX^{th}$ ISPRS Congress 2004. Commission V, WG V/2, 2004.

[3] Csorba M. (1997): Simultaneous Localisation and Map Building. Ph.D. Thesis. Robotics Research Group, Department of Engineering Science, University of Oxford, 1997.

[4] Newman P. M. (1999): On the Structure and Solution of the Simultaneous Localisation and Map Building Problem. Ph.D. Thesis, Australian Centre for Field Robotics, The University of Sydney, 1999.

[5] Journal of Robotics Systems: Volume 21, Issues 1 and 2, January and February 2004, http://www3.interscience.wiley.com/cgi-bin/jhome/35876. Wiley Periodicals, pp 1-94.

[6] Jung I K (2003): Simultaneous Localisation and Mapping in 3D Environments with Stereovision. PhD Thesis. CNRS-LAAS Report N$^{o}$ 04132, France.
[7] Skaloud J., P. Schaer (2003): Towards a More Rigorous Boresight Calibration. ISPRS International Workshop on Theory, Technology and Realities of Inertial/GPS/Sensor Orientation, Commission 1, WG I/5, 2003.
[8] Skaloud J. and P. Viret (2004): GPS/INS Integration: From Modern Methods of Data Acquisition to New Applications, European Journal of Navigation, Nov. 2004: 60-64.