Improvement of Precise Vehicle Location in Urban Areas Using Video-based Photogrammetry

André Pinhal[1,†], José Gonçalves[1,2,*,†]

[1]*University of Porto, Observatório Astronómico, 4430-146 Vila Nova de Gaia, Portugal*

[2]*CIIMAR, Interdisciplinary Centre of Marine and Environmental Research, 4450-208 Matosinhos, Portugal*

Abstract

This paper presents a system under development at the University of Porto, which integrates a Septentrio MosaicGO GNSS receiver, with a helical antenna, associated with a Gopro action camera. Video frames can be processed by a Structure from Motion (SfM) approach to do an alignment of sequential frames, derive relative position of the camera projection centres, and eventually create point clouds of the surrounding environment. The system being developed has the camera and the antenna mounted in a block, which is placed on a rear-view mirror of a car. The camera acquires a video in 4K mode, at 60 frames per second. Frames are extracted from the video at a frequency of between 2Hz to 10 Hz, depending on the car speed. The camera collects GNSS data with its own navigation receiver, allowing to tag all extracted video frames with GPS time and position. Due to the low accuracy of the camera receiver, its positions are discarded and only GPS time is kept. This time is used to synchronize with much more accurate positions obtained by the RTK receiver, connected to a CORS network. Many times, the obstacles present in urban areas do not allow for the high accuracy positioning. A trajectory made within an urban area will have some frames with very precise positions and other with much higher errors. All frames are photogrammetrically processed. Standard deviations of camera positions are considered in the bundle adjustment, allowing for the improvement of the camera positions of lower accuracy. The SfM processing also generates a point cloud of the surrounding objects, which can be densified. Objects identified on this point cloud can be used to assess the location accuracy of the process. Several experiments carried with the system in the city of Porto allowed to confirm that a positional accuracy of 10 cm can be achieved.

Keywords

RTK, ambiguity fix, action camera, MMS, structure from motion, point cloud

1. Introduction

Precise positioning by GNSS RTK (Real Time Kinematics) is now very common in Surveying and GIS data collection for 3D city models. There are many surveying solutions for use in mobile mapping systems (MMS), integrating GNSS positioning, inertial navigation systems (INS) and data acquisition sensors, such as cameras or laser scanners [1]. Popular brands, as Riegl, Leica Geosystems or Trimble, among others, provide solutions that allow for the generation of very dense and accurate point clouds in urban environments, but with costs of several hundred thousand euros, and whose use is restricted to professional markets. Many users will be interested in having cheaper systems, making use of small and low-cost GNSS receivers now available, which can have good performance in urban areas [2]. There are several devices for development, costing less than 1000 euros, which have triple frequency and all the capacity for high-precision differential positioning, in real time or in post-processing. Regardless of the additional sensors associated with the GNSS receiver, it is currently possible to have high-precision kinematic positioning in a motor vehicle with this type of low-cost receivers.

As inevitably happens with GNSS positioning in urban environments, or generally in situations of major obstructions to signal propagation, RTK positioning, with ambiguity fixing ("fix" solution), with errors of a few centimeters, will not be possible in significant parts of trajectories made in this type of environment. "Float" type solutions will often result, with precision on the order of a few

decimeters, or even "single point" solutions (SPP), with errors of a few meters. This limitation is normally overcome with inertial navigation systems, significantly increasing the costs of an MMS. The system under development aims to solve this problem through the use of sequences of images acquired by a small video camera. It is possible to apply to the image sequences, photogrammetric techniques that determine the relative orientation of the images, thus contributing to a more complete positioning solution.

The system is intended to use cameras classified as "action camera". These cameras are compact, rugged, and versatile to capture both video and still images in outdoor conditions. They are very popular among adventure enthusiasts, many times also interested in geolocating images and videos. For this reason, there are several models that incorporate a GPS navigation receiver, allowing to geotag photographs, and video as well. That is the case of cameras of the brand Gopro, since the launch of model Hero 5. These cameras acquire video in format MP4, which can accommodate GPS data, to be later extracted using specific software [3] provided by the camera manufacturer [4]. It is possible to obtain in this way GPS position and GPS time of every individual video frames.

Photogrammetry has seen major developments in recent years, through the incorporation of methodologies derived from computer vision. It was mainly with an algorithm developed by David Lowe [5] – SIFT (Scale Invariant Feature transform) – that made the automatic extraction of common points between images much easier, even for images with variations of orientation and scale. This availability of conjugate points between images of large blocks is combined with the bundle adjustment methods already possible for large blocks. This process is also applicable to cameras for which only approximate internal orientation parameters are known, integrating in the bundle adjustment a self-calibration process. This largely automated methodology of image orientation started to be designated in some communities as "Structure from Motion" (SfM), allowing the processing of aerial or terrestrial images.

In the present system a video is acquired, frames are extracted and are oriented ("aligned", in the terminology associated to SfM) in relative terms. Provided that, for some of the images, coordinates of projection centers are known, all images in the block will have their positions determined. In this way it is possible to complete or correct the camera trajectory.

2. Description of the hardware implemented

The system described in this paper incorporates a Septentrio Mosaic GO X5 GNSS receiver, that has a compact and lightweight design, suitable for integration into platforms such as drones and handheld devices. It supports multiple satellite constellations, including GPS, GLONASS, Galileo and BeiDou, and does phase measurement in three different frequencies. The receiver is mounted in a box, which is attached to the back of the camera, and a helical antenna is placed on top of the box. A small power bank is mounted next to the receiver, which once turned on starts recording a file in the Septentrio Binary Format (SBF). When connected to a CORS (Continuous Operation Reference Station) network the receiver provides reliable real time kinematic (RTK) positioning. The receiver was used in 10 Hz, providing positions with enough density to properly model curved trajectories of a car moving in a city.

The operation of the system requires the use of a smartphone, which connects to the internet. An application runs on the smartphone, which allows for controlling the receiver, receive an NTRIP correction from a CORS network, carry out the differential correction and monitor the receiver's performance. In the present case an Android smartphone, was used, running the SW-Maps application[1], which is free software and includes all those capabilities. Although positions may be recorded in the app, they are retrieved from the receiver memory, in the form of the SBF file and in NMEA format (National Marine Electronics Association), for the RTK positions.

The camera has a fixing piece to be mounted, together with the receiver, on the side rear view mirror of a car. Once the camera and the receiver are turned on, both can be controlled from inside the vehicle,

---

[1]http://softwel.com.np/mobile_products

with the smartphone. Figure 1 shows the block mounted on the right rear-view mirror of a car. The camera optical axis is horizontal, rotated by 15 degrees to the right of the vehicle axis.

Figure 1: Gopro camera and Septentrio Mosaic Go X5 GNSS receiver, mounted in a block on a side rear view mirror of a car.

The camera has a GNSS receiver, which works at a frequency of 18 Hz. After the camera is turned on, the user should check when GPS position is available and only then start the video recording. The receiver must be connected beforehand and should, preferably, have RTK fixed position when the camera starts the video. Note that there is no electronics establishing a connection between the two devices. The way of synchronizing between image and RTK positions is done through the GPS time recorded by the two devices.

The camera acquires video in 4K resolution, with a dimension of 3840 by 2160 pixels, at a rate of 60 fps (frames per second). Action cameras are known for having a large field of view, at the cost of a significant radial deformation. Although images can be used in this way, preference was given to the image mode called "linear", which consists of the application of a general correction model to produce images with a regular central projection. The resulting image has an equivalent focal distance of approximately 1800 pixels, still a very wide angle. Small additional corrections, in the focal distance, principal point position and radial distortion coefficients, characteristic of each camera unit, may be necessary, but these will be handled in the processing step, through a self-calibration incorporated in the SfM bundle-adjustment.

Videos taken from the car are processed to extract navigation information from the camera, which is carried out by programs from the GPMF library[2]. Information is extracted in groups of 60 frames, that is, every 1.001 seconds of the video, and includes video time of each group, position, speed and UTC time, in seconds of the day. Table 1 shows an example of this information. Positions result from the camera navigation receiver and will, in fact, be discarded. The main information is the video time,

---

[2]https://github.com/gopro/gpmf-parser

which is transformed into frame number, and UTC.

Table 1
Sample of GPMF data extracted

Tvideo, s	Latitude, °	Longitude, °	Altitude, m	Speed, m/s	UTC, seconds of day
0.001	41.1550155	-8.6617334	122.159	0.142	52526.179
1.001	41.1550151	-8.6617350	122.110	0.242	52527.169
2.001	41.1550131	-8.6617353	122.226	0.117	52528.159
3.001	41.1550124	-8.6617346	122.295	0.143	52529.204

Frames are extracted from the video, in JPEG format, at a suitable cadence so that the overlaps between consecutive images are adequate for the alignment process to be successful. For the conditions under which the system was operated, the 4Hz cadence proved to be adequate. In situations of higher speed, and especially when there is rotation, the cadence may be increased. In cases where the vehicle is stopped, for example at traffic lights, the corresponding frames may be discarded if the GPMF data has a speed lower than a tolerance, for example 0.5 m/s.

The positions collected by the GNSS receiver are retrieved in the NMEA format, essentially through messages GGA and GST [6], that provide results of UTC time, position, quality of the position (Q), with values of 1 for SPP, 4 for FIX and 5 for FLOAT. Standard deviations are also obtained, which will be of few centimetres in case of Q=4. Surveys were carried with a connection to the Portuguese permanent station network ReNEP ("Rede Nacional de Estações Permanentes"). Although a post-processing can be done over the SBF data, in general the RTK positioning could be obtained in those cases where obstacles allowed for that. Table 2 shows a sample of data extracted from the NMEA files, with a loss of ambiguity fix and a sudden increase of estimated precision.

Table 2
Sample of Septentrio receiver positions, type of position (Q), number of satellites and estimated precision (standard deviations, in meters)

UTC, seconds	Latitude, °	Longitude, °	Altitude, m	Q	N, sat	σ_{LAT}, m	σ_{LON}, m	σ_{ALT}, (m)
53129.4	41.15580621	-8.66189798	64.228	4	6	0.044	0.016	0.098
53129.5	41.15580270	-8.66189887	64.314	4	6	0.045	0.016	0.099
53129.6	41.15579907	-8.66189934	64.177	1	14	6.029	1.598	9.056
53129.7	41.15579563	-8.66189954	64.188	1	14	6.029	1.598	9.056

At this point, a linear interpolation is carried in order to calculate camera position. There is a need of a previous calibration of the relation between frame number and UTC, which is described in the next section. There is also a small offset between the antenna phase centre and the centre of the camera, of approximately 4 cm in the horizontal component and 4 cm in the vertical. In a first approximation this is not being corrected, since it is smaller than what is initially expected for the system accuracy. The camera GPS receiver has a frequency of 18 Hz, i.e., 3.3 times smaller than the video frame rate. Assuming an error of 1.5 frames in the synchronization, i.e., 0.25 seconds, for a car moving at a speed of 30 km/h, the error corresponds to a distance of 20 cm. For that reason, we put the expectations at the level of 1 or 2 decimetres. However, the procedure for correction of the small lever arm is described in the calibration section.

Finally, all the selected frames were aligned in a photogrammetric software that applies the SfM concept, which in our case was Agisoft Metashape [7]. Interpolated positions are provided only for those images that had Q=4. Standard deviation of 0.2 m was considered for the least squares adjustment. As a result of the bundle adjustment positions of all images are obtained. They will be analysed in an independent manner.

3. System calibration

In order to process the image data and RTK positions collected, some calibration steps are necessary, especially regarding the times to be assigned to the frames extracted from the video.

3.1. Time calibration

As seen in Table 2, the UTC time does not correspond to the exact moments of the first frame of each block of 60 frames, since they are not equally spaced. A linear relationship was established between frame number, N_f, and UTC time, through,

$$UTC = A_0 + A_1 N_f, \tag{1}$$

where A_0 is the time of the first frame (starting count at frame zero) and A_1 is the frame rate. This will allow us to determine a more reliable value for the UTC time of the first frame.

An independent validation of this assessment of the time of the first frame, was done with a small flashlight, connected to the GNSS receiver, fired in front of the camera. A precise time of the flash event is recorded on the GNSS receiver, and with a few flashes along a video, the calibration can also be done with the same formula. Very similar results were obtained, with differences in the time of the first frame around 20 ms, i.e., only slightly more than one frame. Figure 2 shows an image of the flash in a frame.

Figure 2: Example of a flash light fired in front of the camera, to assess GPS time of a frame.

3.2. Lever arm between receiver and camera

As referred before, there is a lever arm between the antenna and the camera, which is of only a few centimeters and, in a first approach, is not being considered. There is no attitude assessment, so the coordinate transportation must be done with the orientation of the trajectory. As the vehicle moves approximately in a horizontal plane, the azimuth of the trajectory can be estimated and the corresponding rotation applied to the vector between antenna and camera. In the vertical component there is only a need of subtracting the height difference between the phase center and the camera. The actual projection center position is not known, but since the focal distance of the lens is only 3 mm, it was assumed as the center of the lens, outside the camera.

4. Assessment of system performance

The system was tested in some urban environments, in a first approach, in areas without extreme situations of difficulty in signal capture. A route was taken in an urban residential area of the city of Porto, without tall buildings, but with trees along the streets. The route began and ended at the same

point, it had some crossings and there were some repetitions of some sections along the route, which had a total length of 3.2 km. Images were extracted at a rate of 4 Hz, in a total of 2464 images. Figure 3 shows, on the left side, the path followed, over the Google Maps image base, and on the right, two examples of frames from the video, one in a more unobstructed area and another with more tree cover.

4.1. Image alignment

After interpolation of RTK positions for all images, it was observed that 72% were FIX-type positions. The images were loaded into the Agisoft Mestashape program and a first alignment was made, at this step without coordinates of the projection centers. All were successfully aligned. The coordinates of the images that had FIX were then inserted, with an a-priori accuracy value of 0.2 meters, in the three coordinates. The bundle adjustment was reprocessed, resulting in adjusted coordinates for all images. The effect of trajectory quality improvement is observed in places where the positions were not FIX. Figure 4 shows, on the left, an area where there is an interruption in the FIX positions (blue dots), in a total of nearly 40 images. After aligning the images, the positions were regularized, resulting in a much smoother trajectory that was in line with expectations.

Figure 3: On the left: trajectory over the Google Maps image background; on the right, examples of frames, with less trees (top) and more trees (bottom).
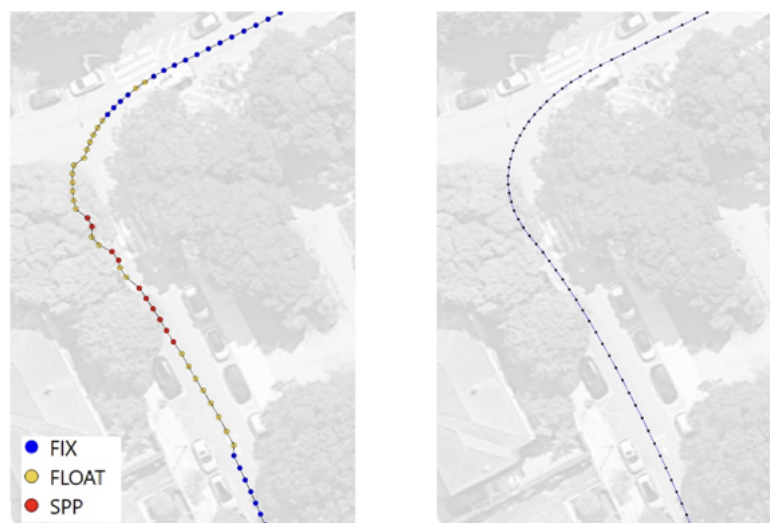
Figure 4: Example of part of the trajectory where some camera positions where not FIX (Float or SPP), on the left. The right image shows the regularized trajectory.

Although visually there is a qualitative improvement in the trajectory, and a very small change in the sections where there was FIX, this assessment has some subjectivity, so it is preferable to have a numerical evaluation of the error.

The simplest way to make this assessment is through checkpoints, whose coordinates can be determined photogrammetrically through the images. A total of 14 points were selected, on well-defined points, which were identified in the final photogrammetric project, on the images where they are observed with the highest quality. This results in coordinates of these points. Subsequently, the points were surveyed on the terrain, with GNSS, and the corresponding three-dimensional errors were evaluated. Figure 5 shows the location of the checkpoints and the location of two of them in the images.

Figure 5: Location of the check points (left) and examples of two points considered (right).

Errors were calculated in the three coordinates (e_x, e_y, e_z), in centimetres, and the corresponding statistics are presented in Table 3. They are: the average error (AVG), the root mean square error (RMSE) and the maximum absolute error (MAX), according to:

$$AVG = \frac{\sum e_i}{n}, RMSE = \sqrt{\frac{\sum e_i^2}{n}}, MAX = max(|e_i|), i = 1, \ldots, n. \tag{2}$$

Table 3
Statistics of the errors found in the independent check points

Coordinate	No. points	AVG, cm	RMSE, cm	MAX, cm
Longitude	14	-1.9	9.3	14.6
Latitude	14	-0.8	11.6	20.8
Altitude	14	-4.4	5.2	12.6

Average errors are small, not evidencing systematic trends. The RMSEs are of the order of one decimetre, agreeing with the initial expectation. These were the first tests, which are quite promising. More performance assessments of the system, in diverse conditions, will be carried in a near future.

5. Conclusions and future work

A positioning system for precise position determination of vehicles in urban environments is under development at the University of Porto. The system integrates video imagery processed by SfM in order to contribute to an integrated trajectory solution. This allows to complement the position gaps in the urban environment. Initial tests point to a possible accuracy at decimeter level. More tests will be carried out in order to assess system performance in a diversity of environments with strong obstructions, both in urban areas as in forested environments.

Several improvements to the system will be developed soon, namely an improvement in the temporal synchronization process, for example using new models of action cameras with higher frame rates. It is also intended to improve the correct reduction of the GNSS position to the camera projection center.

Acknowledgments

References

[1] M. Elhashash, H. Albanwan, R. Qin, A Review of Mobile Mapping Systems: From Sensors to Applications, Sensors 22 (2022) 4262.

[2] V. Hamza, B. Stopar, O. Sterle, P. Pavlovčič-Prešeren, Low-Cost Dual-Frequency GNSS Receivers and Antennas for Surveying in Urban Areas, Sensors 23 (2023) 2861.

[3] K. Petroskey, C. Funk, I. A. Tibavinsky, Validation of Telemetry Data Acquisition Using GoPro Cameras, Technical Report, SAE Technical Paper, 2020.

[4] GoPro, Metadata Format – GPMF, Processing software available at https://github.com/gopro/gpmf-parser (v2.2.1), 2020.

[5] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International journal of computer vision 60 (2004) 91–110.

[6] A. Ardalan, J. Awange, Compatibility of NMEA GGA with GPS receivers implementation, GPS Solutions 3 (2000) 1–3.

[7] Metashape, Agisoft, Agisoft Metashape User Manual, Professional Edition, Version 2.1, Available at https://www.agisoft.com/pdf/metashape-pro_2_1_en.pdf, 2024.