

MOBILE MAPPING SYSTEM BASED ON ACTION CAMERAS

J. A. Gonçalves ^{1*}, A. Pinhal ¹

¹ Science Faculty, University of Porto, Rua Campo Alegre, 4169-004 Porto, Portugal - (jagoncal, apinhal)@fc.up.pt

Commission I, WG I/7

KEY WORDS: mobile mapping, action camera, GNSS, camera calibration, direct georeferencing

ABSTRACT:

Action cameras can operate in outdoor conditions, such as outside a car, and provide good quality imagery that can be exploited to collect geospatial data by photogrammetric means. Recent models include GPS, which can deliver position and time of individual images and video frames. That is the case of the very popular camera, GoPro Hero 5. This paper describes the implementation of a mobile mapping system, based on a GoPro Hero 5 camera mounted on the side rearview mirror of a car. Although the system can be dependent on the camera GPS positions only, it was developed to include a GNSS dual frequency receiver, carried inside the car, on the dashboard. Within good observation conditions, without tall buildings, differential positioning (either RTK or PPK) provides the trajectory with accuracy of a few centimetres. The precise time of individual frames is obtained from the camera GPS and positions are interpolated from the GNSS receiver. Assuming the car moves in a horizontal plane and the camera has no significant tilts, the system is treated in planimetric terms, with camera axis azimuth derived from the vehicle trajectory. Positions of observed objects, such as traffic signs, are derived from consecutive frames. Tests carried out in a sparse urban environment have shown planimetric accuracy better than 40 cm, appropriate for large scale mapping, such as 1:2000. The system can be improved in several forms, through processing techniques, such as structure from motion, but without the incorporation of additional hardware.

1. INTRODUCTION

Topographic data need to be constantly updated in urban areas in order to keep accurate GIS databases for infrastructure management. Aerial photography is still the main source for most of the base map data collection. However, many objects around streets and roads, such as traffic signs and urban equipment, cannot be acquired from aerial images. Field data collection keeps being fundamental for this data acquisition. There are many solutions of mobile mapping systems (MMS) based on different sensors, such as photographic cameras, video or laser scanners, that can be used in these tasks. Most of the commercial systems are relatively complex, since they integrate direct georeferencing equipment (Global Navigation Satellite Systems, GNSS and Inertial Navigation Systems, INS) with other sensors in a synchronised data collection and processing system. Costs are relatively high and economic viability may require large volumes of work. Some papers have been published on "low-cost" MMS (Ellum and El-Sheimy, 2002, Artese, 2007, Madeira et al., 2010), but all involving relatively sophisticated GNSS, INS, and cameras hardware.

Typical MMS are normally classified as survey grade, reaching a positional accuracy of a few centimetres. In many cases that is more than what is needed for GIS databases. In Portugal large scale data are acquired in urban areas, with planimetric accuracy of 20 to 40 cm, corresponding to traditional map scales of 1:1000 or 1:2000. Standard feature catalogue tables (DGT, 2018), established by the national mapping and cadastre authority, include many objects, such as electricity poles, street lamps, different type of walls, and many others object classifications, that have to be done by field data collection. Companies end up using frequently services such as the Google Earth Street View, mainly for verification, but with the risk of

images being out of date for a few years. Simple MMS, based on mobile devices would be of great interest for these data collection, especially if they can reduce costs, be of simple use and keep the required mapping accuracy grade. Recent works by Al-Hamad et al., (2014) and Masiero et al. (2016) use smartphones as the basis of MMS.

Action cameras are capable of working in outdoors, in adverse conditions. They are frequently associated with strong geometric deformations and normally not considered for photogrammetric operations. However, evolution in this field resulted in fast processors that can apply geometric correction models, large frame rates that can reduce rolling shutter effects, include GPS positioning, and other improvements that make them interesting for terrestrial photogrammetry operations. Not many studies have been published on mobile mapping based on action cameras.

This paper describes a MMS based on a camera GoPro Hero 5, which incorporates the features described above. It is operated in a vehicle, which also carries a dual frequency GNSS receiver. The system is extremely simple to operate in the field, involving only sliding the action camera on its adhesive mount, which is stuck the side rear view mirror of a car, turn it on and start driving around. The GNSS receiver will normally work in RTK mode (Real Time Kinematics), with corrections broadcast by a permanent station network

The article describes the composition and data processing of the system and results of some positional accuracy tests carried out. Improvements that can be done, such as the incorporation of structure from motion (SfM) processing, for image orientation improvement and point cloud generation, are also described.

* Corresponding author

2. THE ACTION CAMERA

The camera used in this system is a GoPro Hero 5 Black. It is a small and light camera, very robust, which can be operated in rough conditions. It was used to collect videos from which frames are extracted, for photogrammetric processing. It also incorporates GPS, which allows for tagging frames with position and GPS time. Figure 1 shows the camera in its case, fixed in the rear view mirror of a car. Once mounted its inclination can be adjusted.



Figure 1. GoPro camera in its mounting box, fixed on the side rear view mirror of a vehicle.

2.1 Camera characteristics

The GoPro Hero 5 Black acquires still images at a maximum rate of 3 images every 2 seconds, in a resolution of 12 megapixels. Although this is a relatively high image rate, when moving in a road at a speed of 15 m/s, the spacing between consecutive images would be too large for photogrammetric use. So it was decided to use the video mode.

The camera acquires video in a large variety of resolutions, frame rates and processing modes. The highest resolution, identified as 4K, has a frame size of approximately 8 megapixels, at 30 frames per second (fps). Other resolutions include higher frame rates, up to 240 fps, and processing modes. An important processing mode is called "linear", which consists in removing the radial distortion, typical of wide modes. Figure 2 shows video frames of the standard mode ("wide") and corrected ("linear"). Table 1 has a list of characteristics of some of the video modes: frame size, frame rates and availability of linear mode.



Figure 2. Video frame in wide mode (left) and linear mode (right)

Video mode	Frame size (pixels)	Frame rates (fps)	Linear mode
4 K	3840×2160	30	No
2.7 K	2704×1520	60, 48, 24	Yes
1440 p	1920×1440	80, 60, 48, 30, 24	No
1080 p	1920×1080	120, 90, 80	No
1080 p	1920×1080	60, 48, 30, 24	Yes
960 p	1280×960	120,60	No
720 p	1280×720	240,120,60,30	No

Table 1. GoPro Hero 5 Black video modes

The main considerations in choosing a video mode are: higher frame rates, because of the reduction in the rolling shutter effect, and availability of the linear correction mode. The latter is only available for two of the resolutions, so the choice went for the highest resolution, 2.7K at 60 fps, which corresponds to 4 megapixels.

Another important consideration on the camera operation is to turn off the stabilization mode. Since in this mode the frame is a subsection of a wider image, the principal point may change its position, which would cause strong limitations to the photogrammetric use.

2.2 Frame extraction

Frames were extracted in JPEG format from the MP4 files using program FFMPEG. Time of frame n , in seconds, can be calculated (equation 1) with respect to time of frame 1, taking into account the frame rate. The actual average number of frames acquired per second for mode 60 fps is 59.94 (GoPro, 2018a).

$$t_n = t_1 + (n-1) \frac{1}{59.94} \quad (1)$$

Data blocks with GPS positions and data from other sensors can also be extracted by this program. At the moment this work was done the information about the organisation of these data was scarce. Recently, GoPro released in GitHub the structure of these data blocks (GoPro, 2018b).

2.3 Calibration

Still images acquired by the camera include, in the EXIF header, approximate values of focal distance and pixel size. However, since images of "linear mode" are being used, these values will be changed. Also for the video mode the actual area of the sensor in use may be different, so not much was known about the actual focal distance to be considered. Since the frames cover, in the horizontal direction (2704 pixels), an angle of approximately 85° , the approximate focal distance was calculated as (eq. 2):

$$f = \frac{2704}{2} \cot\left(\frac{85^\circ}{2}\right) \approx 1475 \text{ pixels} \quad (2)$$

Action cameras can be calibrated in a similar manner as other cameras (Balletti et al., 2014). In our case it was done with images of targets marked on a wall, which were rigorously measured. With several images acquired from different angles, and doing an auto calibration with Agisoft Photoscan, this approximate focal distance was improved to 1486 pixels. Additionally it could be verified that for images of the linear mode the principal point can be considered at the image centre and only a residual radial and tangential distortion remains, reaching less than 10 pixels in the image corners. Within the simplifications introduced in the system being implemented, an exact central projection was considered, with a focal distance of 1486 pixels.

2.4 GPS positioning

The camera incorporates a GPS unit, which provides position with standard navigation accuracy, i.e., errors of a few meters. In the case of still images these positions are provided in the EXIF header of JPEG images. In the case of video there are no

standard formats for the storage of positions of the camera trajectory. In the case of MP4 videos acquired by the GoPro, these data are stored in data blocks within the video file. Recently the GoPro company (GoPro, 2018b) has released information about the GoPro Metadata Format (GPMF). At the start of this work there was not much documentation about the organization of these data. A commercial program for use of action cameras in motorised sports, called Race Render, managed to extract the GPS time and position from the video.

The GPS data is acquired at a frequency of 18 Hz, recording time, latitude, longitude, height above the ellipsoid and speed. During the practical tests with the dual frequency receiver it was possible to assess the accuracy of these positions. Table 2 shows the statistics of the errors found. Planimetric accuracy (RMSE, root mean square error) was better than 2 meters, typical of navigation GPS receivers.

Statistic	Long. (m)	Lat. (m)	Height (m)
Minimum	-3.54	-4.52	-5.16
Maximum	3.45	3.28	19.31
Average	-0.30	-0.92	4.72
Std. Deviation	1.05	1.05	3.36
RMSE	1.09	1.40	5.80

Table 2. Statistics of the GoPro GPS positional errors

Relying only on the GoPro camera alone, final accuracy would not be better than this. In order to reach a large scale mapping grade the precision GNSS unit was incorporated in the system. In this case the only information to be used from the GoPro GPS data file will be the initial time. Assuming it corresponds to the first video frame, it will be possible to calculate time of all frames using equation (1).

3. COMPOSITION OF THE MMS

3.1 Precise GNSS positioning

A Trimble R6 GNSS unit was mounted in the vehicle's dashboard. It was fixed with velcro tape, so that the antenna centre is approximately in the vehicle axis. The receiver works in real time kinematic mode (RTK) receiving corrections from ReNEP, a national network of GNSS permanent stations (DGT, 2018) through mobile communications. Later the data was post-processed (PPK) in order to have better accuracy and possibly positions for instants in which RTK was not successful. The RTK has the advantage of giving an audio feedback of ambiguity fixing. In case of loss, the driver may adjust vehicle speed in order to recover fix and obtain a more complete dataset.

In dense urban environments the percentage of ambiguity fixes achieved can be low. For the remaining cases, were not even in post processing, position fixes are not possible, a solution is proposed below, in the suggestion of improvements to this simple MMS. Proposed improvements are based on processing strategies and not in the inclusion of any additional instrumentation.

Since the MMS is intended to be only planimetric, and has no additional hardware to provide camera orientation, the azimuth of the trajectory will have to be calculated. For a set of points obtained by the GNSS receiver, we can calculate for point i , the azimuth α_i of the trajectory using the previous and the next point, as (eq. 3):

$$\alpha_i = \tan^{-1} \left(\frac{E_{i+1} - E_{i-1}}{N_{i+1} - N_{i-1}} \right) \quad (3)$$

where (E, N) are planar coordinates.

3.2 System calibration

The camera and the receiver are mounted in the vehicle. With a total station close to the vehicle, several points are measured, according to figure 3: two points in the vehicle axis (A and B), the GNSS antenna, the camera and a set of points in front of the camera. From these measurements it will be possible to calculate relative positions, in a vehicle reference frame, of the GNSS antenna and the camera (distance b and angle θ). These are used, together with the vehicle azimuth, to transport precise coordinates from the GNSS antenna to the camera, for all instants in which position fix was achieved.

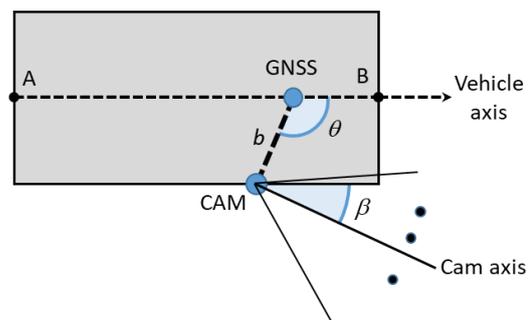


Figure 3. Planar scheme of the vehicle with the GNSS antenna and the camera (CAM), camera axis and surveyed points

The points surveyed in front of the camera, also seen on a video collected by the camera, are used to estimate a point in the centre line of the camera, and so calculate angle β , between the vehicle axis and the trajectory. In the present case the value obtained for β was 26.1° . It will be used to calculate the azimuth of the camera axis for any image.

3.3 Interpolation of positions for video frames

Points were collected by the Trimble receiver at 1 Hz, which were then transported to the camera. Figure 4 shows an example, with the GNSS antenna as black squares and the camera positions, also at 1Hz, as green circles. At a speed of 10 to 15 m/s the point density is relatively small, and a linear interpolation is not appropriate to model the trajectory, especially in curves. For this reason cubic interpolation was used.

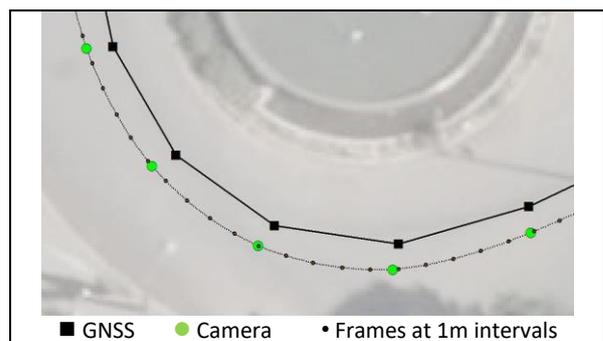


Figure 4. GNSS, camera positions transported from GNSS, and interpolated camera positions at 1 m intervals

All the 60 frames per second were interpolated (smooth trajectory in figure 4). Since they are in an excessive number they were filtered so that distance between consecutive positions is 1 meter (small circles in figure 4). This density provides enough overlap with an acceptable amount of final data. It also solves the problem of irregularity due to speed changes, or of the vehicle being stopped in traffic lights. For all these 1 m spaced images azimuth of the camera axis was calculated as the azimuth of the trajectory plus angle β .

3.4 Triangulation

As referred before, the system is intended to be applied for planar coordinate determination, assuming several conditions, such as the camera being always horizontal, with no tilts. In this condition, vertical objects will appear vertical on the images. For a given object, observed in image k , its x image coordinate, with respect to the image central line (figure 5a), can be transformed into an azimuth, according to equation (4):

$$\alpha_{object} = \alpha_k + \beta + \tan^{-1}\left(\frac{x_{object}}{f}\right) \quad (4)$$

where α_k is the camera axis azimuth of image k . The point can be observed in some other images, leading to a multiple intersection. However, for the sake of simplicity for manual measurement in the tests carried out, only two were considered, with interval of a few images. Figure 5 b shows the same signal in a later frame (9 frame interval). It can be seen that here the signal is not exactly vertical but the difference from top to bottom is small. The assumption of this MMS is that situations like this introduce negligible errors.

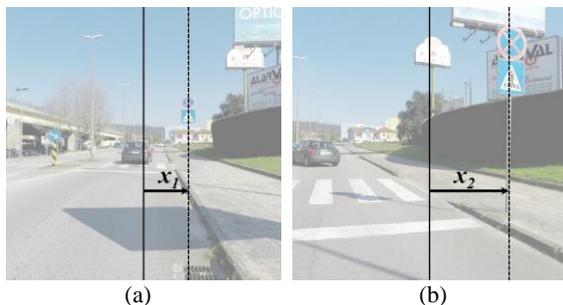


Figure 5. vertical object observed in two frames.
 Only x image coordinates are measured

Two lines (azimuths α_1 and α_2) are defined and the intersection point, P , is calculated (figure 6).

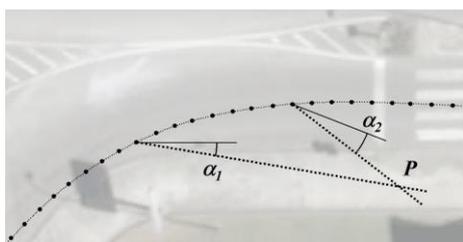


Figure 6. Planar intersection of lines defined for an object observed on two different frames.

4. POSITIONAL ACCURACY ASSESSMENT

A test was carried out in the city of Valongo, a suburban area of the city of Porto, in Portugal. A total of 4.8 km were travelled in an area of 1.5 km². Vector map data with accuracy standards of map scale 1:2000 was available, together with orthoimages. Figure 7 shows part of the area, with those GIS data and positions collected with the MMS.



Figure 7. Sample of the area showing orthoimages, vector data and locations acquired with the MMS (yellow dots).

The streets are relatively large and buildings not very tall. GNSS data, collected at 1 Hz, reached ambiguity fixes in 77% of the time, in RTK mode in the field. However, the post processing (PPK mode) allowed to improve the number of position fixes to 85% of the time. Frames were extracted and filtered in order that distance between consecutive projection centres is 1 m.

Among the available digital map data there was a layer with points corresponding to electric facilities, such as electricity poles. A total of 20 of these points, mainly along the sidewalks, were chosen as check points to assess the positional accuracy of the MMS. Each point was identified in 2 frames and its coordinates were calculated by the process described. Coordinates were compared with the ones present in the cartography, and the statistics of errors were calculated (Table 3). Notice that the check points, coming from the digital map data, also incorporate some error and so the errors may be overestimated.

Statistic	Easting (m)	Northing (m)
Minimum	-0.36	-0.71
Maximum	0.34	0.45
Average	-0.01	0.04
RMSE	0.24	0.37

Table 3. Statistics of the errors found in object coordinates obtained by intersection

The small values of the average might suggest that systematic errors are not present. That is in principle not the case because some of the points were along a road with constant direction and tended to have errors with similar trends. The fact that there are points along streets with different directions, resulted in some cancelation effect.

Anyway these results reveal that this MMS can provide data with an appropriate positional accuracy for map scale 1:2000, or better. In this way, this simple methodology may be applied to complete large scale map data produced by aerial photogrammetry but that require many object classes that cannot be collected from aerial data.

Potential reasons for systematic errors could be the inaccuracy of some of the parameters involved. Further investigations are needed in order to improve the quality of the MMS.

5. IMPROVEMENTS TO BE IMPLEMENTED IN THE SYSTEM

A better estimation of the parameters involved, such as the focal distance, or angle β might have some influence in the quality of the results. An error of 1° in the azimuth angle, which depends on the trajectory azimuth and angle β , introduces a planar error of 17 cm at 10 meter distance. Calibrations should be done with an accuracy better than 1° . A more careful camera calibration, and possibly the consideration of residual radial distortion, might also contribute.

The time synchronisation, based on the first frame time, was acceptable but also here some error may exist: moving at a speed of 15 m/s, an error of 1 frame, at 60 fps, implies a positional error of 25 cm. Methods of time calibration are being studied for further system improvement.

Another possible improvement would be the exploitation of data provided by other GoPro sensors, such as the gyroscope and the accelerometer. Attitude angles would help in the azimuth estimation and also allow for the inclusion of tilts. Anyway, accuracy of 1° or better in the attitude angles would be needed.

The main improvement will come from the inclusion in the frame processing of structure from motion (SfM). If successful it will allow for the determination of position and attitude angles of all the frames. This would allow for the completion of the points for which position fixing was not possible, filling the gaps in the trajectory.

Additionally, SfM would allow for the generation of point clouds. Even with a relatively small success, the extraction of point clouds for some of the objects of interest would be a great help in automating the data extraction. This step is possible and was already tested. Figure 8 shows a sample of a point cloud generated in this way, where some important objects can clearly be seen.



Figure 8. Point cloud generated from a sequence of video frames.

Additionally, if the MMS will become of interest to implement in a production environment it would benefit from the existence of a software tool for data treatment.

6. CONCLUSIONS

The developed MMS is based on the action camera GoPro Hero 5 Black. The system is intended to be very simple and in the simplest form would require only the camera itself, and relying only on the positions provided by the camera GPS. Normally the system will be used with a GNSS receiver. When comparing it with other systems the main difference is that here there are no cable connections. The synchronization of image frames and positions is made by the common GPS time frames.

The system allows for position determination with positional accuracy of around 30 cm, which is appropriate for many high accuracy applications. This MMS is under development and many of the suggested developments will be implemented in a near future.

ACKNOWLEDGEMENTS

Direção Geral do Território for the use of ReNEP, the Portuguese network of permanent GNSS stations.

Valongo City Council, for the provision of digital map data and orthoimages.

REFERENCES

- Al-Hamad, A., Moussa, A., and El-Sheimy, N. (2014). Video based mobile mapping system using smartphones. *Int. Arch. Photogram. Remote Sensing Spatial Info. Sci.*, Volume XL-1, 2014, p. 13-18.
- Artese, G (2007). ORTHOROAD: A low cost Mobile Mapping System for road mapping. In C. V. Tao, & J. Li (Eds.), *Advances in Mobile Mapping Technology - ISPRS Book Series* (pp. 31-41). London, UK: Taylor & Francis Group.
- Balletti, C., Guerra, F., Tsioukas, V. and Vernier, P., 2014. Calibration of Action Cameras for Photogrammetric Purposes. *Sensors*, 2014, 14, pp. 17471-17490.
- DGT, 2018. "Direção Geral do Território" Documentation about Portuguese national mapping standards and facilities for GNSS positioning. <http://www.dgterritorio.pt>
- Ellum, C., and El-Sheimy, N., 2002. Portable Mobile Mapping Systems. National Technical Meeting ION NTM. San Diego, California: The US Institute of Navigation.
- GoPro, 2018a. Hero5 Black User Manual. Available on line: https://gopro.com/content/dam/help/hero5-black/manuals/HERO5Black_UM_ENG_REVC_Web.pdf
- GoPro, 2018b. GPMF - GoPro Metadata Format. Available on-line. <https://github.com/gopro/gpmf-parser>
- Madeira, S., Gonçalves, J.A., Bastos, L., 2010. Photogrammetric mapping and measuring application using MATLAB. *Computers & Geosciences* 36(6), pp. 699-706
- Masiero, A., Fissore, F., Pirotti, F., Guarnieri, A. and Vettore, A. 2016. Toward the use of smartphones for mobile mapping, *Geospatial Information Science*, 19(3), pp. 210-221.