Satellite Markers: a simple method for ground truth car pose on stereo video

Gustavo Gil[1], Giovanni Savino[1,2], Simone Piantini[1], Marco Pierini[1]
1. Dipartimento di Ingegneria Industriale, Università degli Studi di Firenze, Firenze, Italy
2. Monash University Accident Research Centre, Monash University, Clayton, Victoria, Australia

ABSTRACT

Artificial prediction of future location of other cars in the context of advanced safety systems is a must. The remote estimation of car pose and particularly its heading angle is key to predict its future location. Stereo vision systems allow to get the 3D information of a scene. Ground truth in this specific context is associated with referential information about the depth, shape and orientation of the objects present in the traffic scene. Creating 3D ground truth is a measurement and data fusion task associated with the combination of different kinds of sensors. The novelty in this paper is the method to generate ground truth car pose only from video data. When the method is applied to stereo video, it also provides the extrinsic camera parameters for each camera at frame level which are key to quantify the performance of a stereo vision system when it is moving because the system is subjected to undesired vibrations and/or leaning. We developed a video post-processing technique which employs common camera calibration algorithms for the 3D ground truth generation. In our case study, we focus in accurate car heading angle estimation of a moving car under realistic imagery. As outcomes, our satellite marker method provides accurate car pose at frame level, and the instantaneous spatial orientation for each camera at frame level.

Keywords: heading angle, uncalibrated rigs, disparity map, preventive safety, ADAS, ARAS, PTW safety.

1. INTRODUCTION

Vehicular safety is experiencing a fast evolution due to the possibility for part of the vehicles to interpret the traffic situation. Advanced safety systems are often denominated as ADAS (Advanced Driver-Assistance Systems) and ARAS (Advanced Rider-Assistance Systems) in car and motorcycle industry, respectively. These systems have the potential to mitigate the consequences of a crash and in some cases even to avoid it. A single example is the recently proven effectiveness of AEB (Autonomous Emergency Backing) systems in cars [1]–[5]. In motorcycles this safety system is not commercially available but the simulation feasibility assessments are somehow encouraging [6].

Advanced safety systems benefit from artificial vision systems, which play an important role in the perception of the 3D environment. In particular, stereo vision have shown the feasibility for the remote estimation of heading angle of oncoming vehicles [7]–[14].

Aiming to assess how a stereo vision system performs in a specific measuring task (for example, remote heading angle estimation of a moving car), it is necessary to know the measured targets much more accurately than the capability of measurement of the system to assess. For example for heading angle estimation we need to know the "true" heading of the moving car. These kind of referential measures which are used for comparison purposes are known as a ground truth.

The development of autonomous cars have motivated the creation of 3D ground truth data of moving vehicles from custom made sensing platforms. In general terms, large amounts of data are collected from multiple sensors, which are mainly cameras, laser scanners, radars, and geo localization systems. Consequently, the big data need to be fused, facing a range of issues including synchronization between multiple types of measures, data harmonization of different sampling times, in order to obtain exploitable information to be used as ground truth. Good examples of these sensing platforms and types of ground truth are available to the research community [15]–[20].

For motorcycle safety researchers, the utilization of the aforementioned datasets are not suitable because they were acquired from four-wheeled vehicles. All types of sensing platforms developed over a car or van present quite different dynamics compared to a motorcycle. The tilting dynamics of single-track vehicles, such as bicycles, pedelecs, Powered Two Wheelers (which include speed pedelecs, mopeds, e-bikes, scooters, and motorcycles), and Narrow Track Tilting Vehicles (NTTVs), is not present in the datasets that are publically available.

The feasibility of instrumenting a motorcycle with the equipment of four-wheeler sensing platforms is not practical in

terms of power budget and space for all the sensors required. Another important constraint is to maintain the weight distribution in the vehicle in order to preserve a vehicle dynamics similar to the original one of the motorcycle. Our solution to the aforementioned problems is based in stereo vision. In fact, these systems are compact, lightweight and can be implemented in embedded computers with modest power consumption, making them suitable for motorcycle safety.

This motivated the development of this new technique for ground truth generation (only using a stereo vision system), which is an enabler for its utilization on a tilting dynamics vehicle. In our case study, the satellite marker method allows to use in post processing the same stereo video data acquired for the perception system mounted on the motorcycle to generate the ground truth. Another novelty of the satellite marker method is the fact to provide extrinsic information for each camera involved. This allows the quantification of the instantaneous decalibration for each stereo frame acquired in our setup.

The present paper is structured as follows: Section 2 describes the material employed for the development of the research activity. Section 3 explains how is calculated the ground truth. Section 4 presents a case study and the results of the experiments performed with this technique. Section 5 describes the validation of the proposed approach using accurate geo-localized control points. To conclude, the future work and conclusions are presented in Section 6.

2. MATERIALS

A stereo vision system installed in a test scooter, a satellite marker installed on a target car, and the software tool for stereo camera calibration. In this publication we only present the results given for one stereo pair (camera 1-2) of our multifocal stereo rig. For validation purposes we used a RTK (Real Time Kinematic) satellite-based positioning system equipping both the scooter and the target car.

Figure 1. Detail of the cameras used of the multi-focal stereo rig mounted in the front of the PTW sensing platform (foreground). The target car and its satellite marker installed in the frontal view setup is showed behind.

Camera 1 Intrinsics		
Focal length [pix]	1715.2829 +/- 3.5194	1719.6152 +/- 3.5272
Principal point [pix]	923.7204 +/- 3.2798	535.7389 +/- 2.9717
Radial distortion [rad]	-0.2621 +/- 0.0019	0.0772 +/- 0.0040
Camera 2 Intrinsics		
Focal length [pix]	1712.5670 +/- 3.5114	1717.7884 +/- 3.5272
Principal point [pix]	950.4462 +/- 3.1923	512.4862 +/- 2.7443
Radial distortion [rad]	-0.2652 +/- 0.0019	0.0875 +/- 0.0038
Position And Orientation of Camera 2 Relative to Camera 1		
Rotation of camera 2 [rad]		
-0.0106 +/- 0.0005	-0.0056 +/- 0.0010	0.0006 +/- 0.0001
Translation of camera 2 [cm]		
-26.4867 +/- 0.0154	0.1979 +/- 0.0148	0.2166 +/- 0.0719

Table 1. Main characteristics of the imaging system corresponding to the stereo pair formed by camera1-2. The projected baseline of 26.5cm is consistent with the extrinsic value obtained in the static calibration (26.4862±0.0154)cm.

2.1 Stereo vision system

The custom built imagining system was constituted of two stereo rigs which used inexpensive rolling-shutter type camera sensors and pairs of fixed lenses. The stereo rig constituted by camera1-2 was a single stereo pair composed by two action cameras GoPro Hero Black configured in narrow field of view (FoV). The calibration parameters used for this stereo rig are provided in Table 1. The tri-focal stereo rig [13] present similar results according with their focal length and FoV but for simplicity they are not presented in this publication. All cameras recorded video at 30 frames per second (fps) with a resolution of 1920x1080 pixels. In the proposed setup the information of the imagining system was post-processed.

2.2 Satellite marker

The marker chosen was an asymmetric checkerboard pattern. This type of patterns are often used to estimate the intrinsic and extrinsic parameters of cameras with common camera calibration software. We called it 'satellite' marker because it was fixed on top of the target vehicle, thus being practically moving on a "stationary orbit" around the target.

Our satellite marker was built from a planar wood panel of 95x120cm. Black and white squares with sides of 22cm painted on the board constituted a checkboard pattern of 4 by 5 full squares (Figure 1). The checkerboard pattern was extended to the edges of the wood panel (detail in Figure 2) to define symmetric corner features for all the corners belonging to the 4 by 5 full squares pattern. In this way, algorithms for automatic corner localization [21], [22] could be used during the ground truth generation.

2.3 Supporting frame

Due to the fact that the satellite marker selected presents a planar geometry, this marker will be limited to a maximum heading angle change inferior to $\pm\pi$. Thus, a reorientation of the satellite marker is needed to cover the needs for remote car pose estimation. We adopted a rectangular parallelepiped structure (Figure 2) which holds the satellite marker, fixing it to the target car. The orthogonal facets of the structure are used to place the satellite marker in different orientations. In the present experiment the satellite marker was fixed to the car in one of the sides.

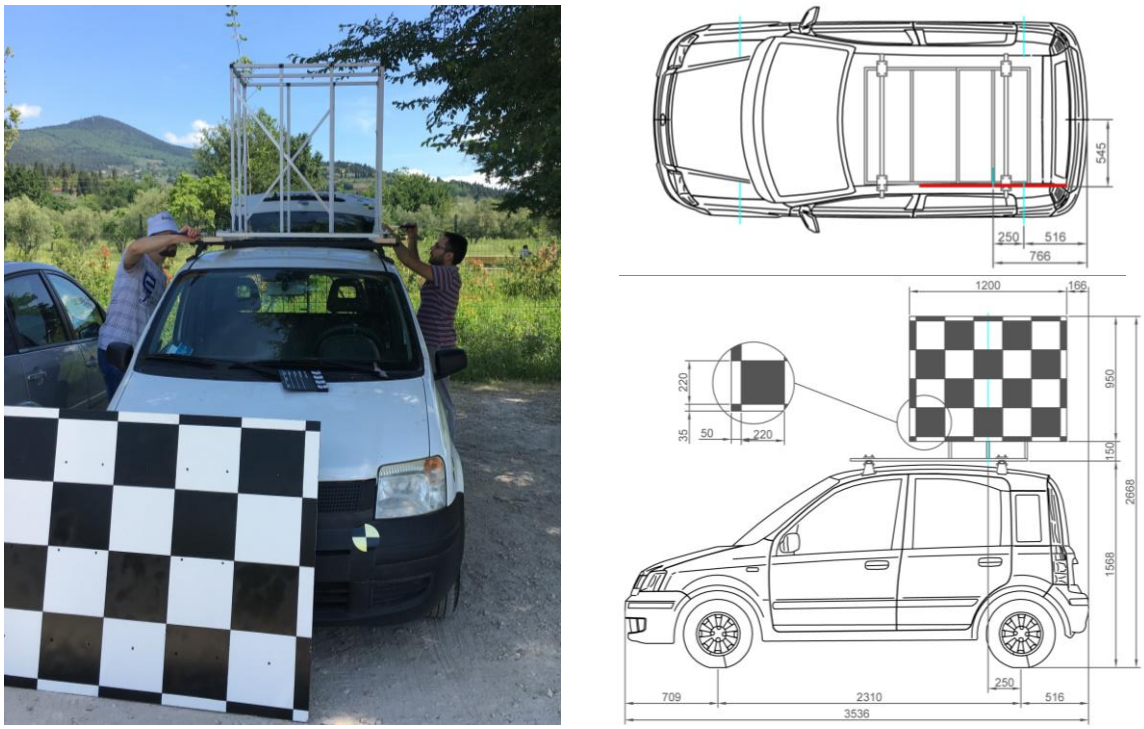

Figure 2. Overview of supporting structure and detailed measurements of the satellite marker and its alignment with respect to the target car. Only the lateral setup (satellite marker in one of the sides) was used in this experiment.

2.4 Stereo camera calibration tool

The camera calibration tool is part of the Matlab Computer Vision System Toolbox. The calibration algorithm [23] uses the pinhole camera model [24] and the lens distortion calculation [25]. The calibration tool is also present in OpenCV.

2.5 RTK system or D-GPS (Differential GPS)

A highly accurate Real Time Kinematic (RTK) satellite navigation system was employed to validate the ground truth generated from the stereo video with the satellite marker method. The D-GPS units (GeoMax Zenith 20) provided the locations of the PTW and the moving target at 20Hz and an accuracy of ±2cm over the ground plane.

3. GROUND TRUTH GENERATION: THE SATELLITE MARKER METHOD

Our proposed strategy is to measure the pose of the target car via measuring the pose of a satellite marker rigidly connected to the target itself. In our application case, the marker was placed on top of the target (Figure 2) in order not to occlude or change in any way the aspect of the target vehicle.

The calculation of the ground truth heading from the stereo videos was post-processed analyzing only the orientation of the marker in the tridimensional space. All the information contained in the stereo video corresponding to the moving target was neglected. The algorithm is a double-step method that first uses the results of the Direct Linear Transformation (DLT) based on the pinhole camera model [24] to initialize the Levenberg-Marquardt optimization method [25].

In order to obtain the desired ground truth with the satellite marker method, first is necessary to create an 'etalon' for the camera system. An etalon for our imaging system as set of stereo frames acquired in static conditions (the camera rig is not moving) which contains enough information to direct the optimization process to converge to very similar intrinsic camera parameters, even if other stereo pairs are added in the calibration process. Therefore calibrating the stereo camera system. Regarding the accuracy, the measurements obtained were controlled maintaining the maximum reprojection error of the etalon set below 0.75 pixel.

Once the etalon was defined, it was set as input of the stereo camera calibration tool together with the stereo frames of the video sequence to be analyzed. For example, our application case consisted of 60 stereo frames for which the first 23 pairs belonging to the etalon.

To obtain the ground truth, it is necessary to remove the extrinsic data corresponding to the etalon set from the obtained results. As a consequence, ground truth files complementing the video containing the pose of the target to measure for each video frame can be created [26]. The information obtained as a ground truth consists in three rotational and three translational values of the satellite marker. As a convention, the 6DoF measured are referred to the optical center of the left camera of the stereo pair.

Regarding to the heading angle reference, the rotational value around Y-axis correspond to the heading angle of the target car. Therefore, once the satellite marker position is defined, obtaining the ground truth heading is straight forward from the stereo camera calibration tool.

4. EXPERIMENTS

Our case study is a single moving vehicle that perform a set of three manoeuvers in front of our imagining system. The experiments were performed in an outdoor car parking in daylight conditions. This setup allowed us to obtain realistic imagery while performing a series of simple and complex maneuvers in a controlled environment.

The stereo system, in the frontal part of the PTW sensing platform (Figure 1), remained static and in upright position during the tests. The manoeuvers performed by the test car are illustrated in Figure 3.

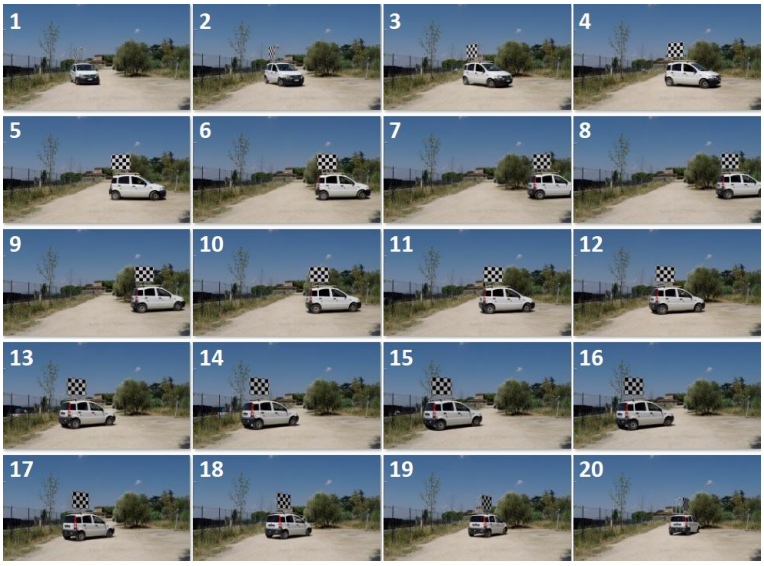

Figure 3. Succession of 20 representative frames to describe the maneuvers for which the heading angle ground truth was obtained. The numbers represent the temporal order of the sequence during the maneuvers.

First, the target car was approaching the PTW from opposite direction and turned in front of the PTW's path. Second, the car stopped and reversed passing again in front of the PTW. Finally, after stopping again, the car moved forward merging in front of the PTW's longitudinal axis.

The experiment was conducted at noon with the sun high in the sky, and the satellite marker was placed in the lateral position (Figure 2) to guarantee a correct sight of view during the execution of the maneuvers. The car driver was instructed to perform the maneuvers with both slow and quick dynamics (not exceeding 45km/h).

As a results of the stereo video sequences acquired, the disparity maps were computed thus generating a 3D reconstruction of the scene (Figure 4). This step allowed us to verify the integrity of the images acquired.

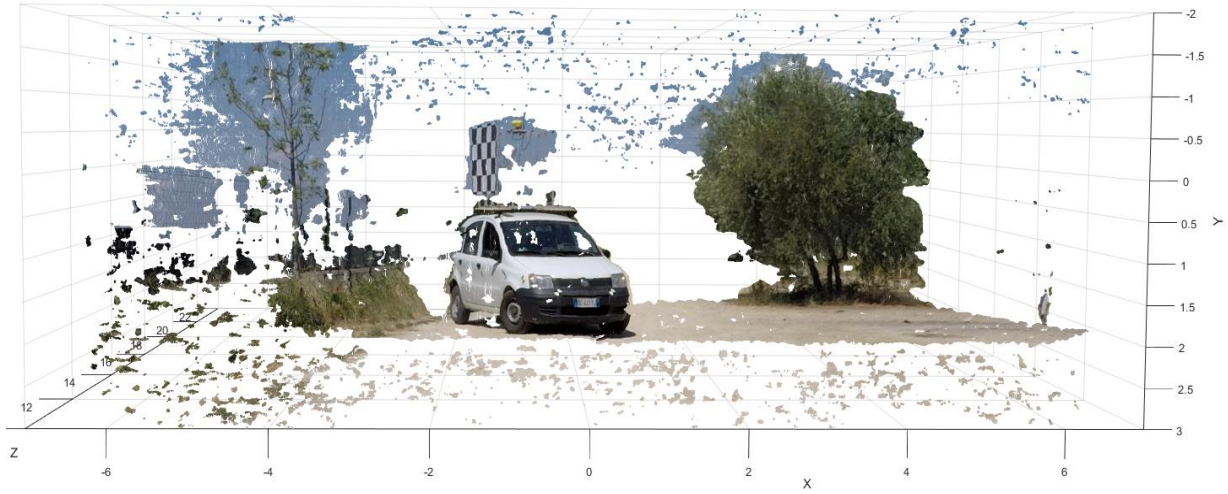

Figure 4. Example of the 3D point cloud generated from the disparity map (all the units are in meters).

Regarding the calculation of the ground truth, the same video sequences were post-processed following the satellite marker method. The raw data from the camera calibration tool is showed in Figure 5, in which two clusters of data sets can be identified: the etalon and the desired ground truth.

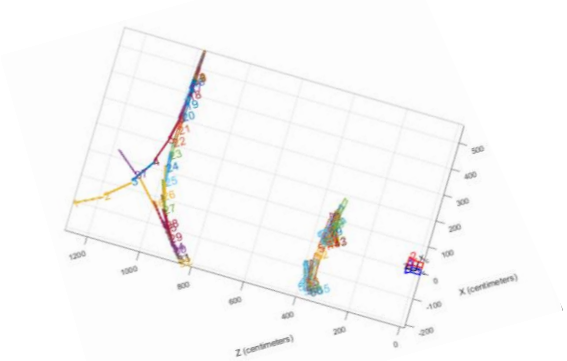

Figure 5. Results obtained from the stereo camera calibration tool. About 300-400cm (3-4m) from the cameras is present the 6DoF information (location and pose) corresponding to the etalon set. Between the 800-1300cm (8-13m) away from the cameras is present the 6DoF information of the satellite marker set.

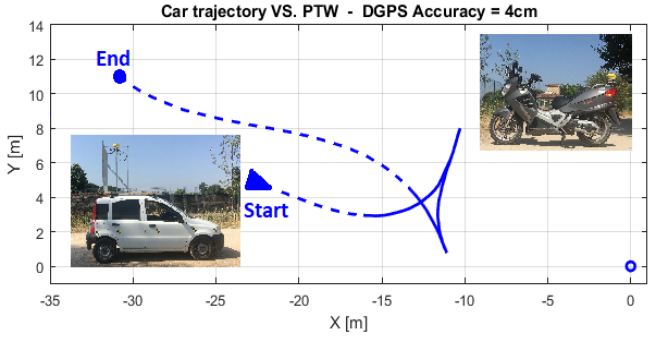

Figure 6. 2D trajectory of the target car during the experiment. The measure indicates the relative position of the two yellow antennas installed in the vehicles. One antenna is positioned in the rear part of the PTW (coordinate origin) and the second antenna on the satellite marker frame.

5. GROUND TRUTH VALIDATION

A representation of the accurate geo-localization of the moving car during our experiment is showed in Figure 6. The dashed trajectory corresponds to the movements of the vehicle where the satellite marker was not visible from the imaging system due to its lateral placement (Figure 4). The solid line represents the vehicle trajectory while the marker was in the line of sight of the stereo camera. Ground truth was generated in all these locations.

The difference between Figure 5 and Figure 6 is consistent with the location of the antennas in both vehicles. In particular, the main contribution to this offset is the distance between the stereo rigs in the front of the PTW with respect to its D-GPS antenna in the back of the PTW (186cm). The second contributor is the distance between the central point of the surface of the marker to the D-GPS antenna (offset of 37cm), which is centered on the top part of the supporting frame.

5.1 Quantification of the error in the ground truth

The ground truth obtained from the stereo video employing the method of the satellite marker provided 6DoF information (pose of the marker). 3DoF corresponds to the translation of the marker in the space, and 3DoF corresponds to the rotation. Our interest corresponds only to the Y-axis rotational component, however the D-GPS system do not provides heading to compare. The D-GPS antennas provides Cartesian coordinates over the ground plane with an error of ±2cm. Consequently, we employ other two components of the ground truth generated from the imaging system to obtain a fair comparison metric which is depicted in Figure 7 .

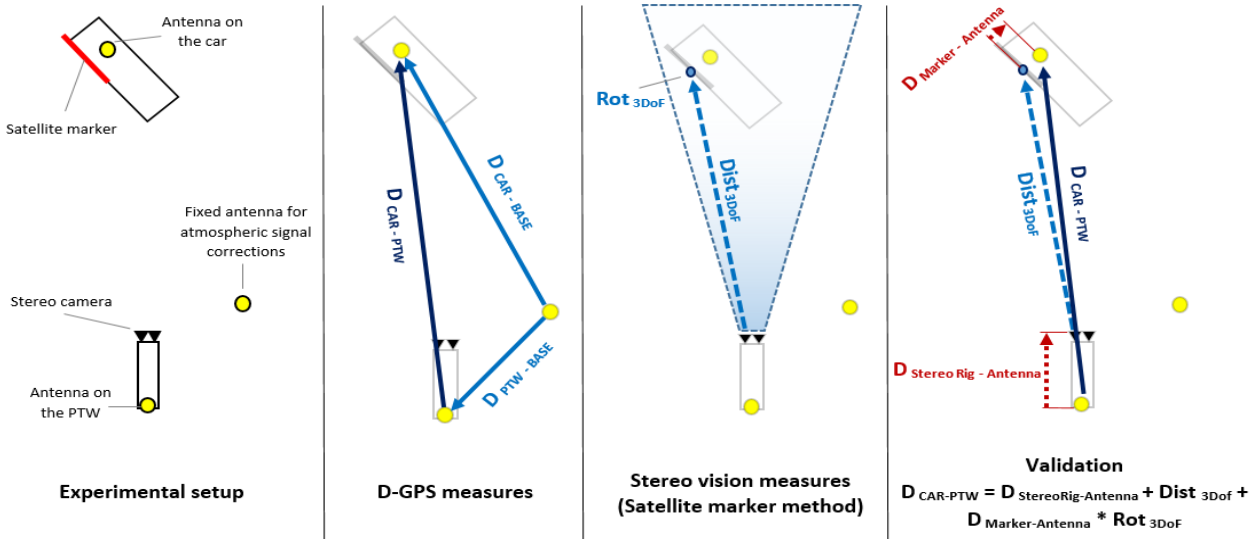

Figure 7. From left to write is illustrated in vectorial form the measurement performed (D-GPS) and the equivalent measure synthesized from three components of the ground truth (imaging system).

Bearing these considerations in mind, we overlaid on the X-Z plane (delivered for the D-GPS) the locations calculated from the ground truth in order to assess it (Figure 8).

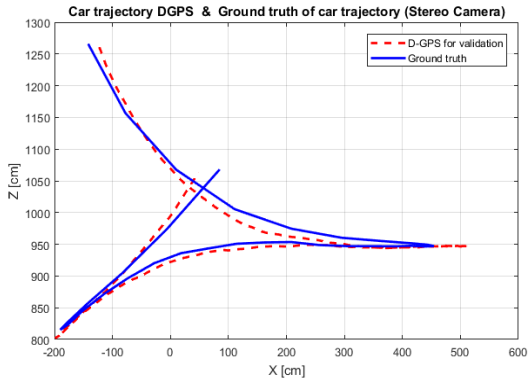

Figure 8. Comparison of the Ground Truth location generated with the satellite marker method and a more accurate reference provided for the D-GPS system.

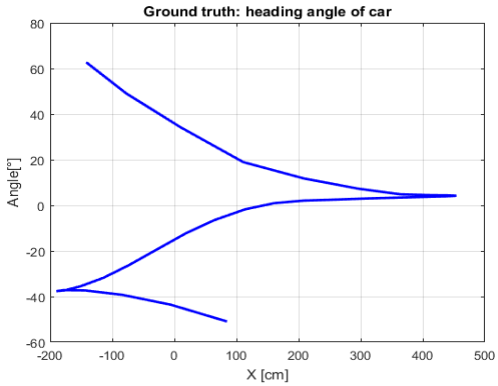

Figure 9. Ground truth heading (satellite marker only). This plot present the variation of the heading angle of the satellite marker which is the same of the car.

The information to the location present a deviation up to 18cm in the section analyzed (Figure 8). The biggest errors appear when the marker have big heading angles (62° and -50°) whit respect the stereo camera in the PTW (Figure 9).

During the analysis of the results of our test, we notice that these pair of cameras of our imagining system tends to overestimate in 3.75% the depth range. For example, traffic cones of 30cm of height aligned with the center of the stereo rig and located at 20m away from it indicated a depth distance of 21.75m in the 3D point cloud reconstruction obtained from the calculation of the disparity map. This is also an error contributor to the heading angle measurement.

6. CONCLUSION AND FUTURE WORK

Our work presents a new way to apply well-known tools which are available in the research community to solve practical problems, adding value in particular applications where ground truth is needed. The ground truth generation from a single vision system is useful to deal with the new challenges that a tilting dynamics vehicles as Powered Two Wheelers (PTWs) and Narrow Track Tilting Vehicles (NTTVs) present.

In our case study we acquired video from a stereo vision system. However, the satellite marker method can be used to measure the 6DoF of one or several satellite markers and also with a single video camera (a monocular camera) or a multi camera system (e.g. 1, 2, 3, 4, 5 … cameras).

The errors that we report in the generation of the ground truth are acceptable for the application in traffic safety. In addition, one advantage of our method is its applicability in urban environments with poor or absence of GPS service due to the buildings and trees along the streets.

For an evaluation of the accuracy employing D-GPS as we did, we recommend to select places whit no presence of tall buildings or trees in order to do not degrade the accuracy of the measurements. Furthermore, the D-GPS systems only provide location of the moving targets, so is not possible to assess directly the 6DoF generated as a ground truth for our method, however the accurate location can be used to obtain control points.

In general terms, the error in the ground truth increases when the satellite marker is oriented with a heading angle larger than ±40°, this effect could be due to the recognition challenge for the corner feature detection algorithms.

Possible improvements of the method may include:

1. A post processing of the ground truth heading analyzing consecutive frames to relate the heading angle to a kinematic car model.
2. A filtering and post processing of the D-GPS location in order to obtain from its derivative a heading vector of the moving target.

The first improvement can be achieved with a simple kinematic model of the target car. The drawback is that one critical parameter of such model, i.e. the wheelbase of the vehicle, is unknown a priori in the typical application of traffic scenario.

The second improvement can apply to non-static conditions. As it can be seen in some example raw data presented in Figure 10, for the static analysis of the geo-localization technique, fictitious and non-negligible jumps in the position appear along the time when the vehicle is not moving. This behavior is expected due to the D-GPS correction process, however such discontinuities make it impossible to get vehicle heading by computing a derivate of the position signal. In dynamic conditions (right charts of the geo-localization data in Figure 10), the use of the derivative of the position of the vehicle appears feasible, with some remarks though. For example, we reported glitches in the D-GPS corrected signals with position signal amplitudes up to 50cm. The source of the glitches requires further investigation.

In conclusion, advanced safety systems (ADAS & ARAS) rely in image understanding and our method is a simple way to enable the “in-house” elaboration of vision datasets with 3D ground truth. Our method quantifies dynamically the orientation for each camera used in the imaging system (stereo rig or multi camera rigs) at frame level, allowing to address the problem of the impossibility to maintain the calibration of the camera rig under the dynamic conditions (mini-bending of the rig structure) imposed for the vibrations in moving platforms. Thus, the satellite marker method is suitable for the evaluation of obstacle detection and trajectory collision prediction systems able to operate in moving tilting platforms.

ACKNOWLEDGEMENTS

This work has been funded for European Community’s Seventh Framework Program through the international consortium called MOTORIST (Motorcycle Rider Integrated Safety) agreement no. 608092.

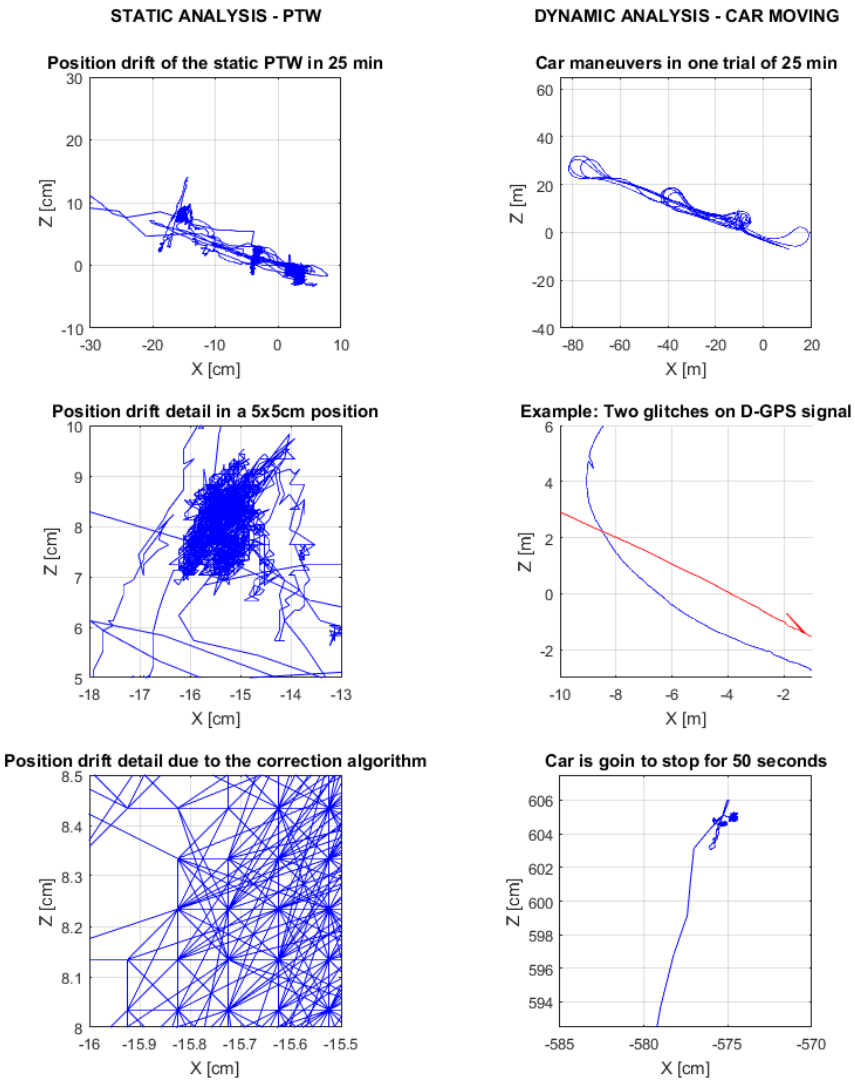

Figure 10. Raw data examples of the trajectories measured with the D-GPS. Static and dynamic data.

REFERENCES

[1] J. B. Cicchino, "Effectiveness of forward collision warning and autonomous emergency braking systems in reducing front-to-rear crash rates," *Accid. Anal. Prev.*, vol. 99, pp. 142–152, Feb. 2017.

[2] B. Fildes *et al.*, "Effectiveness of low speed autonomous emergency braking in real-world rear-end crashes," *Accid. Anal. Prev.*, vol. 81, pp. 24–29, Aug. 2015.

[3] I. Isaksson-Hellman and M. Lindman, "Evaluation of rear-end collision avoidance technologies based on real world crash data," *Proc. Future Act. Saf. Technol. Zero Traffic Accid. FASTzero*, pp. 471–476, 2015.

[4] I. Isaksson-Hellman and M. Lindman, "Evaluation of the crash mitigation effect of low-speed automated emergency braking systems based on insurance claims data," *Traffic Inj. Prev.*, vol. 17, no. sup1, pp. 42–47, Sep. 2016.

[5] M. Kyriakidis, C. van de Weijer, B. van Arem, and R. Happee, "The deployment of advanced driver assistance systems in Europe," 2015.

[6] G. Savino, J. Mackenzie, T. Allen, M. Baldock, J. Brown, and M. Fitzharris, “A robust estimation of the effects of motorcycle autonomous emergency braking (MAEB) based on in-depth crashes in Australia,” *Traffic Inj. Prev.*, vol. 17, no. sup1, pp. 66–72, Sep. 2016.
[7] D. Pfeiffer and U. Franke, “Modeling Dynamic 3D Environments by Means of The Stixel World,” *IEEE Intell. Transp. Syst. Mag.*, vol. 3, no. 3, pp. 24–36, 2011.
[8] M. Coenen, F. Rottensteiner, and C. Heipke, “DETECTION AND 3D MODELLING OF VEHICLES FROM TERRESTRIAL STEREO IMAGE PAIRS,” *ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. XLII-1/W1, pp. 505–512, May 2017.
[9] A. Barth, D. Pfeiffer, and U. Franke, “Vehicle tracking at urban intersections using dense stereo,” in *3rd Workshop on Behaviour Monitoring and Interpretation, BMI*, 2009, pp. 47–58.
[10] B. Barrois and C. Wöhler, “3D pose estimation of vehicles using stereo camera,” in *Encyclopedia of Sustainability Science and Technology*, Springer, 2012, pp. 10589–10612.
[11] U. Franke, C. Rabe, S. Gehrig, H. Badino, and A. Barth, “Dynamic stereo vision for intersection assistance,” in *FISITA 2008 World Automotive Congress, Munich, Germany*, 2008.
[12] F. Engelmann, J. Stückler, and B. Leibe, “Joint object pose estimation and shape reconstruction in urban street scenes using 3D shape priors,” in *German Conference on Pattern Recognition*, 2016, pp. 219–230.
[13] G. Savino, S. Piantini, G. Gil, and M. Pierini, “Obstacle detection test in real-word traffic contexts for the purposes of motorcycle autonomous emergency braking (MAEB),” in *25th International Technical Conference on the Enhanced Safety of Vehicles (2017)*, Detroit, USA, 2017.
[14] A. Barth and U. Franke, “Where will the oncoming vehicle be the next second?,” in *Intelligent Vehicles Symposium, 2008 IEEE*, 2008, pp. 1068–1073.
[15] G. Pandey, J. R. McBride, and R. M. Eustice, “Ford Campus vision and lidar data set,” *Int. J. Robot. Res.*, vol. 30, no. 13, pp. 1543–1552, Nov. 2011.
[16] A. Geiger, P. Lenz, and R. Urtasun, “Are we ready for autonomous driving? the kitti vision benchmark suite,” in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 3354–3361.
[17] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, “Vision meets robotics: The KITTI dataset,” *Int. J. Robot. Res.*, vol. 32, no. 11, pp. 1231–1237, 2013.
[18] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, “1 year, 1000 km: The Oxford RobotCar dataset.,” *IJ Robot. Res*, vol. 36, no. 1, pp. 3–15, 2017.
[19] M. Cordts *et al.*, “The cityscapes dataset for semantic urban scene understanding,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3213–3223.
[20] M. Cordts *et al.*, “The cityscapes dataset,” in *CVPR Workshop on the Future of Datasets in Vision*, 2015, vol. 1, p. 3.
[21] C. Harris and M. Stephens, “A combined corner and edge detector.,” in *Alvey vision conference*, 1988, vol. 15, pp. 10–5244.
[22] J. Shi and C. Tomasi, “Good Features to Track.” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Jun-1994.
[23] J.-Y. Bouguet, *“Camera Calibration Toolbox for Matlab.” Computational Vision at the California Institute of Technology*. .
[24] Z. Zhang, “A flexible new technique for camera calibration,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, 2000.
[25] J. Heikkila and O. Silven, “A four-step camera calibration procedure with implicit image correction,” in *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, 1997, pp. 1106–1112.
[26] G. Gil, G. Savino, and M. Pierini, “First stereo video dataset with ground truth for remote car pose estimation using satellite markers,” presented at the The 10th International Conference on Machine Vision (ICMV 2017), Vienna, Austria, November 13-15.