First stereo video dataset with ground truth for remote car pose estimation using satellite markers

Gustavo Gil[1], Giovanni Savino[1,2], 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

Leading causes of PTW (Powered Two-Wheeler) crashes and near misses in urban areas are on the part of a failure or delayed prediction of the changing trajectories of other vehicles. Regrettably, misperception from both car drivers and motorcycle riders results in fatal or serious consequences for riders. Intelligent vehicles could provide early warning about possible collisions, helping to avoid the crash. There is evidence that stereo cameras can be used for estimating the heading angle of other vehicles, which is key to anticipate their imminent location, but there is limited heading ground truth data available in the public domain. Consequently, we employed a marker-based technique for creating ground truth of car pose and create a dataset* for computer vision benchmarking purposes. This dataset of a moving vehicle collected from a static mounted stereo camera is a simplification of a complex and dynamic reality, which serves as a test bed for car pose estimation algorithms. The dataset contains the accurate pose of the moving obstacle, and realistic imagery including texture-less and non-lambertian surfaces (e.g. reflectance and transparency).

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

1. INTRODUCTION

The usage of PTWs (Powered Two-Wheeler) is growing globally as a means of personal transportation in cities. For example, during the next 15 minutes at least two people will suffer a motorcycle crash in Italy [1]. In the non-fatal cases, the people deal with material losses, medical expenses, rehabilitation time, and consequences for life. PTW crashes at intersections are the more numerous in urban traffic [1] because are the situations most complex to face. The correct interpretation of the heading angle of other vehicle is vital for predict its future location. Thus, a system able to perform a real-time remote estimation of the heading angle of other vehicles is an enabler for ARAS (Advanced Rider-Assistance Systems) in PTW industry and novel ADAS (Advanced Driver-Assistance Systems) in car industry.

Stereo vision systems have showed the potential for heading angle estimation in real traffic and also from mobile platforms or vehicles with a non-leaning dynamics such as cars [2]–[8]. Recently, the feasibility of stereo vision for heading angle estimation was demonstrated using a scooter [9], enabling new promising ARAS for PTW application. Aware of the scarce experimental data in the public domain to address this topic, we designed the present dataset aiming to quantify the accuracy of real-time algorithms able to remotely estimating the pose of a car, with a particular focus in its heading angle. The knowledge of how accurate and robust can be the heading angle estimation of a car moving ahead of the PTW is critical for trajectory prediction. This information can be used for trigger a variety of PTW safety systems such as: wearable air bag [10]–[12], M-AEB (Motorcycle Autonomous Emergency Braking) [13], [14], autonomous emergency steering [15], and inter-vehicle communication systems[16], [17].

This dataset contains a selection of representative urban traffic situations that can lead to PTW crashes in urban environment. The selection of dangerous scenarios for PTW users was based in a prior study that analyzed more than one million of PTW crashes from 2000-2012 [1]. The study presents in its appendix a set of 26 pictograms representing the PTW crashes which covering more than 90% of the total PTW crash configurations including both rural and urban areas. As the dataset contains a reference of the obstacle heading for each video stereo frame, this information can be used as a ground-truth to quantify algorithms' accuracy, speed, and robustness.

This work is organized as follows: Section two present the marker-based method used for the creation of heading ground truth employing a stereo camera as single sensor. Section three contains a detailed description of how the dataset

* *Dataset website: https://github.com/GusRep/HeadingAngleDataset/wiki*

was done and how was organized the content of its files in order to be used for algorithm benchmarking. Section four presents some results of the heading estimation provided by the methodology employed. To conclude, section five highlight the advantages of the method employed for the generation of the dataset and point out activities needed to contribute in the development of advanced PTW safety systems.

2. GROUND TRUTH GENERATION USING THE SATELLITE MARKER METHOD

To produce stereo video data containing ground truth information about the pose of the moving obstacle (or target) we adopted a technique of marker-based pose estimation. The main advantage of a marker-based strategy yields in the fact that no other sensors are used to obtain the ground truth information of the tridimensional scene, simplifying the sensing setup and data synchronization. However common marker-based techniques requires active IR (infra-red) illumination (e.g., body tracking [18]) or deals with the need to partially occlude the targets to sense during the measurement [19]–[21], modifying its morphological features.

In the view of overcome the aforementioned limitations, we used the concept of satellite marker introduced in [22]. The marker selected was a checkerboard placed above the car (Figure 1) in order to preserve the line of sight between the entire vehicle and the cameras, it means not occluding/changing in any way the appearance nor volume of the target car.

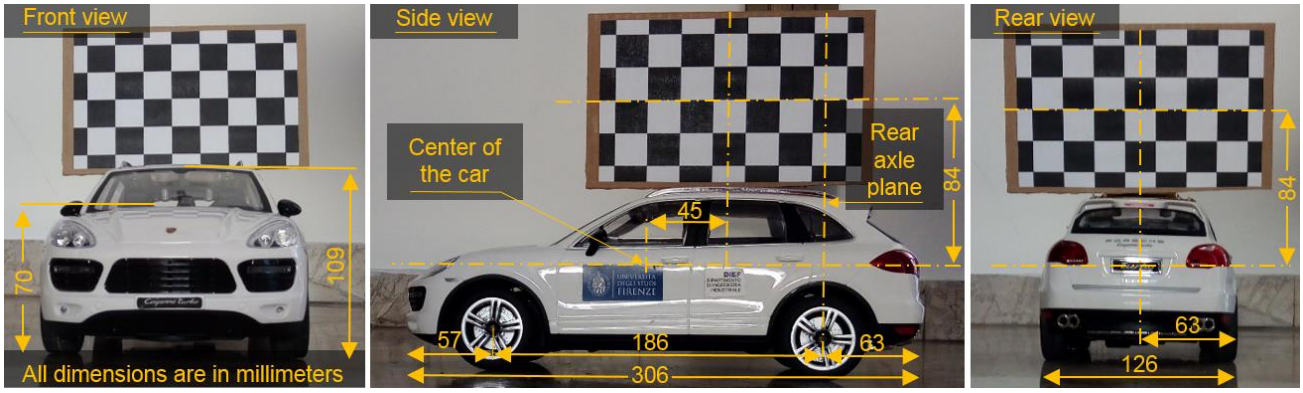

Figure 1. Measures of the target car whit the satellite marker over its roof in three different orientations. In the side view setup the marker is placed parallel to the transversal axis of the car, while in the front and rear view setups, the marker is placed over the rear axle plane but in different orientations.

The marker was placed in two different positions in order to be visible from the cameras viewpoint for each specific maneuver. One position over the rear axle of the car and the second one parallel to the longitudinal axis of the vehicle. Therefore, the translational and rotational measurement of the marker (6DoF – Degrees of Freedom) are correlated to the spatial position and orientation of the target car (i.e., the car toy).

3. PRESENTATION OF THE DATASET

This dataset consist of 32 short and rectified stereo video sequences of a single vehicle moving in front of a stereo camera and 29 video sequences of a moving box. The type of kinematic performed for the vehicle are similar to real pre-crash conditions and/or to crash precipitating events. The dataset provide 6DoF ground truth information about the location (3DoF) and orientation (3DoF) of the moving vehicle for each frame of the stereo video sequences.

The dataset can be download from the website: https://github.com/GusRep/HeadingAngleDataset/wiki

3.1 Materials

The stereo camera used was the DUO3D MXL which presents the following specifications: global shutter sensors; monochrome sensors; maximal FoV 170°; baseline 30.0mm; light spectrum of acquisition: visible + IR; M8 lenses, and 2.0-2.1 focal length. The settings used in the camera sensor during the stereo video acquisition were the following: VGA resolution (640x480); 1x1 binning; 40 stereo frames per second; gain 60%; exposure 46%; and lightening 22%.

The 1:16 scaled model of the car is a radio control toy that was selected bearing in mind to present: a common car kinematics (Ackermann steering geometry); a geometry of a real vehicle (Porsche Cayenne Turbo); and partial texture-less, transparent and reflective surfaces (non-lambertian surfaces).

Figure 2. The five elements used for the generation of the dataset: a stereo camera, a scaled model of a real car, a moving box sized as the scaled car, a scaled model of an intersection of streets, and two checkerboards.

The moving box intends to be a cuboid moving target. It is constituted for the aforementioned toy which was covered by a box of flat facets. The dimension of the moving box is 370x41x10mm. As a remark, the scaled model of an intersection of streets delimited with tape was only designed as a visual aid to conduct the experiments.

The markers are checkerboards constituted for black and white squares sized 20mm with symmetric corner features as defined in [22]. One smaller marker was used in tests performed with the box for practical reasons. The dimensions of the two markers were 7x4 and 5x4 full squares for the toy and box test respectively.

3.2 Generation of the referential heading angle (ground truth)

The ground truth for the stereo videos was generated with the satellite marker method that showed to be accurate for the car pose application (employing a real car) and simple to implement [22]. Regarding the accuracy, the measurements obtained were controlled it maintaining the maximum reprojection error below 0.3 pixel. We generated files containing three rotational and three translational values of the satellite marker, referred to the optical center of the left camera of the imagining system. Therefore the rotational value (around Y axis) corresponding to the heading angle can be directly used because the heading of the marker is equal to the heading of the moving target. For the translational values, as the satellite marker was placed in 3 different ways, they need to be compensated with the following spatial relations (X, Y, Z) with respect to the optical axis of the left camera: fontal view of the car (0, 138.5, 243)mm, lateral view of the car (45, 138.5, 66)mm, and rear view of the car (0, 138.5, 63)mm.

3.3 Utilization of the dataset to test algorithms

Each maneuver trial was recorded simultaneously from two synchronized cameras. We denominate to each couple of rectified videos as a stereo video frame (1280x640 pixels) as is showed in Figure 3. Therefore the videos of each single camera have VGA resolution (640x480 pixels). Additionally, each stereo video sequence was processed to generate the ground-truth corresponding to the location and orientation of the moving target. The ground truth is presented in a couple of XLS files which are paired with the respective stereo video sequence, and also replicated in DAT files.

The ground-truth in the XLS files is organized in three columns (3DoF), for which the row corresponds to the frame number and each column corresponds to X, Y, Z axis for both files. One file define three linear translations [mm] of the satellite marker while the second one define three rotations [rad] around its axes. The same information is provided in a single DAT file containing the same kind of information but in two Matlab variables.

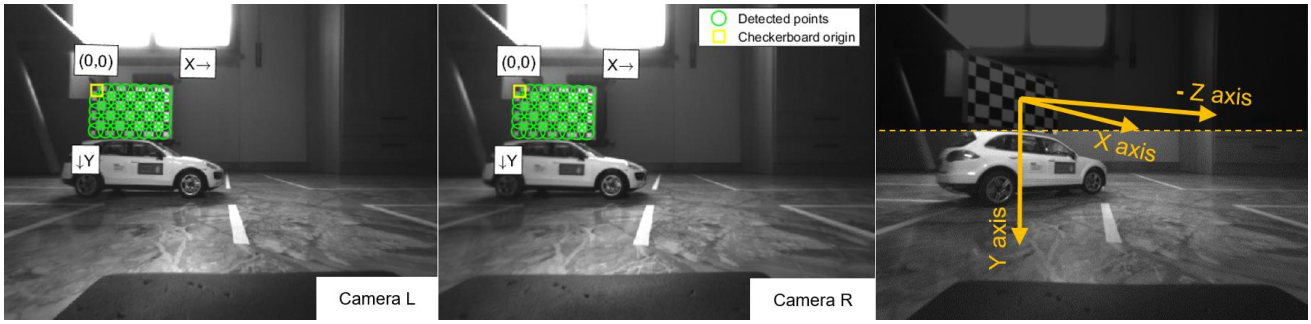

Figure 3. On the left hand a video stereo frame showing the detection of the satellite marker. On the right hand is presented whit a single video frame the two ROIs (Region of Interest) for the different analysis and the reference adopted for the rotational coordinate axes (3DoF).

For the purposes of remote heading estimation we suggest to use the rotational information around Y-axis, however all the information can be used for algorithmic benchmarking purposes. The metrics to calculate the effectivity of remote heading estimation is calculated at stereo frame level by subtracting the estimated value to the true value provided, as it is expressed in equation (1) for which the sub index *k* represents the number of the stereo frame. Other metrics may quantify the transient error response of the estimation about a short period by the integration of the instantaneous error over the time. For our dataset we defined two integration times such as 0.125s and 0.25s that cumulates the error estimation among the last 5 and the 10 frames respectively as it is expressed in equation (2) for which *T* represents the integration time which is associated to the number of frames used.

$$frame\_heading\_error|_k = estimate\_eading|_k - true\_heading|_k \tag{1}$$

$$transit\_heading\_error\_T|_k = \frac{1}{T}\sum_{k}^{k-T} frame\_heading\_error|_k \tag{2}$$

Bearing in mind the correct processing of the 3D geometry processing of the vehicle, it is necessary to define a boundary of the ROI (Region of Interest) for each video sequence at the row 214. This row split the video sequence in to an upper zone in which appear the marker, and in to a lower zone in which appear the target car (Figure 3). The value of this frontier row is consistent for all the video sequences of the car toy. The lower zone is the single zone in which the algorithms for remote heading angle estimation must be tested and evaluated. On the other hand, the part that contains the marker (surface of well-known dimensions and spatial location and rotational orientation) can be used for other purposes such as the assessment of depth accuracy of the estimations on moving targets. For example, the upper zone of the dataset may be used for the assessment of stereo confidences [23]–[28] and accuracy of the stixels [29].

4. RESULTS

The results showed in Figure 4 were obtained employing the satellite marker method for the "toy07" manoeuver. However during the creation of the dataset 32 + 29 maneuvers were analyzed in the same way. A pictographic explanation of all the maneuvers can be found in the wiki-based section of the dataset.

In Figure 4, the number located over each marker represents the stereo frame for which it is associated. The maneuver corresponds to a target car that is approaching to an intersection from a perpendicular lane whit respect to the PTW lane. For this case (see the upper trajectory in Figure 4), the approaching comes from the left of the PTW and the scene finalizes whit the target car turning towards the location of the PTW. Below of this sequence there is another moving sequence but this corresponds to the "etalon" requested for the satellite marker method [22], therefore it will be excluded as ground truth information for all the manoeuvers.

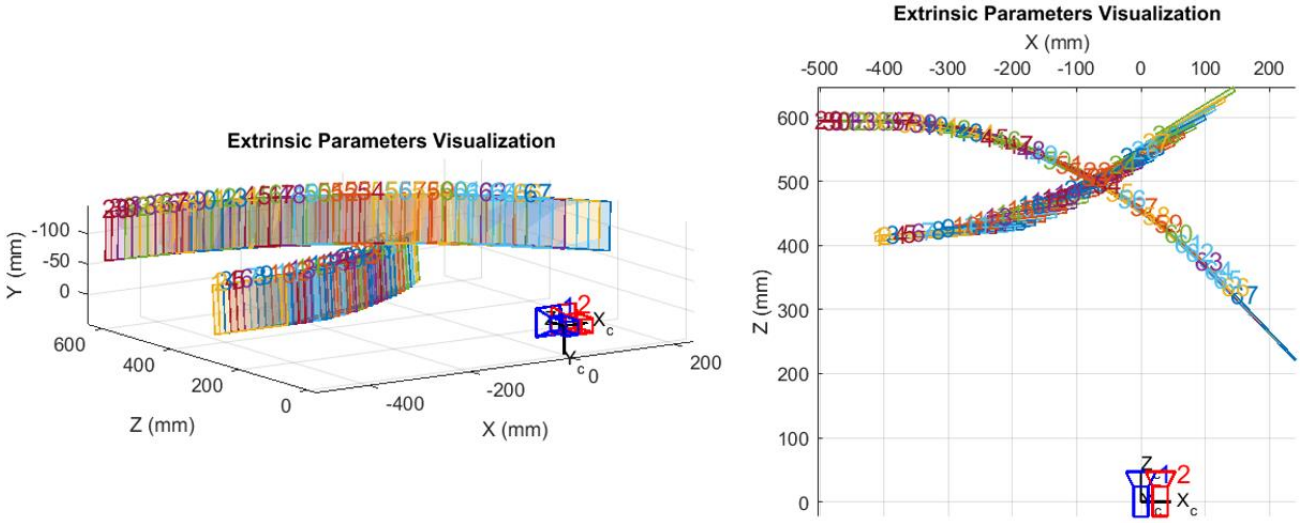

Figure 4. Results obtained from the camera calibration toolbox indicating the 6DoF (3 translations + 3 rotations) of the satellite marker for each frame of the video stereo sequence. On the left hand is the 3D representation of the moving scene and on the right.

In practice, all the marker positioning for the satellite marker are above of the car and the election of an etalon generated at a different height is better for visualization purposes. On the right side of Figure 4, the same traffic scene is visualized from a top view perspective and the ground truth of the heading angle is simple to appreciate. Despite these visualizations, the 6DoF ground truth information is provided in a set of numerical arrays in the dataset as was explained in Section 3.3.

5. CONCLUSION AND FUTURE WORK

The generation of accurate ground truth data of non-synthetic stereo videos is challenging, however the satellite marker method presents a good alternative to generate ground truth from real stereo video. The practical benefits refers to the fact that no additional sensors or specific lighting are required to conduct the activity, and also that common stereo camera calibration tools can be used for obtaining the precious ground truth. All these aspects open the possibilities to perform similar experiments in a variety of application and even in outdoors conditions.

Regarding the dataset generated, which it is unique in his type, it emulates more than 32 variety of PTW pre-crashes or precipitating events that can lead to a PTW crash and provide realistic and accurate heading angle orientation of the opponent vehicle. It contains all the elements to address the challenge of remote car heading angle estimation and consequently, the estimation of its immediate future trajectory. The dataset is available through a GitHub repository.

Finally, an open problem is still the quantification of the precision, consistency and efficiency of the heading angle estimation. Future work need to focus on the development of additional metrics for the evaluation. Additional aspects of improvement may include the analysis including temporal or partial occlusions of the moving obstacle (i.e. overlapping masks in a sequence of the actual video stereo frames), aiming to reproduce commonly situations that happens in urban traffic. In addition to this, the creation of case studies representing real PTW crash scenarios employing real size vehicles are foreseen.

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.

REFERENCES

[1] G. Gil, G. Savino, S. Piantini, N. Baldanzini, R. Happee, and M. Pierini, "Are automatic systems the future of motorcycle safety? A novel methodology to prioritize potential safety solutions based on their projected effectiveness.," *Traffic Inj. Prev.*, no. ja, 2017.

[2] A. Barth and U. Franke, "Where will the oncoming vehicle be the next second?," in *Intelligent Vehicles Symposium, 2008 IEEE*, 2008, pp. 1068–1073.

[3] 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.

[4] U. Franke, C. Rabe, S. Gehrig, H. Badino, and A. Barth, "Dynamic stereo vision for intersection assistance," in *FISITA World Automotive Congress*, 2008, vol. 14, p. 19.

[5] 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.

[6] N. Suganuma and N. Fujiwara, "An obstacle extraction method using virtual disparity image," in *Intelligent Vehicles Symposium, 2007 IEEE*, 2007, pp. 456–461.

[7] 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.

[8] 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.

[9] 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.

[10] K. Fukaya and M. Uchida, “Protection against impact with the ground using wearable airbags,” *Ind. Health*, vol. 46, no. 1, pp. 59–65, 2008.
[11] A. Chawla and S. Mukherjee, “Motorcycle safety device investigation: A case study on airbags,” *Sadhana*, vol. 32, no. 4, pp. 427–443, 2007.
[12] Y. Aikyo, Y. Kobayashi, T. Akashi, and M. Ishiwatari, “Feasibility Study of Airbag Concept Applicable to Motorcycles Without Sufficient Reaction Structure,” *Traffic Inj. Prev.*, vol. 16, no. sup1, pp. S148–S152, Jun. 2015.
[13] G. Savino *et al.*, “Further Development of Motorcycle Autonomous Emergency Braking (MAEB), What Can In-Depth Studies Tell Us? A Multinational Study,” *Traffic Inj. Prev.*, vol. 15, no. sup1, pp. S165–S172, Settembre 2014.
[14] G. Savino, M. Pierini, and N. Baldanzini, “Decision logic of an active braking system for powered two wheelers,” *Proc. Inst. Mech. Eng. Part J. Automob. Eng.*, vol. 226, no. 8, pp. 1026–1036, Aug. 2012.
[15] M. Sieber, K.-H. Siedersberger, A. Siegel, and B. Farber, “Automatic Emergency Steering with Distracted Drivers: Effects of Intervention Design,” 2015, pp. 2040–2045.
[16] M. L. Sichitiu and M. Kihl, “Inter-vehicle communication systems: a survey,” *IEEE Commun. Surv. Tutor.*, vol. 10, no. 2, 2008.
[17] S. Biswas, R. Tatchikou, and F. Dion, “Vehicle-to-vehicle wireless communication protocols for enhancing highway traffic safety,” *IEEE Commun. Mag.*, vol. 44, no. 1, pp. 74–82, 2006.
[18] L. Herda, R. Urtasun, and P. Fua, “Hierarchical implicit surface joint limits for human body tracking,” *Comput. Vis. Image Underst.*, vol. 99, no. 2, pp. 189–209, Aug. 2005.
[19] S. Siltanen, M. Hakkarainen, and P. Honkamaa, “Automatic marker field calibration,” in *Virtual Reality International Conference (VRIC). Laval, France*, 2007, pp. 18–20.
[20] F. Rameau, H. Ha, K. Joo, J. Choi, and I. Kweon, “A Real-Time Vehicular Vision System to Seamlessly See-Through Cars,” in *Computer Vision – ECCV 2016 Workshops*, Cham, 2016, vol. 9914.
[21] A. C. Rice, A. R. Beresford, and R. K. Harle, “Cantag: an open source software toolkit for designing and deploying marker-based vision systems,” in *Pervasive Computing and Communications, 2006. PerCom 2006. Fourth Annual IEEE International Conference on*, 2006, p. 10–pp.
[22] G. Gil, S. Piantini, G. Savino, and M. Pierini, “Satellite Markers: a simple method for ground truth car pose on stereo video,” in *MVR3D 2017: Multiview Relationships in 3D Data 2017 (IEEE International Conference on Computer Vision Workshops)*, Venice, Italy, 2017.
[23] P. Pinggera, D. Pfeiffer, U. Franke, and R. Mester, “Know your limits: Accuracy of long range stereoscopic object measurements in practice,” in *European Conference on Computer Vision*, 2014, pp. 96–111.
[24] A. Spyropoulos, N. Komodakis, and P. Mordohai, “Learning to detect ground control points for improving the accuracy of stereo matching,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1621–1628.
[25] R. Haeusler, R. Nair, and D. Kondermann, “Ensemble Learning for Confidence Measures in Stereo Vision,” 2013, pp. 305–312.
[26] G. Saygili, L. van der Maaten, and E. A. Hendriks, “Adaptive stereo similarity fusion using confidence measures,” *Comput. Vis. Image Underst.*, vol. 135, pp. 95–108, Jun. 2015.
[27] G. Saygili, L. van der Maaten, and E. A. Hendriks, “Stereo Similarity Metric Fusion Using Stereo Confidence,” in *Pattern Recognition (ICPR), 2014 22nd International Conference on*, 2014, pp. 2161–2166.
[28] D. Scharstein and R. Szeliski, “A taxonomy and evaluation of dense two-frame stereo correspondence algorithms,” *Int. J. Comput. Vis.*, vol. 47, no. 1–3, pp. 7–42, 2002.
[29] D. Pfeiffer, S. Gehrig, and N. Schneider, “Exploiting the power of stereo confidences,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 297–304.

AUTHORS' BACKGROUND

Your Name	Title*	Research Field	Personal website
Gustavo Gil	PhD Candidate	Active safety systems and artificial perception	http://www.movingunifi.it/en/gustavo-d-gil/
Giovanni Savino	PhD, Lecturer	Motorcycle integrated safety	http://www.movingunifi.it/en/giovanni-savino/
Marco Pierini	PhD, Ass. Professor	Road vehicles	http://www.movingunifi.it/en/marco-pierini/

*This form helps us to understand your paper better, the form itself will not be published.