The OmniScape Dataset

Ahmed Rida Sekkat, Yohan Dupuis, Pascal Vasseur, Paul Honeine

The OmniScape Dataset

Ahmed Rida Sekkat[1], Yohan Dupuis[2], Pascal Vasseur[1] and Paul Honeine[1].

***Abstract*— Despite the utility and benefits of omnidirectional images in robotics and automotive applications, there are no datasets of omnidirectional images available with semantic segmentation, depth map, and dynamic properties. This is due to the time cost and human effort required to annotate ground truth images. This paper presents a framework for generating omnidirectional images using images that are acquired from a virtual environment. For this purpose, we demonstrate the relevance of the proposed framework on two well-known simulators: CARLA Simulator, which is an open-source simulator for autonomous driving research, and Grand Theft Auto V (GTA V), which is a very high quality video game. We explain in details the generated OmniScape dataset, which includes stereo fisheye and catadioptric images acquired from the two front sides of a motorcycle, including semantic segmentation, depth map, intrinsic parameters of the cameras and the dynamic parameters of the motorcycle. It is worth noting that the case of two-wheeled vehicles is more challenging than cars due to the specific dynamic of these vehicles.**

I. Introduction

Perceiving and understanding the environment is an essential task for a mobile robot or an autonomous vehicle. One of the main issues for the development of these vehicles is the existence of datasets. Among the datasets of pinhole camera images dedicated to the development and study of autonomous vehicles, mention may be made of KITTI [1], Cityscape [2], Berkeley DeepDrive [3], CamVid [4] and Mapillary Vistas Dataset [5]. Omnidirectional cameras can perceive the surrounding environment with a field of view that can reach 360°. For this reason, they are increasingly used in the field of intelligent vehicles, including fisheye cameras due to their compactness and inexpensive design. Several datasets contain fisheye images, such as CVRG [6], LMS [7], LaFiDa [8], SVMIS [9], "Go Stanford" [10], GM-ATCI [11], and RTH Zurich multi-FoV synthetic datasets [12]. However, it is noted that there is a lack of road scenes omnidirectional images datasets embedded in a vehicle dedicated for computer vision applications. Recent work on semantic segmentation of fisheye images of road scenes had been performed on perspective images to which a distortion simulating the fisheye effect is applied [13], [14], [15]. Such deformation induces artefacts in the resulting images. There is a growing need to generate more reliable datasets of omnidirectional images, without the need to go through simple image rectification. Much recent work, especially in deep learning applied on spherical images, have been confirming the need of omnidirectional image databases [16], [17], [18], [19], [20], [21], [22], [23].

In this paper, we propose a framework that can be applied to any simulator or virtual environment to generate omnidirectional images. We present the data acquired from a simulator and describe its use to generate omnidirectional images. We present in detail two well-known types of omnidirectional images, fisheye and catadioptric images. The models can be computed from a calibrated camera, using a large class of models proposed by Geyer and Daniilidis [24], Barreto and Araujo [25], Mei and Rives [26] or Scaramuzza et al. [27].

While the proposed framework is generic, we demonstrate its relevance in two famous simulators: CARLA Simulator and Grand Theft Auto V (GTA V). CARLA is an open-source simulator for urban autonomous driving. It gives the possibility to generate datasets with several ground truth modalities [28]. Grand Theft Auto V (GTA V) is a very high quality AAA video game. Both simulators provide an environment similar to real life, thanks to dynamic weather, seasons, regulated road traffic, traffic lights, signaling, pedestrians, different types of vehicles, ... It is worth noting that there is no support in these simulators or any other simulator for omnidirectional images, which makes the proposed framework of great interest for researchers working on omnidirectional images in robotics and automotive applications.

We show the relevance of this work by releasing the OmniScape dataset[1], which is a dataset of a motorised two-wheeler in the aforementioned simulators. OmniScape comprises stereo fisheye and stereo catadioptric images acquired from the two front sides of a motorcycle, with the corresponding depth maps, semantic segmentation, intrinsic parameters of the cameras and dynamic parameters of the motorcycle, such as velocity, angular velocity, acceleration, location and rotation. See Fig. 1. The OmniScape dataset will be progressively augmented with more omnidirectional data using the same principle with different vehicles, modalities and environments. We have chosen to provide data enquired from a motorcycle because motorcycles present challenging problems that were not addressed before. Indeed, the dynamics of a motorcycle are totally different from the dynamics of cars. As we know, a car is almost all the time parallel to the road. In addition to the distortion in spherical or omnidirectional images, motorcycles undergo rotations yaw, pitch and roll on the three axes, which make the task even harder, due to the inadaptability of classical methods to changes of orientation without a particular learning.

[1]Normandie Univ, UNIROUEN, LITIS, Rouen, France {ahmed-rida.sekkat, pascal.vasseur, paul.honeine}@univ-rouen.fr

[2]Normandie Univ, UNIROUEN, ESIGELEC, IRSEEM, Rouen, France yohan.dupuis@esigelec.fr

[1]https://github.com/ARSekkat/OmniScape

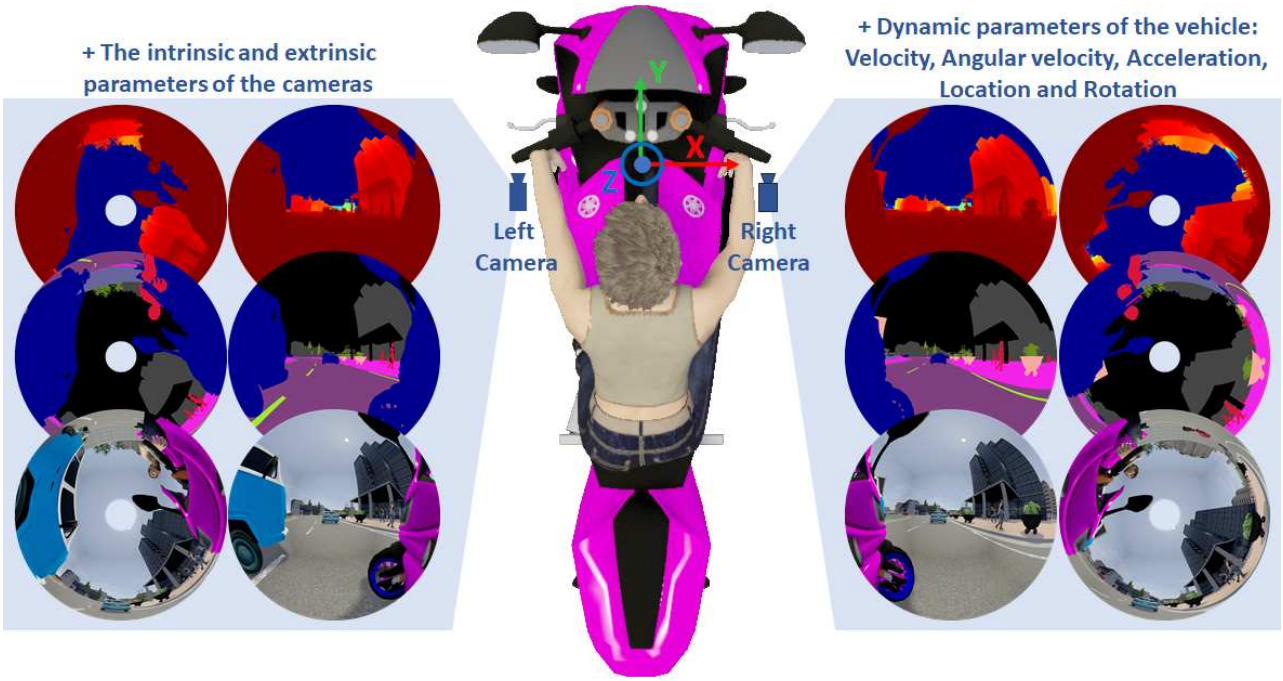


Fig. 1: Recording platform and a representation of the different modalities.

The remainder of this paper is organized as follows. Next section provides a survey on work made using virtual environments as a data source. Section 3 introduces the proposed framework to generate omnidirectional images. Section 4 presents the OmniScape dataset. Finally, Section 6 concludes the paper.

II. Related Work

In the literature, there are several works conducted on virtual environments for the development or validation of autonomous driving systems. Virtual environments have several advantages, mainly the inexpensiveness to generate realistic data, as well as the variety of the nature of the data that can be generated, such as depth maps, semantic segmentation or details of the dynamic properties of the vehicle. These virtual environments allow to simulate different sensors. We also do not have to deal with the problem of data protection and privacy of individuals. Currently several datasets were generated from virtual environments, such as SYNTHIA [29], VEIS [30], and Playing for benchmark [31].

Thanks to advantages offered by the reverse engineering and modding tools, several recent works have been carried out on the generation of data from GTA V. We can mention the method proposed by Doan et al. in [32] for generating perspective images using a virtual camera with six degrees of freedom. In [33], Richter et al. used GTA V to capture pixel-by-pixel semantic segmentation using an open source middleware called renderdoc between the game and the GPU. In [34], Angus et al. also extracted semantic segmentation images by changing the textures and shaders of the game in the game files. Richter et al. generated in [31] a benchmark of several data types from GTA V, all annotated with ground truth data for low-level and high-level vision tasks, including optical flow, instance segmentation, detection and objects tracking, as well as visual odometry.

Johnson-Roberson et al. used in [35] synthetic data generated by GTA V, to show that state-of-the-art algorithms trained only by this data, work better than if they are driven on manually annotated real-world data when tested on the KITTI dataset for vehicle detection. We can note that all these works considered perspective images and, until now, there is neither a dataset for omnidirectional images, nor a dataset for motorcycles or any powered two-wheeler. The present paper seeks to fill this gap, by proposing a framework for omnidirectional data generation from a virtual environment, and generating specifically motorcycles datasets.

III. Proposed Framework

The proposed framework generates omnidirectional images from a virtual environment using 360° cubemap images. To create 360° images, six images are extracted in the six different directions. Using the appropriate omnidirectional camera model, each pixel of the omnidirectional image can be associated with a 3D direction on the unit sphere. We then compute the cube that presents the six images under the cubemap projection. Using ray tracing, we construct a lookup table that stores correspondences between the omnidirectionnal image and the cubemap images. It corresponds to the intersection of the 3D direction associated to each pixel in the omnidirectional image with the cubemap images. Then we just need to affect to each pixel in the omnidirectional image the corresponding relevant information from the cubemap image (RGB, depth or semantic segmentation), as sketched in Fig. 2.

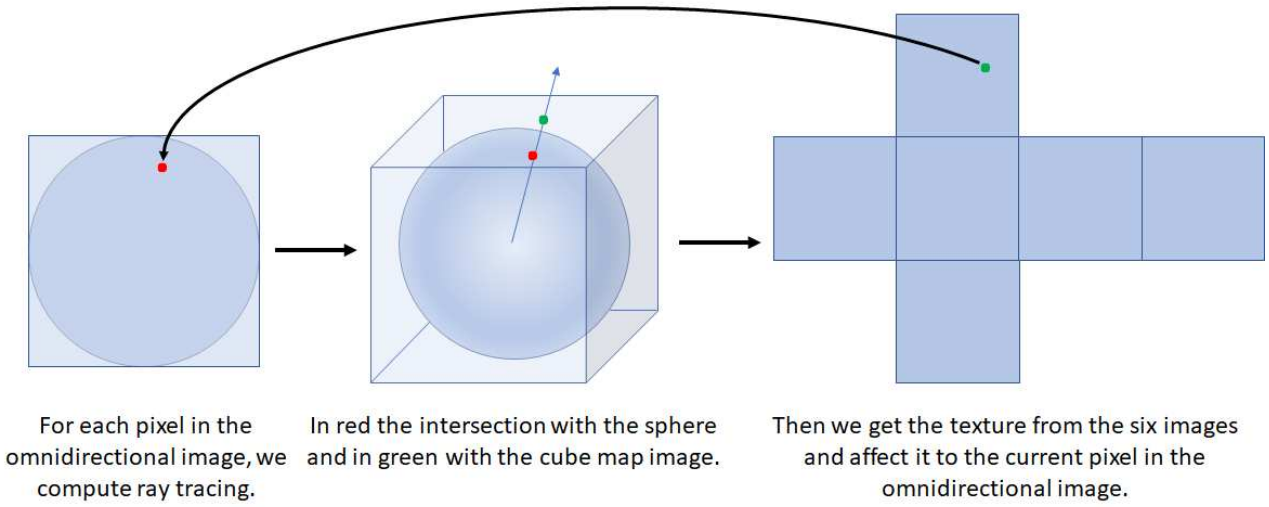


Fig. 2: Lookup table construction to set the omnidirectional image pixel values.

The parameters of the model are calculated from a calibrated camera. To generate omnidirectional images, the proposed framework can use well-known camera models, such as the models proposed by Geyer and Daniilidis [24], Barreto and Araujo [25], Mei and Rives [26] and Scaramuzza et al. [27]. Without loss of generality, we detail in the following the model proposed by Scaramuzza et al. in [27]. It is a calibration model for omnidirectional cameras, considering the omnidirectional imaging system as a compact and unique system composed by a pinhole camera and a mirror. It allows to compute the intrinsic parameters of the omnidirectional camera. This means that it provides the relation between a given 2D pixel and the corresponding 3D vector, from the point of view of the unit sphere, as illustrated in Fig. 3. Let (u, v) be the metric coordinates of a pixel p with respect to the center of the omnidirectional image, and (x, y, z) those of the corresponding 3D vector P, according to

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} u \\ v \\ f(w) \end{bmatrix}, \tag{1}$$

with $w = \sqrt{u^2 + v^2}$. Within this model, the function $f(w)$ is considered to be a polynomial function, namely of the following form

$$f(w) = a_0 + a_1 w + a_2 w^2 + a_3 w^3 + a_4 w^4 + ... \tag{2}$$

The calibration parameters a_i are estimated by the least-squares method on data acquired with a real camera, as described in [36].

Since the above model is general for omnidirectional images, not just fisheye images, we use the same method to generate also catadioptric images. We mean by catadioptric images, in this paper, the images taken by a camera composed by a pinhole camera (perspective camera) and a hyperboloidal mirror [24], [26]. The catadioptric images we generate in this work are made in a way that includes all directions of the road in the images, which means that the pinhole camera is on the top of the mirror or the contrary.

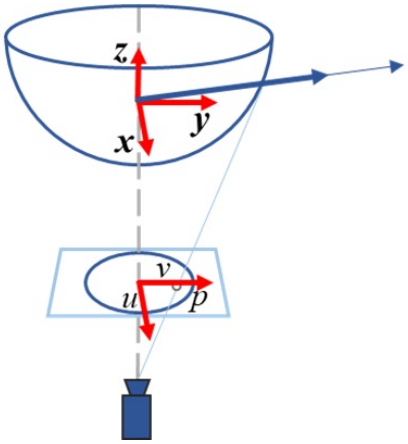
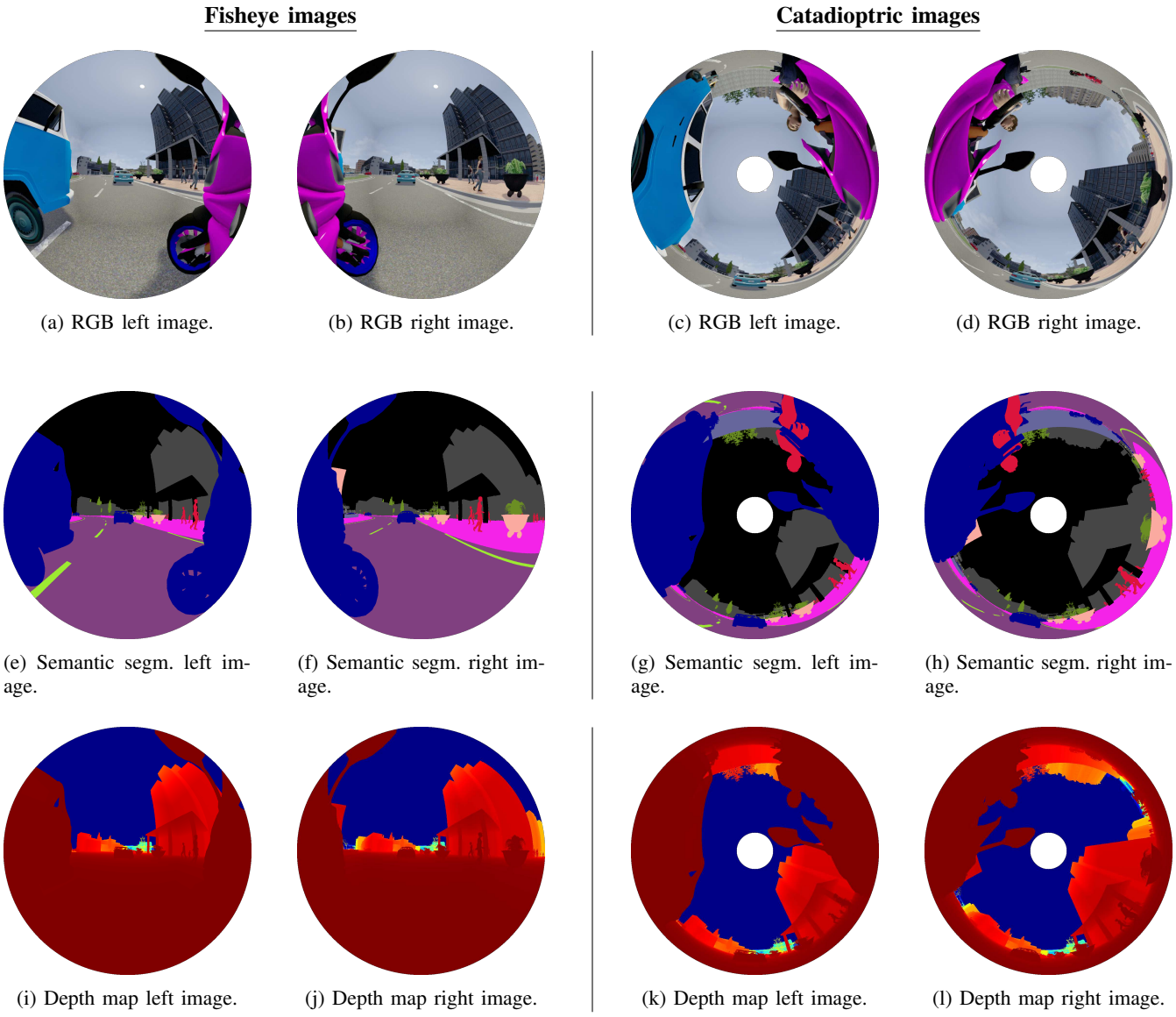


Fig. 3: The omnidirectional camera model proposed in [27].

Fig. 4: Mapping of the six images in the fisheye (left) and catadioptric (right) images.

Fig. 4 shows the mapping of the six images in the fisheye images and the catadioptric images. The colors red, orange, yellow, blue, green and purple represent respectively the six sides, front, back, left, right, up and down.

The computational cost to render one omnidirectional frame does not exceed 4.6ms on Ubuntu 18.04.3 64-bit running on an Intel Core i7-8750H CPU @ 2.20GHz.



(a) RGB left image. (b) RGB right image. (c) RGB left image. (d) RGB right image.

(e) Semantic segm. left image. (f) Semantic segm. right image. (g) Semantic segm. left image. (h) Semantic segm. right image.

(i) Depth map left image. (j) Depth map right image. (k) Depth map left image. (l) Depth map right image.

Fig. 5: Examples of fisheye (left panel) and catadioptric (right panel) images generated from a single capture.

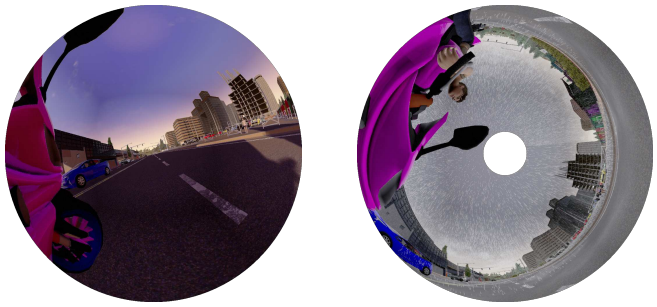
IV. OmniScape Dataset

The OmniScape[2] dataset contains, for each capture, fisheye and catadioptric stereo RGB images from the two front sides of a motorcycle, with semantic segmentation and depth map ground truth, as well as the dynamics of the vehicle with its velocity, angular velocity, acceleration and orientation. See Fig. 1 for an overview. The OmniScape dataset will be progressively augmented with more omnidirectional data using the described framework with different vehicles, modalities and environments. The dataset contains data generated from GTA V and CARLA, and can be extended to other simulators. However, due to space limitation, we present in the following data extracted only from CARLA.

[2]https://github.com/ARSekkat/OmniScape

For more insights on the extraction of data from GTA V, we refer the interested reader to our previous work [37].

To generate the images, we used 5 towns available in CARLA Simulator. An example of images generated from a single capture is given in Fig. 5. The RGB images are available for 14 different weather conditions and time of the day and this for each capture. Fig. 6 shows an example of a capture with 4 different weather conditions in fisheye and catadioptric. CARLA Simulator gives a semantic segmentation into 13 classes, namely Building, Fence, Other, Pedestrian, Pole, Road line, Road, Sidewalk, Vegetation, Vehicle, Wall, Traffic sign, Unlabeled. Fig. 7 shows the distribution of pixel of all images in the dataset for both fisheye images and catadioptric images.



Default

Hard rain sunset

Soft rain noon

Clear noon

Fig. 6: Examples of fisheye and catadioptric images generated from a single capture with four different weather conditions and time. The motorcycle in this capture undergoes rotations.

TABLE I: Statistics concerning the dynamics (yaw, pitch and roll, in degrees) of the motorcycle in a tested route

	mean	std	min	median	max
Yaw	1.15	110.54	-179.99	0.85	179.99
Pitch	-0.15	1.75	-10.26	-0.07	18.94
Roll	0.14	3.62	-24.83	0.00	25.91

In complement to these omnidirectional images, the OmniScape dataset contains also the dynamics of the vehicle at each capture, such as velocity, angular velocity, acceleration and orientation. As we explained before, the case of two-wheelers is more challenging because of the dynamics of these vehicles. We computed statistics concerning these dynamics in CARLA Simulator. As presented in TABLE I, we can see that the roll and the pitch change considerably.

These alterations will surely affect classical tasks such as visual odometry and semantic segmentation. This is due to the fact that most computer vision and machine learning tasks are often trained on perspective data acquired with cars as autonomous vehicles, while these vehicles do not suffer from modifications in these dynamics.

In complement to the images given in this paper, more examples from CARLA Simulator and GTA V can be found in OmniScape GitHub.

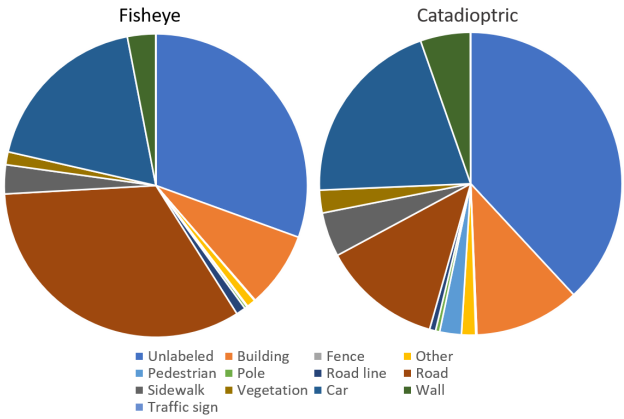


Fig. 7: Percentage of pixels representing each class in the dataset for both fisheye and catadioptric images.

V. Conclusion

This paper presented a general framework to generate datasets of omnidirectional images from virtual environments, and provided the OmniScape dataset. We demonstrated the relevance of this framework by generating fisheye and catadioptric images with depth map, semantic segmentation and dynamic parameters. Two simulators were investigated with success, GTA V and open-source CARLA Simulator.

There are many possible extensions to this application, including the generation of other types of datasets, using different types of omnidirectional camera models and different vehicles like drones. These datasets can be used as evaluation credentials for different vision and deep learning applications, whose algorithms applied to perspective images have limited performance on omnidirectional images. A wide variety of applications include Simultaneous Localization And Mapping (SLAM), visual odometry, depth estimation, object recognition and classification, detection and tracking. Moreover, they can also be used to evaluate or even train semantic segmentation algorithms developed for omnidirectional images.

VI. Acknowledgements

The authors would like to thank Vincent Vauchey for his valuable suggestions on improving the optimisation of the computational cost.

This work was mainly supported by a RIN grant, Région Normandie, France. It was partially supported by ANR CLARA (ANR-18-CE33-0004-02) and DAISI project funded with the support from the European Union with the European Regional Development Fund (ERDF) and from the Regional Council of Normandy.

REFERENCES

[1] J. Fritsch, T. Kuehnl, and A. Geiger, "A new performance measure and evaluation benchmark for road detection algorithms," in *International Conference on Intelligent Transportation Systems (ITSC)*, 2013.

[2] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[3] F. Yu, W. Xian, Y. Chen, F. Liu, M. Liao, V. Madhavan, and T. Darrell, "BDD100K: A diverse driving video database with scalable annotation tooling," *CoRR*, vol. abs/1805.04687, 2018.

[4] G. J. Brostow, J. Fauqueur, and R. Cipolla, "Semantic object classes in video: A high-definition ground truth database," *Pattern Recognition Letters*, vol. 30, pp. 88–97, 2009.

[5] G. Neuhold, T. Ollmann, S. Rota Bulo, and P. Kontschieder, "The mapillary vistas dataset for semantic understanding of street scenes," in *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[6] I. Baris and Y. Bastanlar, "Classification and tracking of traffic scene objects with hybrid camera systems," in *2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, Oct 2017, pp. 1–6.

[7] A. Eichenseer and A. Kaup, "A data set providing synthetic and real-world fisheye video sequences," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Mar 2016, pp. 1541–1545.

[8] S. Urban and B. Jutzi, "Lafida - a laserscanner multi-fisheye camera dataset," *J. Imaging*, vol. 3, p. 5, 2017.

[9] G. Caron and F. Morbidi, "Spherical Visual Gyroscope for Autonomous Robots using the Mixture of Photometric Potentials," in *IEEE International Conference on Robotics and Automation*, Brisbane, Australia, May 2018, pp. 820–827.

[10] N. Hirose, A. Sadeghian, M. Vázquez, P. Goebel, and S. Savarese, "Gonet: A semi-supervised deep learning approach for traversability estimation," *CoRR*, vol. abs/1803.03254, 2018.

[11] D. Levi and S. Silberstein, "Tracking and motion cues for rear-view pedestrian detection," in *2015 IEEE 18th International Conference on Intelligent Transportation Systems*, Sep. 2015, pp. 664–671.

[12] Z.Zhang, H. Rebecq, C. Forster, and D. Scaramuzza, "Benefit of large field-of-view cameras for visual odometry," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, May 2016, pp. 801–808.

[13] A. M. Sweeney, L. M. Bergasa, E. Romera, M. E. L. Guillén, R. Barea, and R. Sanz, "Cnn-based fisheye image real-time semantic segmentation," *2018 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1039–1044, 2018.

[14] L. Deng, M. Yang, Y. Qian, C. Wang, and B. Wang, "Cnn based semantic segmentation for urban traffic scenes using fisheye camera," in *2017 IEEE Intelligent Vehicles Symposium (IV)*, June 2017, pp. 231–236.

[15] L. Deng, M. Yang, H. Li, T. Li, B. Hu, and C. Wang, "Restricted deformable convolution based road scene semantic segmentation using surround view cameras," *CoRR*, vol. abs/1801.00708, 2018.

[16] X. Yin, X. Wang, J. Yu, M. Zhang, P. Fua, and D. Tao, "Fisheyerecnet: A multi-context collaborative deep network for fisheye image rectification," in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham: Springer International Publishing, 2018, pp. 475–490.

[17] Y.-C. Su and K. Grauman, "Learning spherical convolution for fast features from 360°imagery," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 529–539.

[18] N. Perraudin, M. Defferrard, T. Kacprzak, and R. Sgier, "Deepsphere: Efficient spherical convolutional neural network with healpix sampling for cosmological applications," *CoRR*, vol. abs/1810.12186, 2018.

[19] B. Coors, A. P. Condurache, and A. Geiger, "Spherenet: Learning spherical representations for detection and classification in omnidirectional images," in *Computer Vision – ECCV 2018*, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham: Springer International Publishing, 2018, pp. 525–541.

[20] W. Boomsma and J. Frellsen, "Spherical convolutions and their application in molecular modelling," in *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Curran Associates, Inc., 2017, pp. 3433–3443.

[21] C. Esteves, C. Allen-Blanchette, A. Makadia, and K. Daniilidis, "Learning so(3) equivariant representations with spherical cnns," in *The European Conference on Computer Vision (ECCV)*, September 2018.

[22] R. Kondor, Z. Lin, and S. Trivedi, "Clebsch-gordan nets: a fully fourier space spherical convolutional neural network," in *NeurIPS*, 2018.

[23] Q. Zhao, C. Zhu, F. Dai, Y. Ma, G. Jin, and Y. Zhang, "Distortion-aware cnns for spherical images," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*. International Joint Conferences on Artificial Intelligence Organization, 7 2018, pp. 1198–1204.

[24] C. Geyer and K. Daniilidis, "A unifying theory for central panoramic systems and practical implications," in *Computer Vision — ECCV 2000*, D. Vernon, Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 445–461.

[25] J. P. Barreto and H. Araujo, "Issues on the geometry of central catadioptric image formation," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 2, Dec 2001, pp. II–II.

[26] C. Mei and P. Rives, "Single view point omnidirectional camera calibration from planar grids," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*, April 2007, pp. 3945–3950.

[27] D. Scaramuzza, A. Martinelli, and R. Siegwart, "A toolbox for easily calibrating omnidirectional cameras," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Oct 2006, pp. 5695–5701.

[28] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An open urban driving simulator," in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017, pp. 1–16.

[29] G. Ros, L. Sellart, J. Materzynska, D. Vazquez, and A. M. Lopez, "The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 3234–3243.

[30] F. S. Saleh, M. S. Aliakbarian, M. Salzmann, L. Petersson, and J. M. Alvarez, "Effective use of synthetic data for urban scene semantic segmentation," in *ECCV*, 2018.

[31] S. R. Richter, Z. Hayder, and V. Koltun, "Playing for benchmarks," in *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, 2017, pp. 2232–2241.

[32] A. Doan, A. M. Jawaid, T. Do, and T. Chin, "G2D: from GTA to data," *CoRR*, vol. abs/1806.07381, 2018.

[33] S. R. Richter, V. Vineet, S. Roth, and V. Koltun, "Playing for data: Ground truth from computer games," in *European Conference on Computer Vision (ECCV)*, ser. LNCS, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., vol. 9906. Springer International Publishing, 2016, pp. 102–118.

[34] M. Angus, M. ElBalkini, S. Khan, A. Harakeh, O. Andrienko, C. Reading, S. L. Waslander, and K. Czarnecki, "Unlimited road-scene synthetic annotation (URSA) dataset," *CoRR*, vol. abs/1807.06056, 2018.

[35] M. Johnson-Roberson, C. Barto, R. Mehta, S. N. Sridhar, K. Rosaen, and R. Vasudevan, "Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks?" in *IEEE International Conference on Robotics and Automation*, 2017, pp. 1–8.

[36] Y. Dupuis, X. Savatier, J. Ertaud, and P. Vasseur, "Robust radial face detection for omnidirectional vision," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1808–1821, May 2013.

[37] A. R. Sekkat, Y. Dupuis, P. Vasseur, and P. Honeine, "Génération d'images omnidirectionnelles à partir d'un environnement virtuel," in *27-ème Colloque GRETSI sur le Traitement du Signal et des Images*, Lille, France, Aug. 2019.