

Omnidirectional Image Data-set for Computer Vision Applications

Bruno Berenguel-Baeta, Jesús Bermudez-Cameo, José J. Guerrero

Afiliación: Robótica, Percepción y Tiempo Real (RoPeRT)
Instituto de Investigación en Ingeniería de Aragón (I3A)
Universidad de Zaragoza, Mariano Esquillor s/n, 50018, Zaragoza, Spain.
Tel. +34-976762707, e-mail: berenguel@unizar.es

Abstract

In this paper we present an image data-set of different omnidirectional systems. The images include full information of colour, depth, instance segmentation and room layout. This dataset aims to help in the training and test of different neural networks and development of computer vision algorithms.

Introduction

Artificial intelligence and neural networks are the main research topics in the last decade. The use of deep-learning approaches has increased for many different applications. More precisely, the computer vision community is one of the main research fields where deep-learning approaches has made a great impact. The use of neural networks has greatly improved the performance of objects and instance segmentation in images and videos, object detection, facial recognition and many other applications.

However, for a proper training of deep-learning approaches, big image datasets are needed. Even though we can find many different datasets on conventional cameras, as KITTI [1] or SYNTHIA [2], datasets of omnidirectional systems are quite limited. This is because the labelling of these images is quite hard due to the big distortions that present. So, the question is: if omnidirectional images present big distortions that makes them more difficult to handle, why do we use them? The answer is simple: omnidirectional images provide more information of an environment than conventional cameras with a single shot. Figure 1 compares the information that can be obtained with a conventional image and with different omnidirectional systems.

In order to compensate the lack of omnidirectional image dataset, [3] presents a tool to generate different photorealistic images with semantic and depth information from a variety of omnidirectional system. Not only central panoramas, but also different fish-eye projection models, catadioptric systems and non-central projection systems. In this

work we present a data-set of omnidirectional images with semantic and depth information.

Projection models

In this section we resume the projection models of some of the omnidirectional systems proposed in [3]. The models selected are: Equirectangular panorama, Kannala-Brandt's model [4] and the non-central circular panorama [5]. This selection is based on the utility of these images in different computer vision research. We will also provide some examples of the applications for these models.

The equirectangular panorama is a central projection system that maps the environment into a sphere centred in the optical center. The main advantage of this kind of images is that we have a complete, 360° vision around the optical center. However, the distortion that it presents around the poles of the sphere are quite severe, making difficult the use of conventional computer vision algorithms.

The Kannala-Brandt's model [4] is an empiric model to represent fish-eye and central catadioptric systems. This model encapsulates the distortion of the image into a 9th degree polynomial which depends on the incidence ray's angle. Besides, this projection model presents a closed form for the forward and backward projection functions.

The central projection systems mentioned above are characterized by the unique optical center. By contrast, non-central projection systems do not have a unique optical center. This characteristic introduces some problems in the definition of the projection models, which are defined in Plücker coordinates. However, these systems have more information in the image than central projection systems, allowing to obtain more information than in central images. More precisely, from non-central circular panoramas, we can recover the 3D information of the lines in the image if we know the calibration of the system.

Data-set applications

The proposed data-set presents a wide variety of ground truth information from the different projection models. One of the main uses of this information could be the training of different neural networks. Works as [6] and [7] use equirectangular panoramas in order to obtain the layout and a object segmentation respectively from a single image using different deep-learning architectures. In these works, the amount of data used for training is limited in variety, i.e. the layouts have only 4 or 6 walls and most of the images are taken in bedrooms. Our data-set aims to improve the variety, also providing a pixel-wise ground truth information as seen in Fig. 2.

Another approach of computer vision algorithm is the Simultaneous Localization and Mapping (SLAM) implementations. New approaches in this field use omnidirectional images from fish-eye cameras in order to get more information of the environment. With our data-set we provide colour and depth information in the images, see Fig. 2, as well as the location of the camera in the environment, giving perfect information to develop and test SLAM algorithms.

References

- [1]. GEIGER, A., LENZ, P., STILLER, C. and URTASUN, R.. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 2013, vol. 32, no 11, p. 1231-1237.
- [2]. ROS, G., SELLART, L., MATERZYNSKA, J., VAZQUEZ, D., and LOPEZ, A. M. The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. p. 3234-3243.
- [3]. BERENGUEL-BAETA, B.; BERMUDEZ-CAMEO, J. and GUERRERO, J.J. OmniSCV: An Omnidirectional Synthetic Image Generator for Computer Vision. *Sensors*, 2020, vol. 20, no 7, p. 2066.
- [4]. KANNALA, J. and BRANDT, S. A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses. *IEEE transactions on pattern analysis and machine intelligence*, 2006, vol. 28, no 8, p. 1335-1340.
- [5]. MENEM, M. and PAJDLA, T. Constraints on perspective images and circular panoramas. In *British Machine Vision Conference*. 2004. p. 1-10.
- [6]. FERNANDEZ-LABRADOR, C, FACIL, J. M., PEREZ-YUS, A., DEMONCEAUX, C., CIVERA, J., and GUERRERO, J. J. Corners for layout: End-to-end layout recovery from 360 images. *IEEE Robotics and Automation Letters*, 2020, vol. 5, no 2, p. 1255-1262.
- [7]. GUERRERO-VIU, J, FERNANDEZ-LABRADOR, C., DEMONCEAUX, C., and GUERRERO, J. J. What's in my Room? Object Recognition on Indoor Panoramic Images. In *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020. p. 567-573



Figure 1: Comparative of different acquisition systems. On the left, a perspective image; in the middle, a fish-eye image defined by the Kannala-Brandt model; and on the right, an equirectangular panorama.

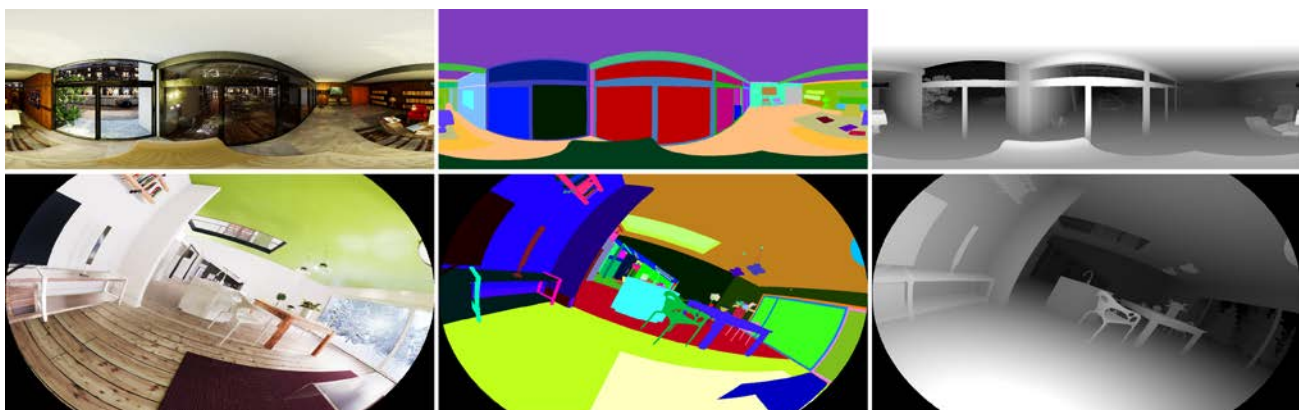


Figure 2: On top, equirectangular panoramas; on bottom, fish-eye images. From left to right: colour image, semantic segmentation and depth information.