Mem. S.A.It. Vol. 93, 91

Memorie della

Omnidirectional people's gathering monitoring by using deep learning algorithms

Claudio Pernechele[1], Nicola Grandis[2], Irene Bondanza[2], Alessandro Capucci[2], Luigi Lessio[1], Lorenzo Paoletti[1], Silviu Sparc[2], Marco Tomasin[1], Renata Yusupova[2]

[1] Istituto Nazionale di Astrofisica–OAPD, Vicolo dell'Osservatorio, 5, 35122, Padova
[2] ASC27 srl, Lungotevere Michelangelo, 9, 00192, Roma

Received: 14 June 2021; Accepted: 3 May 2022

Abstract. It has long been recognized as gathering of people is one of the major risk factor in spreading of viral epidemics. Social distancing is then one of the most simple and powerful system to mitigate the spread of infections. We explore here the possibility of monitoring public people's gathering by using a novel bifocal omnidirectional lens designed by INAF jointly with deep learning-based algorithms. The paper briefly describe how the lens works, the applied deep learning algorithms and the preliminary results of the trials.

1. Introduction

Panoramic cameras have been realized in many different fashions for more than one century. Panoramic images can be created in a variety of ways: the oldest and most popular is one of taking a set of pictures around the observer and stitch them in a post-processing phase. Three main classes of panoramic cameras can be distinguished:

– Panoramic vision systems based on moving cameras;
– Panoramic vision that make use of a set of independent lenses (and sensors) looking at different points on the surrounding space;
– Motionless panoramic cameras (omnidirectional lens).

A pan-tilt mechanism is usually the main part of the first family. A standard perspective camera pans and tilts capturing a number of perspective shots of the scene. A panoramic view is then created composing such images.

The pan-tilt camera requires a significant amount of time and it has to stop while capturing the shots that are used to compose the panoramic image; this represents a great drawback: a moving camera is not able to furnish a simultaneous vision of the surrounding environment and dynamic information are lost. For this reason, the system is useful for static scenery (panoramic photography, art diagnostic, architectural view, etc...). Panoramic sensors with multiple cameras (i. e., the second class) offer a simultaneous vision of the panoramic space and have satisfactory spatial resolution but they are rather expensive, bulky and, usually, require a complicated calibration. To the third class belong the omnidirectional cameras. A motionless panoramic camera has many advantages with respect to the other classes: it makes use of only one detector (on the contrary of class two) and doesn't make use of moving motor (as in the first class); all of these aspects reflects on lower power con-

Table 1. Panoramic cameras classes

Lens Class	Dynamic scene	Complexity	Size	Post Processing	Spatial Resolution
Multiple Lenses	Yes	High	Large	Hard	High
Motorized lens	No	Medium	Medium	Hard	High
Omnidirectional	Yes	Low	Small	Easy	Low

sumption, lower cost, lower weight, smaller bulky and, furthermore, reduce possible failure points. The use of this family is fruitful when the recording of a dynamic scene, combined with lightness, is necessary (as in robotic and anti-collision systems, self-driving, scene monitoring etc...). An evident drawback of this family of panoramic camera is one of furnishing a relative low spatial resolution information of the surrounding environment. These parameters are summarized in the table 1. We describe here how the use of an omnidirectional camera, based on a novel lens designed by the INAF researchers, may be useful in monitoring people's gathering. An omnidirectinal camera has many advantages in solving these tasks:

- It does not use motorized parts (power consuming, possible point failure, maintenence cost);
- It avoids the use of multiple cameras (complex placing and maintenance, post processing image stitching, heavy data stream);
- The apparent weakness of its intrinsic low resolution turn out as an advantage, due to the privacy issue.

2. Bifocal Panoramic Lens

The bifocal panoramic lens (BPL) is an optical device which records a panoramic field of view (360° × 100°) and, simultaneously, an enlarged part of this panoramic field. It has been designed and patented by INAF researchers (Pernechele 2016). A sketch of how it works is shown in the figure 1. The panoramic field is imaged with a donut-like shape, while the center of the donut is filled with the enlarging part of the lens, using a flat folding mirror settled in the lens frontal optics. In this case one obtains a magnification of 3× with respect to the paraxial focal length. We designed the panoramic arm with a focal length of 2 mm (working at F/3.5), while the frontal optics, used for magnifying the field, has a focal length of 6 mm (at F/3.5 too).

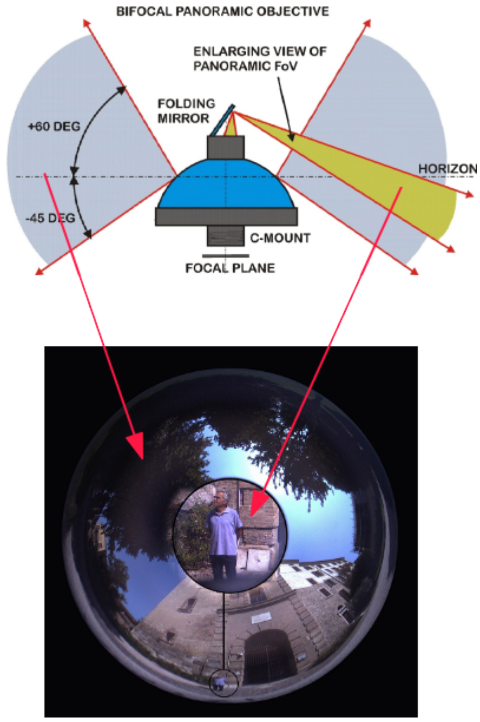

Fig. 1. A graphical explanation of how the BPL works.

2.1. BPL Mapping Function

The way in which the 3D object space is mapped onto the 2D focal plane is described by the lens mapping function. With reference to the figure 2, let be the zenith angle Z measured from zenith, down to the maximum field of view ($Zmax$). The angular field distribution function in the focal plane ψ(Z) is, in general, not linear with respect to Z and an amount of optical distortion would be present. The focal length f is changing along the field. In particular, $f = f(Z)$ (with $f(0)$ the paraxial focal length) and this causes the deformation of the image.

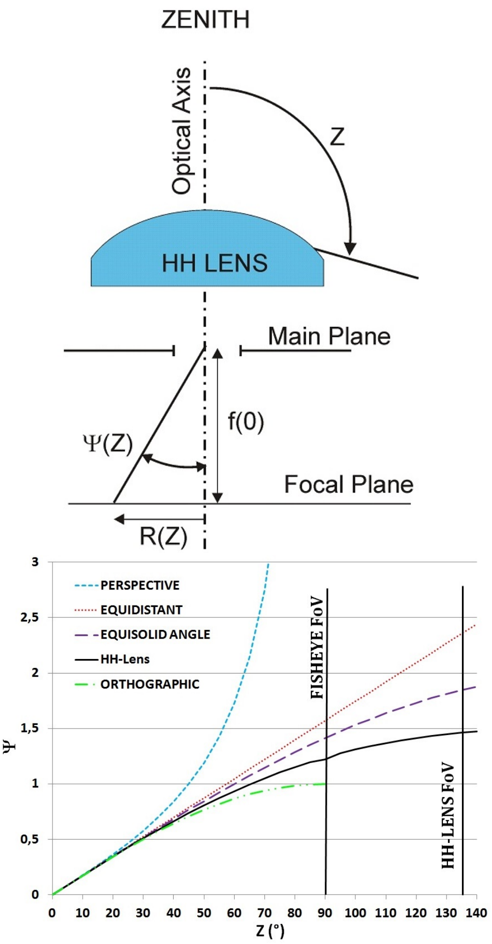

Fig. 2. BPL sky mapping function.

In the case of a fisheye $Zmax = 90°$, while in the BPL it may go up to 135°. Hereinafter we designate with f the paraxial focal length, with $f = f(0)$. The space mapping function $R(Z) = f(Z)$ may have different forms. The "perfect" undistorted map of the object space is one where $R = f tan(Z)$. In this way every point in the space is mapped maintaining the same angular distribution into the focal plane. The lens works like a pinhole camera and object straight lines remain straight (distortion-free) in the focal plane. This function is known as the "perspective projection". It is well known that it has no practical sense for wide angle lenses, because the focal plane would be infinitely extended and the entrance pupil (the pin-hole) would be completely obscured for $Z = 90°$. Equidistant (linear scaled) projection is one where $R = fZ$: it maintains angular distances. This lens projection scheme is also know as F-Theta. The most general projection function has the form:

$$R(Z) = f\psi(Z) = f k_1 sin(k_2 Z). \quad (1)$$

Within this class of mapping functions, high compression of the marginal objects is present. Among those there is the equisolid angle projection ($k_1 = 2$, $k_2 = 0.5$), which maintains surface relations: each pixel in the detector subtends an equal solid angle, i. e. an equal area on the unit sphere. Finally we cite the orthographic projection ($k1 = k2 = 1$) where $R = f sin(Z)$. This mapping function maintains planar illuminance. Here the marginal fields are extremely compressed at the focal plane and it make sense only for $Z < 90°$. The projection function of the BPL lies in the midway between the equisolid angle and orthographic projections (black continuous line in Figure 2), with $f = 2mm$, $k1 = 1.48$ and $k2 = 0.64$. Those parameters are necessary during the image distortion corrections (the unwarping process).

2.2. BPL Applications

Due to its optical multi-functionalities (extremely large field for contextual imaging and simultaneous higher resolution for scientific purposes) jointly with its compactness and motionless fashion, the BPL may find many applications in space applications. We propose its applications as multi-functional tracker for mini- and micro-satellites (Pernechele et al. 2018; Opromolla et al. 2017), stereoscopic auto-driving and simultaneous scientific capabilities for planetary rovers (Pernechele et al. 2020), as an immersive stereo-camera for exploring lunar caves (Borrmann et al. 2021; Pozzobon et al. 2021; Pernechele et al. 2021) and as an ibrid GNSS-optical navigation system for placing seismometers in lunar experiments (Harms et al. 2021).

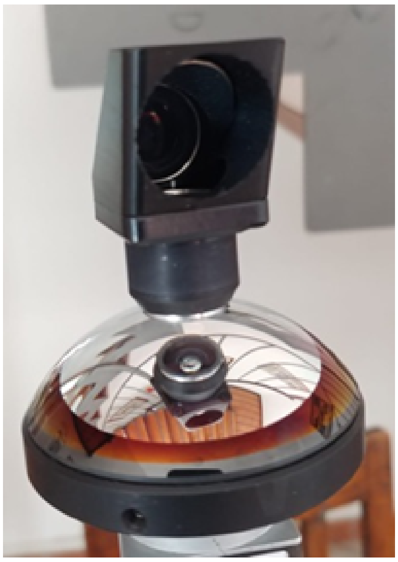

Fig. 3. The manufactured lens.

3. The PANCAM

A remote panoramic monitoring system (PANCAM) based on the bifocal panoramic lens has been assembled following the architecture outlined in figure 4. The image from a CMOS sensor chip (Sony IMX264, 5Mpx, 2/3 inch format) is acquired by an micro-PC (Odroid-XU4, octacore CPUs). The micro-PC also processes the image unwarping, in accordance with the equation (1) of the previous section. The frame transfer travels @USB3 speed. The unwarped images are sent by a client wi-fi toward the remote wi-fi server (master). The ongoing images may be seen by a remote (wi-fi connected) devices, such e.g. smartphones, pc tablets, etc... by means of an dedicated user interface. Images are also saved at the ACQ board-level onto a microSD flash memory at a frame rate of 1 fps. The system is supplied by a DC-DC converter which provides 1A @12V. As an example, the image seen in the central panel of figure 1 has been acquired by this PANCAM system.

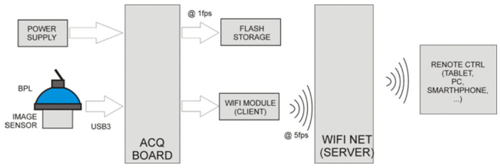

Fig. 4. The PANCAM architecture.

4. System Trial

In order to test the deep learning algorithm on realistic BPL-based PANCAM acquisitions, we record an images stream in an urban environment.

4.1. Image Acquisition

The system has been tested in a public areas (*Piazza dei Signori* in the city of Padova downtown). A "frog view" of the area, with the PANCAM settled in the place, is shown in figure 5.

The site environment shows a high contrast between sun illuminated and shaded areas, typical for extremely wide angle lenses.

Fig. 5. The PANCAM settled in the place.

By running the DL algorithm described in section 4 it was possible to identify the persons in the shaded region, despite of the low level of signal in the scene. Another important result has regarded the identification target algorithm. Indeed, it works also in the original (before to apply the unwarping algorithm) image. If confirmed in future trials, this fact could make not necessary the application of the time-consuming equation (1) in section 2.1. An example is shown in figure 6, where a person has been identified in the shaded area (red and green boxes in, respectively, original and unwarped images), although surrounded by a series of small trees, which are very similar in shape and dimension. Discriminate a person from such similar silhouettes is a valid test showing the goodness of the DL running algorithm.

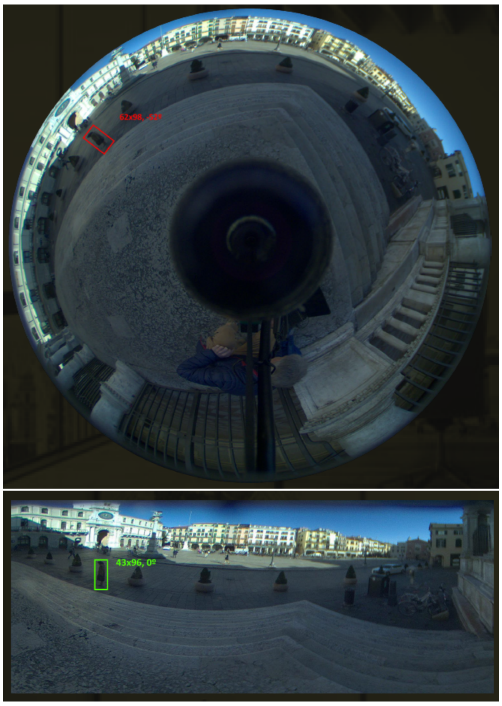

Fig. 6. Example of DL algorithm results.Top: original image; bottom: unwarped image.

4.2. Deep Learning Algorithm

The acquired images are then been processed using the "SafePeople" software developed by the ASC27 company (www.asc27.com). This software, based on artificial intelligence models, has been originally developed for different purposes.

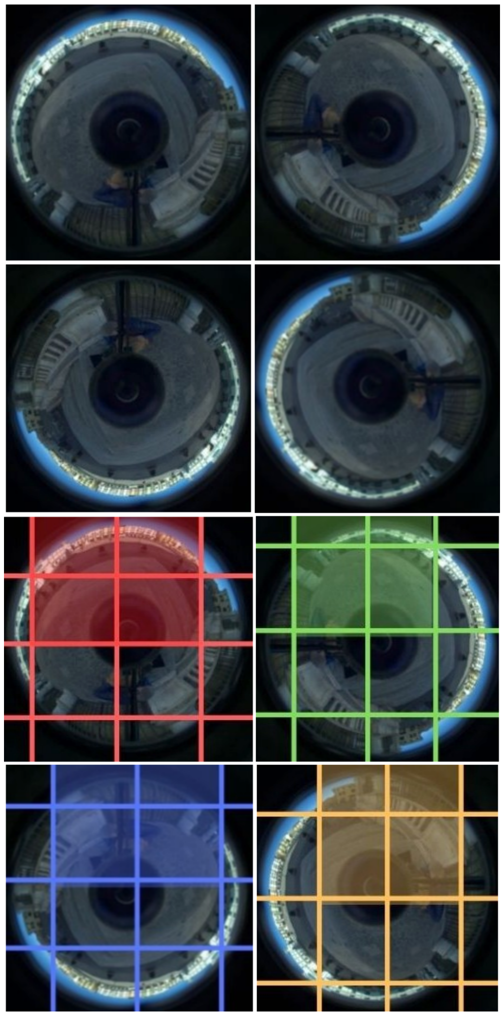

Fig. 7. Image rotated of 90 degrees (upper panels), resulting into four copies and (bottom panels) preprocessing of each images.

In this case we have used models for Object Detection to identify people and models for so-

Fig. 8. Resulting image as the concatenation of the four images of Fig. 7.

cial distancing monitoring and prediction. For people object detection models we adopted two different approaches. In this respect, due to the distortion of the original images, the algorithm need to take into account different angles when detecting targets.

In the first approach, the original image from the lens is used and multiple rotations on the image of 90° degrees each are applied (see upper panels in figure 7). Therefore, for each image, we use four rotated copies of it in oredere to detect objects. Moreover, as in each rotated image the relevant part is the upper one, each rotated image is sliced into sixteen sub-windows and the detection algorithm is applied for the most relevant portions (see bottom panels in figure 7). As some sub-windows may overlap, it is applied a post processing algorithm called Non Maximum Suppression (NMS). This is a standard technique utilized to concatenate all the detected bounding boxes (figure 8).

The second approach is based on a custom SSD (Single Shot MultiBox Detector) network and can be used both with original and unwarped images, by applying the algorithm on the entire image. Regardless of the type of images, an annotator has defined bounding boxes for the targets to be detected, people in this case.

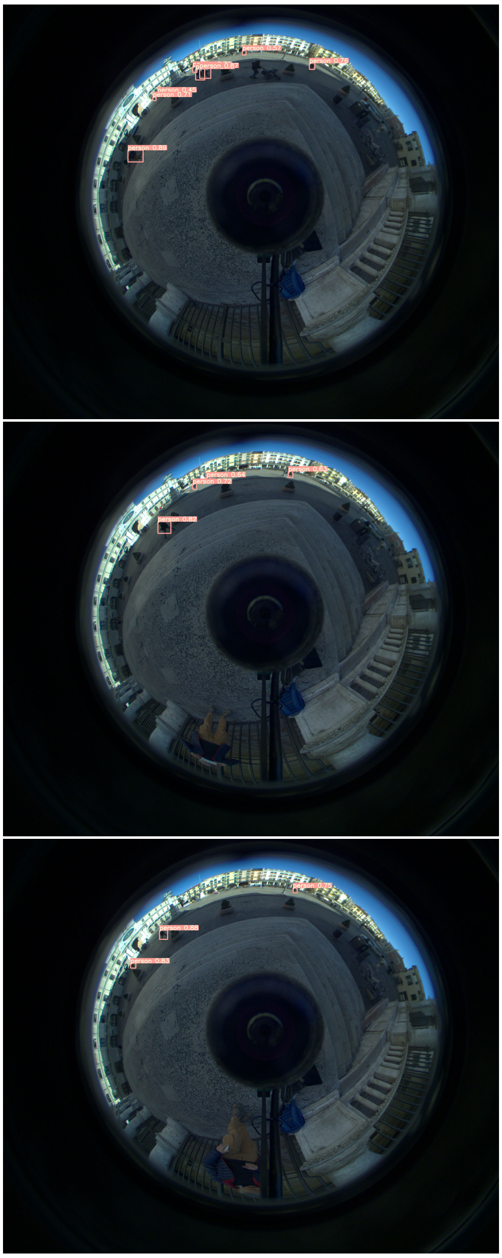

Fig. 9. Original images with people detection.

An important aspect is the unconventional images that the bifocal panoramic lens captures, therefore the annotator must process them taking into account which targets are

Table 2. AP: mean precision of a class, mAP: mean of APs of all classes, Speed V100: averaged speed performances using a V100 GPU machine.

Model	mAPval 0.5:0.95	mAPtest 0.5:0.95	mAPval 0.5	Speed V100 (ms)
MD01A	34.2	35.8	54.9	3.8
MD01B	41.3	43.2	62.8	4.2
MD01C	44.8	47.5	65.5	5.3
MD01D	49.5	50.1	67.2	8.0

more relevant and which are confusing for the algorithm. Furthermore, regarding the original images, objects may not reflect the canonical proportion because of the lens distortion.

Given the labelled dataset, the detection algorithm is trained with a portion of it and then the remaining part of the dataset is used to test the algorithm. In the figure [9] some samples of the detected images are showed, taken from the dataset that contains the original images. Accuracy and performance of the custom object detection models can be improved with further works and training but, actually, they already shown a good quality.

After the identification of people through object detection, the information of the relative bounding boxes are processed by a network for social distancing, in order to identify possible dangerous situations.
In table 2 the performances of the different models, in terms of precision, are shown.

5. Conclusions

In order to test an omnidirectional people's gathering monitoring system based on deep learning algorithms, we developed and tested a panoramic camera (PANCAM) mounting a bifocal panoramic lens (designed by INAF). Trials running on a realistic environment, a square in the Padova downtown, show that, despite of the high contrast of the scene, the system works appropriately in achieving this task. The PANCAM is a compact optical system and may monitor extremely large areas, avoiding the use of any motorized element. This makes the PANCAM system a valid alternative to most common multi-camera and pan-tilt video surveillance systems. Moreover, the intrinsic low resolution of the PANCAM, turn out into a positive fact, saving from privacy issues.

Acknowledgements. We are grateful to INAF and MUR for support this job by means of dedicated covid-contrast funds.

References

Borrmann, D., et al. 2021, LPSC, vol. 2548 (id. 2073).

Harms, J., et al. 2021, *APJ*, doi:10.3847/1538-4357/abe5a7

Opromolla, R., et al. 2017, Acta Astronautica, doi: 10.1016/j.actaastro.2017.08.020, 2017.

Pernechele, C., 2016, Optics Express, doi: 10.1364/OE.24.005014

Pernechele, C., et al. 2018, Adv. in Space Res., doi:10.1016/j.asr.2018.02.025, 2018.

Pernechele, C., et al. 2020, EPSC, doi: 10.5194 /epsc2020-557.

Pernechele, C., et al. 2021, NESF ELS, in press.

Pozzobon, R., et al. 2021, LPSC, vol. 2548 (id. 1886).

Simioni, E., et al. 2020, ISPR, doi: 10.5194/ isprs-archives-XLIII-B3-2020-1151-2020.