[Supplementary Material]: Dense Scene Reconstruction from Light-Field Images Affected by Rolling Shutter

Hermes McGriff [1,3] Renato Martins[1,2] Nicolas Andreff [3] Cédric Demonceaux[1,2]
[1]Université de Bourgogne, CNRS UMR 6303 ICB [2]Université de Lorraine, CNRS, Inria, LORIA
[3]Université de Franche-Comté, CNRS UMR 6174 FEMTO-ST
{hermes.mc-griff,renato.martins,cedric.demonceaux}@u-bourgogne.fr, nicolas.andreff@univ-fcomte.fr

In this "Supplementary Material" to our paper, we provide an introduction to light field theory and light field (plenoptic) cameras. We also discuss the natural synchronization on the SAIs and the creation of RSLF-d from the adaptation of our seminal work [2]. The objective of this "Supplementary Material" is notably to provide key ideas for a better understanding of the unique properties of the Rolling-Shutter Light Fields.

1. The plenoptic camera and the light field: an overview

To facilitate the understanding of the paper, we provide in this section an brief overview of concepts on plenoptic cameras and light field theory.

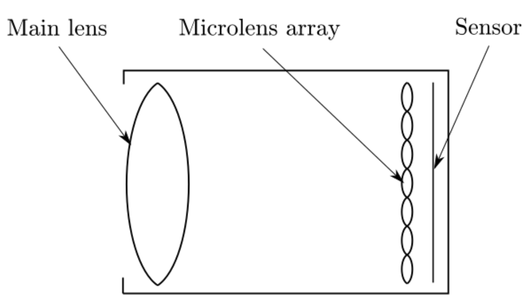

Figure 1. A simplified plenoptic camera scheme. The main difference to a standard camera is the addition of a microlens array between the main lens and the sensor.

Plenoptic camera optical system. In 1992, Adelson and Wang introduced the so-called "plenoptic camera", an optical design capable of capturing a 4D light field in a single acquisition. It was inspired by 1908's integral photography of Lippmann [1]. In the following years, this optical sensor gained popularity in the computer vision community, sometimes under the term "light field cameras" (popularized by the monopolistic company that produces these cameras today). Plenoptic cameras capture the light in 4D spatial dimensions (with an angular resolution superior to one, contrary to monocular cameras). In order to do that, it uses an optical system composed of a classical part ("the main lens") and a micro-lens array in front of the sensor (between the main lens and the sensor). The model of the plenoptic camera is shown in Fig. 1. There are multiple interpretations of the way this imaging system creates images. The easier to understand it, according to us, is to see every single microlens as the end of a unique optical system that images the scene like a tiny camera. The whole array of microlens can then be seen as an array of cameras that have a resolution of a dozen of pixels. Another interpretation is the idea that the microimages (the images behind the microlenses) capture the angular incidence of the light rays at the position of the microlens. That is why one can refer to the supplementary dimensions (in comparison to monocular cameras) as the "angular resolution". The images obtained with such a camera look like the one presented in Fig. 2.

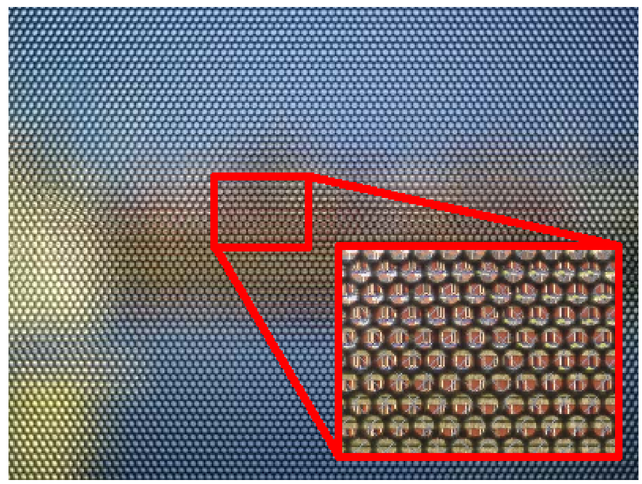

Figure 2. The unprocessed image obtained with a plenoptic camera. We can see in the detail the multitude of micro-images produced by the microlenses. Images from [3].

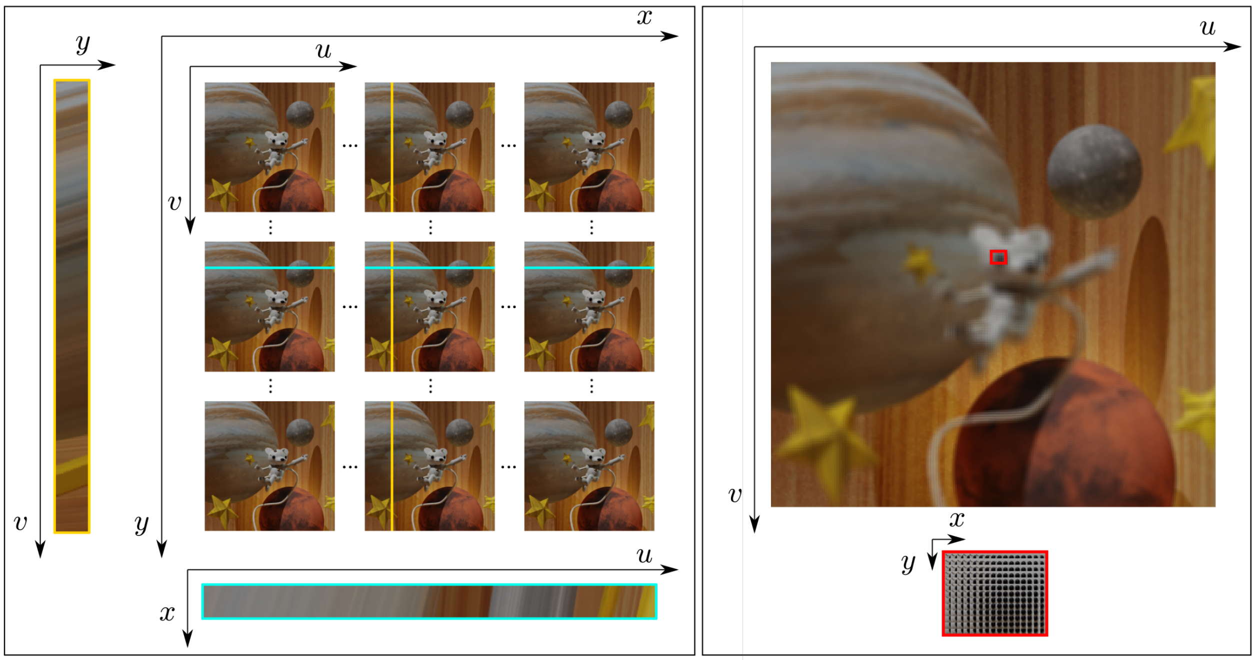

Figure 3. The representations of the 4D light field. In the paper, we use the array (x,y) of SAI (u,v) as an equivalence of an array of cameras.

Light Field representation. In order to exploit such an image we first have to decode the light field out of it. In this context, decoding means segmenting the small circular images called the micro-images and constructing a 2D array of 2D micro-images, thus a 4D tensor, that is typically referred as 4D light field in the context of plenoptic cameras (the light field is in fact a broader concept in radiometry). We name the dimensions of the 4D LF with (x, y, u, v). (x, y) give the coordinates of the "viewpoint". They are related to the coordinates inside the micro-images. These are the aforementioned "angular resolutions". (u,v) give the coordinates in pixels inside an image from a given viewpoint, named the "sub-aperture image" (SAI). (u,v) are the "spatial resolution". These SAI are equivalent to images acquired with an array of cameras with very small baselines. One can also create images by cutting the 4D data in the (x,u) plane or (y,v) plane, called the epipolar plane images. Fig. 3 shows the way the 4D light field can be represented.

2. Time synchronization of the SAIs

We have shown in Sec. 1 of this "Supplementary Material" how the LF is extracted from the plenoptic image. In order to introduce RS we just need to consider that the lines of pixels are acquired sequentially. As the micro-images are very small in comparison to the whole picture, we make the reasonable assumption that all the pixels of a micro-images are acquired at the same time. We can then assume that the lines of micro-images are acquired sequentially, but that they are locally global shutter. What this fair modelization means is that the relation between the time of acquisition of a pixel at coordinates (x,y,u,v) (See Fig. 3) is only related to its v coordinate. As the Sub-Aperture Images are expressed in the same (u,v) coordinate system, they begin to be acquired at the same time (same v) and finish to be acquired at the same time. The SAIs are then naturally synchronized. (Rigorously, the most medium time one can observe between two SAI is the time between the acquisition of the first line of pixel in a micro-image and the acquisition of the last line of pixel in the same micro-image. A few lines. And thus a short time; that we consider negligible in our study and that, in further study would be, at worst, measurable.)

3. The motion information in RSLF

In RS LF, there is no prior needed to compute the instantaneous velocity of each point in 3D. Indeed, we observe every points at different times and from different angles. It is this chore property of the RS LF that we exploit. Fig. 4 shows 4 adjacent SAIs. The blue points are the projections of the same 3D point in these SAI. The red dotted line is the line acquired at the time at which the point is projected in the bottom row of SAI. Because of parallax, the point is not projected in the same line in the upper row. In the GS case, it would be projected in the position of the green dotted cir-

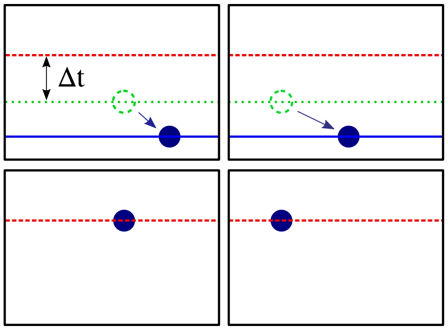

Figure 4. The motion estimation is visible in 3D for every point, since we can observe them from different view points (change in SAI) and at different times (change in vertical coordinate inside the SAI).

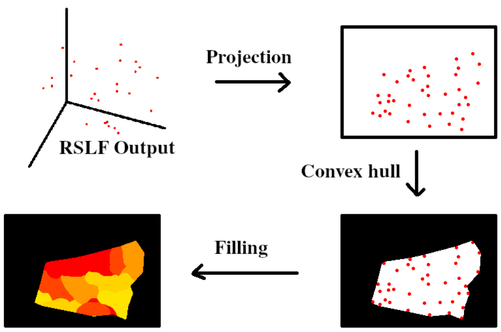

Figure 5. Overview of the densification of the original RSLF result.

cles. But since we are in RS, a time δt has passed between the red dotted line and the green dotted line. The 3D point moved in space during this time. It is then projected in another line (full blue line).

4. RSLF-d: the densified version of RSLF

In order to ease the comparison between the different methods, we propose a new take on [2] by densifying it. Indeed, this method returns a point cloud that is hard to compare with ground truth. It would also need its own metrics, that would add complexity to the comparison. That is why we propose RSLF-d (that can be seen as an annex, minor contribution). By projecting the point cloud obtained with their method (we set it up to have a lot of points) we obtain a very sparse depth map. We fill out the empty areas with the closer value. Also, we compute the convex hull of the 2D projected point cloud in order to restrict the estimated depth map inside the hull of estimated points (we recall that we compare with masked ground truth). The pipeline from RSLF-d is presented in Fig. 5. By doing this, we make sure to stay relatively true to the work of [2], while allowing us to use the evaluation tools we crafted.

References

[1] Gabriel Lippmann. La photographie integrale. *Comptes-Rendus*, 146:446–451, 1908. 1

[2] Hermes McGriff, Renato Martins, Nicolas Andreff, and Cedric Demonceaux. Joint 3D shape and motion estimation from rolling shutter light-field images. In *WACV*, 2024. 1, 3

[3] Mats Wernersson. The camera maker. 1