Use of Epipolar Images Towards Outliers Extraction in Depth Images

Dirce Celorico[1]
dircelorico@isr.uc.pt
Leandro Cruz[1]
lmvcruz@isr.uc.pt
Leandro Dihl[1]
leandro.dihl@isr.uc.pt
Nuno Gonçalves[12]
nunogon@deec.uc.pt

[1] Institute of Systems and Robotics
University of Coimbra
Coimbra, Portugal

[2] Portuguese Mint and
Official Printing Office
Lisboa, Portugal

Abstract

Plenoptic cameras, such as Lytro and Raytrix, has been widely used over the last years. Their main feature is the light intensity acquisition from several viewpoints. From these viewpoint images, we can reconstruct a 3D model of the captured scene by calculating the depth of each pixel by a passive depth estimation, using only one captured image. In this way, the depth denotes the distance between the respective point and the camera.

Although this 3D model can be directly used for several purposes such as refocusing after capture or object segmentation, they are often quite noisy, what is a disadvantage for some applications like 3D visualization, or more complex mesh processing.

In this work, we will present a method for filtering the depth model, reconstructed from light field cameras, based on the removal of low confidence reconstructed values and using an inpainting method to replace them. This approach has shown good results for outliers removal.

1 Introduction

Nowadays, the use of light field (or plenoptic cameras) has been much more common not only because of progress on calibration and decoding but due mainly to depth estimation improvement. These special cameras were introduced by Ng [5]. Their main feature corresponds to the use of an array of microlenses in front of the sensor, which essentially achieves the capture of an array of views simultaneously, in a single shot image, without special patterns projection in the scene.

It allows the user to acquire not only information regarding the intensity of light in the scene but also the information related to the direction of the light rays in the space. This information is gathered from different viewpoints, thus after mathematical manipulation, it allows to extract a 3D model of the scene.

One of the most important parameters obtained by this type of camera is the depth of each point (the distance between the point in the scene and the camera). From a single raw image of a plenoptic camera the depth can be estimated, at least in positions with sufficient local contrast. The ability to determine depth perception and to be able to reconstruct the scene in a 3D environment has become of major importance in tasks such as segmentation, detection or even 3D display.

There are many works about depth estimation [2]. More recently, depth estimation approaches from light-field have handled the epipolar plane images (EPIs) geometry due to this model type has shown robust and correctness results in depth map achievement. In this work, we will keep focus on outliers filtering from a local coherence point removal. After the removal of the points, we fill the produced holes by the usage of an inpainting method.

2 Related Work

For the past few years, the development in the study of light field cameras has provided a range of techniques to accurately estimate the depth. Related works with this kind of cameras are still considered current and has produced such a useful progress as in Monteiro et al. [4]. The use of the epipolar geometry (Epi's) in order to define a depth map is one of the most used techniques, and it presents some better results. For instance, in works like Wanner and Goldluecke [10], Suzuki et al. [9], Lin et al. [6] Si et al. [3] and Zhang et al. [12].

The plenoptic function is represented as a seven-dimensional space (illustrated in equation (1)) and it represents the amount of data in a scene.

$$L = (v_x, v_y, v_z, \phi_x, \phi_y, \lambda, t) \tag{1}$$

It consists of the information of intensity for every 3D point (v_x, v_y, v_z), the corresponding direction (ϕ_x, ϕ_y), wavelength (λ) and time (t). The reduction of variables from 7D to 4D, the so-called Lumigraph, was introduced by Gortler et al. [8]. Taking into account that image acquisition is made by a single shot capture the variable time can be neglected, and it is evenly possible to neglect the variable wavelength by considering a grey scale environment.

We also adopted in this work the two-plane parametrization approach which is widely studied and represents a simple structure where each ray is defined by the intersection with two parallel planes. One corresponding to the camera plane, the other one to the image plane, and both are separated by a distance $z = 1$.

Therefore, a light field can then be represented as a 4D function $f(x,y,s,t)$ where the dimensions (x,y) represent the spatial distribution and the dimensions (s,t) represent the angular distribution. By fixing one of the spatial dimension and one angular dimension we can get a horizontal epipolar plane image S_{t^*,y^*} or a vertical epipolar plane image S_{s^*,x^*}.

Regarding the inpainting approach, it is considered a world-wide known technique to fill lost information in an image, it has been developed accordingly to different algorithms that conduct to the filling of gaps in images [7]. Some of these different approaches rely on Partial Differential Equations (PDE), texture synthesis, exemplar-based search or even wavelet transforms.

3 Depth Estimation

The input of our filtering method is the raw image captured by a light field camera while the output is an RGBD image (the captured colour and the reconstructed depth). From the raw image, we use the epipolar plane images analyses approach so images such as Figure 1 are obtained.

Figure 1: Epipolar plane image example S_{t^*,y^*}.

In an epipolar image, a line corresponds to a point across different viewpoint images. The corresponding slope in the lines can be used to estimate the respective depth value since they are directly related. In order to determine the slope in epipolar images we used the structure tensor, an approach presented by Wanner and Goldluecke [11], and we also defined the reliability assigned to the depth throughout the coherence value. This structure tensor is defined by equation (2). In equation (3) it is represented

(a) (b) (c)

Figure 2: (a) Light field image used; (b) Depth image; (c) Coherence image (with values in the interval [0 1])

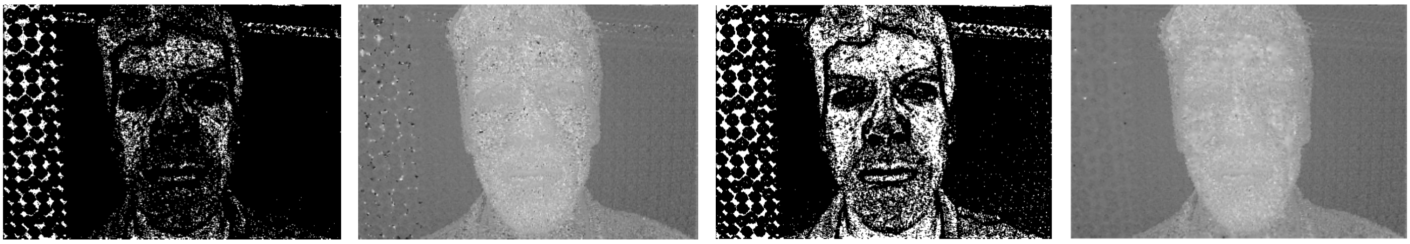
(a) (b) (c) (d)

Figure 3: (a) Mask from values bellow threshold of 0.6; (b) Depth image after inpainting (0.6) ; (c) Mask from values bellow threshold of 0.8 ; (d) Depth image after inpainting (0.8)

the angle to determine the direction of the lines. Finally, equation (4) calculates the reliability parameter. This procedure is implemented for each horizontal (S_{t^*,y^*}) and vertical (S_{s^*,x^*}) plane image and for all viewpoints. For each pixel in the image and depending on the coherence value of the vertical and the horizontal Epi, a selection is performed between them to accomplish a single value for the depth at that pixel.

$$J = \begin{bmatrix} G_\sigma * (S_x S_x) & G_\sigma * (S_x S_y) \\ G_\sigma * (S_x S_y) & G_\sigma * (S_y S_y) \end{bmatrix} = \begin{bmatrix} J_{xx} & J_{xy} \\ J_{xy} & J_{yy} \end{bmatrix} \quad (2)$$

where G represents a Gaussian operator, and S_x and S_y represent the gradients of the Epi in x and y directions respectively.

$$\phi = \frac{1}{2} \arctan\left(\frac{Jyy - Jxx}{2Jxy}\right) \quad (3)$$

$$r_{y^*,t^*} = \frac{\left(J_{yy} - J_{xx}\right)^2 + 4J_{xy}^2}{\left(J_{yy} + J_{xx}\right)^2} \quad (4)$$

An example of one of the image viewpoints is represented in Figure 2(a).

In Figure 2(b) we can see the corresponding depth map achieved with the related coherence image (Figure 2(c)).

In the coherence image we are using a heat colour scheme in order to illustrate the variations of coherence in the image. The red one corresponds to a higher level of reliability (1) and the blue one to a low reliability value (0). As expected, we can notice that the coherence presents lower values in areas with low texture such as the cheeks and forehead.

4 Low Coherence Removal

Following the coherence map creation, we remove all low confidence points. This is performed with the use of a threshold. In This way, coherence values located below the threshold are considered null. The holes created with this removal are filled by using an inpainting method. [1]

This threshold is such an important parameter to achieve superior results. The larger threshold choice the larger the holes, what can imply poor reconstruction. On the other hand, the lower threshold choice the smaller confidence, which implies low noise removal.

For an easier way to illustrate the difference between regions considered for inpainting, we decided to do it with two thresholds values of 0.6 and 0.8. In Figure 3, it is illustrated two masks generated employing these threshold values. For each example, it is also shown the respective depth map result achieved after inpaiting.

In these cases pixels in the depth map that have a coherence value below the threshold are discarded since they represent poorly reconstruction estimation. The second case is a good example in which the threshold trade off was well satisfied so that it was removed most of noisy points, however the resulting holes were not too large, what meets a satisfactory reconstruction.

The inpainting procedure considered [1] solves a Partial Differential Equations (PDE) to fill up the hole by a continuous interpolation of the boundary value (frontier constraint). This method assumes springs (with a nominal length of zero) connect to each node with every neighbour (horizontally, vertically and diagonally). The result is a piece of surface that can be glued to the depth model in a continuous way, obtaining a derivative continuity as well (a smooth surface).

5 Results and Future Work

In this work, we presented a fast outliers filtering method for depth models created using plenoptic cameras. The models obtained by this type of camera often need to be filtered due to the presented high frequency of noise and holes. Furthermore, the proposed filtering can be combined to a low-pass filter to smooth the model (and the outliers removal improve the result of this other filter by reducing the frequency cut).

The proposed method can also be improved in some directions. For instance, in the way we calculate the confidence threshold to be assumed for the inpainting procedure. We have currently been researching a relation between the threshold and the coherence energy (then, how to cut in coherence) in such a way that maximize the coherence average without increasing dramatically the hole size (what can create spurious inpainting reconstruction). To do so, a positive solution might be the usage of a local threshold.

Since this is an ongoing project, one of the following steps is to apply this method to a set of images with depth ground truth. It allows measuring the method accuracy and compare it with other state of the art approaches.

Current results underline the viability of outliers filtering from a local coherence point removal. This method is promising and expands other researches about depth estimation and filtering which are left as future work.

References

[1] J. D'Errico. Inpaint nans. `https://www.mathworks.com/matlabcentral/fileexchange/4551-inpaint_nans`.

[2] A. Jarabo Y. Zhang L. Wang Q. Dai T. Chai G. Wu, B. Masia and Y. Liu. Light field image processing: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):926–954, Oct 2017.

[3] H. Zhu L. Si and Q. Wang. Epipolar plane image rectification and flat surface detection in light field. *Journal of Electrical and Computer Engineering*, 2017, 2017. `https://doi.org/10.1155/2017/6142795`.

[4] J. P. Barreto N. B. Monteiro, S. Marto and J. Gaspar. Depth range accuracy for plenoptic cameras. *Computer Vision and Image Understanding*, 168(C), 2018. ISSN 1077-3142.

[5] R. Ng. Digital light field photography. *Stanford University Ph.D. thesis.*, 2006.

[6] F. Wu P. Lin, J. Yeh and Y. Chuang. Depth estimation for lytro images by adaptive window matching on epi. *Journal of Imaging*, 3(17), 2017.

[7] Mr. Krunal R. Patel R. Suthar. A survey on various image inpainting techniques to restore image. *Int. Journal of Engineering Research and Applications*, 4(2):85–88, 2004.

[8] R. Szeliski S. Gortler, R. Grzeszczuk and M. Cohen. The lumigraph. In *In Proc. ACM SIGGRAPH*, pages 43–54, 1996.

[9] Takahiro Suzuki, Keita Takahashi, and Toshiaki Fujii. Sheared epi analysis for disparity estimation from light fields. *IEICE Transactions on Information and Systems*, E100.D(9):1984–1993, 2017.

[10] S. Wanner and B. Goldluecke. Globally consistent depth labeling of 4d light fields. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pages 41–48, June 2012.

[11] S. Wanner and B. Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(3):606–619, 2014.

[12] Y. Liu H. Wang X. Wang Q. Huang X. Xiang Y. Zhang, H. Lv and Q.i Dai. Light-field depth estimation via epipolar plane image analysis and locally linear embedding. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(4):606–619, 2017.