ISPRS Journal of Photogrammetry and Remote Sensing 144 (2018) 38–47

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Focus model for metric depth estimation in standard plenoptic cameras

Said Pertuz[a,c,*], Edith Pulido-Herrera[b], Joni-Kristian Kamarainen[c]

[a] *Universidad Industrial de Santander, 680002 Bucaramanga, Colombia*
[b] *Universidad Antonio Nariño, 111321 Bogotá, Colombia*
[c] *Tampere University of Technology, 33720 Tampere, Finland*

ARTICLE INFO

Keywords:
Plenoptic camera
Lightfield
Focus
Calibration
Depth estimation

ABSTRACT

In recent years, a lot of efforts have been devoted to the problem of depth estimation from *lightfield* images captured by standard plenoptic cameras. However, most of the metric depth estimation methods in the state-of-the-art leverage pixel disparity only. In this paper, we tackle the problem of focus-based metric depth estimation in standard plenoptic cameras. For this purpose we propose a closed-form model that relates the refocusing parameter with the focus distance of a plenoptic camera in order to allow for metric depth estimation. Based on the proposed model, we develop a calibration procedure that allows finding the parameters of the model. Using measurements of a time-of-flight sensor as ground-truth, experimental validation in a distance range of 0.2–1.6 m shows that focus-based depth estimation is feasible with a root-mean-squared error of less than 5 cm.

1. Introduction

One of the main advantages of plenoptic cameras is to allow a fast capture of the *lightfield* with a single snapshot in a compact device (Lumsdaine and Georgiev, 2009). In turn, depth estimation based on the captured lightfield has many potential applications, such as passive 3D video recording, 3D modeling, augmented reality, and depth-guided scene segmentation and tracking (Strobl and Lingenauber, 2016). However, these applications may often require 2.5D images with *metric* information of the scene. Unfortunately, state-of-the-art metric depth estimation approaches based on standard plenoptic cameras can deliver metric depth estimates based on disparity only, whereas the focus cue is not exploited.

At this point, it is important to distinguish between three main architectures of plenoptic cameras: coded-aperture plenoptic cameras (Veeraraghavan et al., 2007; Marwah et al., 2013), focused plenoptic cameras (Lumsdaine and Georgiev, 2009), and standard plenoptic cameras (Ng et al., 2005).[1] Although the feasibility of coded-aperture plenoptic imaging has been demonstrated (Wetzstein et al., 2011), its main practical limitation is the high computational cost and processing time of the algorithms used to reconstruct the lightfield using compressed sensing theory. In contrast, both standard and focused plenoptic cameras place a microlens array (MLA) in front of the sensor in order to allow for a fast, efficient capture of the lightfield. In the standard plenoptic camera, the sensor is located at the focal length of the MLA. In contrast, in the focused plenoptic camera, the MLA is focused at the focal plane of the main lens. Commercial versions of the standard and focused plenoptic cameras are manufactured by Lytro and Raytrix, respectively.

The main advantage of the focused plenoptic camera is the improved spatial resolution on the sampled light-field. Alternatively, the standard plenoptic camera provides a more compact and flexible design, at a significantly reduced cost. Since the public release of the first commercial version of the standard plenoptic camera, namely the Lytro camera, these features have attracted the interest of the community and fostered research in applications of plenoptic imaging. However, in the literature, most efforts in camera modeling and calibration have been devoted to the disparity cue (Johannsen et al., 2013; Zeller et al., 2016; Strobl and Lingenauber, 2016; Heinze et al., 2016). In contrast, current calibration models for the focus of standard plenoptic cameras are not compatible with commercial devices since they require direct control and knowledge of the internal parameters of the camera (Hahne et al., 2014; Hahne et al., 2015).

This paper tackles the problem of modeling and calibrating the focus for metric depth estimation in standard plenoptic cameras. As previous researchers have shown, the main depth cues in plenoptic images are two: focus and disparity (Tao et al., 2013). Therefore, absolute depth z can be encoded either by means of the refocusing parameter, say ρ, and disparity. Unfortunately, to the best of our knowledge, to date there are not validated models for focus-based

* Corresponding author at: Universidad Industrial de Santander, 680002 Bucaramanga, Colombia.
E-mail address: spertuz@uis.edu.co (S. Pertuz).

[1] In the literature, standard plenoptic cameras have been also referred to as unfocused plenoptic cameras, multi-focus plenoptic cameras or plenoptic camera 1.0. In this work, we have adopted the term *standard plenoptic camera.*

https://doi.org/10.1016/j.isprsjprs.2018.06.020
Received 28 February 2018; Received in revised form 26 June 2018; Accepted 28 June 2018

Fig. 1. Digital refocusing of standard plenoptic camera. Left: refocusing with $\rho = -1.0$. Right: refocusing with $\rho = 1.0$. Without a proper model, the real metric in-focus distance corresponding to each image is unknown.

metric depth estimation, which poses important limitations for its application in different computer vision problems. For illustration purposes, Fig. 1 shows a synthetically refocused lightfield at two different distances. Without a proper model, the real focus distance would remain unknown as a function of the refocusing parameter ρ. [2]

In this work, we propose a closed-form analytical model that relates the absolute depth z of an imaged point with the refocusing parameter ρ. Based on this model, we propose a simple, reproducible calibration method for absolute depth estimation.

The contribution of this paper is twofold:

- We propose a closed-form model that relates absolute depth z with the focusing parameter ρ in standard plenoptic cameras.
- We propose and validate an efficient method for the calibration of the focus cue in standard plenoptic cameras.

Although depth estimation is out of the scope of this work, in addition to the aforementioned contributions, we provide a benchmark with ground-truth data suitable for quantitative assessment of depth estimation in standard plenoptic cameras. The contributions of this work are of interest to the community for future work in the development of metric depth estimation methods, the fusion of multiple cues for metric depth estimation, the comparison of different depth cues, and the objective comparison of depth estimation methods in standard plenoptic cameras.

2. Related work

The most closely related work to our approach can be found in Hahne et al. (2015). In that work, they derived a closed-form solution for the real depth z and a refocusing parameter defined by the authors. With that model, they predicted the depth z of the refocused image for a distance range between 43 and 890 cm approximately. The model proposed by Hahne et al. (2015) has three limitations: first, it requires accurate knowledge of the internal parameters of the camera, such as the focal length of the microlenses, the distance between the MLA and the main lens, and the locations of the principal planes of the main lens system. Secondly, it requires a careful calibration setting with perfectly aligned optics, which can be very difficult to meet in practice. Thirdly, the model in Hahne et al. (2015) has only been validated with simulated images. In contrast, our method does not depend explicitly on internal parameters since they can be found by means of a simple calibration process. In addition, our model is validated using real images.

Calibration methods that do not require explicit knowledge of all the internal parameters of the camera have been studied for focused plenoptic cameras (Zeller et al., 2016; Johannsen et al., 2013; Strobl and Lingenauber, 2016; Heinze et al., 2016). In that scope, calibration of intrinsic camera parameters is also performed similarly as with standard photography cameras (Zhang, 1999). In Zeller et al. (2016) a laser range finder is used to provide metric reference depth measurements for the calibration. In Johannsen et al. (2013), Strobl and Lingenauber (2016) and Heinze et al. (2016) they use planar calibration targets for estimating internal camera parameters that allow mapping the *virtual depth* provided by Raytrix cameras to real depth values. Very recently, a method for the geometric calibration of standard plenoptic cameras has been developed in Bok et al. (2017). All these methods allow for absolute depth estimation based on pixel disparity only. That is, focus is not modeled as depth cue. In contrast to the aforementioned approaches, we propose a focus model tailored for standard plenoptic cameras (Section 3). In addition, based on the proposed model, we derive a calibration method to allow for focus-based metric depth estimation in standard plenoptic cameras (Section 4).

One of the most attracting features of plenoptic cameras is the possibility of exploiting the rich angular and spatial information contained in the lightfield. This has motivated a lot of research efforts in depth estimation based on different concepts, such as EPI representations (Criminisi et al., 2005; Wanner and Goldluecke, 2012), focus stacks (Subbarao and Surya, 1994; Nayar and Nakagawa, 1994; Lin et al., 2015), multiview stereo (Bishop and Favaro, 2012; Sabater et al., 2015), and their combinations (Tao et al., 2013; Jeon et al., 2015; Wang et al., 2016; Tao et al., 2017). However, all these works provide depth estimates in method-specific depth units, which are not linearly related to metric real-world distances. The method presented here tackles this problem and provides an analytical model with a calibration procedure for focus-based metric depth estimation.

3. Proposed focus model

One of the main features of standard plenoptic cameras is to allow digitally refocusing captured images (Ng et al., 2005). That is, the captured image can be synthetically refocused to a different focus distance as a function of a refocusing parameter, say ρ. In general, ρ can take any real value. However, in practice, the amount of refocusing is limited by the resolution of the sampled lightfield. In order to derive a closed-form model that relates the refocusing parameter ρ with the metric focus distance z, we first describe the decoding process of the lightfield in Section 3.1. The refocusing model that relates the metric focus distance z with the refocusing parameter ρ is presented in Section 3.2.

3.1. Lightfield decoding

The light flowing through space can be described in terms of the plenoptic function. The plenoptic function, as described by Adelson and Bergen (Adelson and Bergen, 1991), is a 7-dimensional field $P(x, y, z, \gamma, \theta, \lambda, t)$ that depends on the spatial coordinates (x, y, z), the direction of light rays (γ, θ), the wavelenght λ and time t. In order to understand how refocusing is performed in a standard plenoptic camera, it is important to describe the sampling of the plenoptic function that takes place in the imaging process.

Without loss of generality, using the RGB representation of color and considering static scenes, the wavelength λ and temporal dimension t can be dropped in order to consider a 5-dimensional plenoptic function $P(x, y, z, \gamma, \theta)$. In order to further reduce the dimensionality of the plenoptic function, Levoy and Hanrahan proposed a representation with only four parameters (Levoy and Hanrahan, 1996), namely the lightfield $L(u_1, u_2, s_1, s_2)$. This representation allows to parametrize the plenoptic function within a finite space delimited by two planes in terms of angular coordinates (u_1, u_2) and spatial coordinates (s_1, s_2). For further details on the two-plane representation of the lightfield, we refer the reader to Levoy and Hanrahan (1996) and Dansereau (2014).

In order to sample the lightfield, standard plenoptic cameras place a microlens array (MLA) in front of the sensor of a conventional camera (see Fig. 2). The 2D sensor is located at a distance β behind the MLA, where β is the focal length of the microlenses. The aim is to demultiplex the angular information of the lightfield on the pixels behind each

[2] In this work, the refocusing parameter ρ must not be confused with the blur radius or blurring circle of conventional digital cameras (Pertuz et al., 2015).