Pixel-wise varifocal camera model for handling multilayer refractions

Longxiang Huang✉, Xu Zhao and Yuncai Liu

When observing a scene by a perspective camera through multiple refractive planes (e.g. underwater imaging), distortions and illusions, mainly caused by the refractions occurring at the interfaces between the different mediums, will be brought into the imaging system. Traditional pin-hole camera model becomes invalid in handling such case. A novel general camera model is presented, which is in fact a virtual pixel-wise varifocal model and can be used to encode multilayer refractions effectively. The model is built up based on two important findings. One is that the air between the camera centre and the nearest layer can be modelled as an air layer. The other one is that a ray passing through the air layer and these multiple layers causes only lateral displacement without changing its direction. The proposed camera model guarantees fast calculation of the backward projection. The forward projection equation is also derived and an efficient solving algorithm is proposed. Synthetic and real experiments are designed to verify the effectiveness and efficiency of the proposed theory.

Introduction: A perspective camera, capturing images of objects through multiple refractive planes (Fig. 1*a*), causes distortions owing to refractions. The distortions are related to the scene structure and hard to be corrected [1]. Until recently, most studies have focused on two-layer case (e.g. an underwater camera), where the pin-hole camera looks into water through parallel glasses. The results obtained from the underwater camera [1–5] are difficult to be generalised to multilayer imaging systems. To the best of our knowledge, Agrawal *et al.* [3] have proposed an axial camera model for calibrating of a multilayer setup, and have derived the backward projection (BP) equation and the forward projection (FP) equation. However, they failed to provide an effective and efficient solution for multilayer situation. In this Letter, our objective is to model a practical multilayer camera model for not only deriving but also solving the BP and FP equations.

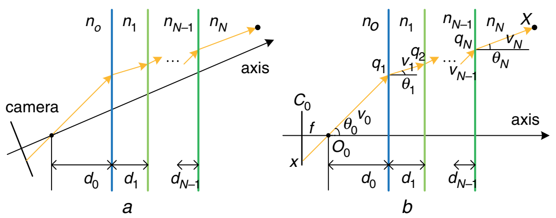

Fig. 1 *Multilayer flat refractive geometry*

a Image plane is non-parallel to interfaces
b By rotation, make image plane be parallel to interfaces

Multilayer camera model: We employ the concept of rotating the camera around the camera centre (Fig. 1*b*). The rotation is bijective; hence, it is easy to be implemented after axis calibration. Furthermore, the problem is converted into modelling the virtual camera, C_0, and deriving the camera model is divided into three stages.

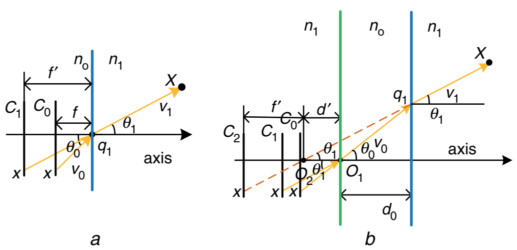

Fig. 2 *One layer geometry*

a Camera centre is on interface
b Camera centre is not on interface

First, assuming that there is only one flat refractive interface and that the camera centre is on the interface (Fig. 2*a*). In this case, similar to the pin-hole camera model analysis, all the light rays projecting on the image plane converge at the camera centre. The idea of a virtual camera is introduced to represent this imaging system. Assuming that a virtual camera, C_1, is designed for the medium in which the object is submerged, there are no refractions between C_1 and the objects. To project onto the same image point, $\boldsymbol{x}$, according to Snell's law, the virtual camera's focal length should be

$$f' = f\frac{\sqrt{n_1^2/n_0^2 - \sin^2\theta_0}}{\sqrt{1-\sin^2\theta_0}}, \quad \sin\theta_0 = \frac{\|\boldsymbol{x}\|_2}{\sqrt{\|\boldsymbol{x}\| + f^2}} \tag{1}$$

where θ_0 is the incident angle in air, n_0 and n_1 are the refractive indices of the corresponding mediums, and f is the focal length of the camera, C_0. As the incident angle is related to the position of the image point (pixel-wise different), C_1 is a pixel-wise varifocal camera.

Next, assuming that there is still one flat refractive interface but the camera centre is not on the flat interface (Fig. 2*b*). Imaging via C_0 through the interface can be considered as the virtual pixel-wise varifocal camera, C_1, derived in the first step, imaging through a parallel air layer. To obtain the same image point without the air layer for the space point, we should translate C_1 from O_1 to O_2 along the camera's optical axis, where the virtual camera, C_2, is located. As per the analysis in [4], the camera (viewpoint) translation (Fig. 2*b*) is

$$d' = d_0\left(1 - \frac{\sqrt{1-\sin\theta_1^2}}{\sqrt{(n_0/n_1)^2 - \sin\theta_1^2}}\right), \quad \sin\theta_1 = \frac{n_0\sin\theta_0}{n_1} \tag{2}$$

For the same reason as the pixel-wise varifocal length of C_1, the viewpoint of C_2 also has a pixel-wise variation. Consequently, C_2 is a virtual camera with a pixel-wise varifocal and viewpoint variation. We can observe that the camera centre on the interface is a special case of the camera centre not on the interface, i.e. $d_0 = 0$.

Third, we extend this to a multilayer situation. Similar to single layer situation, camera C_0 imaging through N layers is equivalent to the virtual camera, C_1, imaging through an air layer and the original N layers (Fig. 3). The pixel-wise focal length of camera C_1 is

$$f' = f\frac{\sqrt{n_N^2/n_0^2 - \sin^2\theta_0}}{\sqrt{1-\sin^2\theta_0}} \tag{3}$$

In addition to the image point, the pixel-wise focal length concerns only with the refractive indices of the mediums that the camera and the objects are submerged in. Then, similar to the step two, camera C_1 is translated to a new position (camera C_2) along the camera axis; consequently, camera C_2 can capture the objects without refractions.

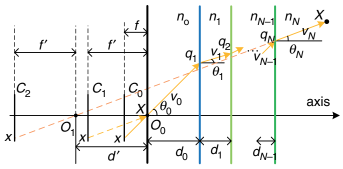

Fig. 3 *Flat refractive geometry with N interfaces*

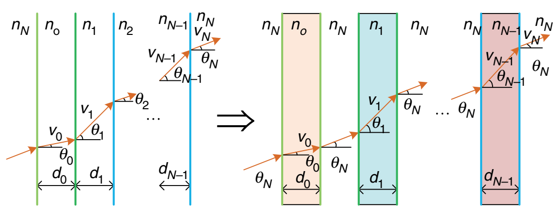

Fig. 4 *Light ray passes through multilayer flat mediums*

In optics, a ray passing through a parallel layer causes only a lateral displacement with no change in its direction; hence, a ray passing through multiple parallel layers can be divided into rays passing through each parallel layer, separately (Fig. 4). The total virtual viewpoint translation of camera, C_2, is presented in a simple form

$$d' = \sum_{i=0}^{N} d_i', \quad d_i' = d_i\left(1 - \frac{\sqrt{1-\sin^2\theta_N}}{\sqrt{(n_i/n_N)^2 - \sin^2\theta_N}}\right) \tag{4}$$

where the subscript, i, represents the correspondence parameters of the ith layer and $n_0 \sin\theta_0 = n_N \sin\theta_N$. It is easy to see that the total viewpoint translation is also a pixel-wise variation.

The proposed pixel-wise camera model has the form of a traditional pin-hole camera model; therefore, we can solve several problems in multilayer vision using the well-developed theory in traditional computer vision. Additionally, the proposed model ensures efficient BP and FP computations.

Backward projection: Similar to the pin-hole-based camera model, BP can be conducted after computing the virtual pixel-wise focal length and camera centre, using (3) and (4). Furthermore, as the computation complexity of our proposed method is linear, parallel computation can be introduced to compute different translation distances according to the different layers, which is absent in the previous studies.

Forward projection: Assuming that the world coordinates coincide with the camera coordinates, the space point, $\boldsymbol{X} = (R_X,\ Z)^{\mathrm{T}}$, is projected at the image point, $\boldsymbol{x} = \left(r_x,\ f\right)^{\mathrm{T}}$, passing through multilayer parallel mediums. As illustrated in Fig. 4, the relationship between the space point and the incident angle in the jth medium is

$$\tan\theta_j = \frac{R_X}{Z - d'} = \frac{\alpha}{\sqrt{1-\alpha^2}} \tag{5}$$

where

$$d' = \sum_{i=0}^{N} d'_i = \sum_{i=0}^{N} d_i\left(1 - \frac{\sqrt{1-\alpha^2}}{\sqrt{\left(n_i/n_j\right)^2 - \alpha^2}}\right), \quad \alpha = \sin\theta_j$$

Then, we select α as variable and obtain a nonlinear function as the FP equation

$$F(\alpha) = \alpha\left(Z - \sum_{i=0}^{N} d'_i\right) - R_X\sqrt{1-\alpha^2} = 0 \tag{6}$$

According to Snell's law, if a ray can pass through these layers, the incident angle in the jth medium should satisfy $0 \le \alpha < \min\left(n_j/n_i,\ 1\right),\ i = 0\ldots N$. In the case of (6)

$$\begin{aligned}\frac{\mathrm{d}F}{\mathrm{d}\alpha} &= \left(Z - d'\right) + \alpha\sum_{i=0}^{N}\frac{\mathrm{d}d'_i}{\mathrm{d}\alpha} + R_X\alpha E_1 \\ \frac{\mathrm{d}d'_i}{\mathrm{d}\alpha} &= d_i\alpha\left(E_1E_2 - E_1^{-1}E_2^3\right)\end{aligned} \tag{7}$$

where $E_1 = (1-\alpha^2)^{-(1/2)}$, $E_2 = ((n_i/n_j)^2 - \alpha^2)^{-(1/2)}$. We select the incident angle in the medium with the smallest refractive index, $n_{\min} = \min(n_i),\ i = 0\ldots N$ and $\alpha = \sin\theta_{\min}$; hence, $E_1 > 0$ and $E_2 > 0$. As $R_X \ge 0$, $n_i \ge n_{\min}$, and $(\mathrm{d}F/\mathrm{d}\alpha) \ge 0$ hold; therefore, $F(\alpha)$ is a monotonic increasing function in the meaningful solution interval, which ensures global convergence using Newton's method for determining α from any value in the meaningful solution interval. After computing α, it is easy to obtain the corresponding image point, $\boldsymbol{x}$.

Experimental results: We test our theory on synthetic data and real images. In the synthetic experiment, we select 10^6 random image points with one/two/three layers to compute the BP lines and FP by using our method and the method in [3]. As the compared FP method did not provide solutions for more than two refractions, we use N/A instead (Table 1). Synthetic experiments show that both the methods obtain exactly the same BP line and the FP image points with the same accurate level (10^{-15} pixels). We also verified that our method is faster for the BP and FP computations than [3] (Table 1).

Table 1: Mean time of computing BP and FP (unit: microseconds). One/two layers are cases 1 and 3 in [3], respectively

Method	BP-1	BP-2	BP-3	FP-1	FP-2	FP-3
Proposed	9.06	12.00	12.71	160.6	196.5	219.5
[3]	9.74	16.62	23.60	183.9	395.6	N/A

In the real experiment, image pairs of the checkerboards in a water tank are captured by a stereo camera (Fig. 5) composed of two action cameras placed in waterproof housings (air–glass–water). First, we capture the checkerboards without water and recover the 3D position of the checkerboard corner as the ground truth. Next, we capture the objects in water with the stereo camera in the same position. Then, we extract the corners of the image and project the 3D points computed in the first step on the image (Fig. 6*a*). Finally, we recover the space point (Fig. 6*b*). The compared method is based on pin-hole camera model, which is represents as single viewpoint perspective (SVP) in Fig. 6. It is easy to find that it will introduce huge errors in FP and reconstruction based on pin-hole camera model. We also find that the proposed theory can be used to compute the FP and to reconstruct the space points with high accuracy.

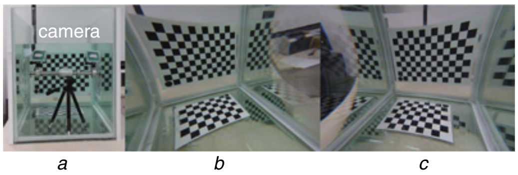

Fig. 5 *Underwater camera for object imaging in water tank*

a Stereo underwater camera
b and *c* Left and right images captured in water

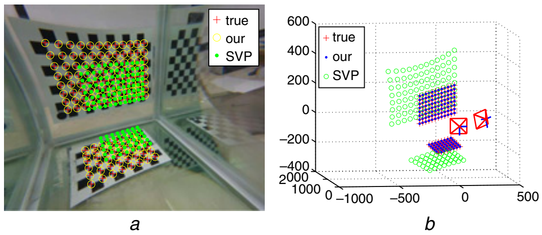

Fig. 6 *Real experiment results*

a Projection
b Reconstruction

The experiments validated the performance of our model. Our proposed methods for computing the BP line and solving the FP equations achieve higher efficiency over the previous method.

Conclusion: In this Letter, we propose a pixel-wise varifocal and variable viewpoint camera model for multilayer imaging systems. The proposed theory can be used in calibration and reconstruction for multilayer imaging systems. Synthetic and real experiments are conducted to verify the effectiveness and efficiency of our proposed theory. In future, we plan to utilise our model to develop a structure-from-motion system with explicit consideration of multilayer refractions, particularly for underwater applications.

Acknowledgments: This work has been supported by NSFC 61375019 and 61673269.

Submitted: *4 May 2017* E-first: *23 June 2017*
doi: 10.1049/el.2017.1648
One or more of the Figures in this Letter are available in colour online.

Longxiang Huang, Xu Zhao and Yuncai Liu (*Department of Automation, Shanghai Jiao Tong University, Shanghai, People's Republic of China*)

✉ E-mail: hlxiang@sjtu.edu.cn

References

1 Treibitz, T., Schechner, Y., Kunz, C., and Singh, H.: 'Flat refractive geometry', IEEE *Trans. Pattern Anal.*, 2012, **34**, (1), pp. 51–65
2 Kang, L., Wu, L., Wei, Y., and Yang, Z.: 'Theory of multi-level refractive geometry', *Electron. Lett.*, 2015, **51**, (9), pp. 688–690
3 Agrawal, A., Ramalingam, S., Taguchi, Y., and Chari, V.: 'A theory of multi-layer flat refractive geometry'. CVPR, Providence, RI, USA, June 2012, pp. 3346–3353
4 Huang, L., and Liu, Y.: 'Camera calibration for plate refractive imaging system'. ICPR, Stockholm, Sweden, August 2014, pp. 4068–4073
5 Yau, T., Gong, M., and Yang, Y.: 'Underwater camera calibration using wavelength triangulation'. CVPR, Portland, OR, USA, June 2013, pp. 2499–2506