Natural Stereo Camera Array using Waterdrops for Single Shot 3D Reconstruction

Akira Furukawa, Fumihiko Sakaue and Jun Sato
Nagoya Institute of Technology, Gokiso, Showa, Nagoya, Japan

Keywords: Natural Stereo Camera System, Water Drop Stereo, Camera Parameters By Optical Aberration.

Abstract: In this paper, we propose a stereo 3D reconstruction from a single image including multiple water drops. Water drops on a surface, e.g. camera lens, refract light rays and the refracted rays are roughly converged to a point. This indicates that water drops can be regarded as approximately small lens. Therefore, sub-images refracted by water drops can be regarded as images taken from different viewpoints. That is, virtual stereo camera systems can be constructed from a single image by using these raindrop characteristics. In this paper, we propose an efficient description of this virtual stereo camera system using water drops. Furthermore, we propose methods for the estimation of the camera parameters and for the reconstruction of the scene. We finally display several experimental results and discuss the validation of our proposed camera model from the results.

1 INTRODUCTION

In field of computer vision, 3D reconstruction is one the most traditional and important aspects, and hence, various methods have been studied extensively(Newcombe et al., 2011; Klein and Murray, 2009; Cheung et al., 2000; Posdamer and Altschuler, 1982; Kolmogorov and Zabih, 2002). In general, 3D reconstruction methods require two or more than two cameras since depth information is lost in a single image. Therefore, multiple camera systems, known as stereo camera systems, are used for 3D scene reconstructions in general.

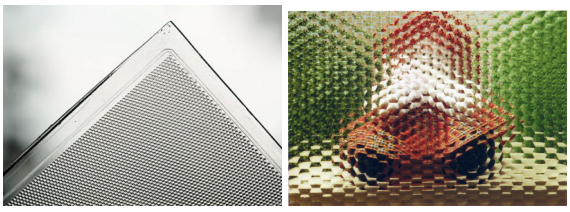
However, the stereo camera system does not always require explicit multiple cameras because the system just requires images that are taken from different viewpoints. Therefore, 3D reconstruction can be achieved by using single camera if the camera can take such images. The most representative 3D reconstruction method that uses a single camera is Structure from Motion (SfM)(Klein and Murray, 2009; Newcombe et al., 2011). In this method, a single camera moves around the target object and takes several images from different viewpoints. From these images, the 3D shape of the target object can be reconstructed by using ordinary stereo reconstruction techniques. Although these methods are considerably convenient since the method can be utilized by using only a single camera, the method cannot be applied when the target object is not rigid. This is because stereo correspondences that are used for stereo reconstruction may be changed when the target object is not rigid, and thus, the stereo constraints for 3D reconstruction are not satisfied in this case.

In order to avoid this problem, another approach that uses a special lens, as shown in Fig.1(a) and 2(a), are proposed. In this approach, multiple images that are taken from different viewpoints are virtually obtained from a single image. These methods can be classified into two techniques based on lens size and lens position.

In the first technique, micro-lens array as shown in Fig 1(a) is used(Levoy and Hanrahan, 1996; Chen et al., 2014). The micro-lens array is equipped on the image plane such as a CCD and it achieves special image photography as shown in Fig. 1(b). This special image includes multiple images taken from different viewpoints, and then, the image can be separated into multiple images. Therefore, 3D scenes can be reconstructed from the images by using an ordinary technique. In general, cameras that are equipped with micro-lens arrays are known as light-field cameras.

The second approach uses large-lens array as shown in Fig 2(b). The array is placed in front of the camera, and then, images that include multiple sub-images from different viewpoints can be taken directly as shown in Fig.2(b). By using the sub-images in the input image, 3D reconstruction can be achieved.

Furukawa, A., Sakaue, F. and Sato, J.
Natural Stereo Camera Array using Waterdrops for Single Shot 3D Reconstruction.
DOI: 10.5220/0007582609010907
In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019)*, pages 901-907
ISBN: 978-989-758-354-4
Copyright © 2019 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved



(a) micro-lens array (b) obtained image

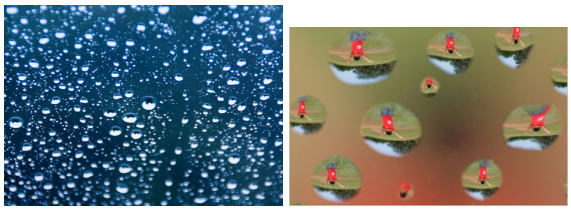
Figure 1: Multiple image photography using micro-lens array equipped on the image plane.

(a) lens array (b) obtained image

Figure 2: Multiple image photographing using lens array placed in front of the main camera.

The approaches using special lens arrays achieve stereo 3D reconstruction from single images. However, the special lens cannot be utilized always. Therefore, the convenience of these methods may become lesser than an ordinary stereo method using multiple cameras. In order to overcome this problem, several methods that use natural optical phenomena have been proposed(Arvind V. Iyer, 2018; You et al., 2016). In these methods, natural optical lenses such as water drops as shown in Fig.3(a) are utilized as lens arrays. When the water drops are placed in front of the camera, the input image includes multiple sub-images as shown in Fig.3(b). Therefore, 3D reconstruction can be achieved from a single image using natural objects alone.

Although these methods are significantly convenient, they require large computational cost for 3D reconstruction as the method has to compute complicated optical refractions by water drops that have complicated shapes. In this paper, we propose an efficient camera model for describing these complicated optical phenomena. In this model, we focus not on the shape of water drop, but rather on the optical refraction by the water drop; moreover, we do not reconstruct the shape of the water drop explicitly. In our model, the optical refraction is described using only a few parameters. Therefore, the computational cost for 3D reconstruction can be drastically reduced. In addition, the robustness of the reconstruction becomes higher as the constraint of the model is powerful. In this paper, we explain this camera model and indicate the calibration and 3D reconstruction method using the model as well.



(a) water drops (b) obtained image

Figure 3: When a water drop is put on the lens, the water drop can be regarded as natural lens array.

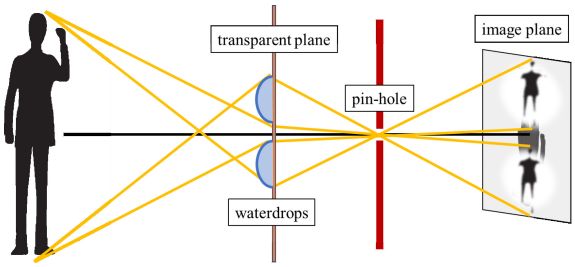


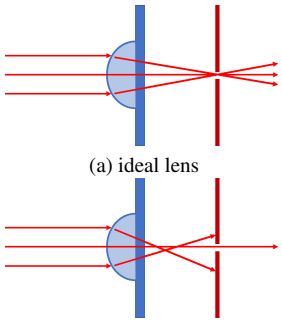
Figure 4: Overview of our proposed camera model.

2 WATER DROP CAMERA MODEL

We first define a camera model that describes the image projection from a 3D scene to the image plane using water drops. Figure 4 shows the overview of our camera model. As shown in this figure, we assume that the main lens of the camera can be approximated by a pinhole and water drops are placed on a transparent plane set in front of the main camera. In this case, input light rays are refracted by the water drops at first. Thereafter, the rays pass through the pinhole and are received by pixels on an image plane.

In general, a surface normal direction is required to compute the refraction of light rays. Specifically, explicit shape description and reconstruction is necessary for describing the behavior of the light rays to the image plane. In addition, light ray refractions by not only the water drops, but transparent planes as well should be considered for describing accurate light rays. When this description is used, the computational cost becomes burdensome if a lot of water drops are placed on the plane. Therefore, we propose an efficient description model that focuses on optical aberration by water drops.

When the water drop is an ideal optical lens, input light rays are converged to a pinhole as shown in Fig.5(a). However, the light rays are not converged into the point owing to inaccuracies such as complicated shapes of the lens in general, as shown in Fig.5(b). This phenomenon can be described by the



(a) ideal lens

(b) ordinary shape

Figure 5: Optical aberration caused by water drops. Figure(a) shows an ideal case where the light rays are converged to a point . Figure(b) shows an ordinary case where the light rays are not converged to a point.

optical aberration model(Tyson, 2010; Geary, 1995; Roddier, 2004) efficiently. In our case, water drops can be regarded as incorrect optical lenses; thereafter the behavior of the light rays can be described by the optical aberration model efficiently.

For this objective, Zernike polynomial is generally used as the model can describe the aberration efficiently. In this model, optical aberration is represented using a few parameters. Therefore, we employ the Zernike aberration model in our proposed method.

Here, we describe the details of the Zernike aberration model. In this model, optical aberration is represented by the linear combination of the Zernike bases and the bases are computed as follows:

$$Z_{nm}(\rho,\theta) = \sum_{s=0}^{\frac{n-m}{2}} \left(\frac{(-1)^s (n-s)! \rho^{n-2s}}{s!\left(\frac{n+m}{2}-s\right)!\left(\frac{n-m}{2}-s\right)!} \right) \begin{matrix} \cos|m|\theta : m \geq 0 \\ \sin|m|\theta : m < 0 \end{matrix} \quad (1)$$

where ρ and θ give the log-polar coordinates representation of the 2D image. Through this base, the optical aberration $W(X,Y)$ at point (X,Y) is represented as follows:

$$W(X,Y) = \sum_{n=0}^{k} \sum_{m=-n}^{n} B_{nm} Z_{nm}(\rho,\theta) \quad (2)$$

where B_{nm} is Zernike coefficient. ∇W can be computed by partial differentiation of W with respect to X and Y as follows:

$$\nabla \mathbf{W} = \begin{bmatrix} \frac{\partial W(X,Y)}{\partial X} \\ \frac{\partial W(X,Y)}{\partial Y} \end{bmatrix} \quad (3)$$

The ∇W represents the extent of the light rays refraction by the water drops directly.

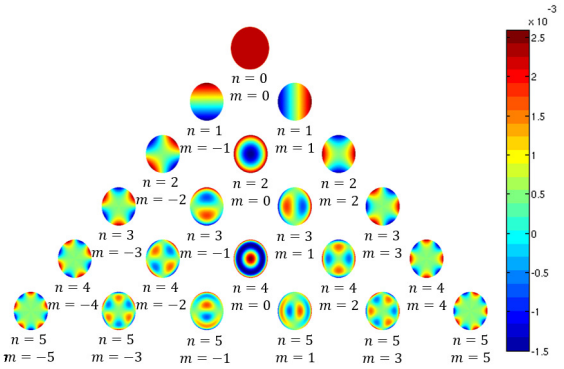


Figure 6: Zernike bases.

Figure 7: Examples of rendered images with different Zernike coefficients.

Figure 6 shows Zernike bases and they correspond to specific optical aberrations such as spherical aberrations. Coefficients of this equation represent the degree of each aberration. In general, ordinary optical aberration can be represented by the combination of a few bases; thereafter, we can describe the behavior of the light rays from water drops by using a few coefficients.

Figure 7 shows the examples of rendered images by our proposed model. In these figures, different Zernike coefficients are used for rendering the images. These figure show that the rendered images change drastically by just changing the Zernike coefficients.

3 CALIBRATION OF CAMERA

Next, we consider the calibration of our proposed camera model, i.e., the estimation of the Zernike coefficients is discussed in this section. In general, a wavefront sensor is utilized for measuring optical aberrations. The sensor is a type of light field camera and it measures the behavior of the wave directly. However, this sensor cannot be used in our case because the aberrated wave should be observed directly for measuring the aberration. To be precise, the sensor should be placed between the water drops and a main lens. It is not realistic a set up. Therefore, we estimate the coefficients by minimizing the image residual of ren-

dered images.

In this estimation, it is assumed that the information of the input scene such as scene shape and texture information is known. In this case, the input image can be virtually rendered again when Zernike coefficients $\mathbf{B}$ are provided. Let $I'(\mathbf{B})$ denote the rendered image by coefficients $\mathbf{B}$ and I denote an input image. In this case, image residual R is computed as follows:

$$R = I - I'(\mathbf{B}) \tag{4}$$

When a parameter $\mathbf{B}$ is equivalent to the correct parameter $\hat{\mathbf{B}}$, the residual R should become small. Therefore, the Zernike coefficient $\hat{\mathbf{B}}$can be estimated by minimizing the residual R as follows:

$$\hat{\mathbf{B}} = \arg\min_{\mathbf{B}} \|R\|^2 = \arg\min_{\mathbf{B}} \|I - I'(\mathbf{B})\|^2 \tag{5}$$

Note that, in general, point light sources that emit spherical waves or plane waves are used for estimating the aberration of the lens. In this case, optical aberration can be measured directly; thereafter, aberration parameters can be easily estimated. In our method, we can estimate the aberrations under the light source easily. In addition, our method can estimate the coefficients even if general light rays are input to the camera because our method focuses on the consistency of the whole input image.

4 SCENE RECONSTRUCTION

Here, we explain the 3D scene reconstruction by using a calibrated camera model. In an ordinary stereo method, the correspondences of feature points such as SIFT (Lowe, 1999) and SURF (Bay et al., 2006) are used for the reconstruction. In this case, the feature points are extracted from the input images at first. Thereafter, correspondences are determined from the feature points. Finally, these correspondences are reconstructed under epipolar constraints.

However, we cannot use the gold standard algorithm in our camera model because images for each camera, i.e., images by water drops do not have enough resolution for extracting feature points. Figure 8 shows the example of an input image. In this image, although a water drop provides images taken from different viewpoints, the provided image does not have enough resolution for extracting feature points.

To overcome this problem, we do not focus on feature points, but rather on whole sub-images provided by the water drop, similar to the calibration process. In this method, whole input images are backprojected to a 3D scene $\mathbf{S}$ and the scene is texture mapped by a sub image. Next, the 3D scene is reprojected to our

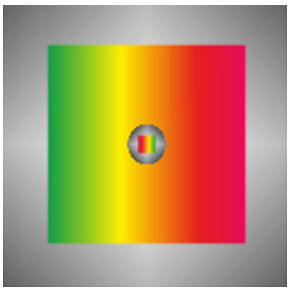


Figure 8: Example of input image with water drops.

camera model. Here, if the 3D scene $\mathbf{S}$ corresponds to a real scene, the reprojected image corresponds to the input image as well. Therefore, the 3D scene can be reconstructed by minimizing the difference of the input image and reprojected image. That is to say, the 3D scene $\hat{\mathbf{S}}$ can be estimated as follows:

$$\hat{\mathbf{S}} = \arg\min_{\mathbf{S}} \|I - P(T(I,\mathbf{S},\mathbf{B}),\mathbf{S},\mathbf{B})\|^2 \tag{6}$$

where P and T represent the projection and backprojection processes of the input image I with shape $\mathbf{S}$ and camera parameters $\mathbf{B}$, respectively.

In order to realize this estimation, the 3D scene $\mathbf{S}$ should be represented by a few parameters. For example, the Bezier curve is a representative parametric shape model. In this model, curves in the scene are determined by control points and the curves represent various 3D shapes by moving the control points. When the Bezier curves are used for scene representation, the 3D scene $\mathbf{S}$ can be described as follows:

$$S(u,v) = \sum_{i=0}^{n}\sum_{j=0}^{n} B_j^n(u)B_j^n(v)\mathbf{q}_{ij} \quad 0 \le u \le 1,\ 0 \le v \le 1 \tag{7}$$

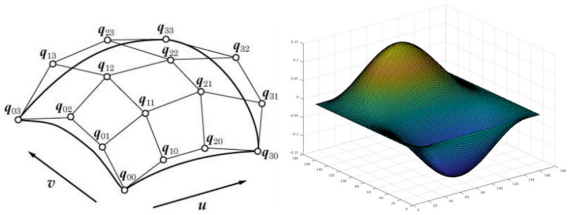
where $\mathbf{q}$ is the control point and B are Bernstein basis functions. The function is computed as follows:

$$B_i^n(u) = {}_nC_i(1-u)^{n-i}u^i \tag{8}$$

$${}_nC_i = \frac{n!}{i!(n-i)!} \tag{9}$$

By using the Bezier curves, whole 3D scenes can be represented and estimated by estimating the 3D position of the control points alone. For example, the 5-th order curve can be estimated by the estimation of 36 control points. Especially, only 36 parameters should be estimated when only the depth of the control points are changed. Figure 9 shows an example of the 5-th order Bezier curve. The curve can be easily changed by changing the position of the control points.

Note that the backprojection of an input image to the 3D scene is not very complicated in our model.



(a) control points (b) Bezier curves

Figure 9: Representation of control points and example of Bezier curves.

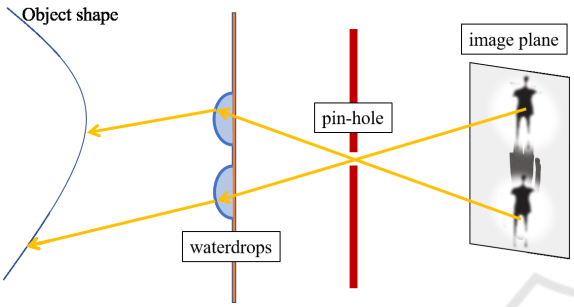


Figure 10: Backprojection of an input image.

As described in the previous section, the main camera of our camera model has a pin-hole, and only a single light ray is received by each pixel. Therefore, the light ray can be backtraced easily as shown in Fig.10. In addition, since the optical aberration model is used for the representation of light ray refractions, the refraction of the traced light ray can be computed directly compared to ordinary light refraction models.

5 EXPERIMENTAL RESULTS

In this section, we present several experimental results by our proposed method. We first explain the experimental environment. In all the experiments, a simulation environment was used for obtaining accurate ground truth. In the computer, an experimental environment was set up and several images were taken by a virtual camera. In front of the camera, a transparent plane was positioned and water drops were placed on the plane. The shape of the each water drop was different from each other. Therefore, light ray inputs to the drops had different behaviors. In the input image synthesis process, each light ray to the image plane was traced based on not our proposed model, but rather the physics rules for validating our proposed model.

In the estimation process, the refractions by water drops were represented by five Zernike coefficients. In front of the camera, a planar surface was placed, and several textures were mapped onto the plane. The plane was taken by the camera. Figure 11 shows the examples of input images and its target object. By using the images, the camera parameters and 3D scenes were estimated respectively.

(a) Target image (b) Input image

Figure 11: Target image and example of input image.

(a) input image (b) calibrated result

Figure 12: Input image and camera calibration result by using Zernike polynomials.

We first show the camera parameters, i.e. Zernike coefficients estimation results. As mentioned above, the input images were synthesized based on physics rules, and then, the validation of our camera model was evaluated in this experiment.

We first extract the region of the water drop by using Hough transformation roughly. After that, Zernike coefficients were estimated for each detected water drop respectively. In order to evaluate validation of our model, the target scene was projected to the image plane by using estimated coefficients. Figure 12 shows an example of the pair of input image and reprojected image using calibrated camera parameters. As shown in this figure, our camera model can synthesize images similar to the input image although they were based on different rules. The results indicate that our proposed camera model can represent the refraction of light rays effectively. In addition, our proposed calibration method can estimate camera parameters from ordinary input images.

We now present scene reconstruction results by using the calibrated parameters. In this estimation, only the depth of the target plane was estimated from a single input image. In order to evaluate the vali-

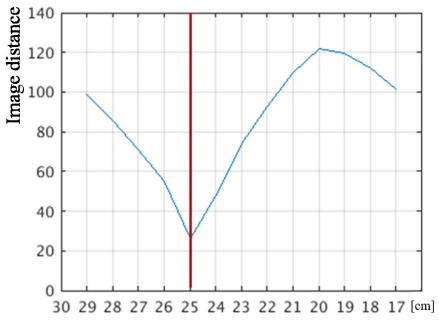


Figure 13: Depth estimation result.

dation of the estimation, we compute the difference of a reprojection image and an input image for each depth. Figure 13 shows the distance of the images at each depth. In this graph, the vertical axis indicates the distance of the images and the horizontal axis indicates the estimated depth. The red line in the figure shows correct depth. In this graph, the distance of the image becomes local minimum at the true depth. This, in fact, indicates that the distance between the reprojection image and input image represents the validation of the estimated depth. Therefore, our method can estimate the depth from a single image.

These experimental results indicate our proposed model can describe the behavior of the light rays efficiently and effectively. In addition, the calibration and reconstruction methods based on the model work efficiently.

6 SIMULTANEOUS ESTIMATION OF CAMERA PARAMETERS AND 3D SCENE

We finally discuss the simultaneous estimation of 3D shape and camera parameters from an input image. In an ordinary stereo method, simultaneous estimation of these parameters, known as bundle adjustment, can be achieved by minimizing the reprojection error of the correspondences. In fact, bundle adjustment in our framework can be achieved in a manner similar to the ordinary method. In our framework, instead of point reprojection error, image reprojection error should be minimized for 3D reconstruction and calibration. Therefore, simultaneous estimation can also be achieved by minimizing the same error.

In this simultaneous estimation, in addition to the parameters of the 3D shape, camera parameters are estimated as well. When the 3D scene is represented by an N-th order Bezier, $(N+1)^2$ parameters are required. In addition, L water drops require $L \times M$ (M is the number of coefficients) parameters for estimating the camera model. Totally, $(N+1)^2+LM$ parameters should be estimated for the simultaneous estimation. This, in fact, indicates that $(N+1)^2+LM$ or more than $(N+1)^2+LM$ constraints are necessary for estimating these parameters. In our proposed estimation, all pixels are used for this estimation, and then, sufficient number of constraints are obtained when the number of pixels are larger than $(N+1)^2+LM$.

7 CONCLUSION

In this paper, we propose 3D scene reconstruction from a single image using water drops. In our proposed method, water drops in the images are regarded as virtual cameras and the 3D shape is reconstructed by using the virtual cameras. For the efficient description of the virtual cameras, we utilize an optical aberration model by Zernike basis. By using the aberration model, complicated light ray refractions can be described via few coefficients. Furthermore, parametric 3D scene description is employed for estimating the 3D scene effectively. We show experimental results in the simulation environment and the results demonstrate the potential of our proposed method. In future work, we extend our proposed method to the simultaneous estimation of camera parameters and 3D scenes.

SCIENCE AND TECHNOLOGY PUBLICATIONS

REFERENCES

Arvind V. Iyer, J. B. (2018). Depth variation and stereo processing tasks in natural scenes. *J. Vis*, 18(6).

Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. In *Computer vision–ECCV 2006*, pages 404–417. Springer.

Chen, C., Lin, H., Yu, Z., Kang, S. B., and Yu, J. (2014). Light field stereo matching using bilateral statistics of surface cameras. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Cheung, G. K., Kanade, T., Bouguet, J.-Y., and Holler, M. (2000). A real time system for robust 3d voxel reconstruction of human motions. In *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*, volume 2, pages 714–720. IEEE.

Geary, J. M. (1995). *Introduction to Wavefront Sensors*. Society of Photo Optical.

Klein, G. and Murray, D. (2009). Parallel tracking and mapping on a camera phone. In *Proc. Eigth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'09)*, Orlando.

Kolmogorov, V. and Zabih, R. (2002). Multi-camera scene reconstruction via graph cuts. In *Computer Vision ECCV 2002*, pages 82–96. Springer.

Levoy, M. and Hanrahan, P. (1996). Light field rendering. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, pages 31–42.

Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, volume 2, pages 1150–1157. Ieee.

Newcombe, R. A., Lovegrove, S. J., and Davison, A. J. (2011). Dtam: Dense tracking and mapping in real-time. In *Proceedings of the 2011 International Conference on Computer Vision*, pages 2320–2327.

Posdamer, J. and Altschuler, M. (1982). Surface measurement by space-encoded projected beam systems. *Computer graphics and image processing*, 18(1):1–17.

Roddier, F. (2004). *Adaptive Optics in Astronomy*. Cambridge University Press.

Tyson, R. K. (2010). *Principles of Adaptive Optics*. CRC Press.

You, S., Tan, R. T., Kawakami, R., Mukaigawa, Y., and Ikeuchi, K. (2016). Waterdrop stereo. In *arXiv*.

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS