

# 3D MEASUREMENT COMBINING MULTI-VIEW AND MULTI-FOCUS IMAGES USING LIGHT FIELD CAMERA

T. Fuse<sup>1,\*</sup>, Y. Kajihara<sup>2</sup>

<sup>1</sup> Dept. of Civil Engineering, University of Tokyo, Hongo 7-3-1, Bunkyo-ku, Tokyo, 113-8656, Japan - fuse@civil.t.u-tokyo.ac.jp

<sup>2</sup> Hitachi, Ltd., Yoshida-cho 292, Totsuka-ku, Yokohama, Kanagawa, 244-0817, Japan - yuki.kajihara.fj@hitachi.com

## Commission II

**KEY WORDS:** Computational Photography, Light Field Camera, 3D Measurement, Multi-View Images, Multi-Focus Images, Structure from Focus, Bundle Adjustment

### ABSTRACT:

In recent years, the demand for inexpensive, simple, and highly accurate 3D measurement has been increasing. Representative methods, photogrammetry, and shape from focus (SfF) have limitations in terms of measurement time and labour. In order to solve them, computational photography (CP) has been proposed. A light field camera, based on CP, has also been developed. It has a feature to acquire multi-view and multi-focus images simultaneously in one shot. It is possible to perform 3D measurements with less time and labour for photographing and calculation processing using these images. In this study, we combined the photogrammetry as applied to multi-view images with the SfF as applied to multi-focus images using a light field camera. We applied the proposed method to a rigid body and verified its accuracy. We confirmed that the proposed method achieved more accurate results than the photogrammetry and the SfF method. Furthermore, we applied the proposed method to screws and cracks on walls of buildings and affirmed its applicability. Finally, we suggested future work on the developed method.

## 1. INTRODUCTION

In recent years, there has been a growing demand for high-precision inspection and quality control that utilises 3D measurement. Similarly, the demand for inexpensive, simple, and highly accurate 3D measurement has also been increasing. Typical examples of the passive measurement method are the photogrammetry and the shape from focus (SfF) method. The photogrammetry can perform high-precision measurements regardless of the brightness of the objects. However, when a regular pattern is seen on the surface, incorrect correspondence between the images may occur. The SfF method uses multiple images with different focal length. 3D measurement is performed by specifying the image with the smallest defocus at each pixel (Pertuz, 2013). The SfF method can perform high-precision measurement regardless of the regularity of the surface pattern. However, when the brightness of the object is too high, an error will occur in the evaluation of defocus. Thus, the photogrammetry and the SfF method have complementary properties. Additionally, an application of the methods requires multiple camera shots, thereby needing time and effort to calculate the relative position between images.

Computational photography (CP) attempts to exploit the cheaper and faster computing to overcome the physical limitations of a camera, such as dynamic range, resolution or depth of field, and extend the possible range of applications. The computational techniques encompass methods from modification of imaging parameters during capture to modern image reconstruction method from the captured samples (Raskar and Tumblin, 2006). CP aims to redefine the process of image processing technology using a conventional digital camera and acquiring more light information in a 3D space from an image sensor through computation (Adelson, Bergen, 1991; Adelson, Wang, 1992). CP is based on plenoptic function,

where, a 3D space is considered as an environment (light field) filled with light rays in an infinite number of directions and the environment is described using seven variables. CP captures the data of the image sensor as an intermediate product and obtains an image by post-processing that data.

A light field camera, composed of many small lenses (micro lens array) placed between the main lens and the image sensor, was developed. The light field camera can obtain information on the variables of the plenoptic functions with only one shot. The light reflected by the object during imaging is imaged by the micro lens array after passing through the main lens, decomposed into light beams, and recorded on the sensor. By processing the information recorded in the sensor, it is possible to acquire multiple images with different viewpoints (multi-view images) and different focuses (multi-focus images) simultaneously (Tao et al., 2013; Tao et al., 2015). The photogrammetry was applied using the only multi-view image, and 3D measurement was performed (Yang et al., 2016).

In this study, a 3D measurement method that integrates both the photogrammetry and SfF method using imaging data obtained from a light field camera was proposed.

## 2. COMPUTATIONAL PHOTOGRAPHY

### 2.1 Light Field

As mentioned before, the CP is based on plenoptic function  $P$ , which has seven variables:

$$I = P(X, Y, Z, \theta, \phi, \lambda, t) \quad (1)$$

where  $X, Y, Z =$  coordinates in 3D space

\* Corresponding author

$\theta, \phi$  = direction of a light ray  
 $\lambda$  = wavelength  
 $t$  = time

Conventional cameras can be interpreted as devices that record the luminance distribution for a range of  $(\theta, \phi, \lambda, t)$  in the light field. The role of the lens is to redirect the incident light towards the position of the image sensor.

The light field is represented by the passing point of a plane  $(x, y)$  and the angle of the ray  $(u, v)$ . The  $x-u$  plane records the coordinates  $(x, u)$ , when a ray on the  $x-z$  plane passes through the  $x$  axis (Figure 1). At this time, parallel light is represented as a vertical line, and light passing through the same point is represented as a horizontal line on the  $x-u$  plane.

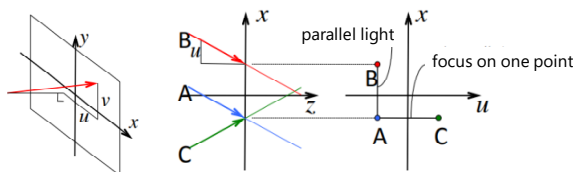


Figure 1. Representation of light field

A ray matrix is a linear transformation expressing the relationship between the light field and lens (Figure 2). When the original coordinates are  $(x, u)$  and the coordinates of the light ray approaching the lens by a distance  $d$  are  $(x', u')$ , the relationship between them is expressed using the following equation:

$$\begin{bmatrix} x' \\ u' \end{bmatrix} = \begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix}. \quad (2)$$

The coordinates  $(x', u')$  at the time when the light ray reaches the lens is expressed as follows by using the focal length  $f$ .

$$\begin{bmatrix} x' \\ u' \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -1/f & 1 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix}. \quad (3)$$

The coordinates of the refracted ray are obtained in the same way as when the ray reflected by the object reaches the lens.

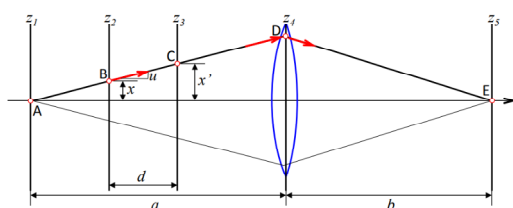


Figure 2. Relationship between ray matrix and lens

## 2.2 Multi-View and Multi-Focus Images Generation

Figure 3 shows a concept of multi-view images generation. It shows the trajectory of the light rays reflected by the target. The light rays are imaged by each micro lens, decomposed into light rays, and recorded in the sensors. In the situation, each micro lens collects light from each viewpoint to create sub-aperture image. When the number of sensors per micro lens is  $n$ ,  $n$  sets of multi-view images with parallax are acquired.

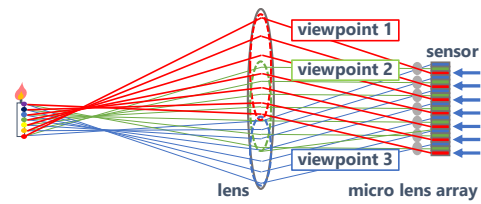


Figure 3. Multi-view images generation

Figure 4 shows a concept behind multi-focus images generation. It shows the same situation described in Figure 3. In the situation, the light ray reflected by an object between the target and the camera is recorded on a sensor covered by a different micro lens. For every light ray reflected by the object, the sensor where the ray is recorded is specified. The pixel values on the sensor are averaged, so that an image with a different focus can be obtained.

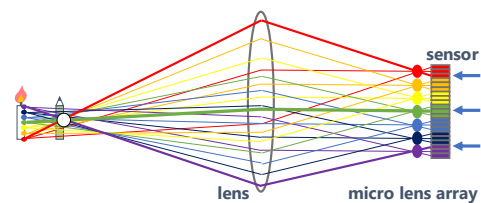


Figure 4. Multi-focus images generation

## 3. DEVELOPMENT OF INTEGRATING METHOD

### 3.1 Framework of the Proposed Method

A 3D measurement method integrating the photogrammetry and SfF method is proposed in this study. Figure 5 shows the framework of the proposed method. First, original binary data is converted into multi-view and multi-focus images by decoding input light field data.

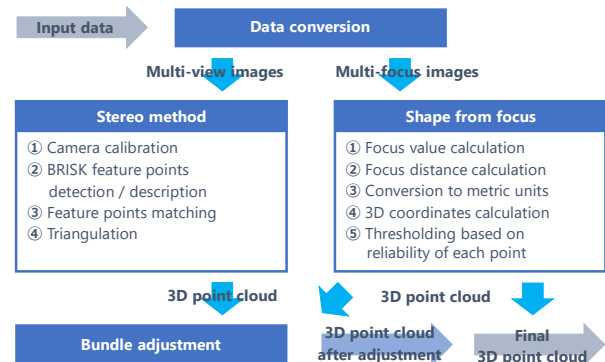


Figure 5. Framework of the proposed method

In the photogrammetry, the camera is calibrated for multi-view images. Feature points are detected and then matched using the images. A 3D measurement is performed by using the matched feature points.

In the SfF method, the focus values of every pixel are calculated for multi-focus images. The largest focus value is selected as the focused image at each pixel, and then the value is converted to the focus distance in metric units. From the focus distance values, 3D coordinates corresponding to the pixels can be calculated. Additionally, the reliability of each coordinate is

estimated, and coordinates with low reliability points are discarded.

In order to integrate photogrammetry and SfF methods, bundle adjustment is applied. The 3D coordinates of the feature points are obtained using the weighted sum of coordinates of the photogrammetry and SfF methods, accounting for the reliability from the SfF method. The 3D coordinates of the feature points and camera parameters are adjusted through the minimising re-projection errors.

### 3.2 Conversion of Light Field Data

To generate multi-view and multi-focus images, it is necessary to estimate the centre coordinates of the sensors corresponding to each micro lens (light field camera calibration) (Dansereau et al., 2013). A Bayer filter is placed on the surface of the sensors of the light field camera, and it is necessary to interpolate values (linear interpolation) to obtain RGB values (demaosaicing) (Malvar et al., 2004). According to the light field camera calibration, resampling is applied to the raw image (Tao et al., 2013). (Figure 6)

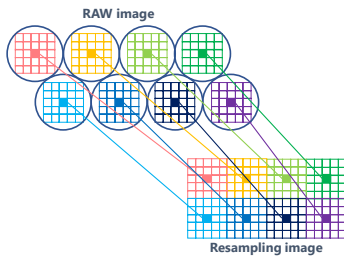


Figure 6. Resampling from raw image

**3.2.1 Multi-View Image Generation:** For the multi-view image generation, the relative coordinates  $(u, v)$  in the micro lens are set. By arranging the  $(u, v)$  coordinates of each micro lens according to the arrangement of the micro lens, an image at image coordinates  $(i, j)$  of the viewpoint  $(u, v)$  can be obtained. Figure 7 shows an example where  $(u, v) = (-1, -2)$ . The coordinates of the image with viewpoint  $(u, v)$ , namely  $(I_{aspect(u, v)}C_x(i, j), I_{aspect(u, v)}C_y(i, j))$ , can be expressed using the centre coordinates of the micro lens in the resampling image  $(I_{resample}C_x(i, j), I_{resample}C_y(i, j))$ :

$$\begin{aligned} I_{aspect(u, v)}C_x(i, j) &= I_{resample}C_x(i, j) + u \\ I_{aspect(u, v)}C_y(i, j) &= I_{resample}C_y(i, j) + v \end{aligned} \quad (4)$$

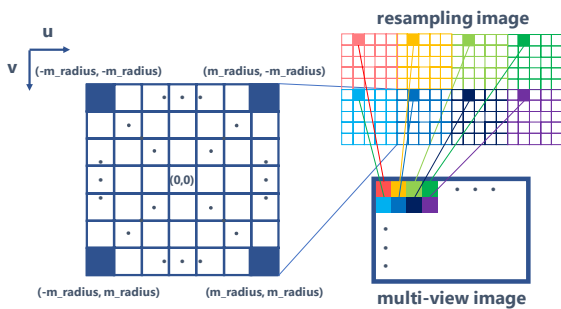


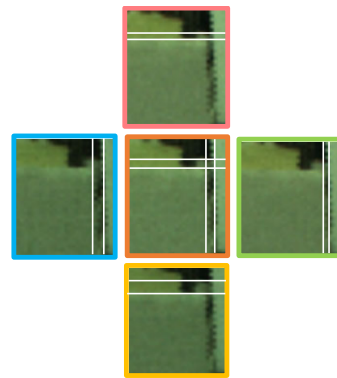
Figure 7. Multi-view image generation

Figure 8 shows an example of a generated multi-view image (a), and an enlarged view (b). The centre image in the figure is the

image at the viewpoint  $(u, v) = (0, 0)$ , and the upper, lower, left, and right images are  $(0, -m\_radius)$ ,  $(0, m\_radius)$ ,  $(-m\_radius, 0)$ , and  $(m\_radius, 0)$ , respectively. In the enlarged view, a white line indicating the corner of the same object is drawn to aid in comparing each image. Comparing the intervals between the white lines in the row and column directions for the five images, it can be seen that the appearance of the scene differs slightly depending on the image.



(a) generated images



(b) enlarged views

Figure 8. Example of multi-view image

**3.2.2 Multi-Focus Image Generation:** Figure 9 shows a concept of multi-focus image generation. In multi-focus image generation, the focal length at the time of shooting with a light field camera is set as  $f_{original}$ . The plane parallel to the image, whose distance from the image plane is  $f_{original}$ , is defined as the focus plane  $P_{original}$ . The plane parallel to  $P_{original}$ , which can generate a multi-focus image by refocusing, is defined as the new focus plane  $P_{refocus}$ . Let the distance between the image and  $P_{refocus}$  be  $f_{refocus} = \alpha f_{original}$ .

When the focus of the image is changed from  $P_{original}$  to  $P_{refocus}$ , a micro lens, which records the  $(u, v)$  light ray of the micro lens at the  $i$ -th row and  $j$ -th column of the resampling image, is specified. The coordinates of the image, namely  $(I_{refocus(\alpha)}C_x(i, j, u, v), I_{refocus(\alpha)}C_y(i, j, u, v))$ , can be expressed as follows:

$$\begin{aligned} I_{refocus(\alpha)}C_x(i, j, u, v) &= u(1-1/\alpha) + i \\ I_{refocus(\alpha)}C_y(i, j, u, v) &= v(1-1/\alpha) + j \end{aligned} \quad (5)$$

As the coordinates are not always integer values, bilinear interpolation is applied. The pixel value for all  $(u, v)$  is calculated. The average pixel value of each micro lens is the value of the micro lens in  $i$ -th row and  $j$ -th column after refocusing. In principle, viewpoint  $(0, 0)$  is the viewpoint of the multi-focus image at the centre of the micro lens.

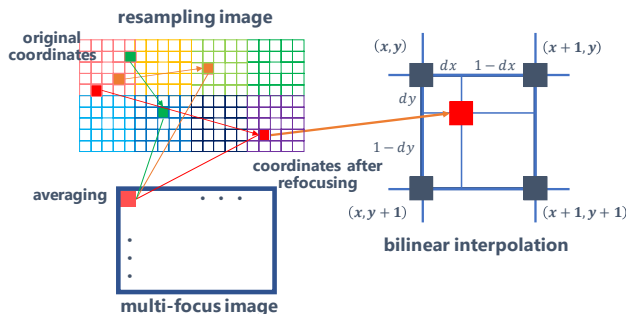


Figure 9. Multi-focus image generation

Figure 10 shows an example of a generated multi-focus (below), and an enlarged view (above). From the left, images with  $\alpha = 0.2$ ,  $\alpha = 1.0016$ , and  $\alpha = 2.0$  are shown.



Figure 10. Example of multi-focus image

### 3.3 Photogrammetric Method

In the photogrammetry, two sets of multi-view images are used. The first set is an image pair of the upper left viewpoint  $(-m\_radius, -m\_radius)$  and the lower right viewpoint  $(m\_radius, m\_radius)$  of the micro lens, and the second set is the lower left viewpoint  $(-m\_radius, m\_radius)$  and the upper right viewpoint  $(m\_radius, -m\_radius)$ .

**3.3.1 Camera Calibration:** First, the camera's interior orientation elements, exterior orientation elements, and the relative positions between images are estimated through camera calibration using Bouguet's method (Bouguet, 2004). Camera models include Zhang's pinhole camera model (Zhang, 2010) and Heikkila's lens distortion model (Heikkila, Silven, 1997). Camera parameters are represented by the following camera matrix:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = K [R | t] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (6)$$

$$= \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

where  $s = \text{scale}$   
 $u, v = \text{image coordinates}$   
 $K = \text{intrinsic matrix}$   
 $R = \text{rotation matrix}$   
 $t = \text{translation vector}$   
 $X, Y, Z = \text{world coordinates}$   
 $f_x, f_y = \text{focal length}$   
 $c_x, c_y = \text{principal point coordinates}$

In this study, 30 checkerboard patterns are imaged from different positions and angles using a light field camera, and a viewpoint image used for the photogrammetry is generated from each image data. After generating the multi-view images, the interior orientation elements of each image and the exterior orientation elements of the other image are calculated and optimised.

**3.3.2 Feature Points Detection and Matching:** In this study, typical detectors, such as FAST (Rosten, Drummond, 2006), minimum eigenvalue (Shi, Tomasi, 1994), Harris-Stephens (Harris, Stephens, 1988), binary robust invariant scalable keypoints (BRISK) (Leutenegger et al., 2011), SURF (Bay et al., 2008), KAZE (Alcantarilla, 2012), and MSER (Matas et al., 2004), are applied, and their performances are compared. Ultimately, Leutenegger's BRISK is adopted. BRISK is a corner detection and description method that expresses feature values in binary code. For feature point detection, a pyramid image obtained by reducing the input image in multiple scales is used to achieve scale invariance. In describing the feature values, binary codes are generated from the brightness differences between two points using 60 pixel values concentrically sampled at equal intervals, so that rotation invariance is guaranteed.

After applying feature point detection and description on each image, feature point matching of the image pair is performed. For feature point matching, Muja's method (Muja, Lowe, 2012), which is a binary feature point matching method, is used. In this method, a hierarchical tree is created from a set of input feature points, and the nearest neighbour search for the corresponding points of each feature point is performed to calculate the Hamming distance. The set which minimises the sum of squares of the Hamming distance of the feature point pair is returned as the output.

The result of the matching contains incorrect correspondence. We use M-estimator sample and consensus (MSAC) (Torr, Zisserman, 2000), which is a robust estimation, to eliminate false correspondence. MSAC is a variant of Random Sample Consensus.

**3.3.3 Intersection:** According to the result of the previous section, intersection (Hartley, Zisserman, 2003) is applied to calculated 3D coordinates. For the first image pair (upper-left and lower-right viewpoint images), 3D coordinates in the coordinate system of the upper-left viewpoint image are obtained. For the second image pair (lower-left and upper-right viewpoint images), 3D coordinates in the coordinate system of the lower-left viewpoint image are obtained. In order to integrate these 3D coordinates, the coordinate system of both pairs is adjusted to the camera coordinate system of the viewpoint  $(0, 0)$  using the calibration result of each image and the image of the viewpoint  $(0, 0)$ . The coordinates obtained by the intersection are converted into the coordinates of the viewpoint  $(0, 0)$  using equation (6).

### 3.4 Structure from Focus

The input of the SfF method includes all multi-focus images generated at different parameter  $\alpha$  values. The number of multi-focus images is denoted as  $n_{\text{refocus}}$ , and the  $\alpha_0$ th image is denoted as  $I_{\alpha_0}$ .

**3.4.1 Focus Value and Distance Calculation:** Focus value is an index that indicates how much of a pixel is in focus. For the first  $n_{\text{refocus}}$  images, the focus value of each pixel is calculated. The focus value is calculated using the Max-Min method. The method calculates the focus value as the difference between the maximum and the minimum points of the RGB values of the pixel of interest and the eight neighbouring pixels for all pixels of all images (Figure 11). It is based on the idea that the difference between the maximum and minimum RGB values is large in the in-focus area, and small in the blurred area.

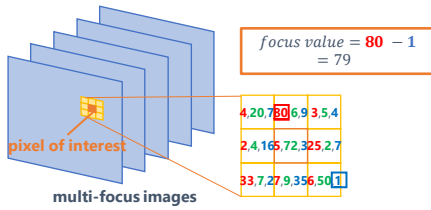


Figure 11. Focus value calculation

We focus on each pixel and compare the focus values in the first  $n_{\text{refocus}}$  images. The image with the largest focus value is regarded as the image that is best focused on that image, and the parameter  $\alpha$  value used when generating the image is defined as the focus distance (Figure 12).

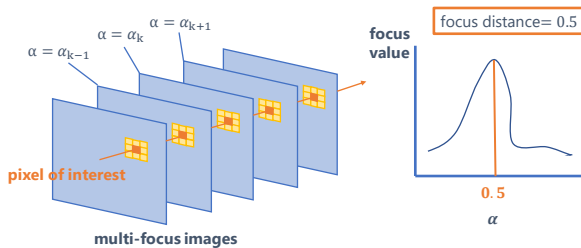


Figure 12. Focus distance calculation

**3.4.2 Conversion to Metric Units:** The focus distance in  $\alpha$  units obtained for all pixels is converted to meters by applying linear regression. For the linear regression, the object to be measured is placed at different distances from the light field camera, and images are taken at each distance to calculate the focus distance in  $\alpha$  units.

**3.4.3 3D Coordinates Calculation:** The 3D coordinates are calculated using the focus distance and the camera calibration results described in section 3.3.1. As the viewpoint of the multi-focus image is basically the same as the image of the viewpoint (0, 0), the interior orientation elements of the viewpoint (0, 0) are used. Let the focal length in pixel units be  $f$  and the image coordinates of the principal points in pixel units be  $(c_x, c_y)$ . The 3D coordinates  $(x(i, j), y(i, j), z(i, j))$  of the pixel  $(i, j)$  with viewpoint (0, 0) are acquired as follows:

$$\begin{aligned} x(i, j) &= \frac{j - c_x}{f} \cdot mI_{\alpha}(i, j) \\ y(i, j) &= \frac{i - c_y}{f} \cdot mI_{\alpha}(i, j) \\ z(i, j) &= mI_{\alpha}(i, j) \end{aligned} \quad (7)$$

where  $mI_{\alpha}(i, j)$  = focus distance

The 3D coordinates can be obtained for all pixels using the SfF method; however, the accuracy depends on the variance of the focus value. For example, in an area without texture, the focus value is small regardless of the choice of multi-focus image, and the difference between the images is also considered small. In this study, we defined the reliability of the 3D coordinates of each pixel and treated the points whose reliability is above a certain level as the final result obtained by the SfF method. The reliability of the 3D coordinates of each pixel is defined as the standard deviation of the focus values of the first  $n_{\text{refocus}}$  images of the multi-focus image.

### 3.5 Integration through Bundle Adjustment

Of the feature points measured using the photogrammetry, those that are also measured by the SfF method are integrated by weighting the 3D coordinates. The weights are determined by reliability of the 3D coordinates in SfF method.

$$\begin{aligned} P_{new(x,y,z)} &= (1 - w \cdot Conf(x, y)) \cdot P_{stereo(x,y,z)} \\ &+ w \cdot Conf(x, y) \cdot P_{sff(x,y,z)} \end{aligned} \quad (8)$$

where  $P_{new(x,y,z)}$  = 3D coordinates after integration

$P_{stereo(x,y,z)}$  = 3D coordinates obtained by photogrammetry

$P_{sff(x,y,z)}$  = 3D coordinates obtained by SfF method

$Conf(x,y)$  = reliability of coordinates in SfF method

$w$  = adjustment parameter

Once the initial value of the orientation elements and 3D coordinates of the feature points are acquired, bundle adjustment can be applied (Luhmann *et al.*, 2014). The feature points have coordinates  $(X_i, Y_i, Z_i)$ . Each image has a 3D coordinates  $(X_j, Y_j, Z_j)$  as the image position. At image  $j$ , the feature point  $i$  has camera coordinate system  $(x_{ij}, y_{ij})$ . A transformation between the camera coordinate and the world coordinate systems represents collinearity equation.

$$\bar{x}_{ij} = x_{ij} - \Delta x_{ij} = -c \frac{a_{11}(X_i - X_j) + a_{12}(Y_i - Y_j) + a_{13}(Z_i - Z_j)}{a_{31}(X_i - X_j) + a_{32}(Y_i - Y_j) + a_{33}(Z_i - Z_j)} \quad (9)$$

$$\bar{y}_{ij} = y_{ij} - \Delta y_{ij} = -c \frac{a_{21}(X_i - X_j) + a_{22}(Y_i - Y_j) + a_{23}(Z_i - Z_j)}{a_{31}(X_i - X_j) + a_{32}(Y_i - Y_j) + a_{33}(Z_i - Z_j)}$$

where  $c$  = focal length

$\Delta x_{ij}, \Delta y_{ij}$  = factors of interior orientation

$a_{kl}$  = factors of rotation matrix

The position and rotation updates are computed iteratively by minimising an objective function of the re-projection error. In order to solve the bundle adjustment problem, Levenberg-Marquardt method (Hartley, Zisserman, 2004) is applied. The objective function  $E$  is approximated using the following formula:

$$E(\mathbf{x} + \delta \mathbf{x}) \approx E(\mathbf{x}) + \mathbf{g}^T \delta \mathbf{x} + \frac{1}{2} \delta \mathbf{x}^T H \delta \mathbf{x} \quad (10)$$

$$(H - \lambda I) \delta \mathbf{x} = \mathbf{g} \quad (11)$$

where  $\mathbf{g} = \left. \frac{dE}{d\mathbf{x}} \right|_{\mathbf{x}}$  (gradient)

$H = \left. \frac{d^2E}{d\mathbf{x}^2} \right|_{\mathbf{x}}$  (Hessian)

$\lambda$  = dumping factor

The variable  $\mathbf{x}$  expresses parameters vector, not image coordinates.

#### 4. EXPERIMENTAL RESULTS

##### 4.1 Light Field Camera and Setting for Evaluation

The light field camera used in this study is Lytro ILLUM developed by Lytro. The relevant structural feature of this camera is the micro lens array placed between the main lens and the sensor. Table 1 shows the specification of the camera. The GSD is about 1 mm on average. As mentioned in 3.3, two sets of stereo images, namely 4 images, are used.

Camera	Lytro ILLUM
Lens	Focal length: 30–250 mm, Aperture: f/2.0
Zoom	8 × zoom (optical)
Image sensor	1/1.2 inch CMOS
# Light field sensor	40 million light rays
Record format	Light field RAW (.lfr)
# Output pixels	4 million (2450 × 1634)

Table 1. Specification of the light field camera

Figure 13 shows a target object and reference points for evaluation. Accuracy evaluation is performed with the true value of the distance between reference points as posterior. The distances are not used as constraint in the bundle adjustment. The distances between reference point 1 and reference points 2 to 8 are 20 mm, 20 mm, 10 mm, 20 mm, 10 mm, 20 mm, and 30 mm, respectively. The reference points are set and the distances are measured manually with accuracy of 0.1 mm.

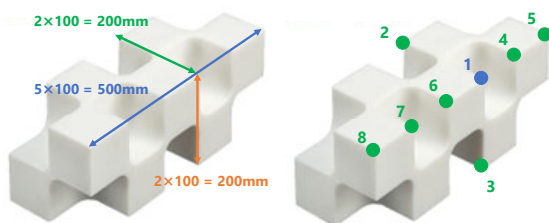


Figure 13. Target object for evaluation

In order to convert the focus distance into meters, linear regression is applied. The distance between the light field camera and target object is 20 mm to 300 mm, and the measurement is performed while changing the distance by 10 mm. The regression equation is estimated as  $d = 182.1079\alpha - 75.9289$  (Figure 14).

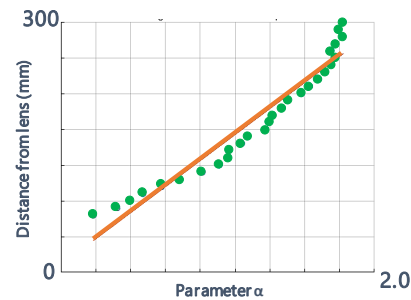


Figure 14. Linear regression between distance and  $\alpha$

Next, reliability in the SfF method is investigated. Figure 15 shows the histogram of standard deviation of the focus value for all pixels. In the histogram, the top 5% reliability (value above the red line in the figure) is used as the threshold.

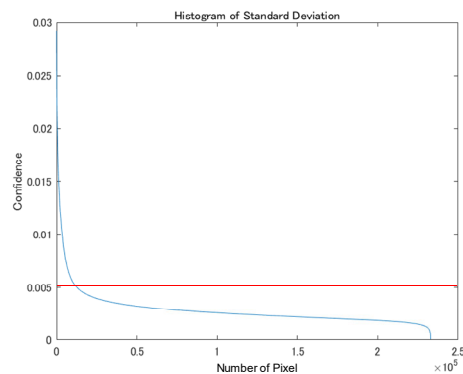


Figure 15. Histogram of standard deviation of focus value

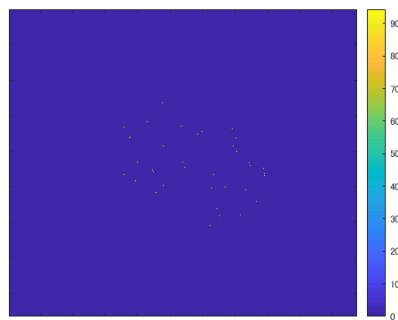
##### 4.2 Accuracy Evaluation

The proposed method is applied with the abovementioned setting. The weights for integrating 3D coordinates are tested as follows:  $w = 0$  (only photogrammetry),  $w = 0.5$  (proposed method), and  $w = 1$  (only SfF method). Table 2 shows the root mean square error (RMSE) (mm) of the results, and Figure 16 depicts a depth map of the results. The number of measured feature points are 64 (photogrammetry), 10570 (SfF method), and 10582 (proposed method). The proposed method achieves the highest accuracy and density measurement value. It is 2 mm and 5.5 mm higher than those of photogrammetry and SfF methods, respectively.

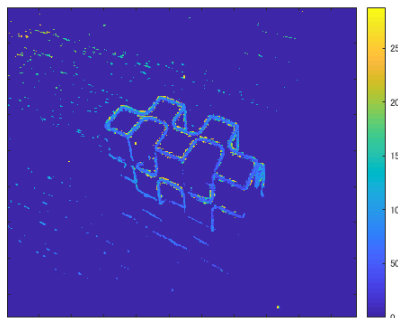
	Weights $w$		
	0	0.5	1
RMSE (mm)	5.67	3.62	9.22

Table 2. RMSE (mm) of the results

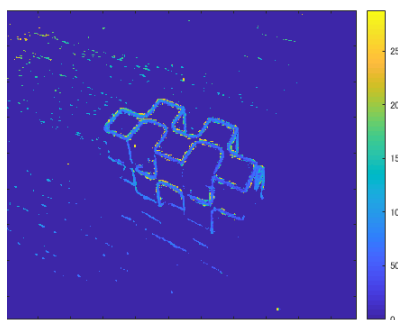
The accuracy depends on weight  $w$ . The sensitivity analysis against the weight is conducted. Figure 17 shows the result of the analysis. When  $w$  is equal to 0.453, the RMSE reaches minimum at 3.5847. Additionally, the weight is changed according to the reliability of SfF at each feature point. However, this has no discernible effect on the RMSE of the results.



(a) only photogrammetry



(b) only Sff



(c) proposed method

Figure 16. Experimental results

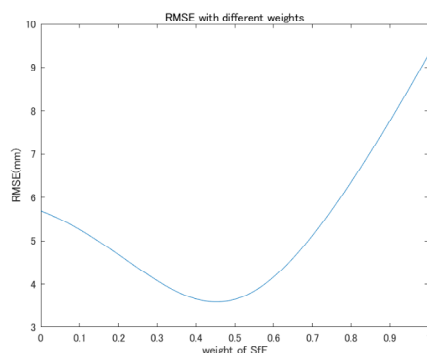
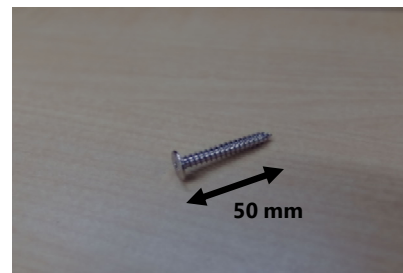


Figure 17. Sensitivity analysis of weight

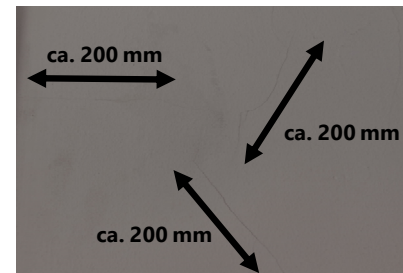
### 4.3 Applications

The proposed method is used to measure a screw and cracks on the wall. The size of screw is 50 mm, and one of cracks is 200 mm (Figure 18). Figure 19 shows the depth map obtained using the proposed method. It is confirmed that the depth of the object reflect shapes of the screw and cracks, although noise is present

in some areas of the map. In the applications, the accuracy of the depth map is not verified.

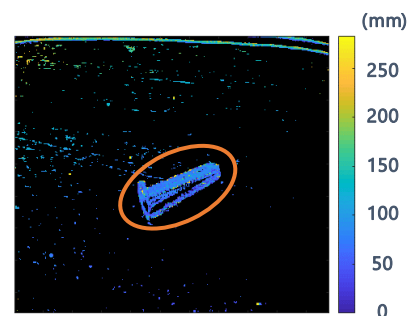


(a) screw

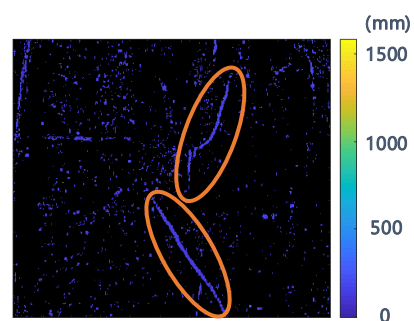


(b) cracks

Figure 18. Targets of applications



(a) depth map of screw



(b) depth map of cracks

Figure 19. Applications of the proposed method

## 5. CONCLUSIONS

In this study, a method of 3D measurement using a light field camera was developed. The proposed method combines both method, the photogrammetry to a multi-view image and the SfF method to a multi-focus image. The proposed method was applied to a rigid body with sharp edges, and its accuracy was evaluated. Through experiments, it was determined that the proposed method achieved higher accuracy and denser

measurements (RMSE: 3.58 mm, # feature points: 10582) compared to the photogrammetry (RMSE: 5.67 mm, # feature points: 64) and SfF method (RMSE: 9.22 mm, # feature points: 10570) used individually. As an application, we measured a small screw and a crack on a building wall using the proposed method.

Although the developed method was proven effective, there exist limitations that may be subject of future work. Accuracy may be improved by analysing the reliability of points measured using the photogrammetry. Additionally, noise removal methods such as smoothing for the SfF method may also contribute to accuracy. Furthermore, the application of the developed method to various objects and circumstances is suggested. The method can be modified based on the characteristics of the objects to be measured and other circumstances. Moving objects may also be used as subjects of the developed method. These endeavours will contribute to the applicability of the light field camera.

#### ACKNOWLEDGEMENTS

This research was funded by the Japan Society for the Promotion of Science (JSPS), Grant-in-Aid for Scientific Research (B) 18H01554.

#### REFERENCES

Adelson, E.H., Bergen, J.R., 1991: The plenoptic function and the elements of early vision. *Computational Models of Visual Processing*, 3-20.

Adelson, E.H., Wang, Y.A., 1992: Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis & Machine Intelligence* (2), 99–106.

Bouguet, J-Y., 2004. Camera Calibration Toolbox for Matlab. [http://www.vision.caltech.edu/bouguetj/calib\\_doc/index.html](http://www.vision.caltech.edu/bouguetj/calib_doc/index.html).

Bay, H., Ess, A., Tuytelaars, T. Vangool, L., 2008: Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3), 346-359.

Dansereau, D.G., Pizarro, O., Williams, S.B., 2013. Decoding, calibration and rectification for lenselet-based plenoptic cameras. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1027-1034.

Alcantarilla, P.F., Bartoli, A., Davison, A.J., 2012. KAZE features. *Proceedings of European Conference on Computer Vision*, 214-227.

Harris, C., Stephens, M., 1988. A combined corner and edge detector. *Proceedings of Alvey Vision Conference* 15, 147-151.

Hartley, R., Zisserman, A., 2003: *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Heikkila, J., Silven, O., 1997. A four-step camera calibration procedure with implicit image correction. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1106-1112.

Leutenegger, S., Chli, M., Siegwart, R.Y., 2011. BRISK: Binary robust invariant scalable keypoints. *Proceedings of IEEE International Conference on Computer Vision*, 2548-2555.

Luhmann, T., Robson, S., Kyle, S., Boehm, J., 2014: *Close Range Photogrammetry and 3D Imaging* [2nd Edition]. Walter de Gruyter GmbH, Berlin, Germany.

Malvar, H.S., He, L., Cutler, R., 2004. High-quality linear interpolation for demosaicing of Bayer-patterned color images. *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing* 3, iii–485.

Matas, J., Chum, O., Urban, M., Pajdla, T., 2004: Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing* 22(10), 761-767.

Muja, M., Lowe, D.G., 2012. Fast matching of binary features. *Proceedings of Ninth Conference on Computer and Robot Vision*, 404-410.

Pertuz, S., Puig, D., Garcia, M.A., 2013: Analysis of focus measure operators for shape-from-focus. *Pattern Recognition* 46(5), 1415-1432.

Raskar, R., Tumblin, J., 2006. Computational Photography, SIGGRAPH 2006 Course 15 Notes.

Rosten, E., Drummond, T., 2006. Machine learning for high-speed corner detection. *Proceedings of European Conference on Computer Vision*, 430-443.

Shi, J., Tomasi, C., 1994. Good features to track. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 593-600.

Tao, M.W., Hadap, S., Malik, J., Ramamoorthi, R., 2013. Depth from combining defocus and correspondence using light-field cameras. *Proceedings of the IEEE International Conference on Computer Vision*, 673-680.

Tao, M.W., Srinivasan, P.P., Malik, J., Rusinkiewicz, S., Ramamoorthi, R., 2015. Depth from shading, defocus, and correspondence using light-field angular coherence. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1940-1948.

Torr, P., Zisserman, A., 2000: MLESAC: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding* 78(1), 138-156.

Yang, P., Wang, Z., Yan, Y., Qu, W., Zhao, H., Asundi, A., Yan, L., 2016: Close-range photogrammetry with light field camera: from disparity map to absolute distance. *Applied Optics*, 55(27), 7477-7486.

Zhang, Z., 2010: A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330-1334.