

# Accurate Camera Calibration Robust to Defocus using a Smartphone

Hyowon Ha    Yunsu Bok    Kyungdon Joo    Jiyoung Jung    In So Kweon  
Robotics and Computer Vision Lab., KAIST

{hwha, ysbok, kdjoo, jyjung}@rcv.kaist.ac.kr

iskweon@kaist.ac.kr

## Abstract

We propose a novel camera calibration method for defocused images using a smartphone under the assumption that the defocus blur is modeled as a convolution of a sharp image with a Gaussian point spread function (PSF). In contrast to existing calibration approaches which require well-focused images, the proposed method achieves accurate camera calibration with severely defocused images. This robustness to defocus is due to the proposed set of unidirectional binary patterns, which simplifies 2D Gaussian deconvolution to a 1D Gaussian deconvolution problem with multiple observations. By capturing the set of patterns consecutively displayed on a smartphone, we formulate the feature extraction as a deconvolution problem to estimate feature point locations in sub-pixel accuracy and the blur kernel in each location. We also compensate the error in camera parameters due to refraction of the glass panel of the display device. We evaluate the performance of the proposed method on synthetic and real data. Even under severe defocus, our method shows accurate camera calibration result.

## 1. Introduction

Camera calibration is a process to estimate the transformation between the image coordinate of a camera and the real world coordinate. It is an essential step in computer vision fields and its accuracy highly influences the quality of various computer vision applications such as 3D reconstruction [1, 26], depth estimation [25, 19], and robot navigation [20]. To achieve an accurate camera calibration result, acquiring an exact set of correspondences between images and the real world is extremely important [14, 4]. Therefore, recent researches on camera calibration have focused on developing patterns having distinctive feature points that can be accurately localized in the images [27, 8, 16].

Among various patterns, a planar checkerboard has been one of the most famous patterns because of the ease of de-

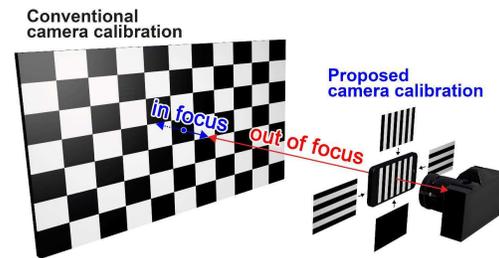


Figure 1: The proposed camera calibration that is robust to defocus using a smartphone display compared with a conventional camera calibration using a checkerboard.

tection, low cost and simple structure [27, 12, 24, 17]. A traditional way of producing a calibration pattern is to print out the pattern on a piece of paper and attach it on a planar board so that the pattern maintains the plane assumption.

Recently, the calibration methods using display devices such as monitors, tablets or smartphones have come to the forefront. Compared to the printed patterns, the displayed patterns have the advantages of adjustable brightness, guaranteed flatness and known pixel sizes [29, 22, 11]. The devices can display programmable patterns for higher level features such as time-coded pattern which can help the camera calibration to identify the feature location [6]. However, since a display device is usually equipped with a glass panel in front of the LCDs, the observed feature location in the image may deviate from the actual feature location displayed by the LCDs due to refraction by the glass panel.

For all sorts of planar pattern based calibration methods, printed or displayed, an important condition is required to the images used for the calibration: All images have to be in focus for accurate feature detection. Defocused images yield incorrect correspondences which lead to performance degradation of the calibration. This requirement brings natural inconveniences to the data capture process. For example, if we want to calibrate a camera focusing at a very short distance to capture a small object, the calibration pattern has to be small as well. On the other hand, if we mount a camera on a vehicle to capture street views for autonomous driving [23] or urban scene reconstruction purposes [18], the camera is set to focus in the range from a few meters

This research is supported by the National Research Foundation, Korea, under the NRF-ANR joint research programme (No. 2011-0031920).

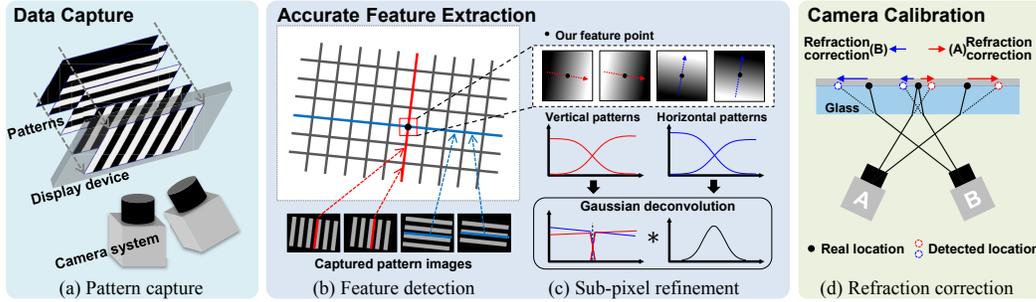


Figure 2: The overall procedure of the proposed camera calibration.

to a few hundred meters. Then, the pattern has to be large enough for the features to be detected in the focused region, and either the pattern or the camera has to move in a large range to obtain the images in various poses.

In this paper, we present a novel camera calibration method that is robust to defocus of the images using a planar display device. Assuming that defocus blur is modeled as a convolution with a Gaussian point spread function (PSF), we propose to capture a series of unidirectional patterns shown in Fig. 3 consecutively in order to estimate the size of the convolved Gaussian kernel. Since our calibration method is highly accurate in severe defocus, we can use a display device as small as a mobile phone to display the patterns to calibrate a camera focusing at a far distance, even at infinity. The data capture process can be done in a small area regardless of the focus distance of the camera (see Fig. 1).

We use a display device because it enables the camera to capture a series of patterns at the same viewpoint. When calibrated in close distances for the convenience of the data capture, error due to refraction of the glass panel has a non-negligible effect on the calibration performance. We correct the error due to refraction in consideration of Snell's law (the law of refraction) and obtain highly accurate camera parameters. To the best of our knowledge, this is the first work that corrects the error due to refraction of the panel on a display device for camera calibration.

The contributions of this paper is summarized as follows:

- A set of multiple unidirectional binary patterns to be displayed consecutively on a display device is proposed for accurate feature detection that is robust against severe image defocus.
- Subpixel refinement for the proposed set of unidirectional patterns is proposed to refine the detected feature locations in a sub-pixel level of accuracy.
- Refraction correction that corrects the errors in camera parameters due to refraction of the glass panel is presented for camera calibration using a display device.

The overall procedure of the proposed defocus-robust camera calibration is illustrated in Fig. 2. The method con-

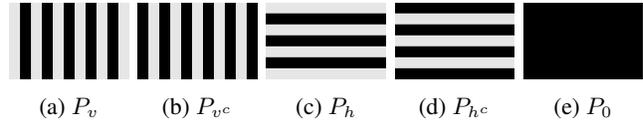


Figure 3: The proposed set of complementary binary patterns. (a)-(d) Two pairs of vertical and horizontal binary patterns. (e) The black pattern is used to capture ambient illumination.

sists of three stages. First, the proposed patterns are displayed on a screen, and the camera system captures the set of patterns at various viewpoints regardless of its focus. For each viewpoint, the same set of patterns are captured consecutively. Second, initial feature detection is carried out by extracting and labeling the proposed features, which are the intersection points of edges, followed by the refinement of the feature points in sub-pixel level accuracy. Finally, the camera parameters are estimated using the feature point locations and then refined by compensating the error due to refraction of the glass panel of the display device. To emphasize the wide applicability of the proposed method, we use an iPhone5 as the display device throughout the paper.

## 2. Accurate Feature Extraction

In this section, we describe how we extract an accurate set of features from severely defocused pattern images. A complementary set of unidirectional binary patterns is proposed in Sec. 2.1 and the initial feature detection by computing edginess images and a grid mask from the set of pattern images is presented in Sec. 2.2. The feature refinement in sub-pixel level accuracy is described in Sec. 2.3.

### 2.1. Complementary Binary Patterns

The feature extraction in defocus can be regarded as a 2D Gaussian deconvolution problem to estimate the size of Gaussian blur and the feature location that can be used to synthesize the sharp pattern image as long as the shape of the pattern is known. However, in case of severe defocus, the size of Gaussian blur becomes too large to be estimated using the conventional deblurring approaches such as [28]. Thus, we simplify the 2D problem into a 1D problem with

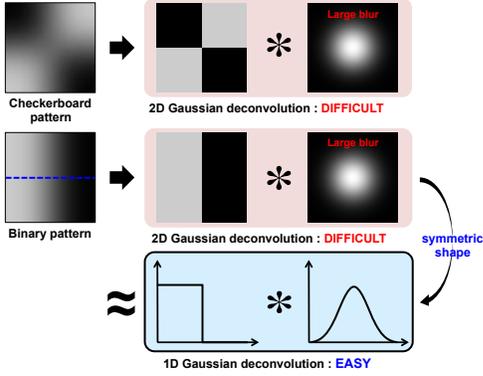


Figure 4: The principle of simplifying 2D Gaussian deconvolution to a 1D Gaussian deconvolution problem using the symmetric shape of unidirectional binary patterns.

multiple observations using the proposed set of patterns that consists of two pairs of unidirectional complementary binary patterns, as shown in Fig. 3.

To achieve successful camera calibration using images in severe defocus, we introduce a set of complementary binary patterns. These patterns are designed to have three major characteristics: (1) unidirectional and binary, (2) paired in complementary patterns, and (3) generating a sufficient number of feature points for camera calibration.

Because of the symmetric shape of the proposed patterns and the blur kernel, the 1D Gaussian deconvolution problem along the axes of symmetry approximates the 2D localization problem of an intersection point of edges, as described in Fig. 4. Multiple observations of unidirectional binary patterns instead of a single observation of a checkerboard facilitate such design of the 1D problem. A pair of complementary patterns helps to localize the edge adaptively to the non-uniform brightness of the pattern images. The patterns are designed to produce  $6 \times 10$  features, which are considered as a sufficient number in common camera calibration.

**Composition** The proposed patterns are composed of five binary patterns: two horizontal complementary patterns  $P_h$  and  $P_{h^c}$ , two vertical complementary patterns  $P_v$  and  $P_{v^c}$ , and one black pattern  $P_0$  (see Fig. 3). The black pattern is used to capture the ambient illumination.

For simplicity, the set of complementary binary patterns is described as  $\mathcal{B} = \{h, h^c, v, v^c\}$  and a subscript  $b \in \mathcal{B}$  is adopted to represent an arbitrary one in the set  $\mathcal{B}$ .

**Preprocessing** Given a set of captured complementary pattern images, we first eliminate the ambient illumination by subtracting the black pattern image from four complementary pattern images. We blur the pattern images with a small Gaussian kernel to compress the error due to image noise. Then we obtain the preprocessed set of pattern images  $\{I_b\}$ . In this preprocessing, our assumption that the defocus blur follows Gaussian distribution is not violated and even satisfied in well-focused cases, since the convolution of two Gaussian kernels becomes another Gaussian.

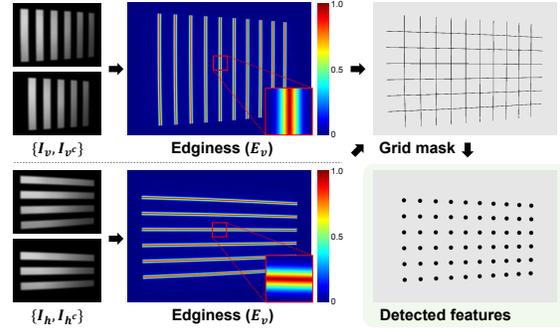


Figure 5: Feature detection procedure.

## 2.2. Initial Feature Detection

The initial feature detection is an important step for calibration, since it provides crucial cues to automatically label the feature points with 3D world coordinates and guarantee a search range for subpixel refinement. Our initial feature detection utilizes the vertical edges from  $I_v$  and  $I_{v^c}$ , and the horizontal edges from  $I_h$  and  $I_{h^c}$  in Fig. 5.

For the set of pattern images  $\{I_b\}$ , we calculate two edginess images  $E_v$  and  $E_h$  for vertical and horizontal patterns, respectively.

$$E_v = \begin{cases} 1 - \frac{\max(I_v, I_{v^c}) - \min(I_v, I_{v^c})}{0.5 \sum I_b}, & \text{if } \sum I_b > \alpha, \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

$$E_h = \begin{cases} 1 - \frac{\max(I_h, I_{h^c}) - \min(I_h, I_{h^c})}{0.5 \sum I_b}, & \text{if } \sum I_b > \alpha, \\ 0, & \text{otherwise,} \end{cases}$$

where  $\sum I_b$  is the sum image of all the pattern images and  $\alpha$  is a threshold to filter out non-pattern regions. For  $\alpha$ , we use the intensity of 0.1.

From the edginess images  $E_v$  and  $E_h$  that have wide ranges around the edges as shown in Fig. 5, two masks for the vertical and horizontal edges are robustly obtained by thresholding the edginess values. We then apply the skeleton extraction algorithm [15] on each mask to leave only the skeletons having 1 pixel thickness. By adding the separately calculated skeleton masks, the initial mask for our feature grid is obtained. To find initial feature points on this grid mask, we utilize the checkerboard detection algorithm proposed in [17]. This detection algorithm finds initial feature points on the given mask of feature point grid by checking the numbers of neighbor pixels and label the feature points according to their connected relationship as a graph.

## 2.3. Sub-pixel Feature Refinement

In order to refine the initially detected feature locations to be more accurate in sub-pixel level, we first estimate the axes of symmetry, vertical and horizontal, in the image using the gradient direction of the intensity. Then, we sample the pixels along each axis to obtain blurred intensity profiles. Using the blurred intensity profiles, we syn-

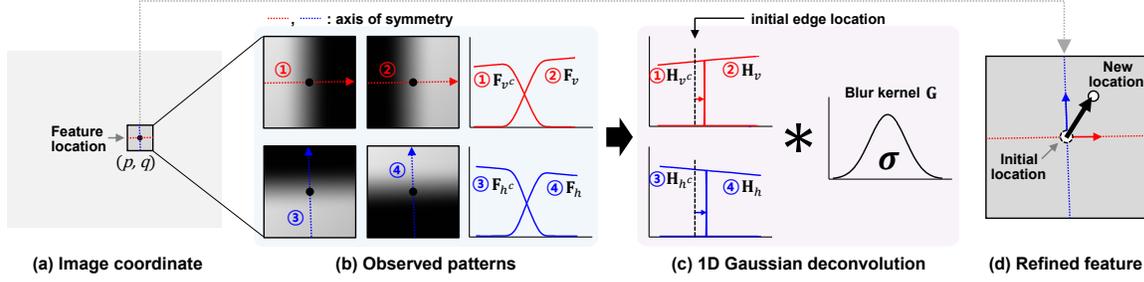


Figure 6: The procedure of our feature refinement. The feature locations are refined based on 1D Gaussian deconvolution.

thesize the original sharp intensity profiles assuming they are in the shape of a step function. The standard deviation of the Gaussian blur kernel is initially estimated by a simple Gaussian fitting on the gradient of one blurred intensity profile. Finally, for each feature, its location and the size of the blur kernel are refined through nonlinear optimization. The procedure of feature refinement is described in Fig. 6. Please note that each feature point is refined as the form of point itself in this refinement step.

**Axes of symmetry** The proposed feature refinement is carried out to localize each feature in sub-pixel level by solving a 1D Gaussian deconvolution problem on the set of pattern images. Since the simplification of the problem from the 2D Gaussian deconvolution is based on the symmetric shape of the patterns, we first estimate the axis of symmetry in each pattern image using the gradient direction of the image intensity. By filtering the pattern images with  $3 \times 3$  Scharr operators [21], the gradient direction  $\theta_b$  is calculated for each pixel  $(p, q)$  in the pattern image  $I_b$ .

$$\bar{\theta}_b(p, q) = \tan^{-1} \left( \frac{\sum_{i=-1}^1 \sum_{j=-1}^1 \sin(\theta_b(p+i, q+j))}{\sum_{i=-1}^1 \sum_{j=-1}^1 \cos(\theta_b(p+i, q+j))} \right), \quad (2)$$

where  $\bar{\theta}_b$  is the average of  $\theta_b$  in a local  $3 \times 3$  patch for robustness to image noise. As each pair of complementary patterns shares the axis of symmetry, but has opposite gradient directions, we estimate the angle  $\phi_v$  of the vertical axis by averaging the opposite angles from the complementary vertical patterns as:

$$\phi_v = \tan^{-1} \left( \frac{\sin \bar{\theta}_v + \sin(\bar{\theta}_{v^c} - \pi)}{\cos \bar{\theta}_v + \cos(\bar{\theta}_{v^c} - \pi)} \right), \quad (3)$$

where  $\phi_v$  is the angle of the axis of symmetry in  $I_v$  and  $I_{v^c}$ . We estimate the angle  $\phi_h$  in  $I_h$  and  $I_{h^c}$  analogously. The axis of symmetry is illustrated in Fig. 6-(b).

**Blurred intensity profiles** For each feature location, we extract four intensity profiles from the pattern images. A blurred intensity profile is a vector of image intensities along the axis of symmetry in each pattern image. We sample a line crossing the feature point in one pixel width.

$$\begin{aligned} \mathbf{F}_v[x|p, q] &= I_v(p+x \cos \phi_v(p, q), q+x \sin \phi_v(p, q)), \\ \mathbf{F}_{v^c}[x|p, q] &= I_{v^c}(p+x \cos \phi_v(p, q), q+x \sin \phi_v(p, q)), \end{aligned} \quad (4)$$

where  $\mathbf{F}_v$  and  $\mathbf{F}_{v^c}$  are the blurred intensity profiles extracted from vertical pattern images  $I_v$  and  $I_{v^c}$  along the

axis of angle  $\phi_v$ . The blurred intensity profiles  $\mathbf{F}_h$  and  $\mathbf{F}_{h^c}$  are analogously extracted from horizontal pattern images  $I_h$  and  $I_{h^c}$  along the axis of angle  $\phi_h$ .

**Sharp intensity profiles** The sharp intensity profiles are synthesized by using the blurred intensity profiles in the assumption that the original intensity profiles of the focused pattern images are in the shape of a step function. The sharp intensity profiles  $\mathbf{H}_v$  and  $\mathbf{H}_{v^c}$  for vertical complementary patterns are estimated as:

$$\begin{aligned} \mathbf{H}_v[x|p, q] &= \begin{cases} \mathbf{F}_v[x|p, q] + \mathbf{F}_{v^c}[x|p, q], & x > 0, \\ \frac{1}{2} (\mathbf{F}_v[0|p, q] + \mathbf{F}_{v^c}[0|p, q]), & x = 0, \\ 0, & x < 0, \end{cases} \\ \mathbf{H}_{v^c}[x|p, q] &= \begin{cases} 0, & x > 0, \\ \frac{1}{2} (\mathbf{F}_v[0|p, q] + \mathbf{F}_{v^c}[0|p, q]), & x = 0, \\ \mathbf{F}_v[x|p, q] + \mathbf{F}_{v^c}[x|p, q], & x < 0, \end{cases} \end{aligned} \quad (5)$$

where  $\mathbf{H}_v$  and  $\mathbf{H}_{v^c}$  are the sharp intensity profiles synthesized from blurred intensity profiles  $\mathbf{F}_v$  and  $\mathbf{F}_{v^c}$ . The sharp intensity profiles  $\mathbf{H}_h$  and  $\mathbf{H}_{h^c}$  for horizontal complementary patterns are estimated in the same way using  $\mathbf{F}_h$  and  $\mathbf{F}_{h^c}$ . This process is described in Fig. 6-(c).

In Eq. 5, it is implied that the sum of two complementary pattern images would be the same as a white pattern image. The sharp intensity profiles are synthesized by cutting the sum of two complementary pattern images at the current feature location. In this way, the sharp intensity profiles estimate the original step-shaped patterns adaptively in non-uniform brightness of the images. Even though the sum of two complementary intensity profiles is the blurred intensity profile of the white pattern image, it can be approximated to the sharp intensity profile by assuming that the white pattern image contains only low-frequency components.

**Gaussian blur kernel** The Gaussian blur kernel to be estimated is a normalized Gaussian function such as:

$$\mathbf{G}[x|\sigma] = \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/2\sigma^2}, \quad (6)$$

where  $\sigma$  is the standard deviation of the normalized Gaussian function. For each feature point, we initially estimate the  $\sigma$  by fitting a Gaussian function on the gradient of one blurred intensity profile, which is roughly the integral of the Gaussian blur kernel.

**Optimization** Finally, for each feature point, its location  $(p, q)$  and the size of the Gaussian blur kernel  $\sigma$  are refined by minimizing the differences between the blurred intensity profiles and the convolution of the sharp intensity profiles with the blur kernel as follows:

$$(p^*, q^*, \sigma^*) = \arg \min_{p, q, \sigma} \sum_{b \in \mathcal{B}} \|\mathbf{F}_b - \mathbf{H}_b * \mathbf{G}\|^2, \quad (7)$$

where  $\mathbf{F}_b$ ,  $\mathbf{H}_b$  and  $\mathbf{G}$  follow Eq. 4, Eq. 5 and Eq. 6. By using the Levenberg-Marquardt (LM) method [10, 13], the parameters  $(p, q, \sigma)$  are iteratively updated until convergence to solve this problem.

### 3. Refraction Correction

We estimate the intrinsic and extrinsic camera parameters based on Zhang’s calibration method [27] using the estimated feature points from the previous section. Since we utilize a display device to show multiple patterns at the same viewpoint, the detected feature locations contain the error due to the light refraction by the glass panel of the device.

In this section, we present the refraction correction of the camera parameters when a display device is used in camera calibration. Since a display device is usually equipped with a transparent panel made of glass in front of LCDs, the light is refracted on the surface of the panel. Thus, the observed feature positions through the panel are shifted from the actual feature position displayed by the LCDs. Physically, this shift of position cannot be larger than the thickness of panel, which is normally small and ignorable compared to the camera-display distance. However, as the proposed method covers a wide range of image capture distances including extremely close distances, we need to consider the small feature shifts due to refraction to achieve accurate camera calibration.

Fig. 7 describes the refraction model of the display device. By the effect of refraction, an actual point located at  $p$  is observed at an apparent position  $p'$  on the line along the incident ray direction  $\mathbf{l}$ . To compensate this difference of positions while preserving projective geometry, we calculate  $\mathbf{c}$ , which is the 3D vector to move the point at  $p$  to  $q$  where the plane and the line of incidence are intersecting. As long as  $\mathbf{n}$  and  $\mathbf{l}$  are known unit vectors, their dot product gives  $|\mathbf{n} \cdot \mathbf{l}| = \cos \theta_1$ . When  $n_1$  and  $n_2$  are the refractive indices of the air and the medium of the panel respectively, and  $\theta_1$  and  $\theta_2$  are the angles of incidence and refraction measured from the normal of the plane, we know  $n_1 \sin \theta_1 = n_2 \sin \theta_2$  by the Snell’s law. Then, the vector  $\mathbf{c}$  for refraction compensation can be calculated as:

$$\mathbf{c} = D \left( \frac{1}{|\mathbf{n} \cdot \mathbf{l}|} - \frac{1}{\sqrt{n_2^2 - 1 + (\mathbf{n} \cdot \mathbf{l})^2}} \right) (\mathbf{l} - (\mathbf{n} \cdot \mathbf{l}) \mathbf{n}), \quad (8)$$

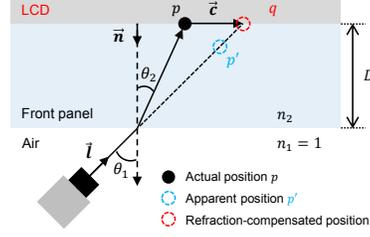


Figure 7: The light refraction model of the panel and the refraction correction of our method.

where  $D$  is the thickness of the panel.

Then, for each  $j$ -th feature point in the  $i$ -th image, the projection of the corresponding 3D point  $\mathbf{X}_j$  to the camera coordinate system is represented as:

$$\lambda \mathbf{x}_{ij} = \mathbf{R}_i \mathbf{X}_j + \mathbf{t}_i + \mathbf{c}_{ij}, \quad (9)$$

where  $\mathbf{X}_j = [X_j, Y_j, Z_j]^T$  is the world coordinate of the  $j$ -th feature point,  $\mathbf{x}_{ij} = [x_{ij}, y_{ij}, 1]^T$  is its normalized camera coordinate, and  $\lambda$  is a scaling factor.  $\mathbf{R}_i$  and  $\mathbf{t}_i$  are the rotation matrix and the translation vector of the  $i$ -th image with respect to the world coordinate system.  $\mathbf{c}_{ij}$  is individually calculated according to Eq. 8 using  $\mathbf{n} = \mathbf{R}_i [0, 0, 1]^T$  and

$$\mathbf{l} = \frac{\pi^{-1}(\mathbf{k}, \mathbf{p}, \mathbf{K}^{-1} \tilde{\mathbf{u}}_{ij})}{\|\pi^{-1}(\mathbf{k}, \mathbf{p}, \mathbf{K}^{-1} \tilde{\mathbf{u}}_{ij})\|_2}. \quad (10)$$

The function  $\pi^{-1}$  is the inverse function of the lens distortion model, which returns an undistorted camera coordinate of the given distorted camera coordinate  $\mathbf{K}^{-1} \tilde{\mathbf{u}}_{ij}$  using the radial distortion parameters  $\mathbf{k}$  and the tangential distortion parameters  $\mathbf{p}$ .  $\tilde{\mathbf{u}}_{ij}$  denotes the image coordinate of the  $j$ -th feature point in the  $i$ -th image, and  $\mathbf{K}$  denotes the camera intrinsic matrix.

Finally, the previously estimated camera parameters  $\mathbf{K}$ ,  $\mathbf{k}$ ,  $\mathbf{p}$ ,  $\mathbf{R}_i$ , and  $\mathbf{t}_i$  are refined to minimize the following:

$$\sum_i^m \sum_j^n \|\tilde{\mathbf{u}}_{ij} - \mathbf{u}_{ij}(\mathbf{K}, \mathbf{k}, \mathbf{p}, \mathbf{R}_i, \mathbf{t}_i, D | n_2, \mathbf{X}_j)\|^2, \quad (11)$$

where  $\mathbf{u}_{ij} = \mathbf{K}[\pi(\mathbf{k}, \mathbf{p}, \mathbf{x}_{ij})]$ . In our implementation, this optimization is also solved by using the LM method.

Since the refractive index  $n_2$  and the thickness of the glass panel  $D$ , are usually not provided for a display device, the convergence and the possibility of overfitting in this step have been tested for two cases. Unfortunately, it is found out that refining both parameters in the proposed optimization has a serious overfitting problem. The reprojection error reaches to zero by abnormally changing the parameters. On the other hand, refining either one of the two parameters converge to a reasonable value and has shown good results.

Therefore, we fix the refractive index  $n_2$  and refine the thickness  $D$  because the refractive index of glass is typically in a small range from 1.52 (crown glass) to 1.62 (flint

glass) and also has less influence than the thickness  $D$  in the proposed refraction correction model. The reliability in selecting the refractive index value has been investigated in our experiments (Sec. 4.2).

## 4. Experimental Results

In this section, we investigate the performance of the proposed camera calibration in severely defocused conditions. Our experiments include a synthetic experiment and three real experiments. In the synthetic experiment, the sub-pixel accuracy of the proposed feature refinement algorithm is evaluated with various Gaussian blur conditions. In real experiments, the accuracy and the validity of the proposed method are verified in various perspectives.

**Methods for comparison** The proposed feature refinement algorithm is compared with three existing methods: the Harris corner detector [7] which is one of the most representative corner detectors, and the two recent corner refinement methods by Geiger *et al.* [5] and Placht *et al.* [17]. Given an initial feature location, Geiger *et al.* [5] estimate the corner by finding the local maximum point of corner-ness. Placht *et al.* [17], so-called ROCHADE, finds the saddle point of the polynomial around the corner. Also, a simple method, the zero-crossing detection of the difference image of the proposed patterns, is compared.

**Implementation** All methods including the proposed method have been implemented in MATLAB. For fair comparison, we utilize the open sources for the existing methods. We also try to set a reliable parameter for each method: We set 41 pixels as the intensity profile length for the proposed method and as the patch size for the Harris corner. We set 9 pixels as the patch size for [5] and [17]. These parameters have been used in all synthetic and real experiments. For the zero-crossing detection, the zero-crossing point is refined to minimize the sum of squared differences between the complementary patterns.

### 4.1. Synthetic Experiment

**Experimental setup** For synthetic dataset, we first create the proposed unidirectional binary patterns and a checkerboard pattern to have only one feature point at the center. The proposed patterns are used to evaluate the proposed method, and the checkerboard is used to evaluate three other methods. We generate multiple images of each pattern at 300 random viewpoints within  $\pm 30^\circ$  from the front view. The image resolution is set to  $300 \times 300$  pixels. Then, we

Bouguet’s camera calibration toolbox, [http://www.vision.caltech.edu/bouguetj/calib\\_doc](http://www.vision.caltech.edu/bouguetj/calib_doc)

MATLAB and Computer Vision System Toolbox Release 2015a, The MathWorks, Inc., Natick, Massachusetts, United States.

ROCHADE: Robust Checkerboard Detection, <http://www.metrilus.de/blog/portfolio-items/rochade/>

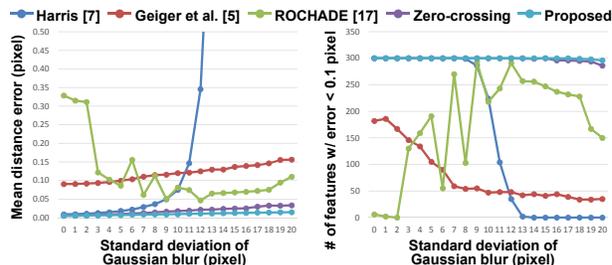


Figure 8: Accuracy of the feature refinement methods according to blur sizes. (Left) The mean distance error of the refined features from the ground truth location. (Right) The number of refined features with error less than 0.1 pixel.



Figure 9: The stereo camera system in the real experiments.

synthesize the defocused pattern images by convolving all the images with the same Gaussian kernel. We change the standard deviation of the Gaussian kernel from 0 to 20 pixels, which result in a maximum blur size of 120 pixels in diameter. Finally, we add signal-dependent noise [9] with 1% intensity variance. As the initial feature location, a random location within 3 pixels from the ground truth location is used for all methods.

**Accuracy of feature refinement** The accuracy of the five methods are shown in Fig. 8. The left graph shows the mean distance errors of refined features from the ground truth location according to the different Gaussian blur sizes. The right graph shows the number of refined features with error less than 0.1 pixel. Harris corner [7] shows excellent performance under small blur conditions. As the size of the blur gets larger, Geiger *et al.* [5] and ROCHADE [17] show better performances than Harris corner, but still they show lower accuracy than the zero-crossing detection and the proposed method. And the proposed method shows slightly better performance while the synthetic images have uniform brightness which is the prerequisite of the zero-crossing detection to be mathematically true.

### 4.2. Real Experiments

**Experimental setup** For real experiments, the proposed patterns (Fig. 3) and a checkerboard pattern that are designed to have the same feature locations are captured at 50 viewpoints within 15cm distance by the stereo camera system focused at infinity. The  $6 \times 10$  features are uniformly distributed with the interval of 92 pixels which is 7.1681mm as the pixel density of iPhone5 is 326 pixels per inch. The auto-brightness of the display and the auto-

Table 1: Mean reprojection errors (pixel) from the calibration of the stereo camera system (Fig. 9) focused at infinity. Two single camera calibrations and one stereo camera calibration are evaluated using different methods.

Camera Setup	Without refraction correction					With refraction correction				
	[7]	[5]	[17]	Zero-crossing	Proposed	[7]	[5]	[17]	Zero-crossing	Proposed
Left camera	1.1432	0.6447	0.6429	0.2125	<b>0.1159</b>	1.2457	0.6354	0.6341	0.1999	<b>0.0896</b>
Right camera	1.2343	0.7405	0.7401	0.2310	<b>0.1293</b>	1.3758	0.7304	0.7300	0.2184	<b>0.1079</b>
Stereo(Left+Right)	1.2178	0.7728	0.7735	0.3646	<b>0.3406</b>	1.1565	0.7014	0.7024	0.2272	<b>0.1158</b>

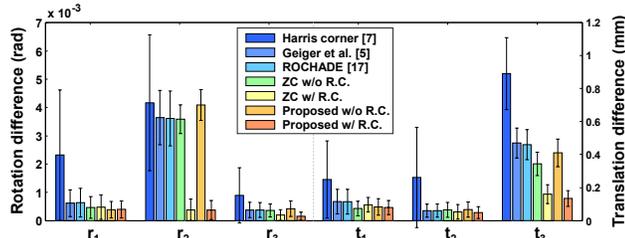


Figure 10: The differences of the extrinsic parameters between single and stereo camera calibrations. R.C. stands for refraction correction.  $r_1$ ,  $r_2$  and  $r_3$  are the rotation parameters, and  $t_1$ ,  $t_2$  and  $t_3$  are the translation parameters.

settings of the camera are turned off. Two Grasshopper3 cameras ( $3376 \times 2704$  pixels) mounted together on a rig are used as in Fig. 9. The refractive index of 1.52 has been used in refraction correction, unless stated otherwise.

For evaluation, we perform two single camera calibration on left and right cameras independently and a stereo camera calibration on both cameras.

#### 4.2.1 Performance Evaluation

The three checkerboard-based methods, the zero-crossing detection and the proposed method are compared by calibrating the left, right and stereo cameras. The initial feature locations are extracted by the proposed feature detection algorithm described in Sec. 2.2 because other methods usually fail in severe defocus. Then the feature refinement of five methods are evaluated with and without the proposed refraction correction.

The mean reprojection errors of the five methods are shown in Table. 1. Without the refraction correction, the proposed method presents high performance with errors around 0.12 pixel for single camera calibration and 0.34 pixel for stereo camera calibration, while the checkerboard-based methods show relatively poor performances with large errors around 0.72 pixel. With the refraction correction, the proposed method exhibits considerable improvements. Especially, the superior performance in the stereo camera calibration is established by the error of 0.12 pixel. Meanwhile, the zero-crossing detection shows quite lower accuracy than the proposed method in contrast to the result of the synthetic experiment because of the non-uniform brightness of the images.

In another aspect, the extrinsic parameters estimated by

Table 2: The mean reprojection errors (pixel) of a full camera calibration using the checkerboard corners extracted by each method (Original) and of extrinsic-parameter-only calibration with our intrinsic parameters (Validation).

Corner method	Left camera		Right camera	
	Original	Validation	Original	Validation
Harris [7]	0.8733	0.9615	0.8823	0.9527
Geiger <i>et al.</i> [5]	0.8616	0.9423	0.8617	0.9274
ROCHADE [17]	0.8617	0.9416	0.8602	0.9248

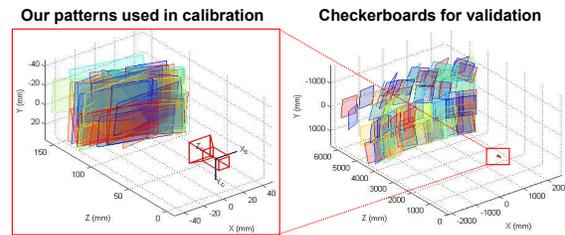


Figure 11: Visualization of extrinsic parameters. (Left) The camera and the proposed patterns used in the estimation of our intrinsic parameters. (Right) The camera and the checkerboard patterns used in the validation experiment.

the single and stereo calibrations are compared. The 50 sets of relative extrinsic parameters are calculated by applying the chain rule between the synchronous left and right camera poses from the single camera calibrations, and compared with one set of parameters estimated by the stereo calibration in Fig. 10. Even though the parameters estimated by the stereo camera calibration are not the ground truth for the relative extrinsic parameters between two cameras, the consistency of the parameter estimation is clearly shown.

Table. 1 and Fig. 10 show the accuracy of the features refined by the proposed method and the effectiveness of the proposed refraction correction.

#### 4.2.2 Validation on estimated intrinsic parameters

The proposed calibration method allows the use of small-sized patterns captured at close distances. To validate the intrinsic parameters estimated in such setup using the proposed method, we perform the following experiment.

We capture 107 images of a  $0.7m \times 0.9m$ -sized checkerboard from  $5m$  distance. The images are slightly defocused because the cameras are focused at infinity. We extract corners from the images using the conventional methods and

Table 3: The mean projection errors (M.R.E.) [pixel] and the focal lengths ( $f_x$ ,  $f_y$ ) [pixel/mm] obtained without or with refraction correction using different refractive index values.

$n_2$	Camera	Method	M.R.E.	$f_x$	$f_y$
N/A	Left	Single	0.1159	3370.9	3371.1
1.52	Left	Single	0.0896	<b>3365.7</b>	<b>3365.7</b>
1.62	Left	Single	0.0897	<b>3366.1</b>	<b>3366.0</b>
N/A	Left	Stereo	0.3406	3385.7	3386.4
1.52	Left	Stereo	0.1157	<b>3367.6</b>	<b>3367.6</b>
1.62	Left	Stereo	0.1155	<b>3367.7</b>	<b>3367.7</b>
N/A	Right	Single	0.1293	3371.4	3370.7
1.52	Right	Single	0.1079	<b>3367.5</b>	<b>3366.6</b>
1.62	Right	Single	0.1080	<b>3367.9</b>	<b>3367.0</b>
N/A	Right	Stereo	0.3406	3392.3	3392.8
1.52	Right	Stereo	0.1157	<b>3367.1</b>	<b>3366.1</b>
1.62	Right	Stereo	0.1155	<b>3367.2</b>	<b>3366.2</b>

perform the camera calibration. For validation, we perform the same camera calibration using the same set of features with the intrinsic parameters fixed as the estimates using the proposed method. Then, we compare the mean reprojection errors of the conventional calibration methods and those of the validation calibration. In Table. 2, the former are listed as "Original" and the latter are listed as "Validation".

Though the mean reprojection errors for validation slightly increase, the differences are marginal considering the cameras are focused at infinity and the image resolution is 10M pixels. Moreover, considering the compelling size differences of the patterns for the proposed method and the checkerboard-based methods as in Fig. 11, these small differences in errors are meaningful results to show the possibility of calibration setup free from the camera focus.

#### 4.2.3 Self-comparison on refractive index values

The dependency of the proposed method to different refractive index values has been investigated to show that the fixed value 1.52 of the refractive index is reasonable.

The calibrations are performed (1) without the refraction correction, and with refraction correction using (2) the refractive index of 1.52 and (3) 1.62. For each case, the mean reprojection errors (MRE) and the focal lengths from the calibrations of the left, right and stereo cameras are shown in Table. 3. Insignificant differences are observed in both MREs and the focal lengths when the refractive index value changes from 1.52 to 1.62. The focal lengths obtained with the refraction correction using the different refractive index values are very similar to each other while those obtained without the refraction correction show visible changes.

In this experiment, it is shown that the proposed refraction correction improves the accuracy with little dependency to the refractive index values in the typical range, which reflects the practicality of the proposed method that needs no additional information about the display device.

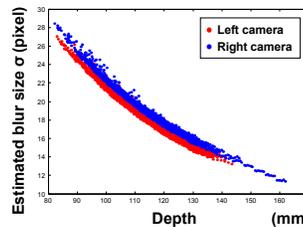


Figure 12: The relationship between the depth of the scene and the size (standard deviation) of the defocus blur obtained by the proposed method.

## 5. Discussions

**Application** The relationship between the depth of the scene and the size of the defocus blur may be explored by the proposed method. Since the size of the defocus blur for each feature is estimated in our feature refinement and the depth of the feature is estimated in our camera calibration, the depth-defocus information is obtained (see Fig. 12). Defocus-deblurring or depth-from-defocus may be strong applications.

**Limitations** The proposed method can handle the defocus as large as the interval between the features. To handle larger defocus, we can simply display the binary patterns with sparser features on the same device. The images captured with auto-settings of a camera such as auto-exposure do not ensure the best performance of the proposed method because our feature extraction assumes the images captured in the same illumination condition. Due to the similar reason, the JPEG compression of images is not recommended.

**Conclusion** In this paper, a novel framework for accurate camera calibration that is robust to defocus using a display device has been proposed. To handle the image blurs due to defocus, the proposed patterns for display are designed to be a complementary set of unidirectional binary patterns. The features are defined as the intersection points of their edges. By utilizing the patterns, it has been shown that a severe defocus blur can be estimated together with the accurate feature location by the proposed feature extraction algorithm. The feature localization problem is modeled as the Gaussian deconvolution given the set of 1D observations. Moreover, the errors in the camera calibration due to refraction of the panel are corrected by refining the camera parameters in consideration of the refraction model.

The proposed method presents a solution to the conventional difficulties in using defocused images for camera calibration. Even though the method requires a set of patterns to be captured in the same physical location, it can calibrate cameras regardless of their focus distances. This advantage allows the use of defocused images and the calibration between cameras focusing at different distances which becomes practicable with high accuracy by the proposed method. The effectiveness of the proposed method has been emphasized in several real experiments using a compact display device such as a smartphone.

## References

- [1] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski. Building rome in a day. *Communications of the ACM*, 54(10):105–112, 2011. **1**
- [2] M. Baba, M. Mukunoki, and N. Asada. A unified camera calibration using geometry and blur of feature points. In *IEEE International Conference on Pattern Recognition (ICPR)*, volume 1, pages 816–819, 2006.
- [3] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, (6):679–698, 1986.
- [4] A. Datta, J.-S. Kim, and T. Kanade. Accurate camera calibration using iterative refinement of control points. In *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 1201–1208. IEEE, 2009. **1**
- [5] A. Geiger, F. Moosmann, O. Car, and B. Schuster. Automatic camera and range sensor calibration using a single shot. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 3936–3943, 2012. **6, 7**
- [6] M. Große, M. Schaffer, B. Harendt, and R. Kowarschik. Camera calibration using time-coded planar patterns. *Optical Engineering*, 51(8):083604–1, 2012. **1**
- [7] C. Harris and M. Stephens. A combined corner and edge detector. 1988. **6, 7**
- [8] J. Heikkilä. Geometric camera calibration using circular control points. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 22(10):1066–1077, 2000. **1**
- [9] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel. Adaptive noise smoothing filter for images with signal-dependent noise. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, (2):165–177, 1985. **6**
- [10] K. Levenberg. A method for the solution of certain problems in least squares. *Quarterly of applied mathematics*, 2:164–168, 1944. **5**
- [11] Y. Liu and X. Su. Camera calibration with planar crossed fringe patterns. *Optik-International Journal for Light and Electron Optics*, 123(2):171–175, 2012. **1**
- [12] L. Lucchese and S. K. Mitra. Using saddle points for sub-pixel feature detection in camera calibration targets. In *Circuits and Systems, 2002. APCCAS'02. 2002 Asia-Pacific Conference on*, volume 2, pages 191–195. IEEE, 2002. **1**
- [13] D. W. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, 11(2):431–441, 1963. **5**
- [14] K. Nakano, M. Okutomi, and Y. Hasegawa. Camera calibration with precise extraction of feature points using projective transformation. In *IEEE International Conference on Robotics and Automation (ICRA)*, volume 3, pages 2532–2538, 2002. **1**
- [15] C. W. Niblack, P. B. Gibbons, and D. W. Capson. Generating skeletons and centerlines from the distance transform. *CVGIP: Graphical Models and Image Processing*, 54(5):420–437, 1992. **3**
- [16] Y. Oyamada, P. Fallavollita, and N. Navab. Single camera calibration using partially visible calibration objects based on random dots marker tracking algorithm. In *IEEE and ACM International Symposium on Mixed and Augmented Reality, Workshop on Tracking Methods and Applications (TMA)*, Atlanta, USA, 2012. **1**
- [17] S. Placht, P. Fürsattel, E. A. Mengue, H. Hofmann, C. Schaller, M. Balda, and E. Angelopoulou. Rochade: Robust checkerboard advanced detection for camera calibration. In *European Conference on Computer Vision (ECCV)*, pages 766–779. Springer, 2014. **1, 3, 6, 7**
- [18] M. Pollefeys, D. Nistér, J.-M. Frahm, A. Akbarzadeh, P. Mordohai, B. Clipp, C. Engels, D. Gallup, S.-J. Kim, P. Merrell, et al. Detailed real-time urban 3d reconstruction from video. *International Journal of Computer Vision (IJCV)*, 78(2-3):143–167, 2008. **1**
- [19] C. Rhemann, A. Hosni, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3017–3024, 2011. **1**
- [20] E. Royer, M. Lhuillier, M. Dhome, and J.-M. Lavest. Monocular vision for mobile robot localization and autonomous navigation. *International Journal of Computer Vision (IJCV)*, 74(3):237–260, 2007. **1**
- [21] H. Schar. *Optimal operators in digital image processing*. PhD thesis, 2000. **4**
- [22] Z. Song and R. Chung. Use of lcd panel for calibrating structured-light-based range sensing system. *IEEE Transactions on Instrumentation and Measurement*, 57(11):2623–2630, 2008. **1**
- [23] H. Surmann, A. Nüchter, and J. Hertzberg. An autonomous mobile robot with a 3d laser range finder for 3d exploration and digitalization of indoor environments. *Robotics and Autonomous Systems*, 45(3):181–198, 2003. **1**
- [24] Z. Wang, W. Wu, X. Xu, and D. Xue. Recognition and location of the internal corners of planar checkerboard calibration pattern image. *Applied mathematics and computation*, 185(2):894–906, 2007. **1**
- [25] K.-J. Yoon and I. S. Kweon. Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 28(4):650–656. **1**
- [26] F. Yu and D. Gallup. 3d reconstruction from accidental motion. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3986–3993, 2014. **1**
- [27] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 22(11):1330–1334, 2000. **1, 5**
- [28] S. Zhuo and T. Sim. Defocus map estimation from a single image. *Pattern Recognition*, 44(9):1852–1858, 2011. **2**
- [29] Z. Zongqian. Camera calibration based on liquid crystal display (lcd). *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37. **1**