Improving the Reliability of Phase Detection Autofocus

Chin-Cheng Chan and Homer H. Chen; National Taiwan University

Abstract

The goal of autofocus is to enable a digital camera to capture sharp images as accurately and quickly as possible in any lighting condition without human intervention. Recent developments in mobile imaging seek to embed phase-detection sensor pixels into the image sensor itself because these phase-detection sensors are able to provide information for controlling both the amount and the direction of lens offset and thereby expedite the autofocus process. Compared to the conventional contrast-detection autofocus algorithms, however, the presence of noise, the lack of contrast in the image, and the spatial offset between the left and right phase sensing pixels can easily affect phase detection. In this paper, we propose to address the issue by characterizing the relation between phase shift and lens movement for various object depths by a statistical model. Experiments are conducted to show that the proposed method is indeed able to improve the reliability of phase-detection autofocus.

1. Introduction

The proliferation of smartphone has made phone camera a predominant imaging device for most people. It is commonly agreeable that the tiny phone camera is an iconic element of the smartphone. Considering that the camera has to capture sharp images as fast as possible in any condition without human intervention, it is fair to say that the most challenging image processing task of a phone camera is autofocus. An autofocus algorithm usually consists of two components: focus measurement and search strategy. The former measures how sharp an image is or how close the in-focus position is during an autofocus process, and the latter determines how to reach the in-focus lens position given current and previous focus measurements.

Autofocus algorithms that use image contrast (equivalently, image sharpness) as focus measurement are collectively called contrast-detection autofocus (CDAF). In CDAF, the image contrast is first calculated in either the image domain [1], [2] or the frequency domain [3]. Then a lens movement is made according to the search strategy [3], [4]. As the lens sweeps across its active motion range, the image contrast forms a "hill-shaped" trajectory with the peak at the in-focus position. The trajectory is referred to as focus profile in this paper. Because image contrast as a focus measurement is relatively reliable, CDAF has been a useful technique. However, when the image has low contrast, especially when the image is blurry, CDAF may fail to give a good estimate of lens movement, resulting in poor autofocus performance.

Autofocus algorithms that use image phase shift as focus measurement are collectively called phase-detection autofocus (PDAF). Such algorithms are developed for sensors with phase sensing elements that are normally classified into left and right sensing elements. Depending on the design considerations, these phase sensing elements can be placed sparsely or densely on the image sensor, and the left and right pixels may or may not be colocated. In theory, when an object is in-focus, the left image formed by the left pixels is aligned with the right image formed by the right pixels. When the object is out-of-focus, the two images

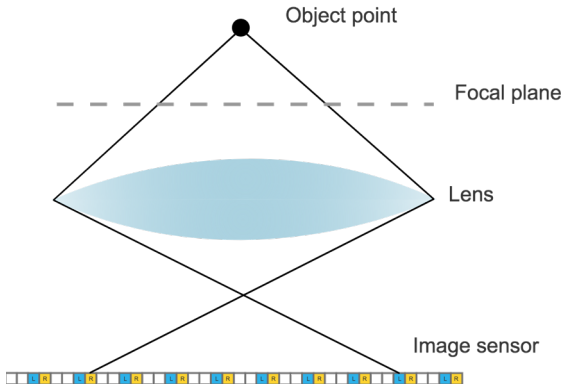


Figure 1. An illustration of the formation of left and right phase images for PDAF. (a) The left light ray coming off an object point behind the focal plane reaches a left (blue) pixel sensor, and the right light ray reaches a right (yellow) pixel sensor. (b) The responses of all such left (right) pixel sensors form a left (right) phase image.

are out of alignment [5]. The sign of the shift between the two images reflects the order of the focal plane with respect to the object. If the object is behind the focal plane, the left image has a positive shift with respect to the right image. On the other hand, if the object is in front of the focal plane, the left image has a negative shift with respect to the right image. While the sign of the phase shift determines the direction of the lens movement, the magnitude of the phase shift indicates the distance between the object and the focal plane. The larger the shift is, the larger the distance is [5].

If the phase information is noise-free, the phase shift between the two phase images can be precisely computed without error, so can the travel distance of the lens to the in-focus position. In practice, however, the phase shift can seldom be error-free due to, for example, image noise, sparsity of phase sensors, and spatial offset between corresponding left and right pixels. Our previous study showed that, although a Gaussian filter can be applied to improve the quality of phase shift measurement [9], it is still insufficient to guarantee a successful autofocus process. More work is needed.

The objective of this work is to investigate how to make a reliable estimate of lens movement from noisy phase information.

We resort to a statistical approach to characterize the relation between phase shift and the distance to the in-focus position. The statistical model is then exploited to determine the lens travel distance given that the object to be focused on can be at any possible depth. Since the resulting PDAF algorithm is to be integrated with a CDAF algorithm [6], [13], [14] that requires three or more data points located on both sides of the hill-shaped focus profile, we need to move the lens to the hillside (slope region) of the focus profile as fast as possible by PDAF and let CDAF take over the remaining autofocus process.

The rest of the paper is structured as follows. The principle of phase detection is described in Section 2. Then our proposed PDAF algorithm is described in Section 3, followed by experimental results in Section 4. The experimental results are discussed in Section 5, followed by the conclusion of the paper in Section 6.

2. Background

Consider a camera pointing at an object behind the focal plane of the camera as shown in Fig 1(a). Part of the light emitted from the object point enters the camera and forms an image on the image sensor. For the image sensor, masks are placed above the left (right) pixels to block light coming from the right (left) and thereby prevent it from reaching the photodiodes underneath the masks. In this way, the left (right) pixels can only detect the light coming from the left (right) side of the pixels. The responses of the left (right) pixels to an object point form the left (right) point spread functions (PSFs) of the camera. A conceptual sketch of left and right PSFs is shown in Fig. 1(b).

The responses of the left (right) pixels to all scene points form the left (right) phase images, which in theory can be obtained by convolving the scene with the left (right) PSF. Specifically, let $s(x,y)$ denote the scene and $h_L(x,y)$ and $h_R(x,y)$ the left and right PSFs, respectively. (Note that, for simplicity, Fig. 1(b) only shows one-dimensional responses of the left and right pixels, but the two dimensional responses can be easily extended from the one-dimensional responses by treating each row of the two-dimensional responses as a separate one-dimensional response.) The left and right phase images, denoted by $l(x,y)$ and $r(x,y)$, respectively, can be described by the following equations:

$$l(x,y) = s(x,y) * h_L(x,y), \tag{1}$$

$$r(x,y) = s(x,y) * h_R(x,y). \tag{2}$$

In general, when the object point is behind the focal plane, the left phase image is a right-shifted version of the right phase image; the further away the object point is, the larger the shift is. On the other hand, when the object point is in front of the focal plane, the left phase image is a left-shifted version of the right phase image; the closer the object point is, the larger the shift is.

2.1 Phase Shift Estimation

For an out-of-focus object, the phase shift between the left and right phase images can be modeled by a one-dimensional translation,

$$r(x,y) = l(x + \Delta x, y), \tag{3}$$

where the sign and the magnitude of Δx indicate the direction and the amount of the phase shift, respectively. Then Δx can be estimated by phase correlation [8]. Specifically, let L and R denote the 2D Fourier transform of $l(x,y)$ and $r(x,y)$, respectively. The correlation matrix $p(x,y)$, whose peak location represents the phase shift, is given by the following formula:

$$p(x,y) = \mathcal{F}^{-1}\{\frac{L \circ \bar{R}}{|L \circ \bar{R}|}\}, \tag{4}$$

where $\mathcal{F}^{-1}\{\cdot\}$ denotes the inverse two-dimensional Fourier transform, "$\circ$" denotes entry-wise multiplication, and "$\bar{}$" denotes complex conjugate.

Ideally, the correlation matrix should have a single peak, from which the phase shift can be determined. However, due to the presence of image noise, the correlation matrix is usually noisy and may contain false peaks. Chan et al. [9] alleviated the impact of noise on phase shift calculation by smoothing the correlation matrix with a Gaussian kernel g(x),

$$p_f(x) = p_r(x) * g(x), \tag{5}$$

where $p_r(x)$, called the correlation curve in this paper, is a row of $p(x,y)$, and $p_f(x)$ represents the filtered result. Note that, even though the correlation matrix is two dimensional, one dimensional filtering is performed in practice since the phase shift is one-dimensional. The effect of Gaussian filtering is shown in Fig. 2. The blue curve represents the correlation curve before the Gaussian filtering is performed, and the red curve represents the filtered version. It can be seen that the unfiltered correlation curve is very noisy, from which it is difficult to determine the phase shift. After filtering, the correlation curve becomes smoother, making it possible to estimate the phase shift by finding the peak of the curve,

$$\Delta x = \arg\max_x p_f(x). \tag{6}$$

Note that only integer phase shifts can be determined from Eq. 6. In practice, the phase shift is usually between ± 2; therefore, only a total of five different lens travel distances is resulted. To make lens movements finer, a subpixel accuracy for phase shift estimation is required. We adopt an interpolation-based method proposed by Tian et al. [10] to obtain subpixel phase shifts.

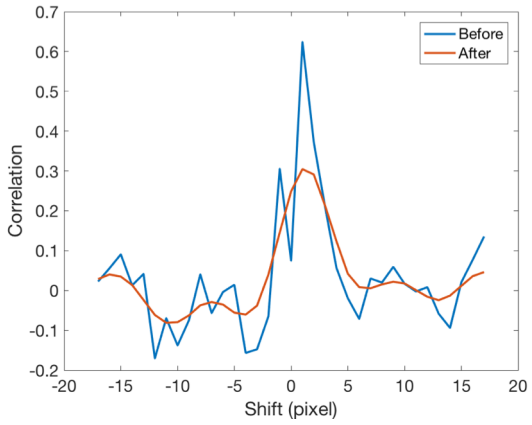
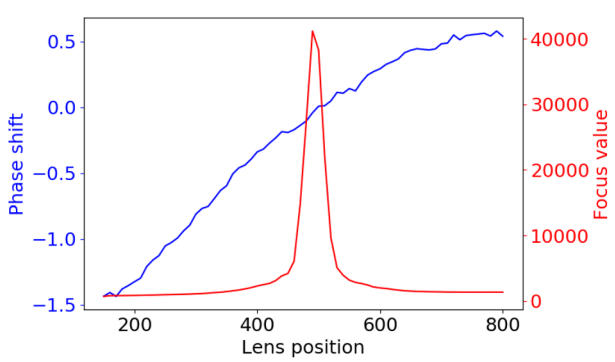
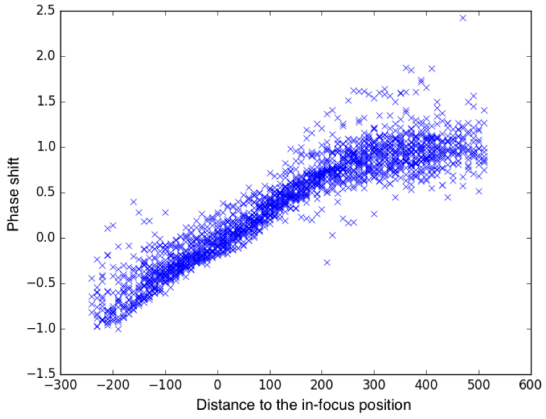


Figure 2. An example correlation curve before and after the Gaussian filtering is performed. The blue curve represents the correlation curve before performing the Gaussian filtering, and the red curve represents the filtered result.



(a)



(b)

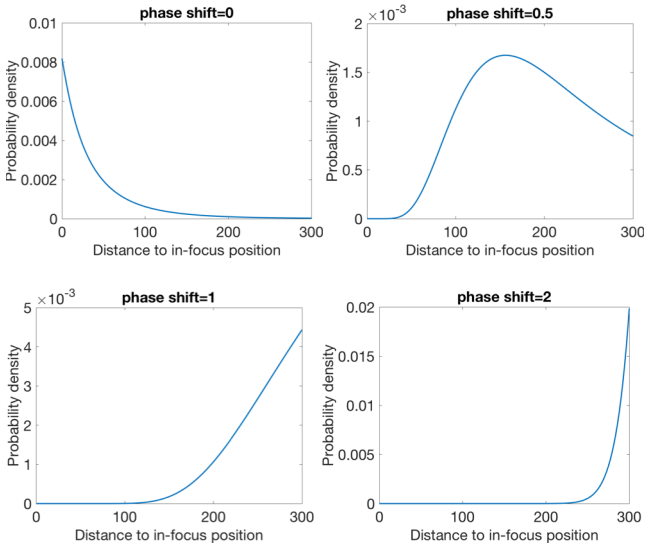
Figure 3. (a) An example focus profile (red) and phase shift profile (blue) with in-focus position at 500. (b) Distribution of phase shifts plotted against the distance to in-focus position.

3. Determine Lens Movement from Phase Shift

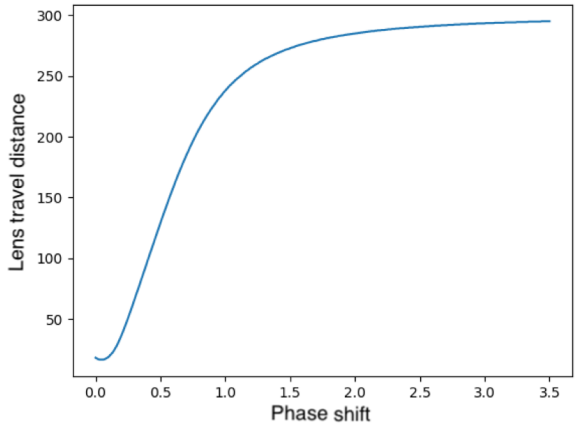
Determining lens movement from phase shift at each step of an autofocus process entails an estimation of the travel distance of the lens to the in-focus lens position. This operation is highly sensitive to noise due to the fact that the phase data are not noise-free. We address the noise issue by using a statistical model to characterize the relation between the phase shift and the lens movement for various object depths. This is done offline with sufficient data. Then, the resulting statistical model is used in the autofocus process to determine the lens travel distance from phase shift.

3.1 Data Preparation

In the data preparation step, the data for creating the statistical model were collected by capturing the focal stacks of 28 different scenes with various depths. For each image in a focal stack, we estimated the phase shift between the left and right phase images. The corresponding distance to the in-focus position was known a priori in the setup. Thus we can obtain the distribution of phase shifts as a function of the distance to in-focus lens position. The resulting plot is called phase shift profile in this paper. Fig. 3(a) shows an example phase shift profile and the corresponding focus profile obtained in this data preparation step. It can be seen that phase shift is roughly proportional to the lens position and close to zero at the in-focus lens position. Fig. 3(b) shows the phase shift distribution for all the 28 scenes.



(a)



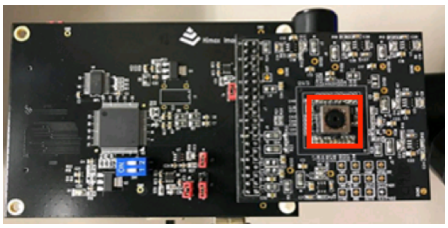
(b)

Figure 4. (a) Probability density function of the distance to in-focus position when phase shift is equal to 0, 0.5, 1, and 2. (b) Lens travel distance calculated by the proposed method for different phase shifts.

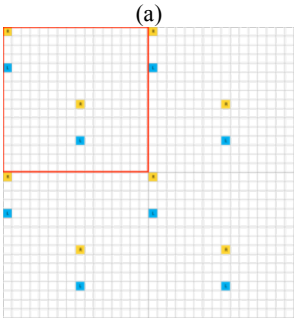
3.2 Mathematical Formulation

Let P denotes the phase shift between the left and right images and X denotes the distance to in-focus lens position. We model the distribution of the phase shift given the distance to in-focus lens position as a Gaussian distribution,

$$f_{P|X}(p|x) = \frac{1}{\sqrt{2\pi\sigma(x)^2}} e^{-\frac{(p-\mu(x))^2}{2\sigma(x)^2}}, \qquad (7)$$



(a)



(b)

Figure 5. (a) Our PDAF development platform. The red box indicates the location of the camera module. (b) Phase sensor pattern. Yellow cells represent the right pixels, blue cells represent the left pixels, and the rest are regular pixels. The red box indicates the basic phase sensor placement pattern that repeats throughout the sensor.

where $\mu(x)$ and $\sigma(x)$, respectively, denote the sample mean and variance of the phase shift calculated at lens position x.

We then employ the Bayes' theorem to obtain the distribution of the distance to in-focus position given the phase shift as follows:

$$f_{X|P}(x|p) = \frac{f_{P|X}(p|x) \cdot f_X(x)}{\sum_{x \in S_x} f_{P|X}(p|x) \cdot f_X(x)}, \tag{8}$$

where S_x denotes the range of lens positions. Assuming the in-focus lens position is uniformly distributed (hence $f_X(x)$ is a uniform distribution), we have

$$f_{X|P}(x|p) = \frac{\frac{1}{\sigma(x)} e^{-\frac{(p-\mu(x))^2}{2\sigma(x)^2}}}{\sum_{x \in S_x} \frac{1}{\sigma(x)} e^{-\frac{(p-\mu(x))^2}{2\sigma(x)^2}}}. \tag{9}$$

Example distributions of the distance to the in-focus position are shown in Fig. 4(a). We can see that, as the phase shift becomes larger, the distribution shifts to the right, meaning that the distance to the in-focus position is larger.

Finally, the lens movement can be determined by

$$\max x \quad \text{subj.} \quad F_{X|P}(x|p) \le \mathrm{c}, \tag{10}$$

where $F_{X|P}(x|p)$ is the cumulative distribution function of $f_{X|P}(x|p)$, and $\mathrm{c} \in [0, 1]$ controls the aggressiveness of the lens movement—the higher c is, the larger the lens movement for the same phase shift. Fig. 4(b) shows the lens travel distance plotted against the phase shift when c is set to 0.6. It can be seen that the larger the phase shift is, the longer the distance. When the phase shift is larger than 1.5, the lens travel distance increases very slowly and almost reaches saturation. We make a conservative lens movement decision when the phase shift is relatively large, because larger phase shift is less reliable.

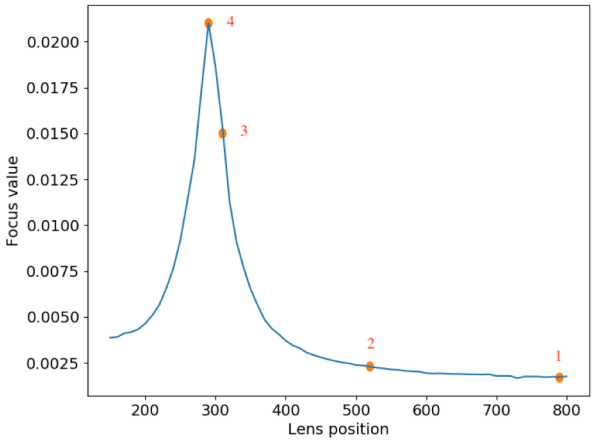


Figure 6. An example lens trajectory in an autofocus process. The blue curve represents the focus profile. The orange dots represent the lens positions in a sequence of movements, with the number represents the order of movement. The initial lens position is 790 and the in-focus lens position is 290 in this case.

Table 1. The probability that the lens reaches the hill side of the focus profile after each movement

# movements	Distance between initial position and in-focus position				
	100	200	300	400	500
1	0.73	0.70	0.25	0.00	0.00
2	0.87	0.80	0.71	0.79	0.83
3	0.90	0.85	0.89	0.92	0.91
4	0.94	0.86	0.93	0.93	1.00

4. Experiment

4.1 Setup

The PDAF platform including the image sensor used in our experiment is shown in Fig. 5. The sensor size of the camera is 3280×2464 pixels, and the size of the left and right phase images are both 410x154 pixels. The range of lens movements is uniformly partitioned into 600 units. In other words, the lens moves at an integer number of units, and the total number of possible lens positions is 600.

We tested our PDAF algorithm on a number of scenes with objects at various depths. We launched the autofocus process 140 times, each time from a different initial lens position, and terminated the process after four lens movements were made or after the in-focus position was reached, whichever came first. A total of 28 scenes were used as test data in the experiment.

4.2 Performance Evaluation

An example lens trajectory of an autofocus process is shown in Fig. 6, for which the initial lens position was 790, and the in-focus position was 290. Our algorithm made large lens movements at first and gradually decreased the lens travel distance as the lens

moved closer to the in-focus position. A desirable PDAF algorithm should get out of the flat area (reach the hillside) of the focus profile as fast as possible while avoiding back-and-forth lens movements; therefore, it should be capable of adjusting the lens travel distance according to the distance to in-focus position.

It is commonly agreeable that CDAF is more accurate for the estimation of in-focus lens position than PDAF when a sufficient number of data points near the in-focus position are available, and that PDAF is able to quickly move the lens to the hillsides of the focus profile. Therefore, it is reasonable to launch PDAF first and let CDAF take over the rest of the autofocus process afterwards. We evaluated the performance of the proposed PDAF algorithm by counting the percentage of times that it moves the lens to the hillside of the focus profile. In this experiment the hillside region for each focus profile is manually determined.

The results are shown in Table 1. It can be seen that when the lens is at a distance of 500 from the in-focus position, at least two lens movements are needed for the lens to reach the hillside of the focus profile. This is the case where the lens is farthest from the in-focus position in the experiment. In the other cases, the lens may have a chance to reach the in-focus position in just one movement. Such behavior is expected, because we take a conservative approach to prevent large lens movements so that a poor estimate of the travel distance would not result in a catastrophic effect in the autofocus process. From the user experience point of view, a smooth ride to the sharpest image with as little bouncing as possible is more desirable than a bumpy ride that generates video frames of rapid alternating sharpness. It can also be seen from Table 1 that the probability for the lens to reach the hillside of the focus profile generally increases with the number of movements. After the third movement, it is higher than 85 percent. The results show the effectiveness of the proposed method.

5. Discussions

This paper describes part of our efforts toward a complete autofocus system employing both CDAF and PDAF. A preliminary result of the integrated autofocus using the development board described in Section IV can be found online [11]. The demo video was obtained by testing our algorithm on four different scenes, including the standard ISO chart, a low textured object, and two outdoor scenes. It can be seen that, in average, our algorithm is able to bring the lens to the in-focus lens position in 6 steps, or equivalently 0.2 seconds at 30fps video rate.

It was brought to our attention that some recent image sensors with built-in PDAF are able to complete the autofocus process in as short as 0.03 seconds [12]. It seems the performance of such sensors is attributed to high-speed and high-density collocated phase sensors that can generate phase shift data at a rate much higher than 30fps. The on-board autofocus capability also saves data communication between the sensor and the image processor. However, it should be noted the proposed approach can be implemented as part of the on-board processing. Furthermore, as noisy phase information is common to PDAF sensors and algorithms, the proposed statistical method can be used by these image sensors to make reliable estimates of lens movement from phase shift data. That is, our method can be combined with cutting-edge imaging hardware to further increase the autofocus performance.

6. Conclusion

The phase data obtained for PDAF are intrisically noisy, which severly affects the performance of PDAF. In this paper, we have presented a statistical approach to address the issue. The proposed method is able to make the lens reach the slope of a focus profile in three moves in 85% of the test cases, allowing for a nice leverage of the complementary nature of PDAF and CDAF. Once the lens is at such position, it becomes relatively easy for CDAF to bring the lens to the in-focus position with precision.

7. References

[1] M. Subbarao, T. Choi, and A. Nikzad, “Focusing Techniques,” J. Opt. Eng., vol. 21, pp. 2824–2836, 1993.

[2] J. M. Tenenbaum, “Accommodation in computer vision,” Stanford University, 1971.

[3] K. S. Choi, J. S. Lee, and S. J. Ko, “New autofocus technique using the frequency selective weighted median filter for video cameras,” IEEE Trans. Consum. Electron., vol. 45, no. 3, pp. 820–827, Aug. 1999.

[4] E. P. Krotkov, Active Computer Vision by Cooperative Focus and Stereo. Springer-Verlag New York, 1989.

[5] M. Levoy, “Autofocus: phase detection.,” 2010. [Online]. Available: http://courses.cs.washington.edu/courses/cse131/13sp/applets/autofocusPD.html.

[6] D.-C. Tsai and H. H. Chen, “Reciprocal Focus Profile,” IEEE Trans. Image Process., vol. 21, no. 2, pp. 459–468, 2012.

[7] M. Hamada, “Imaging device including phase detection pixels arranged to perform capturing and to detect phase difference,” US20130088621, 2013.

[8] C. D. Kuglin and D. C. Hines, “The Phase Correlation Image Alignment Method,” in Proc. IEEE Conf. Cybernetics and Society, 1975, pp. 23–25.

[9] C.-C. Chan and H. H. Chen, “Enhancement of phase detection for autofocus,” in IEEE Proc. of Int. Conf. Image Processing, 2017, pp. 41–45.

[10] Q. Tian and M. N. Huhns, “Algorithms for subpixel registration,” Comput. Vision, Graph. Image Process., vol. 35, no. 2, pp. 220–233, Aug. 1986.

[11] C.-C. Chan, “Autofocus Using Phase and Contrast Information,” 2017. [Online]. Available: https://youtu.be/0lLnIIFYdWA.

[12] “Sony Announces a New Type 1/2.6 22.5 Megapixel Exmor RSTM, the Industry’s First Stacked CMOS Image Sensor with Built-in Hybrid Autofocus and 3-Axis Electronic Image Stabilization,” 2016. [Online]. Available: https://www.sony.net/SonyInfo/News/Press/201602/16-013E/.

[13] D.-C. Tsai and H. H. Chen, “Autofocus System,” US 8571403, Oct. 2013.

[14] D.-C. Tsai and H. H. Chen, “Autofocus Method,” US 8254774, Aug. 2012.

8. Author Biography

Chin-Cheng Chan is currently a senior undergraduate student in National Taiwan University. He was a summer intern at MediaTek, Taiwan in 2016. His research interests are in signal processing, especially image processing, and his research topics include medical image processing and autofocus for smartphone cameras.

Homer H. Chen is an IEEE Fellow. His professional career has spanned industry and academia. Since August 2003, he has been with the College of Electrical Engineering and Computer Science, National Taiwan University, where he is Distinguished Professor. Prior to that, he held various R&D management and engineering positions with U.S. companies, including AT&T Bell Labs and Rockwell Science Center, over a period of 17 years.