On the Distinction between Phase Images and Two-View Light Field for PDAF of Mobile Imaging

Chi-Jui (Jerry) Ho and Homer H. Chen

Abstract

Abstract—*A phase detection autofocus (PDAF) algorithm iteratively estimates the phase shift between the left and right phase images captured in an autofocus process and uses it to determine the lens movement until the estimated in-focus lens position is reached. Such phase images have been assumed to be equivalent to a two-view light field. If the assumption is true, then the phase shift between the two phase images can be obtained by stereo matching or similar techniques. In this paper, we argue that it is a wrong assumption and provide insights into the distinctions between phase images and two-view light field from the autofocus perspective. We also support our argument by conducting an experiment to show that both stereo matching and optical flow result in inferior PDAF performance than the phase correlation technique and the AF-Net technique that specifically target phase images.*

1 Introduction

Autofocus is a key function for mobile imaging [8]–[14]. It normally takes more than one iteration to find the in-focus lens position and hence the sharpest image. In the iterative process, the first image acquired is usually a blurry image. Then, the autofocus system subsequently makes decision according to the information obtained in the previous step. This iterative process continues until the in-focus position is found. In general, the autofocus performance is evaluated by how fast it completes an autofocus process and how close the final lens position is to the actual in-focus lens position.

A typical autofocus technique is called phase detection autofocus (PDAF), which requires a special sensor that provides phase information [1]. On the sensor plane, some regular sensor elements are replaced by phase detectors. Unlike the regular image sensor that captures the light from all directions, the phase detector only takes the light from a certain direction. In PDAF, the left and right phase detectors take the light from left and right, respectively. These phase detectors are embedded in the image sensor. The image formed by the left phase detectors are called the left phase image. Similarly, the image formed by the right phase detectors are called the right phase image. In each iteration of the autofocus process, the lens movement is determined from the phase images [11]–[13].

By comparing the pixel value of left and right phase images, we can find an offset between the two phase images. This offset is called phase shift, which corresponds to the relative position between the object and the focal plane. The larger the distance, the larger the magnitude of phase shift. The phase shift is positive when the object is behind the focal plane and negative when the object is in front of the focal plane. Since the position of focal plane varies with the lens position, the phase shift between the left and right phase images indicates the optimal lens movement that is required to align the focal plane with the object plane.

Phase shift estimation of PDAF may seem similar to disparity estimation of stereo images or optical flow estimation of a dynamic image sequence. The rationale behind this viewpoint is that all such algorithms take a pair of images as input and outputs the displacement between the two images. This viewpoint is also supported by the observation that both phase shift and disparity relate to object depth.

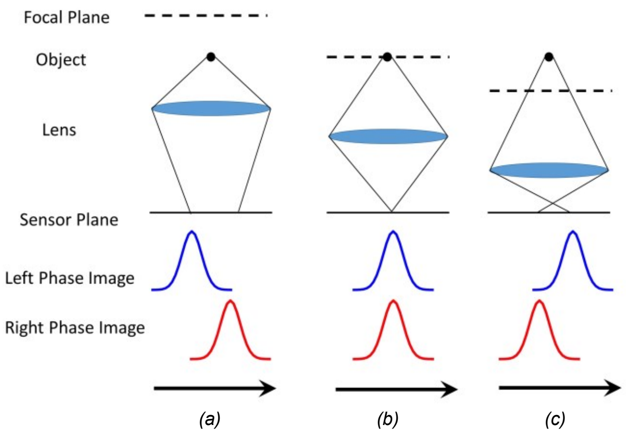


Figure 1. *Illustration of phase images for three cases: The focal plane is (a) behind, (b) at, and (c) in front of the in-focus position.*

We argue that such a viewpoint is incorrect and that displacement estimation algorithms developed for two-view light field (or stereo images) are inappropriate for PDAF. To support our argument, we conduct an experiment to show that classical disparity estimation and optical flow estimation result in either unstable phase shift estimation or inferior autofocus performance than techniques that are specifically designed for phase images.

2 Background

In this section, we first review the basic principle of PDAF and two-view light field. Then we discuss the related methods.

2.1 Phase Detection Autofocus

Suppose the camera specifications are known. Then we may determine the focal plane position from the lens position. An autofocus algorithm iteratively estimates the position of the object plane and move the lens accordingly until the lens reaches the in-focus position (at which a sharp image can be captured). The sharpness of an image is related to the distance between the object plane and the focal plane [15]. Traditional algorithms estimate the distance between the focal plane and object plane according to image contract. Such algorithm is called contract detection autofocus (CDAF). It is usually accurate but slow.

Phase detection autofocus (PDAF) emerged as a replacement of CDAF due to its speed advantage. Phase detectors for PDAF are embedded on the image sensors. Each phase detector detects the light coming from a distinct direction. For example, left phase detectors only capture light coming from the left direction, and right

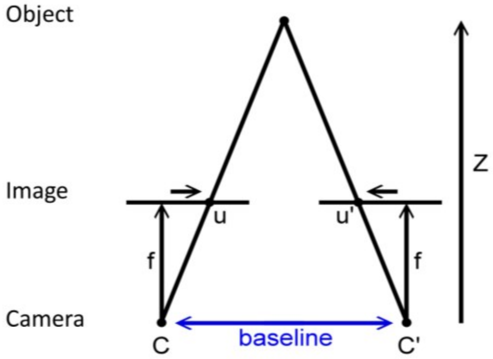


Figure 2. *Illustration of the configuration of a pair of stereo images.*

phase detectors only capture light coming from the right direction. The pixel values of these right (left) phase detectors are assembled as right (left) phase images. As shown in Fig. 1, when the focal plane and object plane are aligned, the left and right phase images are identical. On the other hand, when the two planes are not aligned, a phase shift between the left and right phase images is resulted. The phase shift is positive (negative) when the focal plane is behind (in front of) the object. Moreover, the magnitude of phase shift increases as the focal plane moves away from the object in either direction. If we plot the phase shift against lens position, a nearly linear profile called phase shift profile is obtained.

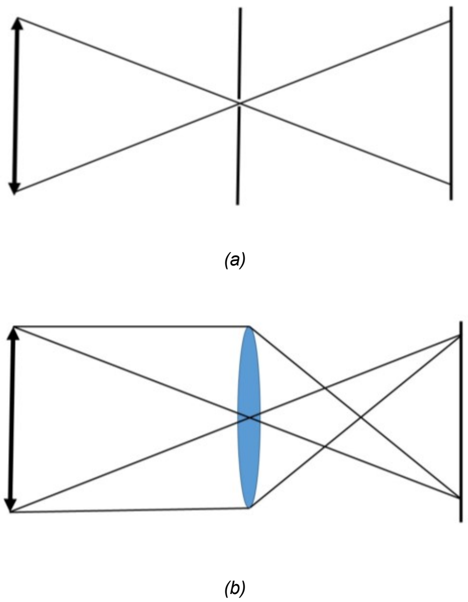
PDAF is faster than CDAF for two reasons. First, the target of PDAF is zero phase shift, which is invariant for any given scene. In other words, it is a fixed target to search. In contrast, the target of CDAF is the highest contrast, which varies with scenes and hence is a variable target to search. As a result, a CDAF algorithm has to sample more points and hence more iterations to find the in-focus lens position. Second, PDAF can determine whether the focal plane is behind the object from the sign of phase shift. This is not the case for CDAF, which takes at least two frames to determine the correct direction of lens movement. That is, PDAF can make the lens to move toward the in-focus position at the kickoff of an autofocus process, but CDAF cannot.

Typically, a PDAF algorithm first computes the phase shift using phase correlation [11]–[13] and then determine the lens movement. The phase correlation is performed in, say, the x-direction. Only a one-dimensional correlation is required. The phase shift corresponds to the peak of correlation curve, which is computed as follows:

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

where $F^{-1}\{\cdot\}$ denotes the inverse 2D Fourier transform, L and R denote the 2D Fourier transform of left and right images, respectively, and the symbols “ $\circ$ ” and “‾” denote element-wise multiplication and complex conjugate, respectively.

Then, a statistical [11] or reinforcement learning method [13] is applied to characterize lens movement from phase shift. However, such algorithms are noise sensitive. To address the issue, a CNN-based model called AF-Net [2] has been proposed. The AF-Net directly determines the lens movement from phase images. An attractive feature of this approach is that it can reach the in-focus lens position in two lens movements on average even for noisy phase data.


(a)

(b)

Figure 3. *(a) Pin-hole (b) thin-lens model of image formation.*

2.2 *Two-view Light Field*

Two-view light field technique is typically used for depth estimation. In this technique, a static scene is captured by two identical cameras at different positions. The placement of the cameras is illustrated in Fig. 2. Since the baseline between the two cameras is nonzero, an object appears at pixel position $u = (x, y)$ in the left image will appear at pixel position $u' = (x - d, y)$ in the right image. The offset d is called disparity. Give the pixel correspondence between the image pair, the object depth z is computed as follows:

$$z = \frac{B \times f}{d}, \tag{2}$$

where B denotes the baseline between two cameras and f denotes the focal length of cameras.

A typical algorithm for disparity estimation consists of four pipelined steps: matching cost computation, cost aggregation, optimization, and refinement [17]. In recent years, CNN has been introduced to this pipeline. Zbontar and LeCun [3] were among the first to introduce a deep Siamese network (MC-CNN) for matching cost computation. Subsequently, many extensions were developed to replace the whole pipeline with an end-to-end network. Mayer et al. adopted an encoder-decoder architecture for depth map estimate [4]. Chang et al. [5] further improved the accuracy by aggregating the context features in different scales.

3 Distinctions

In this section, we discuss the distinctions between phase images and two-view light field from two standpoints: image formation and physical property.

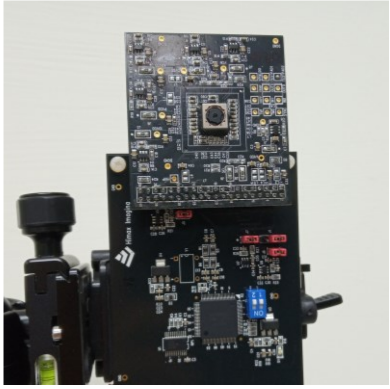


Figure 4. *Our PDAF platform.*

3.1 *Image Formation Process*

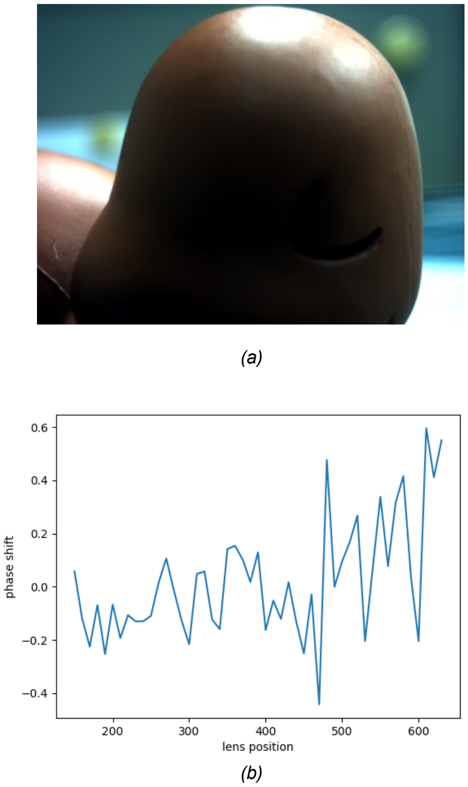
The formation of phase shift and disparity is illustrated in Figs. 1 and 2, respectively. Phase images and two-view light field are captured by different image formation processes. Most two-view light field algorithms use the pinhole model shown in Fig. 3(a) to describe the image formation process. Under this model, each pixel in the left image has a corresponding pixel in the right image on the epipolor line.

It should be noted that two-view light field algorithms assume that all the inputs are sharp images. However, the images for autofocus are not necessarily sharp. In fact, the images are sharp only when the lens is at the in-focus position. It becomes obvious that the pinhole model is not applicable to autofocus, and a more realistic model like the thin-lens model illustrated in Fig. 3(b) is required. This model can describe the phenomenon where the lens is out of focus and a circle of confusion is resulted on the image plane. As the lens goes away from the in-focus position in either direction, the area of the circle of confusion increases.

3.2 *Physical Property of Phase Shift and Disparity*

Consider an in-focus object, the phase shift between the left and right phase images is zero. That is, the object appears at the same pixel position in the left and right phase images. In contrast, an out-of-focus object is not collocated in the left and right phase images; there is an offset between the pixels corresponds to the object in the left and right phase images. The sign of the offset depends on whether the object is in front of the focal plane or not. However, this is not the case for two-view light field, for which the object is not collocated in two stereo images unless the object is at a distance. Therefore, when we alternatively display the left and right phase images on a monitor, every pixel moves except those on an in-focus object. This is not the case for stereo images, for which every pixel moves except those on objects beyond the hyperfocal distance.

A special case of stereo imaging occurs when the camera placement is verging. In such case, zero disparity corresponds to a finite depth rather than infinity. However, because the epipolar line is tilted, the disparity has a vertical component. This is not the case for PDAF, where the phase shift only has horizontal component since the left and right phase sensors are on the same plane. Therefore, we only consider non-verging stereo images.



(a)

(b)

Figure 5. *(a) In-focus image and (b) phase shift profile of a focal stack removed from our dataset.*

4 Experimental Setup

We conducted an experiment to support our argument that phase images obtained in PDAF cannot be approximated by two-view light field. In this section, we describe the details of our experiment, including the data collection, the policy of autofocus, and the metrics for evaluation.

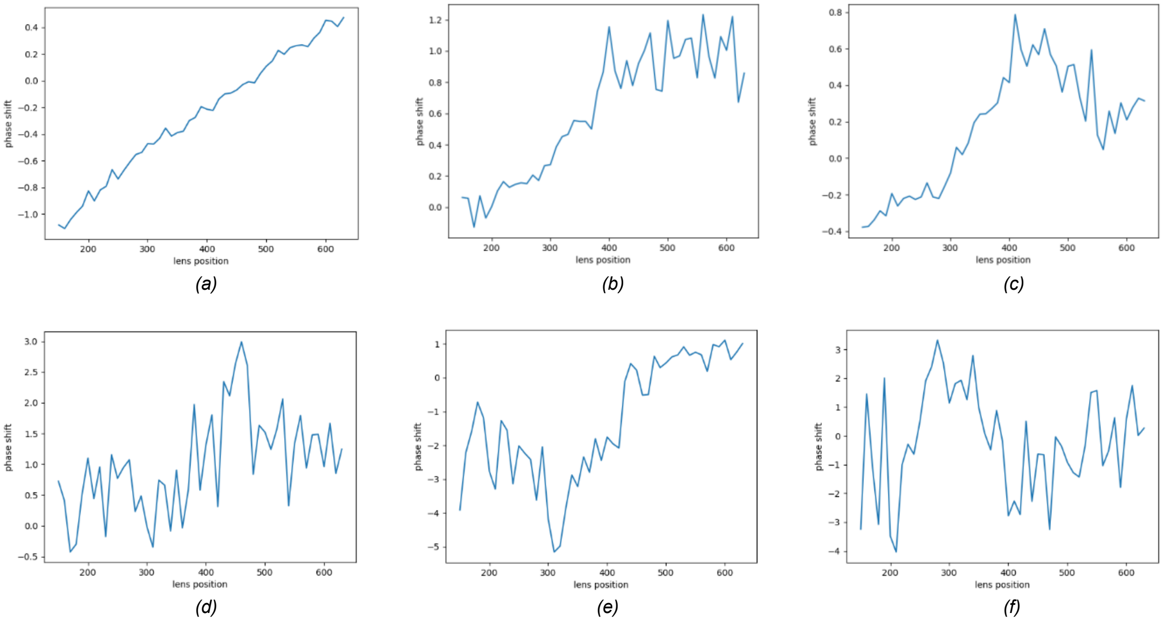
4.1 *Data Collection*

We use the camera shown in Fig. 4 to collect data. For each scene, we sweep the lens along all the 49 available positions of the lens and capture corresponding images. The group of images forms a focal stack. When sweeping the lens, the focal plane goes from zero to the hyperfocal distance. We determine the in-focus image of a focal stack according to the image contrast. That is, the lens position corresponds to the peak image contrast is labeled as the in-focus position, which is the target for autofocus.

Note that many factors such as reflection, backlighting, and over-exposure may affect the quality of a focal stack . Because of their poor image quality, focal stacks suffered from these factors should be excluded from the dataset. In this work, we determine the quality of a focal stack by measuring how the phase shift profile fluctuates. An example of poor-quality phase shift profile is shown in Fig. 5.

4.2 *Policy of Autofocus*

We fired up a number of autofocus processes for each test scene. The lens starts at a different initial position in each autofocus process. The initial distance to the in-focus lens position ranges from -30 to



(a) (b) (c)

(d) (e) (f)

Figure 6. *The phase shift profiles obtained from (a)-(c) phase correlation and (d)-(f) census cost.*

Table 1. Performance of Different PDAF Methods

Model	Success Rate (%)	Lens Position Error	Number of Lens Movements
AF-Net [2]	**95.98**[*]	**1.094**	**2.07**
Reinforcement Learning [13]	84.95	2.870	2.43
Statistical [11]	47.69	5.057	2.66

[*]The highest performance is shown in boldface.

30. We terminate an autofocus process if the maximum iteration is reached or the estimated distance to the in-focus lens position is within a nearness threshold. After termination, an autofocus process is claimed successful if the distance of the lens to the in-focus position is within the threshold. In our experiment, the maximum iteration is 5 and the nearness threshold is 3.

4.3 *Metrics*

We use three metrics to evaluate the PDAF performance. The first is success rate, which measures the ratio of the number of successful autofocus processes to all autofocus processes. The second is average iteration, which is the average number of lens movements of successful autofocus processes. The third is lens position error, which measures the average distance between the terminating lens position and the in-focus lens position.

5 Results and Discussions

In this section, we first compare the performance of phase correlation with AF-Net. Then, we show the performance of two-view light field for autofocus.

Table 2. Performance comparison of AF-Net and FlowNet 2.0

Model	Success Rate (%)	Lens Position Error	Number of Lens Movements
AF-Net [2]	**95.98**	**1.094**	**2.07**
FlowNet2.0 [7]	42.24	8.384	2.92

5.1 *Comparison of Phase Correlation and AF-Net*

We compare the AF-Net with two typical methods that determine the lens movement from phase shift using statistical model [11] and recurrent neural network (RNN) agent [13]. Specifically, the RNN agent learns the lens movement by the reinforcement learning technique.

As shown in Table 1, AF-Net has superior performance in terms of accuracy and speed. We also observed that the AF-Net performs stably even in the presence of noisy phase data. However, this is not the case for the other two methods, which work well on the nearly linear phase shift profile such as Fig. 6(a). For phase shift profiles with high fluctuation shown in Figs. 6(b) and (c), a slow or inaccurate autofocus is obtained due to noisy phase shift.

5.2 *Phase Correlation and Census Cost*

We compare two methods for phase shift estimation: phase shift correlation and Census cost. The latter is typically used to compute the matching cost between stereo patches [16]. The former is augmented with a refinement step [12] to enhance its robustness.

Fig. 6 shows the experimental results of three selected scenes. Figs. 6(a)–(c) are the results of the phase correlation method, and Figs. 6(d)–(f) are the results of Census cost. As we can see, the phase shift profiles generated by phase correlation are approximately linear around zero phase shift. However, the three phase shift profiles generated by Census cost are very noisy. These results support our argument that two-view light field algorithms are not the right tools for phase shift estimation because phase images for PDAF are not two-view light field.

5.3 *Comparisons between AF-Net and FlowNet2.0*

FlowNet 2.0 [7] is a CNN-based method for optical flow estimation; it works well for multiscale displacement estimation. Here we applied the pre-trained FlowNet 2.0 to PDAF as follows:

- Generate a flow map between the two phase images
- Take the horizontal component of average flow value in the flow map.
- Convert average flow value to lens movement by a transformation.

We compare the performance of AF-Net and FlowNet 2.0 and show the results in Table II. As we can see, the AF-Net outperforms the FlowNet 2.0 in all metrics. One possible reason is that the second step of the baseline is problematic; the focus window may contain information unrelated to the focused object, such as background. As a result, directly taking the average horizontal flow value in the focus window may lead to erroneous lens movement.

5 Conclusions

Phase detection autofocus is an important technique for digital imaging. However, phase images for PDAF can be easily confused with the two-view light field in stereo vision, leading to the misconception that the phase shift for PDAF can be accurately computed by a two-view light field technique. In this paper, we have argued that they are distinctively different in image formation and physical properties and that one should not confuse them when designing a PDAF system. We have also provided extensive experimental results using a mobile imaging platform to show that when a two-view light field algorithm is applied to PDAF, either an inaccurate or slow autofocus will be resulted. This work is part of a research project that aims at advancing the computational technique for mobile imaging [18].

References

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

[2] C. J. Ho, C. C. Chan, and H. Chen, "AF-Net: A convolutional neural network approach to phase detection autofocus," accepted by *IEEE Trans. on Image Process.*

[3] J. Zbontar and Y. LeCun, "Computing the stereo matching cost with a convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2015, pp. 1592–1599.

[4] N. Mayer, E. Ilg, P. Häusser, P. Fischer, D. Cremers, A. Dosovitskiy, T. Brox, "A Large Dataset to Train Convolutional Networksfor Disparity, Optical Flow, and Scene Flow Estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2016, pp. 4040–4048.

[5] J-R Chang and Y-S Chen, "Pyramid Stereo Matching Network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 5410–5418.

[6] P. Fischer, A. Dosovitskiy, E. Ilg, P. H¨ausser, C. Hazirbas, V. Golkov, P. van der Smagt, D. Cremers, and T. Brox, "FlowNet: Learning optical flow with convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.,* 2015, pp. 2758–2766.

[7] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, "FlowNet 2.0: Evolution of optical flow estimation with deep networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1647–1655.

[8] D.-C. Tsai and H. H. Chen, "Focus profile modeling," in *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 818-828, 2016

[9] D.-C. Tsai, Z.-M. Tsai, and H. H. Chen, "A Simulation model for continuous autofocus design," *IEEE Trans. Consumer Electron.*, vol. 59, no. 4, pp. 731-737, 2013

[10] D.-C. Tsai and H. H. Chen, "Reciprocal focus profile," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 459 - 468, Feb. 2012

[11] C.-C. Chan and H. H. Chen, "Improving the reliability of phase detection autofocus," *Proc. IS&T Electronic Imaging*, Jan, 2018, pp. 241-1–241-5.

[12] C.-C. Chan, S.-K. Huang, and H. H. Chen, "Enhancement of phase detection for autofocus," in *Proc. IEEE Int. Conf. Image Process.*, 2017, pp. 41–45.

[13] C.-C. Chan and H. H. Chen, "Autofocus by deep reinforcement learning," in *Proc. Electronic Imaging*, 2019.

[14] N. Wadhwa, R. Garg, D. E. Jacobs, B. E. Feldman, N. Kanazawa, R. Carroll, Y. Movshovitz-Attias, J. T. Barron, Y. Pritch, and M. Levoy, "Synthetic Depth-of-Field with a Single-Camera Mobile Phone," *ACM Trans. Graph.*, vol. 37, no. 4, 2018.

[15] Y. Schechner and N. Kiryati, "Depth from defocus vs. stereo: How different really are they?" in *International Journal of Comput. Vis.,* vol. 39, no. 2, pp. 141-162, 2000.

[16] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondance," in *European Conf. Comput. Vis.*, 1994, pp. 151–158.

[17] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," in *International Journal of Comput. Vis.*, vol. 47, no. 1, pp. 7–42, 2002.

[18] C.J. Ho and H. H. Chen, "Demo video: Comparison between the AF-Net and iPhone7," 2019, [Online]. Available: https://youtu.be/ApXMDT774aA

Author Biography

Chi-Jui Ho *was born in Taipei, Taiwan, in 1996. He received his bachelor's degree in in electrical engineering from National Taiwan University in 2019. He was a summer intern in Mediatek, Hsinchu, Taiwan, in 2018. His research interests include image processing, computer vision and machine learning. His research topic is 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 the US companies, including AT&T Bell Labs, Rockwell Science Center, iVast, and Digital Island over a period of 17 years. He was a General Chair of 2019 IEEE ICIP. Currently, he serves on the Awards Board of IEEE Signal Processing Society and the Senior Editorial Board of IEEE JSTSP.*