DOI: 10.2352/ISSN.2470-1173.2016.18.DPMI-254

Using Disparity Information for Stereo Autofocus in 3-D Photography

Shao-Kang Huang, Cheng-Chieh Yang, Kuang-Tsu Shih, and Homer H. Chen; National Taiwan University, Taipei, Taiwan

Abstract

Most stereo cameras are equipped with lens modules of fixed focus to capture images. To provide sharp images over a wide range of object distances, stereo cameras with adjustable focus lens modules have been developed. Although such stereo cameras can be controlled by performing existing monocular autofocus for each individual camera independently, the lack of coordination between the stereo cameras and the underutilization of depth information embedded in the image data make the approach not as efficient as it can be. In this paper, we propose a novel stereo autofocus approach that exploits the disparity of stereo images to control lens movement. Experimental results show that the proposed approach can bring the lenses to the peak zone of the focus profile within two lens movements in most (92.3%) cases even if the lenses are initially far from the in-focus position.

Introduction

Existing autofocus (AF) methods for single cameras can be classified into two categories: active and passive. Active methods [1], [2] exploit an external device emitting infrared (or ultrasound) to estimate the object distance and the in-focus lens position for taking a photo of the object. Although active methods are powerful for handling dark scenes where the image contrast is extremely low, it may easily fail in the presence of moving objects or if there is a transparent object between the camera and the scene of interest (e.g. looking out through a glass window). In contrast, passive methods [3]–[5] only rely on the image captured by the camera and normally consist of two basic operators: measuring image sharpness [6]–[9] and searching the in-focus lens position. The in-focus lens position is determined by examining the relation between the focus value (the sharpness of image) and the lens position. The direction and the distance to the in-focus lens position are estimated according to the focus data available at each sampling time.

However, it is unlikely for a single camera to accurately estimate the in-focus lens position instantly because the object distance is not known. It normally requires a few samples to make a sensible estimation. For this reason, directly applying conventional contrast detection autofocus (CDAF) techniques for monocular cameras to each of the stereo camera may results in a sluggish start. By utilizing the depth information embedded in the stereo images, it is possible to speed up the autofocus process. Therefore, we are motivated to take an integral approach to stereo autofocus by exploiting the stereo images.

Stereo cameras mimic human binocular vision and enable the reconstruction of 3-D object structure from 2-D image data. The estimated object distance serves as a good basis to determine the in-focus lens position. Therefore, in theory, autofocus for stereo cameras should be able to reach in-focus faster than that for monocular cameras. In practice, however, the first pair of stereo images is often out of focus (and thus blur) because the lens is not at the in-focus position yet, resulting in inaccurate disparity estimate. Although several methods [10], [11] have been proposed to solve the problem by deblurring the defocused images, the time-consuming deblurring operation significantly slows down the autofocus process. A more effective method is needed.

Given that depth information is useful for a wide range of applications, the goal of this work is to develop an effective approach that can generate sharp stereo images of 3-D scenes in a fast and robust manner regardless of the object distance. If stereo images are always sufficiently sharp, accurate depth estimation can be obtained, yielding fast and accurate autofocus. However, in practice, the object of interest can be at an arbitrary position with respect to the initial lens position, so the stereo autofocus process often starts with blur images. This creates an inherent dilemma: Sharp stereo images cannot be captured unless the lens of each camera is at the in-focus position, which cannot be obtained if the stereo images are blur. Therefore, it is essential for stereo autofocus to resolve the dilemma effectively.

There are methods that attempt to detect the blur region in the image and propagate the disparity computed at confident (or sharp) pixels in the image to neighboring blur pixels for depth map reconstruction [12]. The additional computational cost introduced by this approach is a major concern. For the purpose of autofocus, the stereo matching can be confined to the image area within a small focus window, and the scene captured within the focus window is represented by a single depth. Of course, this uniform depth assumption is not necessarily true. Therefore, how the depth variation impacts the accuracy of disparity estimation has to be carefully examined. In principle, a stereo autofocus process should be resistant to the depth variation so that robust performance can be ensured.

To investigate how disparity can be used to speed up CDAF, we are motivated to first analyze the performance of the (disparity-based) stereo autofocus. One of the critical issues concerns how fast, in term of the number of video frames required, the disparity estimate can converge with the lens at an arbitrary initial position. Another critical issue is about the accuracy of stereo autofocus. It is required that the lens finally moves to the in-focus position. For most applications of stereo cameras, the cameras have to be able to capture sharp images such that accurate depth estimation can be made. It would be desirable if the disparity-based lens movement decision can effectively guide the lens to reach the peak zone of the focus profile no matter how blurry the initial images are. The peak zone is a small region around the peak of the focus profile [3]. This would resolve the sluggish start issue of most CDAF methods.

Unlike other methods, the disparity-assisted autofocus proposed in this paper does not require a computationally expensive deblur process to sharpen the images. This greatly reduces the computational cost. A correct decision of the lens movement direction holds the key to a successful autofocus. These are the novelties of the proposed approach.

DOI: 10.2352/ISSN.2470-1173.2016.18.DPMI-254

Overview of the Proposed Method

The flow chart of the proposed stereo AF algorithm is shown in Fig. 1. Basically, the algorithm works as follows. First, a pair of stereo images is captured at the initial lens position. Next, the disparity of the image region within the focus window is computed. Since the pixel values in one image may differ from the pixel values in the other image by a offset, zero-mean sum of absolute differences (ZSAD) is applied to estimate the disparity. Finally, the lens movement is estimated according to an empirically pre-determined mapping table. The mapping table records the relation between the in-focus lens position and the disparity value, as depicted in Fig. 2. It is created by applying regression to the pre-captured stereo image data. The regression only needs to be done offline once. The three steps are repeated until the estimated lens movement is zero.

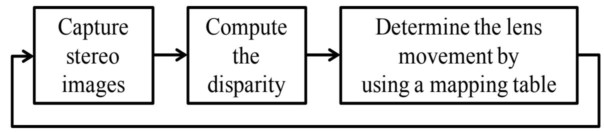

Figure 1. Flow chart of our stereo AF algorithm.

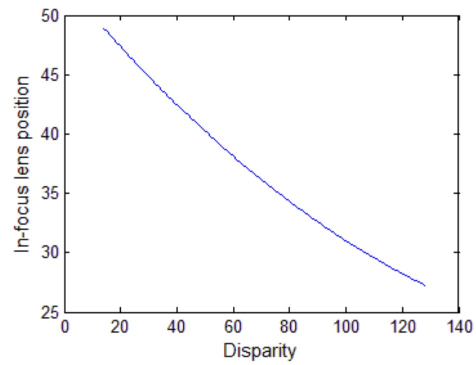

Figure 2. Mapping of disparity to in-focus lens position determined empirically.

Disparity-based Lens movement

When a pair of stereo images is captured by the cameras, stereo matching is applied to estimate the disparity between the two adjacent views of the scene. In autofocus, this time-consuming operation is not applied to the whole image but only to the image area within the small focus window. We convert the color space of the image from RGB to grayscale. Denote the left and right intensity image by I_1 and I_2 respectively. For each pixel $p(i, j)$ of the focus window, the cost of stereo vision at disparity δ is computed by

$$C(\delta) = \sum_{\forall(i,j)} \left| I_1(i,j) - l - I_2(i,j+\delta) + r \right| \quad (1)$$

where l and r are the average intensity values in the left and right focus windows, respectively. The disparity between the stereo views can then be obtained by

$$\delta^* = \arg\min_{\delta} C(\delta) \quad (2)$$

Finally, the lens movement is determined by the mapping table using the disparity (an integer) as an input index.

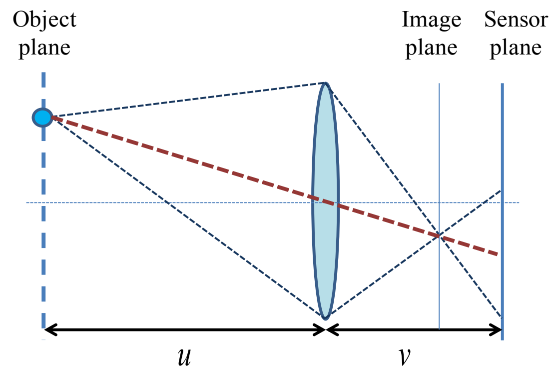

Figure 3. The image formation model of a thin lens camera.

The Mapping Table

Now, we characterize the in-focus lens position as a function of the disparity computed from (2) for stereo autofocus. Fig. 3 shows the image formation of a thin lens model. With a thin lens model, the distance between the lens and the sensor plane of the camera can be defined by

$$v = \frac{uf}{u - f} \quad (3)$$

where f is the focal length and u is the distance between the lens and the object plane. According to the geometry of stereo vision, the depth u can be represented by

$$u = \frac{fb}{\delta} \quad (4)$$

where b is the baseline between the two cameras. Substituting (4) to (3), we have

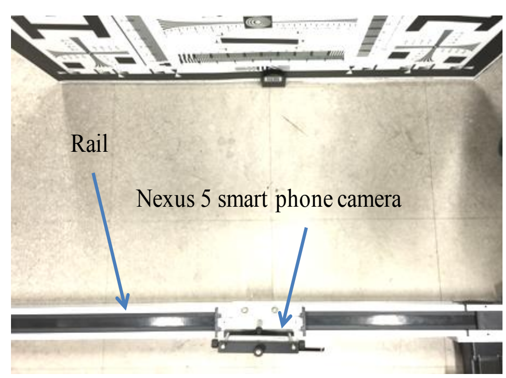

Figure 4. Top view of the experimental setup for generating stereo data by moving a phone camera on a rail.

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.18.DPMI-254

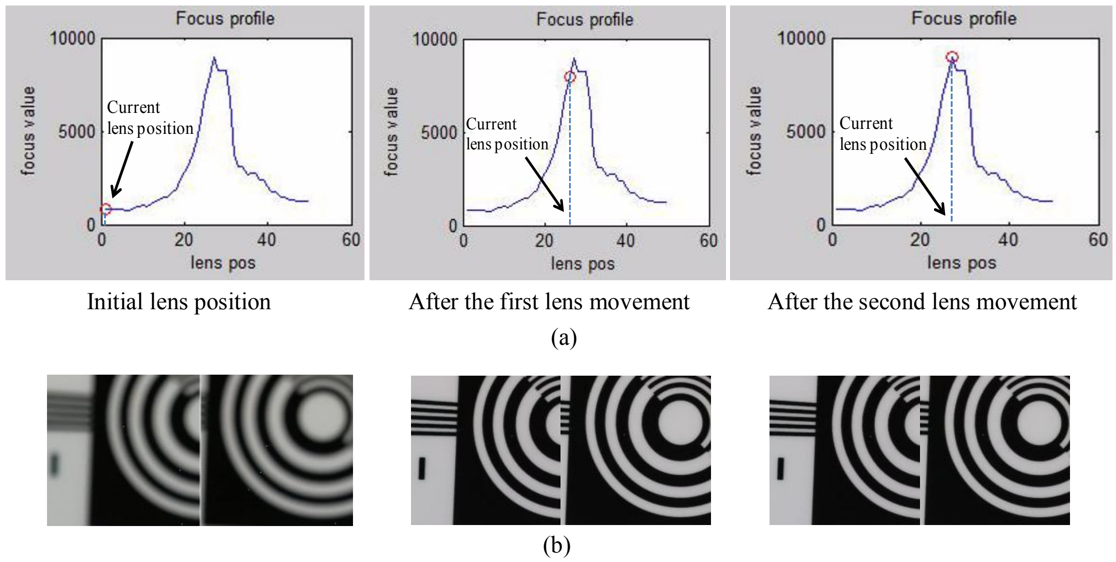

Figure 5. (a) The focus profile of the test scene and the focus value in the searching process. (b) The corresponding stereo views.

$$v = \frac{f^2 b}{fb - f}\frac{1}{\delta} \tag{5}$$

We can see that the in-focus position is in inversely proportional to the disparity. However, in practice, even we have the accurate disparity, the estimation of v may not be correct. This is because most camera parameters are approximate values [13]. If we use (5) directly for stereo autofocus, the approximation error may fool the autofocus system and make it capture a blurry image. To prevent the error from affecting the in-focus decision for stereo cameras, we build a mapping table relating the disparity to the in-focus position.

The setting is shown in Fig. 4. We place a planar object with sufficient features in front of the camera at position u and collect stereo images of the static scene. We collect 50 images of the scene by sweeping the lens across its entire dynamic range at a constant interval. To eliminate the effect of lens aberration, the focus window is selected near the center region of the image. Then, Tenegrad method [14] is applied to measure the image sharpness. The image is convolved with a horizontal filter and a vertical gradient filter. The sharpness s of the image is computed by

$$s = (\sum_{\forall(i,j)} g_v(i,j)^2 + \sum_{\forall(i,j)} g_h(i,j)^2)^{1/2} \tag{6}$$

where $g_h(i, j)$ and $g_v(i, j)$ are the resulting gradient image of horizontal and vertical gradient filter, respectively.

Each distance v associated with the sharpest stereo pairs and its corresponding disparity δ computed using (2) form a data point on the curve represented by the mapping table. We move the planar object to various positions and repeat the process described above to collect at least 10 data points. Then, the mapping of disparity to in-focus lens position is constructed using the least square fitting operator of MATLAB software. Fig. 2. shows the resulting curve.

The stereo disparity is related to the depth of an object. Given the depth of a target object, we can move the camera lens to obtain a clear image of the object. However, in autofocus, the first pair of stereo images is often defocused as discussed in the introduction section. Blurry stereo pairs cannot guarantee that accurate disparity can be computed by (2). Nevertheless, as long as the image gets sharper as the searching process continues, the accuracy of disparity is likely to improve. In this work, we emphasize on the use of disparity for lens movement rather than depth map construction. Whether the disparity of blur stereo images can be used to guide the lens to a correct direction can be examined by experiment.

Experimental Results

An experiment is conducted to evaluate the proposed stereo autofocus approach. The experimental setup is shown in Fig. 4, where a Google Nexus 5 smart phone is mounted on a rail. By moving the smartphone to a different position, a stereo-camera setup can be simulated with one single camera in a controlled environment for the purpose of experiment. At both camera positions, a set of images (aka focal stack) are captured by sweeping the lens across its dynamic range at a constant interval. We compute the focus value (sharpness) of the image captured at each lens position and form a focus profile from the set of focus data. The peak of the focus profile represents the corresponding image is sharpest, which is achieved when the lens is at the in-focus position.

Two scenarios, a planar scene and a complex scene with non-uniform depth, are considered in the experiment to evaluate the effectiveness of stereo autofocus. The initial lens position is set to be far away from the in-focus lens position. That means we launch the stereo autofocus with blurry stereo images. For a good visualization of the autofocus process, the position of the lens as it moves is overlaid on the focus profile. The top raw of Fig. 5 shows the first three consecutive lens positions for the planar scene, and the bottom raw shows the corresponding stereo images. We can see that the lens is very close to the in-focus position after the first

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.18.DPMI-254

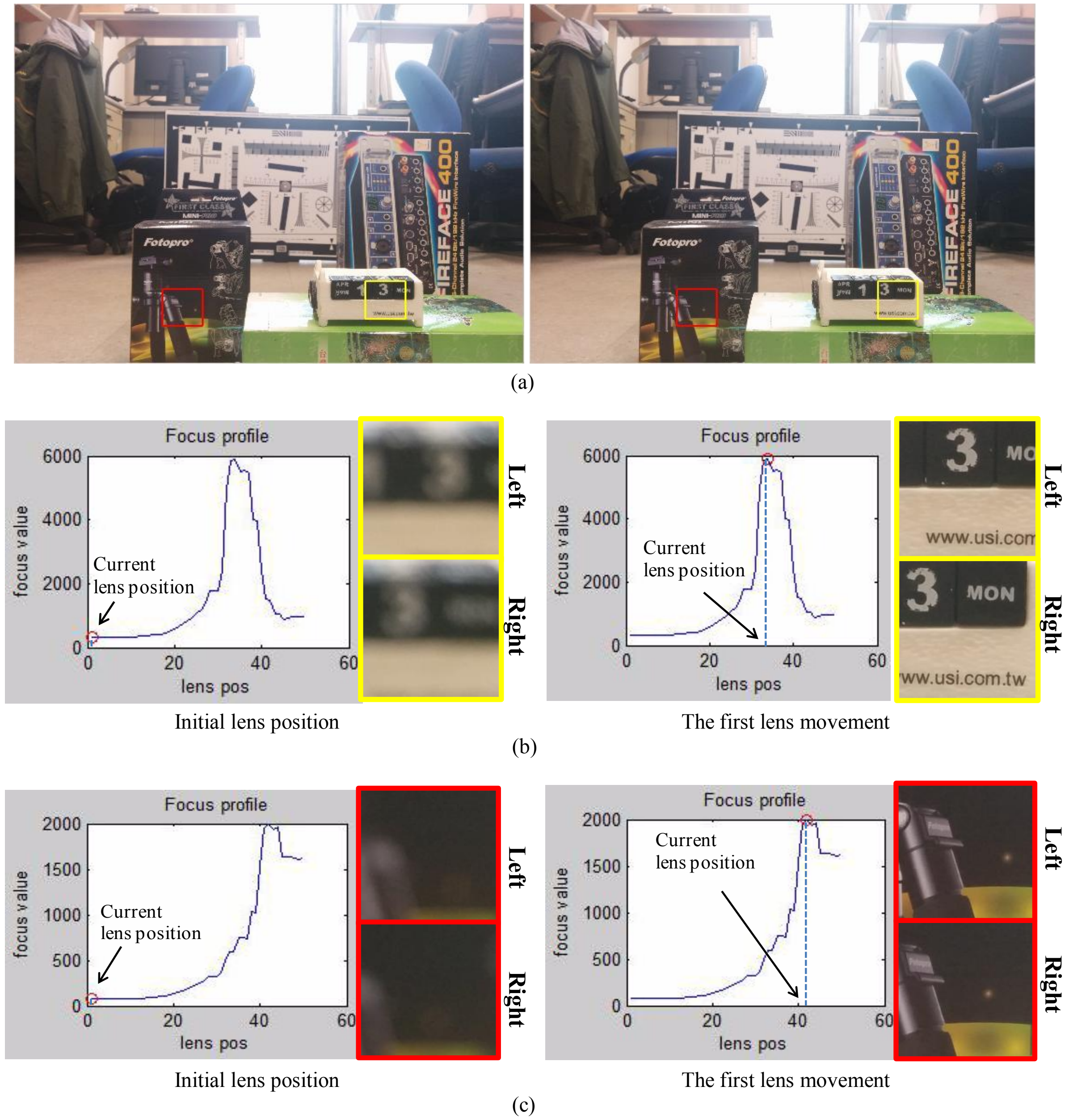

Figure 6. (a) The left and right images of a scene with non-uniform depth. Two focus windows are shown with two colored rectangles. The yellow window contains a near object and the red window a farther object. (b) The focus profiles, lens positions in the AF process, and the corresponding stereo images for the yellow window. (c) The same for the red window.

movement and that the peak of the focus profile is reached after the second movement. Compared with most CDAF-based monocular autofocus, which normally takes more than four lens movements, the stereo autofocus has the speed advantage in this case. We test the proposed stereo autofocus approach on complex scenes as well. Fig. 6(a) shows a stereo image pair with two focus windows that are placed at different depths of the complex scene. Hence, the two corresponding focus profiles have different peaks. A separate stereo autofocus process is launched for each focus window. The estimated lens position is shown in Fig. 6(b). As we can see, the peak can be reached with only one lens movement in both cases.

To evaluate the speed of the proposed stereo autofocus, we count the average number of the lens movements in the autofocus process. All possible initial positions within the entire motion range of the lens are considered. In the experiment, five objects are placed at various pre-selected distances (10cm, 20cm, 40cm, 60cm, and 80cm) in front of the camera as shown in Fig. 7. Fig. 8. shows that the search process converges with one or two lens movements even the initial lens position is far away from the in-focus position. We also find that the final lens position after the search process converges is located at the peak zone of the focus profile in 92.3 percent of the separate stereo autofocus tests. All of the above

DOI: 10.2352/ISSN.2470-1173.2016.18.DPMI-254

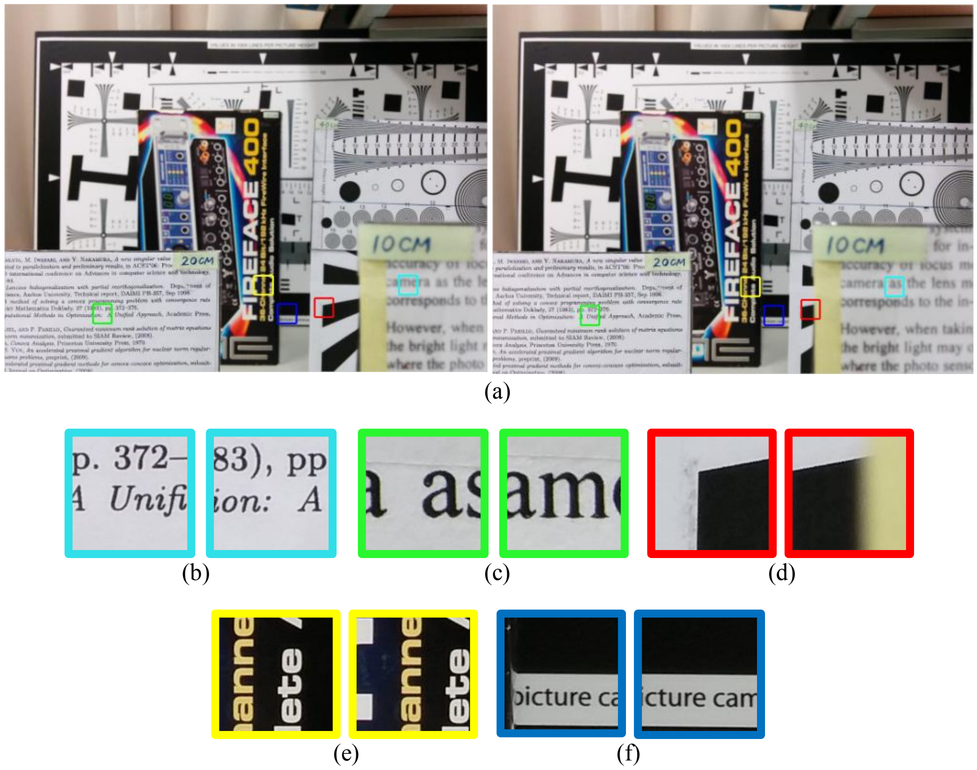

Figure 7. (a) The left and right image of a scene with objects placed at various depths. Five focus windows marked by five colored rectangles (cyan, green, red, yellow, and blue) are shown. From (b)−(f) are the corresponding stereo views focusing on the objects placed at 10cm, 20cm, 40cm, 60cm, and 80cm in front of the camera respectively.

results conclude that our approach can bring the lens to reach the peak zone within two lens movements in most cases regardless of where the initial lens position is. To prove the effectiveness of our mapping table, we empirically compare the accuracy of disparity-based stereo autofocus with and without using our mapping table according to the error between the in-focus position and the final lens position. Through our experiment, the average error decreases from 5.15 lens steps to 4 lens steps when using the mapping table. The lens step is the unit of lens position.

We have observed through experiments that the performance of stereo AF relies on reliable disparity estimation. Take the scene shown in Fig. 9 as an example. The images in the focus window (the red rectangular) only have one edge for disparity estimation. Therefore, any disparity estimate along the edge is a valid estimate. In this example, the overall performance of stereo AF is similar to that of single-camera AF, and there is no obvious gain by using stereo information. In practice, the focus window should be set properly to guarantee there are sufficient edges (or textures) of different orientations for reliable disparity estimation.

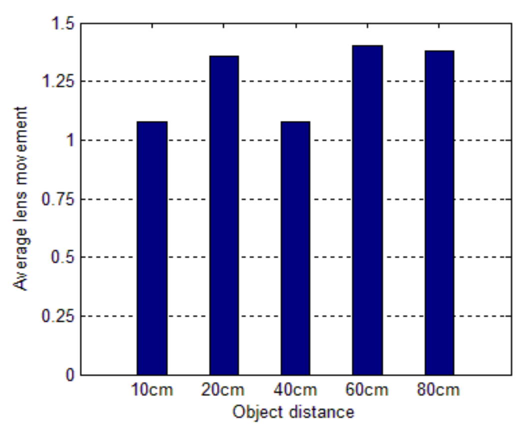

Figure 8. The average number of lens movement taken by the proposed stereo autofocus until the search process converged from an arbitrary initial lens position.

Discussion

Depth information is helpful for autofocus because it allows the in-focus lens position to be directly determined. In this work, the disparity-based stereo autofocus approach is developed under the assumption that there is no matching ambiguity when the stereo images are in-focus. However, matching error does occur if the stereo images are lack of sufficient textures or contain repetitive patterns. In this circumstance, the erroneous disparity estimate would mislead the autofocus system to an incorrect in-focus decision. As a result, to utilize the disparity efficiently for stereo autofocus, the quality of disparity should be examined carefully for reducing the impact of matching ambiguities.

Compared with the disparity, using the sharpness of image for autofocus does not have ambiguity problem on image with repetitive patterns since it indicates the gradient energy of image. However, given the image of scene contains sufficient textures, the focus profile normally has a flat out-of-focus region where the

DOI: 10.2352/ISSN.2470-1173.2016.18.DPMI-254

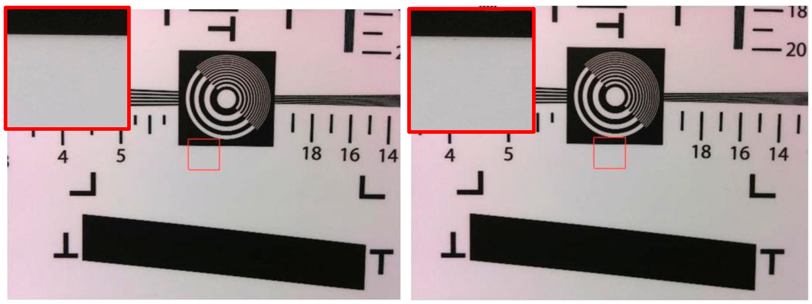

Figure 9. The left and right images of a test scene and the close-up of the image inside the focus window (red rectangle).

corresponding sharpness measure of defocused images does not provide a good clue to make reliable autofocus decision. If the autofocus process is trapped in the flat region, it takes longer time to find the in-focus position. In comparison, the disparity-based autofocus is able to point out the right reliable direction as the search process sets off. From this work, we find that both the disparity-based and the CDAF methods are complementary to each other and should be part of an integral solution.

Conclusion

In this paper, we have investigated the use of disparity for autofocus. The disparity-based stereo autofocus can guide the lens to reach the peak zone of the focus profile within two lens movements in most cases even the initial lens position is far away from the in-focus position. The computational efficiency and the accuracy of this approach are enhanced by using a mapping table that relates the disparity to in-focus lens position. It is our plan to integrate disparity-based autofocus with contrast detection autofocus and compare the integrated approach with the phase detection autofocus technique in future work.

References

[1] Y. Yoshida, S. Shinohara, H. Ikeda, K. Tada, H. Yoshida, K. Nishide, and M. Sumi, “Control of lens position in auto-focus cameras using semiconductor laser range finder,” *Proc. 34th Midwest Symposium Circuits Systems*, vol. 1, pp. 359–398, 1991.

[2] N. Kehtarnavaz and H. Oh, “Development and real-time implementation of a rule-based auto-focus algorithm,” *Real-Time Imaging*, vol. 9, no. 3, pp. 197–203, 2003.

[3] D. C. Tsai and H. H. Chen, “Reciprocal focus profile,” *IEEE. Trans. Image Process*., vol. 21, no. 2, pp. 459–468, 2012.

[4] 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, 1999.

[5] S. Y. Lee, Y. Kumar, J. M. Cho, S. W. Lee, and S. W. Kim, “Enhanced autofocus algorithm using robust focus measure and fuzzy reasoning,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 9, pp. 1237–1246, Sep. 2008.

[6] M. Subbarao, T. Choi, and A. Nikzad, “Focusing techniques,” *J. Optical Eng*., vol. 32, no. 11, pp. 2824−2836, Nov. 1993.

[7] J. Kautsky, J. Flusser, B. Zitova, and S. Simberova, “A new wavelet-based measure of image focus,” *Pattern Recognit. Lett.*, vol. 23, no. 14, pp. 1785−1794, Dec. 2002.

[8] S. Pertuz, D. Puig, M. A. Garcia, and A. Fusiello, “Generation of all-in-focus images by noise-robust selective fusion of limited depth-of-field images,” *IEEE Trans. Image Process.*, vol. 22, no. 3, pp. 1242−1251, Mar. 2013.

[9] S. K. Nayar and Y. Nakagawa, “Shape from focus,” *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 16, no. 8, pp. 824–831, Aug. 1994.

[10] A. N. Rajagopalan, S. Chaudhuri, and U. Mudenagudi, “Depth estimation and image restoration using defocused stereo pairs,” *IEEE Trans. Pattern Anal. Mach. Intell*., vol. 26, no. 11, pp. 1521–1525, Nov. 2004.

[11] J. Yang and D. Schonfel, “Virtual focus and depth estimation from defocused video sequences,” *IEEE Trans. Image Process*., vol. 19, no. 3, pp. 668–679, Mar. 2010.

[12] C. H. Chen, H. Zhou, and T. Ahonen, “Blur-aware disparity estimation from defocus stereo images,” *IEEE Int. Conf. Comput. Vis*., pp. 855–863, 2015.

[13] S. Pertuz, M. A. Garcia, and D. Puig, “Efficient focus sampling through depth-of-field calibration,” *Int. J Comput. Vis*. vol. 112, no. 3, pp. 342–353, 2015.

[14] J. M. Tenenbaum, *Accommodation in Computer Vision*, 1970: Stanford University.

Author Biography

Please submit a brief biographical sketch of no more than 75 words. Include relevant professional and educational information as shown in the example below

Shao-Kang Huang received the M.S. degree in electrical engineering from the National Taiwan Normal University, Taipei, Taiwan (2013), and he is currently pursuing the Ph.D. degree in the Graduate Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan. His research interests are in stereo matching and image processing.

Cheng-Chieh Yang received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan (2014). His research interests include stereo matching and image processing.

Kuang-Tsu Shih received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan (2009). He is currently pursuing the Ph.D. degree in the Graduate Institute of Communication Engineering. His research interests include color image processing, color science, and computer vision.

Homer Chen has been with the College of Electrical Engineering and Computer Science at National Taiwan University, Taipei, Taiwan, since he returned to Taiwan in August 2003, where he is a Distinguished Professor. He was a U.S. delegate of the ISO and ITU standards committees and contributed to the development of many new interactive multimedia technologies that are now part of the MPEG-4 and JPEG-2000 standards. His current research interests include multimedia processing and communications.