High-sensitivity Imaging using a Multi-aperture Camera Based on Image Synthesis with Disparity Compensation

Kuniyuki Kugenuma, Takashi Komuro
Graduate School of Science and Engineering
Saitama University
Saitama, Japan

Bo Zhang, Keiichiro Kagawa, Shoji Kawahito
Research Institute of Electronics
Shizuoka University
Hamamatsu, Japan

***Abstract*— In this paper, we propose a method that reduces the noise of an image sensor by image synthesis based on a disparity map using a multi-aperture camera. The multi-aperture camera consists of multiple pairs of a lens and an image sensor and virtually realizes a fast lens, which cannot be realized by a single aperture, using aperture synthesis. We used a global method to estimate a disparity map from low SNR images which were captured in a low light condition. From the experimental result, we confirmed that it is possible to estimate a disparity map even from images under low illumination. Using the estimated disparity map, we succeeded in obtaining a clear synthetic image by synthesizing the multiple images with disparity compensation.**

***Keywords*— Aperture synthesis; low SNR images; MRF model; Belief Propagation**

I. INTRODUCTION

If a camera that can capture images without motion blur even in dark environment is realized, it is possible to take high quality images of any scene, and the convenience of the camera becomes dramatically improved. The sensitivity of a camera can be improved by using a fast large-aperture lens to increase the number of photons that are received by a pixel of the sensor. However, as we can see from general large-aperture lenses, a fast lens is not only large and heavy but also has large residual aberration. Therefore, the spatial resolution becomes low with an open aperture and also considerable vignetting occurs. Furthermore, the depth of field becomes extremely shallow, which makes it difficult to use the camera.

We have developed a *multi-aperture camera* in which small CMOS image sensors are arranged in a two-dimensional array [1] as an imaging system which satisfies both high sensitivity and small size. By synthesizing multiple images captured by the multi-aperture camera, it is possible to reduce noise in the images. However, there is a problem that blur occurs when the images of a scene with large depth changes are synthesized with only simple registration.

In this paper, we propose a method for estimating a disparity map of the scene from low-SNR multiple images captured by the multi-aperture camera. The block matching method which is often used in the acquisition of the disparity map has a problem that erroneous correspondences often occur in texture-less regions, regions having discontinuous depths and occlusion regions. In particular, this problem becomes more serious when using low SNR images captured in a low light condition, which is the target of this study. Therefore, we use a global method [2] for estimating the disparity map. Global methods are the technique to estimate an image (or a disparity map) by formulating general properties and noise characteristics of natural images using a probabilistic function or an energy function and by solving the optimization problem. Furthermore, it is shown that the estimation results by global methods were better than those by local methods in the experiment using noisy images [3]. For the above reasons, we formulate the estimation of a disparity map as maximum a posterior (MAP) estimation problem using a Markov random field (MRF) model. Then, we use Belief Propagation (BP) to solve the optimization problem [2] [4] [5] and to estimate the disparity map. Using the estimated disparity map, we can obtain a clear synthetic image by synthesizing the multiple images with disparity compensation.

II. MULTI-APERTURE CAMERA

Figure 1 shows the architecture of the multi-aperture camera. While a single-aperture camera uses a single large-aperture lens and a single image sensor, the multi-aperture camera consists of multiple pairs of a lens and an image sensor and virtually realizes a fast lens, which cannot be realized by a single aperture, using aperture synthesis.

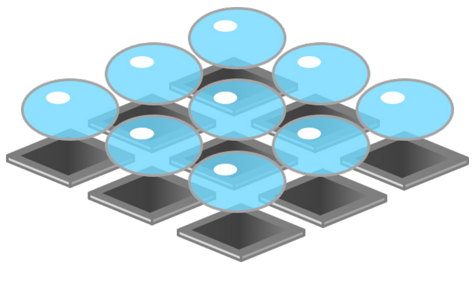


Fig. 1. Architecture of the multi-aperture camera

There has been research on camera array systems for generating free-viewpoint video, super-resolution, and refocusing [7-9]. However, this system aims to realize a fast lens having a very small f-number just by increasing the number of apertures. Comparing to light field cameras [10], in which multiple lenses are placed in front of an image sensor, this system does not require a main lens and realizes a smaller f-number.

Equation (1) shows the SNR of the multi-aperture camera. M is the number of apertures, n_{ph} is the number of photons that are received by a pixel of the sensor per image acquisition, σ^2 is the variance of sensor noise per pixel. In addition to sensor noise, photon shot noise having a variance that is proportional to the number of incident photons occurs.

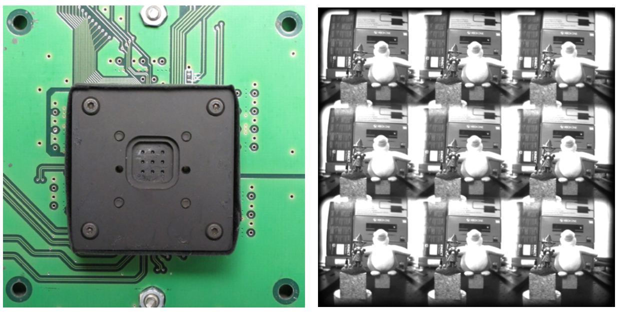
$$\mathrm{SNR}_{\mathrm{MA}}[\mathrm{dB}] = 20\log_{10}\frac{Mn_{\mathrm{ph}}}{\sqrt{M\sigma^2 + Mn_{\mathrm{ph}}}} \tag{1}$$

On the other hand, in a single-aperture camera whose lens has a $\sqrt{M}$ times smaller f-number (M times faster) than the lens of the multi-aperture camera, the number of incident photons increases M times, and the SNR of the single-aperture camera becomes as shown in Eq. (2).

$$\mathrm{SNR}_{\mathrm{SA}}[\mathrm{dB}] = 20\log_{10}\frac{Mn_{\mathrm{ph}}}{\sqrt{\sigma^2 + Mn_{\mathrm{ph}}}} \tag{2}$$

The noise level of current commercial low-noise image sensors is about 1-2 electrons RMS, and that in research level is under 0.3 electrons RMS [6]. Therefore, even in a low light condition, photon shot noise becomes dominant compared to sensor noise ($\sigma^2 << n_{\mathrm{ph}}$). Then, the SNRs of both cameras are determined almost only by photon shot noise, and also the SNRs of both a multi-aperture camera having M apertures and a single-aperture camera whose lens has a $\sqrt{M}$ times smaller f-number become nearly equal. However, it is difficult in practice to manufacture a lens with such a small f-number when M increases. Such a fast lens can be realized only virtually by a multi-aperture system.

In the multi-aperture camera used in this study, the same number of lenses as apertures are arranged on a single image sensor. Since an image that is captured by the multi-aperture camera includes multiple sub-images, we crop the sub-images from the captured image and use them for synthesis. The number of pixels of the image sensor is 1280×1024, and the noise level is about 1 electron RMS. The number of apertures is 3×3 and the f-number of each lens is *f*/3. Figure 2 shows the appearance of the multi-aperture camera and an image captured by the multi-aperture camera.



(a) (b)

Fig. 2. Multi-aperture camera: (a) appearance, (b) captured image

III. Image Synthesis with Disparity Compensation

A. Model of a Multi-aperture Camera

Let I_0 be the reference image captured in the center aperture, and I_i be an image captured in an aperture i (i=1, …, M-1) around the center aperture. The coordinates (x_i,y_i) of the image I_i that correspond to the coordinates (x,y) of the image I_0 is written as

$$x_i = x + \frac{ft_x^{(i)}}{Z(x,y)} = x + \tilde{t}_x^{(i)} d(x,y) \tag{3}$$

$$y_i = y + \frac{ft_y^{(i)}}{Z(x,y)} = y + \tilde{t}_y^{(i)} d(x,y) \tag{4}$$

Here, $t_x^{(i)}, t_y^{(i)}$ are translations of an aperture i relative to the center aperture, f is the focal length of the lenses, $Z(x,y)$ and $d(x,y)$ are the depth and disparity in the coordinates (x,y) of the center aperture image, $\tilde{t}_x^{(i)}, \tilde{t}_y^{(i)}$ are the values that express a direction obtained by dividing $t_x^{(i)}, t_y^{(i)}$ by the baseline b, and $\tilde{t}_x^{(i)}, \tilde{t}_y^{(i)} \in \{-1,0,1\}$. The relation between a disparity d and a depth Z is $d=fb/Z$.

Equations (3), (4) assume that each aperture is located in an ideal position. However, in the actual multi-aperture camera, each aperture has different internal camera parameters, and also there are misalignment and tilt. Therefore the above equations cannot be applied as they are. Equation (3), (4) have to be modified to include compensation based on the camera parameters. The modified equations that consider the camera parameters, which are derived from the geometric relationship based on the perspective projection model, are written as

$$x_i = f_x^{(i)} \frac{X'}{Z'} + c_x^{(i)} \tag{5}$$

$$y_i = f_y^{(i)} \frac{Y'}{Z'} + c_y^{(i)} \tag{6}$$

$$X' = \frac{r_{11}^{(i)}}{f_x^{(0)}}\left(x - c_x^{(0)}\right) + \frac{r_{12}^{(i)}}{f_y^{(0)}}\left(y - c_y^{(0)}\right) + r_{13}^{(i)} + \frac{t_x^{(i)}}{fb} d(x,y) \tag{7}$$

$$Y' = \frac{r_{21}^{(i)}}{f_x^{(0)}}\left(x - c_x^{(0)}\right) + \frac{r_{22}^{(i)}}{f_y^{(0)}}\left(y - c_y^{(0)}\right) + r_{23}^{(i)} + \frac{t_y^{(i)}}{fb} d(x,y) \tag{8}$$

$$Z' = \frac{r_{31}^{(i)}}{f_x^{(0)}}\left(x - c_x^{(0)}\right) + \frac{r_{32}^{(i)}}{f_y^{(0)}}\left(y - c_y^{(0)}\right) + r_{33}^{(i)} + \frac{t_z^{(i)}}{fb} d(x,y) \tag{9}$$

Here, $f_x^{(i)}, f_y^{(i)}$ are the horizontal and vertical focal lengths of each aperture, $c_x^{(i)}, c_y^{(i)}$ are the image center of each aperture, $r_{11}^{(i)} \sim r_{33}^{(i)}$ are the elements of the rotation matrix of each aperture, $t_x^{(i)}, t_y^{(i)}, t_z^{(i)}$ are translations of each aperture, and f and b are the focal length and baseline of the ideal aperture.

These equations are used to estimate a disparity map and to synthesize the multiple images described later.

B. Disparity Estimation

We use a global method using an MRF model to estimate a disparity map even from low-SNR images. The disparity map can be estimated by solving the MAP problem that the posterior of the observed image becomes maximum. Let I be the observed image, and D be the estimated disparity map. The MAP problem is written as

$$\hat{D} = \underset{D}{\arg\max}\, P(D|I) \tag{10}$$

The posterior $P(D|I)$ is written using the likelihood $P(I|D)$ and prior $P(D)$ as

$$\begin{aligned} P(D|I) &\propto P(I|D)P(D) \\ &= \prod_{s=1}^{N} \Phi(d_s) \prod_{s=1}^{N} \prod_{t \in N_s} \Psi(d_s, d_t) \end{aligned} \tag{11}$$

$$\Phi(d) = \prod_{i=1}^{M-1} \exp\left(-\alpha \left| I_0(x, y) - I_i(x_i, y_i) \right|\right) \tag{12}$$

$$\Psi(d_s, d_t) = \exp\left(-\beta \left| d_s - d_t \right|\right) \tag{13}$$

Here, s and t are the neighboring pixels in the image, d_s and d_t are disparities in s and t, respectively. N is the number of all pixels, and N_s is a set of the neighbor pixels to the pixel s. α and β are the constants indicating parameters to the likelihood and the prior. We solve this MAP problem by maximizing the posterior using the max-product method of Loopy Belief Propagation (LBP) [2], which is a variation of BP.

C. Image Synthesis

We obtain a synthetic image using the obtained disparity map. The coordinates of the pixel in each image I_i that corresponds to a pixel in the center aperture image I_0 are calculated using Eqs. (5), (6) and the pixel values in all the images are averaged. However, there is a problem of occlusion. The pixel in an image I_i that corresponds to a pixel in the image I_0 sometimes also corresponds to other pixels in the image I_0. In such a case, only the front pixel that has the largest disparity is used for synthesis since the other pixels are thought to be occluded by the front pixel.

IV. IMAGE SYNTHESIS WITH DISPARITY COMPENSATION

Using the estimation method described above, we conducted an experiment to estimate a disparity map and obtain a synthetic image from the multiple images captured by the multi-aperture camera. Figure 3 shows the arrangement of the objects and the background. The distances from the camera to the objects were 15 cm and 40 cm, and that to the background was 80 cm. We captured four images with changing illuminance using an ND (Neutral Density) filter. The illuminances on the surface of the front object were 38.7 lx, 3.87 lx, 1.161 lx, and 0.387 lx. The theoretical SNRs of the images, which were calculated from the numbers of electrons that were converted from the signal values, were 36.3 dB, 26.2 dB, 21.4 dB, and 17.8 dB, respectively. We set the parameters of the likelihood and prior to α = 0.01 and β = 1, and the number of iterations in LBP was 50.

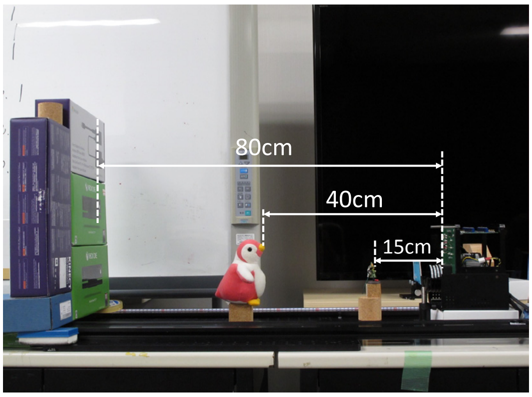


Fig. 3. Subject arrangement

Figure 4 shows the captured images in the center aperture, synthetic images and estimated disparity maps for different illuminances. The size of each aperture image is 220×220. The noise was reduced in the synthetic images for all illuminances. The estimated disparity map became worse as the illuminance decreases. However, the regions in which there was much estimation error were originally those with little texture, and the error did not affect the quality of the synthetic images. There was little blur in the synthetic images and the details of the images such as characters, which were hidden by noise, became clearer.

V. CONCLUSION

We proposed a method that reduces the noise of an image sensor by image synthesis based on a disparity map using a multi-aperture camera. We used a global method for estimating the disparity map from low SNR images which were captured in a low light condition. We formulated the problem as MAP estimation using an MRF model and used BP to solve the optimization problem. From the experimental result, we confirmed that the disparity map and the synthetic image with little blur can be obtained even from low SNR images.

Future work includes modeling of the sensor-specific noise to obtain better estimation in a lower light condition, and performing the quantitative evaluation.

ACKNOWLEDGEMENTS

A part of this work was carried out under the Cooperative Research Project Program of the Research Institute of Electronics, Shizuoka University.

36.3 dB | 26.2 dB | 21.4 dB | 17.8 dB

Fig. 4. Captured images in the center aperture (top), synthetic images (middle) and estimated disparity maps (bottom)

REFERENCES

[1] B. Zhang, K. Kagawa, T. Takasawa, M. Seo, K. Yasutomi, and S. Kawahito, "RTS Noise and Dark Current White Defects Reduction Using Selective White Defects Reduction Using Selective Averaging Based on a Multi-Aperture System," Sensors, Vol.14, No.1, pp.1528-1543, 2014.

[2] J. Sun, H. Shum, and N. Zheng, "Stereo matching using belief propagation," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.25, No.7, pp.787-800, 2003.

[3] H. Hirschmüller and D. Scharstein, "Evaluation of cost functions for stereo matching," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp.1-8, 2007.

[4] P. Felzenszwalb and D. Huttenlocher, "Efficient belief propagation for early vision," International Journal of Computer Vision, Vol.70, No.1, pp.41-54, 2006.

[5] M. Tappen and W. Freeman, "Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters," Proc. Ninth IEEE International Conference on Computer Vision, pp.900-906, 2003.

[6] M. Seo, S. Kawahito, K. Kagawa, and K. Yasutomi, "A $0.27e^-_{rms}$ read noise 220-µV/e^- conversion gain reset-gate-less CMOS image sensor with 0.11-µm CIS process," IEEE Electron Device Letters, Vol. 36, No. 12, pp. 1344-1347, 2015.

[7] N. Fukushima, T. Fujii, Y. Ishibashi, T. Yendo, and M. Tanimoto, "Real-time free viewpoint image rendering by using fast multi-pass dynamic programming," Proc. 3DTV-Conference, 2010.

[8] Y. Taguchi, T. Koike, K. Takahashi, and T. Naemura, "TransCAIP: A Live 3D TV System Using a Camera Array and an Integral Photography Display with Interactive Control of Viewing Parameters," IEEE Transactions on Visualization and Computer Graphics, Vol.15, No.5, pp. 841-852, 2009.

[9] K. Venkataraman, D. Lelescu, J. Duparré, A. McMahon, G. Molina, P. Chatterjee, R. Mullis, and S. Nayar, "PiCam: An Ultra-Thin High Performance Monolithic Camera Array," ACM Transactions on Graphics, Vol. 32, No. 6, 2013.

[10] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light Field Photography with a Hand-held Plenoptic Camera," Stanford Tech Report CTSR 2005-02, 2005.