Depth Map Estimation and Restoration of an Improved-quality Image from Multi-aperture Images

Kuniyuki Kugenuma[1], Takashi Komuro[1], Bo Zhang[2], Keiichiro Kagawa[2], Shoji Kawahito[2]
1 Graduate School of Science and Engineering, Saitama University
255 Shimo-okubo, Saitama, 338-8570 Japan
2 Research Institute of Electronics, Shizuoka University
3-5-1 Johoku, Hamamatsu, 432-8011 Japan
E-mail:komuro@mail.saitama-u.ac.jp

Abstract In this study, we propose a method for estimating a depth map from images captured by the multi-aperture camera, which places CMOS image sensors and lenses in a two dimensional array. By employing a global estimation method, we succeeded in estimating a depth map even from noisy images which are captured in a low light condition. We also succeeded in obtaining an improved quality image by synthesizing the multi-aperture images with disparity compensation.

Keywords: Multi-camera stereo, low SNR images, belief propagation, image synthesis

1. 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 relatively 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 depth map of the scene from multiple images captured by the multi-aperture camera. We use a global method [2] for estimating the depth map. Global methods are the technique to estimate an image (or a depth 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. We formulate the estimation of a depth 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] and to estimate the depth map. Using the estimated depth map, we can obtain a clear synthetic image by synthesizing the multiple images with disparity compensation.

2. 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 divides a lens and an image sensor to small ones and virtually realizes a fast lens, which cannot be realized by a single aperture, using aperture synthesis.

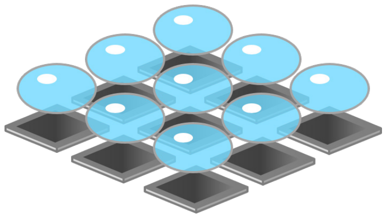
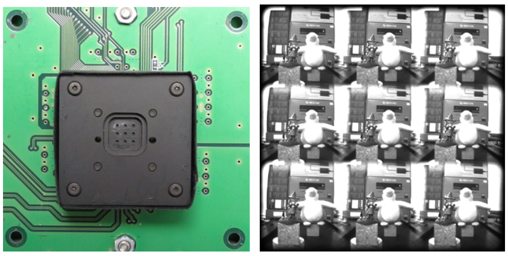


Fig. 1. Architecture of the multi-aperture camera

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

3. Depth map estimation

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

$$x_i = x + \frac{f t_x^{(i)}}{Z(x,y)} \tag{1}$$

$$y_i = y + \frac{f t_y^{(i)}}{Z(x,y)} \tag{2}$$

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, $Z(x,y)$ is the depth in the coordinates (x,y) of the center aperture image. Equations (1), (2) 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, we modified the above equations to include compensation based on the camera parameters and used them for estimation.

We use a global method using an MRF model to estimate a depth map. The depth 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 Z be the estimated depth map. The MAP problem is written as

$$\hat{Z} = \arg\max_{Z} P(Z|I) \tag{3}$$

Here, the problem of estimating a depth map is equivalent to estimating a disparity map. They can be converted to each other using the following relation.

$$d(x,y) = \frac{fb}{Z(x,y)} \tag{4}$$

where b is the baseline. In this study, we estimate a disparity map D instead of a depth map Z. The posterior $P(D|I)$ is written using the likelihood $P(I|D)$ and prior $P(D)$ as

$$\begin{aligned} P(D|I) &\propto P(I|D)P(D) \\ &= \prod_{s=1}^{N} \Phi(d_s) \prod_{s=1}^{N} \prod_{t \in N_s} \Psi(d_s, d_t) \end{aligned} \tag{4}$$

$$\Phi(d) = \prod_{i=1}^{M-1} \exp\left(-\alpha \left| I_0(x,y) - I_i(x_i, y_i) \right|\right) \tag{5}$$

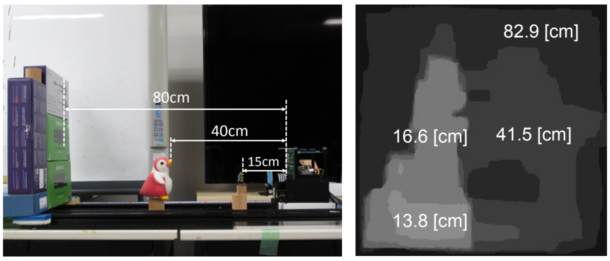
$$\Psi(d_s, d_t) = \exp\left(-\beta \left| d_s - d_t \right|\right) \tag{6}$$

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.

4. Experiment

Using the estimation method described above, we conducted an experiment to estimate a depth map and obtain a synthetic image from the multiple images captured by the multi-aperture camera. Figure 3(a) shows the arrangement of the subjects and the background. We captured an image using an ND (Neutral Density) filter to control the illuminance. The illuminance of the front of the object was 3.87 lx and the theoretical SNR of the image, which were calculated from the numbers of electrons that were converted from the signal values, was 26.2 dB. We set the parameters of the likelihood and prior to $\alpha = 0.01$ and $\beta = 1$, and the number of iterations in LBP was 50. Figure 3(b) shows the estimated depth map and Fig. 4(a) and Fig. 4(b) show the captured image in the center aperture and the synthetic image, respectively.

From the result, we can see that the depth map was correctly estimated. Also, a clear synthetic image was obtained by synthesizing the multi-aperture images with disparity compensation. There were little blur in the synthetic images and the details of the images such as characters, which were hidden by noise, became clearer.



(a) (b)

Fig. 3. (a) arrangement of subjects, (b) estimated depth map

(a) (b)

Fig. 4. (a) captured image in the center aperture, (b) synthetic image

5. Conclusion

We proposed a method for estimating a depth map from images captured by the multi-aperture camera, and for reducing noise of an image sensor by image synthesis with disparity compensation. We used a global method using an MRF model in the depth map estimation from noisy images which were captured in a low light condition. From the experimental result, we confirmed that the depth map and the synthetic image with little blur can be obtained even from noisy images. Future work includes modeling of the sensor-specific noise to obtain better estimation in a lower light condition, and performing the quantitative evaluation.

References

[1] B. Zhang, K. Kagawa, T. Takasawa, M. Seo, K. Yasutomi, S. Kawahito: "RTS noise and dark current 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] P. F. Felzenszwalb, D. P. Huttenlocher: "Efficient belief propagation for early vision", International Journal of Computer Vision, Vol.70, No.1, pp.41-54 (2006)

[4] M. F. Tappen, W. T. 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).