DOI: 10.2352/ISSN.2470-1173.2016.21.3DIPM-405

Depth Estimation Algorithm for Color Coded Aperture Camera

Ivan Panchenko, Vladimir Paramonov and Victor Bucha; Samsung R&D Institute Russia; Moscow, Russia

Abstract

In this paper we present an algorithm of depth estimation from a single frame captured with color coded aperture camera. Our algorithm provides continuous depth and is more robust to lack of texture comparing to the state-of-the-art. The main contributions of this work comparing to prior-art algorithms are: (1) robust metric for depth determination in a pixel, (2) depth map sub-pixel estimation, (3) depth propagation to low-textured areas, (4) depth map edge restoration, (5) depth quality enhancement, (6) raw data processing. We also made an efficient algorithm implementation which processes FullHD frame for 50 ms on GeForce GTX 780 and requires 15 seconds on Qualcomm Andreno 330 GPGPU for the same frame.

Introduction

Single frame passive depth sensors allow extracting depth of a moving object outdoors which is not possible with active sensor and structure-from-motion techniques. The stereo camera is a good choice but it increases camera power consumption and adds additional space and cost which can be critical for handheld devices. That is why single-lens single frame passive depth sensors based on coded aperture seems to be a good choice. A number of researchers in this area [1, 2, 3, 4] developed the technology proposed in [5] but the depth estimation quality as well as the algorithm timing performance is not satisfactory.

We have chosen a color-coded aperture [1] over binary-coded aperture [2] due to the following reasons:

1. it allows differentiating defocused and smooth object;
2. it allows differentiating if the object is closer than focus or beyond it;
3. has lower liability to diffraction (due to bigger size of a smallest element);
4. has higher light-efficiency (2 times for the same lens);
5. has much faster timing performance (500 times faster for the same image resolution and comparable depth quality);.

Our goal was a development of a more robust, more precise and faster algorithm of disparity estimation for color coded aperture. This was a part of our work on developing a single-lens single-frame passive depth sensor with minor hardware changes made to conventional camera.

Disparity Map Estimation

Overview

The pipeline of depth extraction is shown in Figure 1. It is similar to [5, 1]. We capture an image (1), compute a cost volume (2), filter it (3), extract preliminary depth (4), and make depth enhancement (5). However, we made a significant modification of all the algorithm parts. (1) Processing RAW image instead of compressed one can lead to significant depth quality enhancement on several scenes. (2a) We compute a mutual correlation of color channels with different candidate disparity values (shifts). Mutual correlation metric is similar to the color lines metric [1], though it is more robust to texture lack. (2b) We use an exponentially-weighted window (Figures 2(d)-2(f)) as it gives more weight to closer pixels and increases depth quality in low-textured areas. Convolution with this window can be efficiently implemented as described in [6]. (3) We filter cost volume to propagate the information to low-textured areas. We use a joint-bilateral filter approximation in which Gaussian function is changed to exponential function [6]. We take a middle color channel as a reference for this filtering procedure. (4) We use a sub-pixel estimation to extract continuous depth with parabola fitting technique. (5) We use a joint-bilateral filtering to restore depth on the edges (reference image is a middle color channel image). These modifications are discussed in the following subsections in more details.

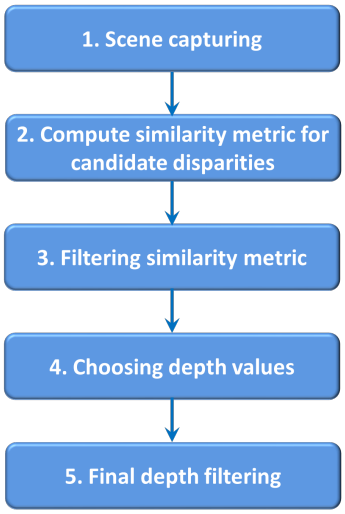


Figure 1. *Depth extraction algorithm pipeline.*

Mutual Correlation Estimation

Let $\{I_i\}_1^n$ represent a set of n captured color channels of the same scene from different viewpoints, where I_i is the $M \times N$ frame. We form a conventional correlation matrices $\mathbf{C}_d$ for the $\{I_i\}_1^n$ set and candidate disparity values d:

$$\mathbf{C}_d = \begin{pmatrix} 1 & \dots & corr(I_1^d, I_n^d) \\ \vdots & \ddots & \vdots \\ corr(I_n^d, I_1^d) & \dots & 1 \end{pmatrix}, \quad (1)$$

DOI: 10.2352/ISSN.2470-1173.2016.21.3DIPM-405

where superscript $(*)^d$ denotes parallel shift in the corresponding channel. The determinant of the matrix $\mathbf{C}_d$ is a good measure of $\{I_i\}_1^n$ mutual correlation. Indeed, when all channels are in strong correlation, all the elements of the matrix are equal to one and $det(\mathbf{C}_d) = 0$. On the other hand, when data is completely uncorrelated, we have $det(\mathbf{C}_d) = 1$. To extract a disparity map using this metric one should find disparity values d corresponding to the smallest value of $det(\mathbf{C}_d)$ in each pixel of the picture.

Here, we derive another particular implementation of the generalized correlation metric for $n = 3$. It corresponds to the case of aperture with three channels. The determinant of the correlation matrix is:

$$\begin{aligned} &det(\mathbf{C}_d) = \\ &= 1 - corr(I_1^d, I_2^d)^2 - corr(I_2^d, I_3^d)^2 - corr(I_3^d, I_1^d)^2 + \\ &\quad + 2corr(I_1^d, I_2^d) corr(I_2^d, I_3^d) corr(I_3^d, I_1^d). \end{aligned} \tag{2}$$

Again,

$$\begin{aligned} &\underset{d}{\operatorname{argmin}}\, det(\mathbf{C}_d) = \\ &= \underset{d}{\operatorname{argmax}} \left[\sum corr(I_i^d, I_j^d)^2 - 2 \prod corr(I_i^d, I_j^d) \right]. \end{aligned} \tag{3}$$

This metric is similar to the color lines metric [1], though it is more robust. The extra robustness appears when one of three channels does not have enough texture in a local window around a point under consideration. In this case the color lines metric cannot provide disparity information even if the other two channels are well defined. The generalized correlation metric avoids this disadvantage and allows the depth sensor to work similarly to a stereo camera in this case.

Exponentially Weighted Window

Disparity map estimation is based on the measurement of correspondence between color channels. The prior art approach [1] uses the color lines metric in a square local moving window (Figures 2(a-c)) for similarity estimation, but this leads to a large amount of errors in non-textured areas. This made us to modify prior-art approach.

To mitigate errors in non-textured areas and to preserve the advantage of low computational complexity, we propose to estimate conventional mutual correlation metric in a weighted local window. We use an exponentially-weighted window (Figures 2(d-f)) as it gives more weight to closer pixels. Convolution with this window can be efficiently implemented.

We use an implementation of a recursive separable convolution with exponential function proposed in [6]. It significantly reduces the number of arithmetical operations required for per-pixel computation.

Fast Recursive Filter

Recursive separable convolution with exponential kernel is implemented via a 1^{st} order Infinite Impulse Response (IIR) filter. The equation for one half of the exponential function is:

$$I_f(n) = I(n) \cdot (1 - \alpha) + I_f(n-1) \cdot \alpha, \tag{4}$$

where $I(i, j)$ and $I_f(i, j)$ are respectively the input and output images in pixel n, and α is a coefficient which is responsible for the exponential function attenuation. The convolution with the second half of exponential function is applied the same way but in reverse pixel order.

Four filters need to be applied to process an image with a 2D separable IIR filter: two in the X-direction and two in the Y-direction.

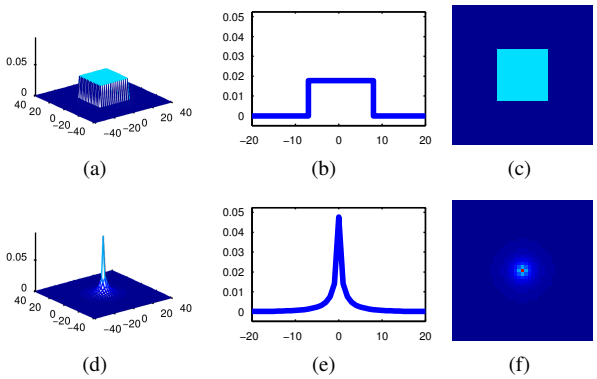


Figure 2. *Local support window for disparity map estimation. Conventional approach (e.g.,[1]): (a) 3D axes, (b) zero-cross section and (c) XY projection; Our approach: (d) 3D axes, (e) zero-cross section and (f) XY projection.*

Disparity Map Enhancement

Usually, passive sensors provide sparse disparity maps. However, dense disparity maps can be obtained by propagating disparity information to non-textured areas. The propagation can be efficiently implemented via joint-bilateral filtering of mutual correlation metric cost C. This can also be efficiently approximated with a 1^{st} order IIR filter:

$$C_f(n) = C(n) \cdot (1 - \alpha(n)) + C_f(n-1) \cdot \alpha(n), \tag{5}$$

where all variables have the same meanings as in (4) and α varies with respect to similarity in intensities of a point and its neighborhood in the color-range domain and is defined as follows:

$$\alpha(n) = exp(-\sigma_{sp}) \cdot exp(-\sigma_r \cdot (I(n) - I(n-1)), \tag{6}$$

where σ_{sp} and σ_r are the smoothing parameters of joint-bilateral filter in spatial and range domain respectively. Furthermore, we enhance the disparity map resolution using sub-pixel estimation via quadratic polynomial interpolation for each pixel:

$$d_{sp} = d_{min} - \frac{C^{d_{min}+1} - C^{d_{min}-1}}{2C^{d_{min}-1} - 4C^{d_{min}} + 2C^{d_{min}+1}}, \tag{7}$$

where index $C^{d_{min}}$ is the cost value in layer d_{min}, d_{min} is the disparity corresponding to the minimum of cost function and d_{sp} is the disparity after sub-pixel estimation.

Results

Algorithm Modifications

First, we present the impact of algorithm modifications discussed above (see Figure 3). You can see that all of the proposed depth extraction algorithm modifications positively affect depth map quality.

DOI: 10.2352/ISSN.2470-1173.2016.21.3DIPM-405

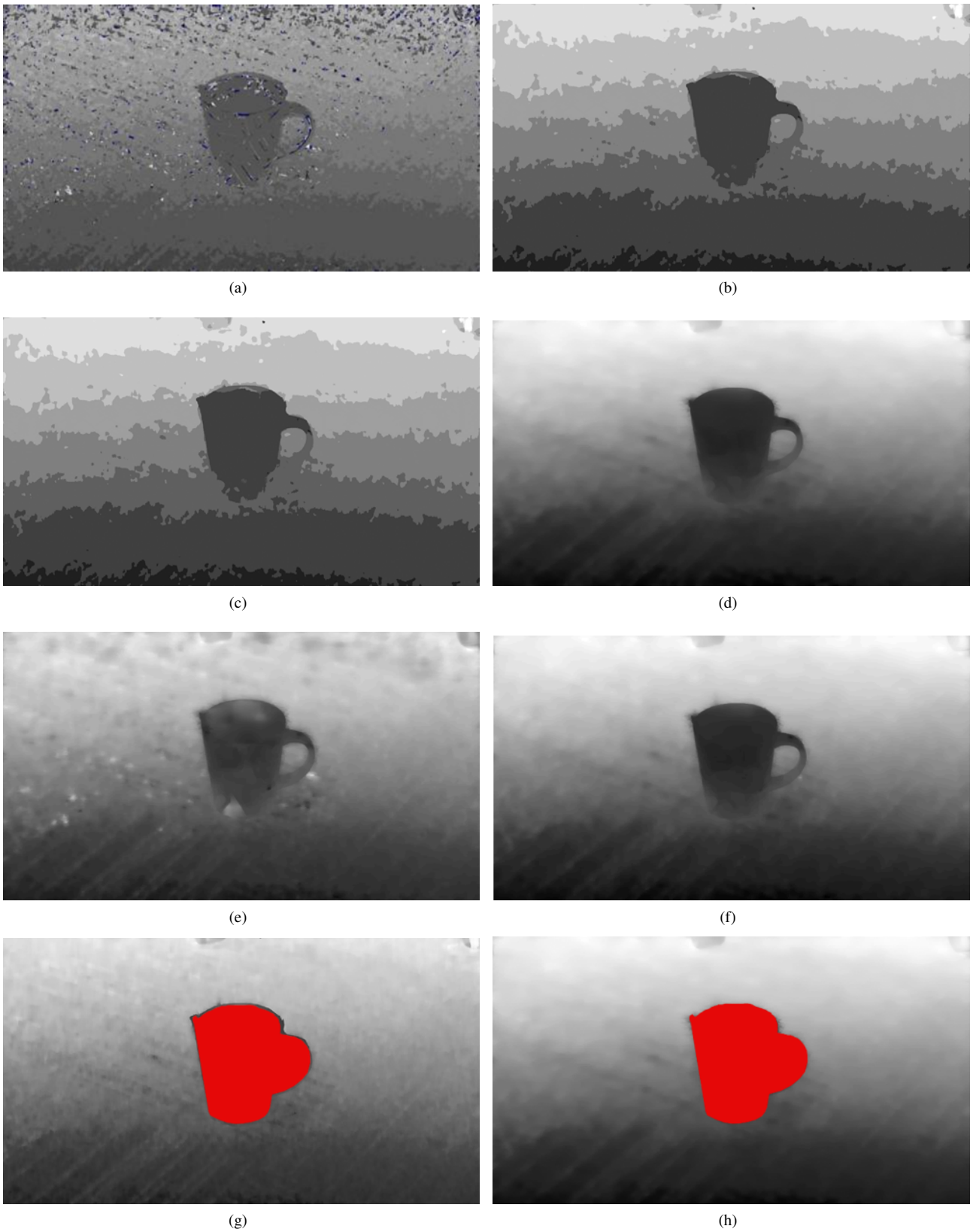
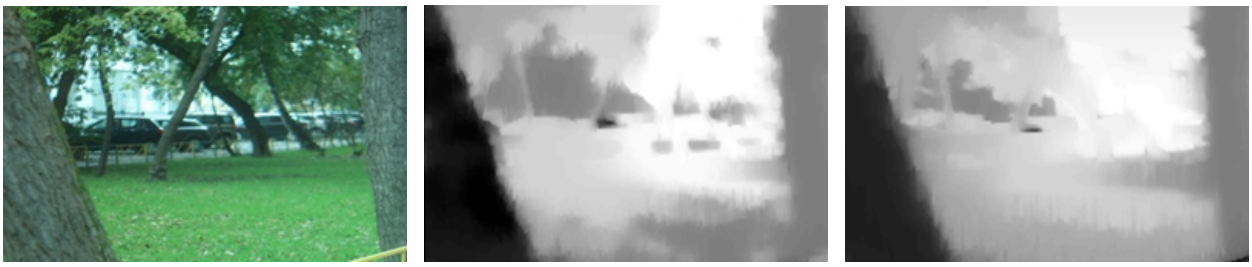


Figure 3. *Algorithm modifications impact in overall depth quality increase comparing to prior-art implementation available online [1]. Line 1: (a) color lines metric and (b) mutual correlation metric; Line 2: (c) layered depth map and (d) continuous depth map with sub-pixel estimation; Line 3: (e) depth map without cost filtering and (f) with cost filtering using joint-bilateral filter; Line 4: (g) depth map without edge restoration and (h) with edge restoration.*

DOI: 10.2352/ISSN.2470-1173.2016.21.3DIPM-405



(a) (b) (c)

Figure 4. *Input image compression may lead to depth artifacts: (a) input image, (b) depth map extracted from JPEG image, (c) depth map extracted from RAW image.*

Our algorithm implementation requires about 50 ms for FullHD resolution on Nvidia GeForce 780 Ti GPU and about 15 seconds on mobile device (see Table 1).

Performance on different target platforms

Platform	Time
PC CPU (Intel Core i7-2600), Matlab	6 s
PC GPU (NVidia GeForce 780 Ti), OpenCL	48 ms
Qualcomm Adreno 330, OpenCL	15 s

Proposed algorithm produces strip artifacts due to non-smoothness of bilateral filter approximation. However, these artifacts were not critical for our applications.

Prototype Evaluation

We have implemented a numerical simulator for image formation as well as the prototype for depth extraction based on Canon 60D camera and Canon EF 50mm f/1.8 lens. We have made a comparison of our algorithm with prior art for the same aperture as well as with commercial plenoptic camera Lytro which has a size comparable to prototype lens [10] (see Figure 5).

Next we compare our depth estimation algorithm with a highly light-efficient solution proposed in [4] (see Figure 6). Please, note that we captured images with different exposure times to overcome an issue of different light-efficiencies. Each depth map is scaled from its minimum to its maximum.

Depth Accuracy

Actually, measuring the correspondence between pixels allows to compute a disparity map (not a depth map). However, most of researchers in this area use disparity and depth as synonyms in this context, so we do. We differentiate these terms only in this sub-section. To calculate a depth map having a disparity map one should use a following equation:

$$\frac{1}{z_1}+\frac{1+Disp/R}{z_2}=\frac{1}{f}, \tag{8}$$

where: f is a lens focal length, R is a lens radius, $Disp$ is a disparity value, z_1 is a lens-object distance, z_2 is a lens-sensor distance.

We made a numerical fitting of z_1 and z_2 in (8) for disparity-to-depth conversion equation for a single lens and estimated depth accuracy in near focus area. We used a highly-textured synthetic scene to minimize the impact of wrong disparity map estimation.

The results in Figure 7 show that in ideal condition depth accuracy of proposed approach is close to Microsoft Kinect.

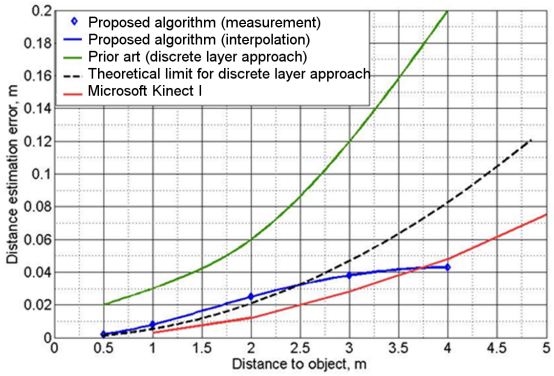


Figure 7. *Depth accuracy analysis in a center point of an image.*

Discussion

Here we discuss the limitations of the proposed color-coded aperture depth sensor compared to the most popular passive depth sensor, i.e. a stereo camera. We analyzed the theoretically achievable depth accuracy for different sizes of color-coded aperture cameras with respect to the distance to the object. This analysis is based on (8) and is in close agreement with experimental results shown in Figure 7.

In Figure 8 we show the distance between depth layers corresponding to disparity values equal to 0 and 1. The layered depth error is two times smaller than this distance. The sub-pixel refinement reduces the depth estimation error by half again (see Figure 7). That gives an accuracy better than 15 cm on the distance of 10 m and better than 1 cm on the distance below 2.5 m for the color-coded aperture equivalent baseline of 20 mm.

These results are in good agreement with plenoptic camera [7] and stereo camera [8] accuracies.

However, a color-coded aperture depth sensor has a number of limitations:

1. A working range of an color-coded aperture depth sensor is mostly limited by the its equivalent baseline. For example, for conventional smartphone cameras the working range is limited to 1 m.

DOI: 10.2352/ISSN.2470-1173.2016.21.3DIPM-405

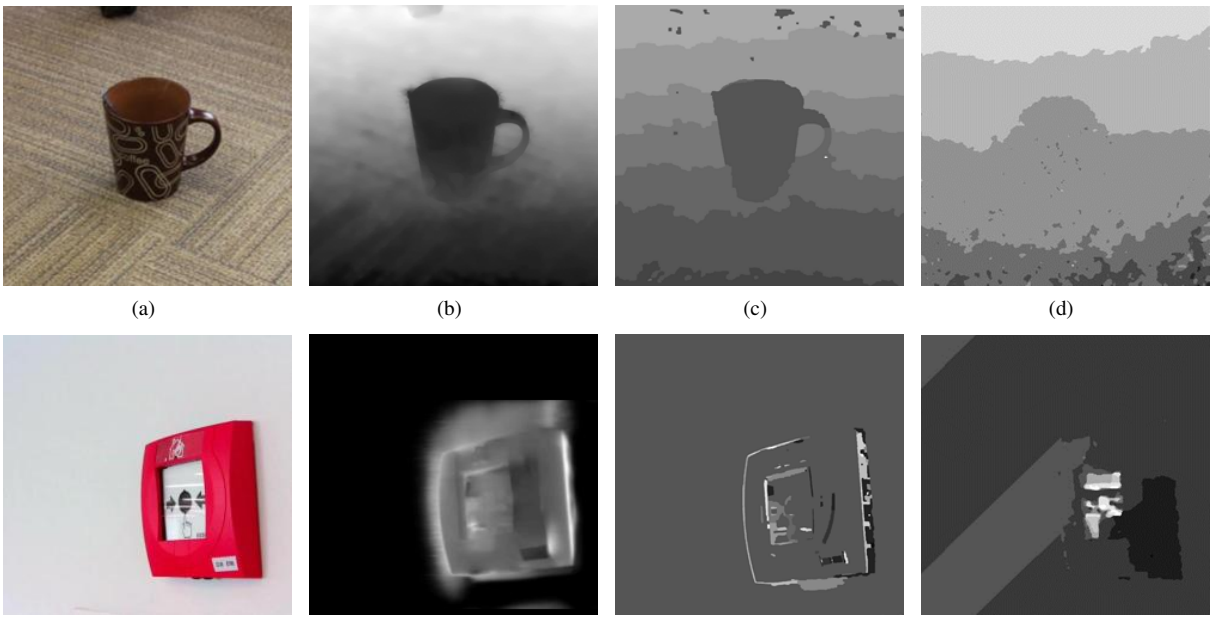


Figure 5. *Proposed algorithm provides better depth quality for both low- and highly-textured scenes. Line 1, highly-textured scene: (a) scene image, (b) proposed depth, (c) prior-art depth [1], (d) plenoptic camera depth [10]. Line 2, low-textured scene: (e) scene image, (f) proposed depth, (g) prior-art depth [1], (h) plenoptic camera depth [10].*

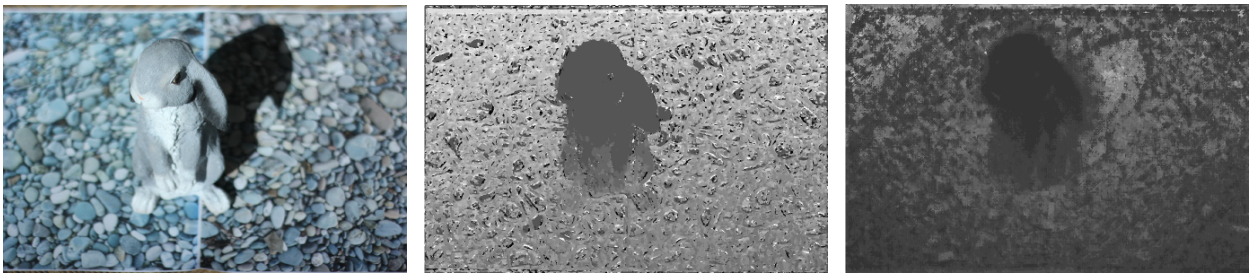


Figure 6. *Depth quality extracted with proposed algorithm is better than recent results in this area [4]. (a) Scene image, (b) proposed depth, (c) prior-art depth [4].*

2. All passive depth sensors require texture information for depth extraction. For a color-coded aperture depth sensor this requirement is stronger than for a stereo camera, as good texture should be present in each color channel.

3. The accuracy of our depth sensor is low in strongly defocused areas. Strong blur leads to low texture in these areas and therefore to disparity estimation accuracy degradation.

4. A color-coded aperture depth sensor requires computational restoration to get a sharp image, because it needs low f-number lens for disparity estimation. A stereo camera does not have this disadvantage.

Nevertheless, if these limitations are taken into account, a color-coded aperture depth sensor can be used in applications which require a single-lens single-frame depth sensor, e.g. 3D endoscope [9].

Conclusion

We have made a number of algorithm enhancements comparing to prior-art solutions [1]. They lead to more robust depth maps of a better quality. Moreover, we showed that in case of highly-textured scene near-focus depth accuracy of proposed approach is close to Microsoft Kinect. However, the color-coded aperture based depth sensor still suffers the lack of light-efficiency and is dependent of a texture quality even more than a stereo camera. This seems to be a promising direction of a future research.

DOI: 10.2352/ISSN.2470-1173.2016.21.3DIPM-405

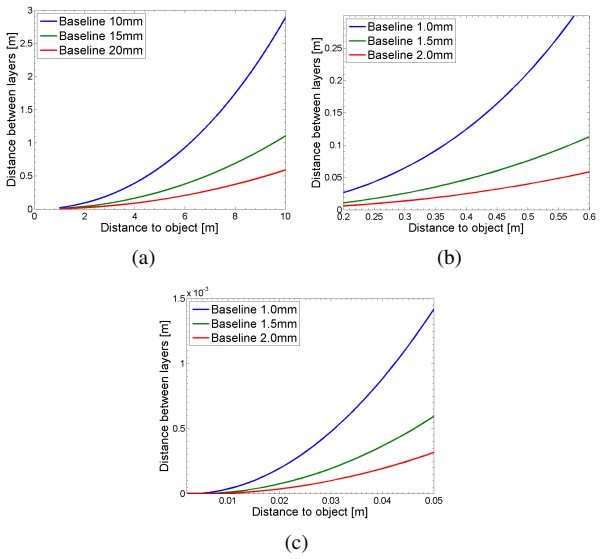


Figure 8. *Depth sensor accuracy analysis for different aperture baselines: (a) full-size camera with f-number 1.8 and pixel size 4.5 μm; (b), (c) compact camera with f-number 1.8 and pixel size 1.2 μm.*

References

[1] Y. Bando, B.-Y. Chen, and T. Nishita. Extracting depth and matte using a color-filtered aperture. *ACM Trans. Graph.*, 27(5):134:1–134:9, 2008.

[2] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Trans. Graph.*, 26(3), July 2007.

[3] E. Lee, W. Kang, S. Kim, and J. Paik. Color shift model-based image enhancement for digital multifocusing based on a multiple color-filter aperture camera. *Consumer Electronics, IEEE Transactions on*, 56(2):317–323, 2010.

[4] A. Chakrabarti and T. Zickler. Depth and deblurring from a spectrally-varying depth-of-field. *Proceedings of the European Conference on Computer Vision*, 2012.

[5] Y. Amari and E. Adelson. Single-eye range estimation by using displaced apertures with color filters. In *Proc. Int. Conf. Industrial Electronics, Control, Instrumentation and Automation*, pages 1588–1592, 1992.

[6] I. Panchenko and V. Bucha. Hardware accelerator of convolution with exponential function for image processing applications. *International Conference on Graphic and Image Processing (ICGIP 2015) in Society of Photo-Optical Instrumentation Engineers (SPIE)*, 2015.

[7] N. Zellera, F. Quintb, and U. Stillac. Calibration and accuracy analysis of a focused plenoptic camera. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1:205–212, 2014.

[8] M. Kytö, M. Nuutinen, and P. Oittinen. Method for measuring stereo camera depth accuracy based on stereoscopic vision. In *IS&T/SPIE Electronic Imaging*, pages 78640I–1–9. International Society for Optics and Photonics, 2011.

[9] Y. Bae, and H. Manohara, and V. White, and K. V. Shcheglov, and H. Shahinian. Stereo Imaging Miniature Endoscope. *NASA. Tech Briefs 35.6*, 2011.

[10] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz Light field photography with a hand-held plenoptic camera *Computer Science Technical Report CSTR 2.11*, 2005.

Author Biography

Ivan Panchenko has got his B.S. (2009) and M.S. (2011) degrees from St. Petersburg State Electrotechnical University "LETI" where he is now pursuing his Ph.D.. Ivan joined Samsung R&D Institute Russia as Research Engineer in 2011. His research interests include Signal and Image Processing, Computer Vision, Embedded Systems and High Performance Computing.

Vladimir Paramonov received his M.S. in Mechanics from Lomonosov Moscow State University in 2009 where he is now pursuing his Ph.D.. Vladimir joined Samsung R&D Institute Russia as Research Engineer in 2012. His research interests include Applied Mathematics, Computational Methods and Numerical Modeling.

Victor Bucha has got his Ph.D. from United Institute of Informatics Problems of National Academy of Sciences of Belarus (2006) and M.B.A. from The Open University (2013). Victor joined Samsung R&D Institute Russia in 2007 as a Senior Research Engineer. His research interest include Signal Processing, Image Processing, 3DTV and Data Science.