Vol.2010-CVIM-174 No.7
2010/11/18

F-number control for Accurate Depth Estimation with Color-Filtered Aperture

Hiroki Yazawa[†1] and Shinichi Nakajima [†1]

Accurate depth map estimation is one of the hot topics in the field of computer vision as well as computer graphics, because of its wide range of applications. The color-filtered aperture (CFA) approach has shown to be promising with its handy hardware installation and simple estimation algorithm. However, it does not always perform well for various scene situations, if all the optical parameters are fixed. In this paper, we show that controlling f-number based on the geometrical optics effectively cope with variety of scene situations. More specifically, we first propose a way to estimate the optimal f-number from layout of the camera and objects, e.g., distance between the camera and objects, and relative distances between objects. We also propose a depth estimation from multiple images with different f-numbers, which widens the applicable range of scene variation with high accuracy. We demonstrate the performance gain of our approaches. Our approach can be realized without additional hardware modification if we install the CFA on a lens equipped with an adjustable pupil aperture.

1. Introduction

Depth map estimation is a useful application that enables wide variety of post-exposure processing, e.g., image segmentation, refocusing, and partial modification of brightness and contrast. Studies in this area can be roughly divided into two categories, *active* and *passive* methods. *Active* methods includes stereo-camera technique[12], structured light methods[20], with which relatively high accuracy can be achieved. However, additional light source or second camera is typically needed. On the other hand, *passive* methods including depth from defocus (DFD)[13] is typically realized with a smaller hardware modification.

Levin et al. (2007)[2] proposed a method to estimate the depth from a single image, with a uniquely patterned aperture, called *coded aperture* (CA), at the pupil plane. The pattern is designed to have a special frequency components that depend on the amount of defocus, deconvolution with pre-measured point spread function (PSF) can clearly distinguish the image points from one depth point to another[2),3),8)]. CA can be made also with phase masks[14),15)]. One of the key advantages of CA approaches is that it can be installed with minimal modifications to a conventional camera lens.

Another direction of CA research requires multiple image acquisition. In general, one can relatively easily estimate the depth by comparing degree of image blur or shift between multiple images captured by different coded apertures[7),9),10)]. Ng et al. (2005) proposed the *light-field camera* where light rays are separated accordingly to their incident angles by placing a microlens array in front of the image sensor. The obtained multi-view images (captured through a single lens though) enable accurate depth estimation[4)]. This can be realized with an attenuation mask put on the image sensor[3)], or with the multiple apertures splitting light rays[5),6)]. This approach needs relatively complex optical system.

Recently, the *color-filtered aperture* (CFA) approach has been proposed and shown its excellent performance[1)]. CFA uses multiple color channels as light rays with different incident angles. The depth estimation is based on the *color lines model*[2)], and does not need deconvolution with PDFs. This is advantageous because we do not need to pre-measure the PDFs that typically depend on the image position. Also its simple algorithm greatly reduces the calculation time, compared with typically heavy deconvolution calculation. The installation of CFA is very handy, which is another big advantage.

In this paper, we propose new methods that enhance the performance of CFA. Since Bando et al. used CFA with a fixed f-number, accurate depth estimation cannot always available in various scene situations. To cope with this, we first propose a way to estimate the optimal f-number for depth estimation based on the distances of objects from the camera. Since f-number control is readily available if one installs the CFA on exchangeable lens equipped with adjustable aperture, our approach widens the applicable range of scene variety without additional hardware modification. We also propose to use multiple images captured with different f-numbers, which realizes finer resolution and wider dynamic range at

†1 Optical Research Laboratory, Nikon Corporation, 1-6-3, Nishi-ohi, Shinagawa-ku, Tokyo 140-8601 Japan

Vol.2010-CVIM-174 No.7
2010/11/18

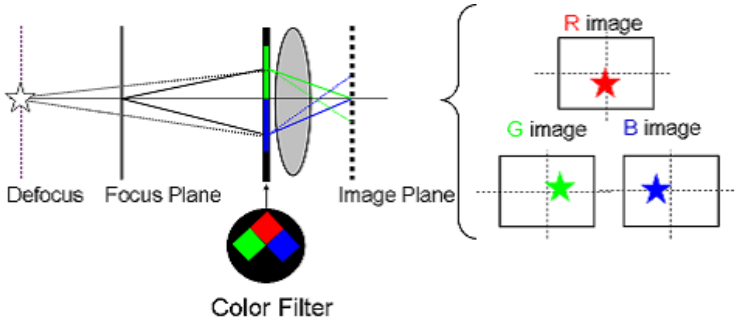

Fig. 1 Color-filtered aperture (CFA) approach proposed by Bando et al. (2008)[1].

the same time. We experimentally demonstrate the performance gain of our approaches.

2. Color-Filtered Aperture (CFA) Approach

In this section, we review the Color-Filtered Aperture (CFA) approach proposed by Bando et al. (2008)[1].

In the CFA approach, an image is captured through the color-filtered aperture, shown in Figure 1. Since a shift at the pupil plane corresponds to a tilt of rays on the image plane, the defocus, i.e., the distance between the focal plane and an object, causes shifts between R, G, and B images on the image plane.

Let $I_r(x,y)$, $I_g(x,y)$ and $I_b(x,y)$ be the R, G, and B images. Let $d = d(x,y)$ be the distance between the camera and the visible face of object at image position (x,y). The R, G, and B images shift upward, rightward and leftward, respectively, according to the defocus (see the color filter shown in Figure 1). Therefore, the shift-corrected images can be written as $I_r(x, y-\epsilon)$, $I_g(x+\epsilon, y)$ and $I_b(x-\epsilon, y)$. Here, $\epsilon = \epsilon(d)$ is the degree of parallax.

One of the advantages of the CFA approach is that it does not require point spread functions, unlike other coded aperture approaches. Instead, it relies on the *color lines model*[11]; the tendency of colors in natural images to form elongated clusters in the RGB space. We assume that pixel colors within a local window $w(x,y)$ around (x,y) belong to one cluster, which is *elongated*. The degree of elongation introduced in the following is adopted as the *color alignment measure*.

Let

$$S_I(x,y;d) = \{(I_r(x, y-\epsilon(d)), I_g(x-\epsilon(d), y), I_b(x+\epsilon(d), y)) \mid (s,t) \in w(x,y)\}$$

be the set of colors of the pixels within the local window $w(x,y)$, plotted in the RGB color space. Here, ϵ is the parallax that depends on the *hypothesized* distance $d(x,y)$. We minimize the following criterion over d for each image position (x,y):

$$L(x,y;d) = \frac{\lambda_0\lambda_1\lambda_2}{\sigma_r^2\sigma_g^2\sigma_b^2}, \tag{1}$$

where λ_0, λ_1, and λ_2 $(\lambda_0 \geq \lambda_1 \geq \lambda_2 \geq 0)$ are the eigenvalues of the covariance matrix Σ of the set $S_I(x,y;d)$ of color points, and σ_r^2, σ_g^2, and σ_b^2 are the variances along R-, G-, and B-axes (i.e., the diagonal elements of Σ). Finally, we convert ϵ to the object distance d.

$L(x,y;d)$ is the product of the variances of the color distribution along the principal axes, normalized by the product of the variances along the RGB axes. It gets small when the cluster is elongated (i.e., $\lambda_0 \gg \lambda_1, \lambda_2$) in an oblique direction with respect to the RGB axes, meaning that the RGB components are correlated. Thus, we can evaluate the true depth d thorough the smallest value of $L(x,y;\epsilon)$. See Bando et al. (2008)[1] for further details and illustrative explanation.

3. Proposed Methods

In this section, we propose new methods that provide accurate depth map estimation with CFA for various scene situations. Our methods can be available simply with an *adjustable* aperture, placed next to the CFA (see Figure 2). (Note that exchangeable lenses are usually equipped with it.)

By controlling the f-number, we can balance the important optical properties for depth estimation, namely, the dynamic range and the resolution. For example, if we adopt a small f-number, say $f/1.8$,[⋆1] high depth resolution is available

⋆1 By convention, we denote the f-number as $f/1.8$, where 1.8 is the f-number. The variable f denotes the focal length of the lens, and $N = f/1.8$ corresponds to the diameter of the aperture at the pupil.

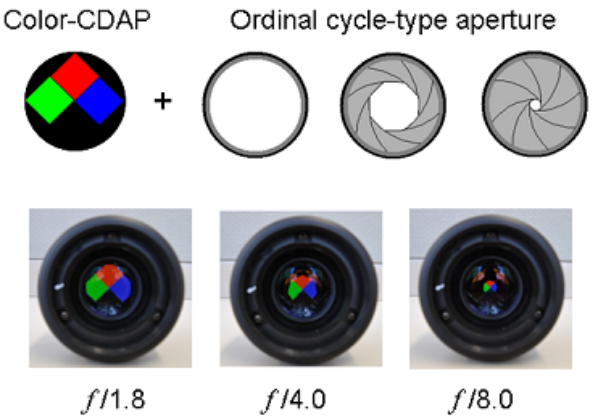

Fig. 2 Our approach can be realized by adjusting the circular aperture, placed next to the color filter. The photos in the lower row show actual color filters bounded by the aperture ($f/1.8$, $f/4.0$, and $f/8.0$).

because of large parallax, at the expense of small dynamic range because of short depth-of-field. Too much image blurring rubs out the texture of objects far from the focal plane disables the color alignment measure from extracting the depth information. On the other hand, a large f-number provides a wide dynamic range with lower depth resolution. Furthermore, f-number also controls the brightness of the image; we need to adopt a large f-number with dark illumination.

Our first method gives a way to estimate the optimal f-number, or acceptable range, for various scene situations, based on the distance information of the target objects from the camera. Our second method extends both of the dynamic range and the resolution by using multiple images captured with different f-numbers. This works well especially when objects has (psudo) periodical structure and the single image CFA fails.

3.1 Proposed Method 1: CFA with Optimal F-number

In our first proposed method, we set the f-number in the following range:

$$\frac{1}{40}\frac{1}{(f+s)}\left[f^2(1-\frac{s}{s_f})\right] < F < \frac{1}{4p}\frac{f^2}{(f+s)}\left[\frac{s}{s_f}\frac{\Delta d}{(s_f-\Delta d)}\right], \qquad (2)$$

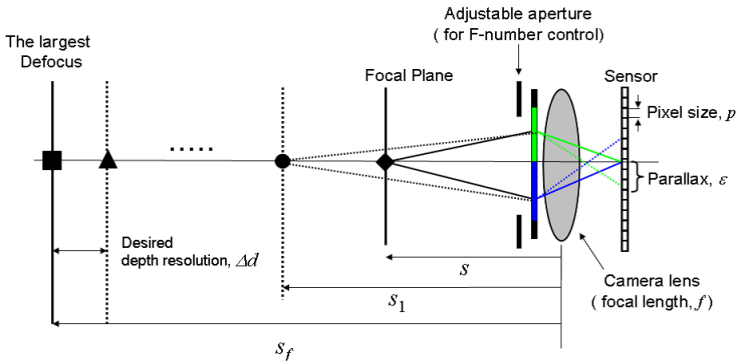

Fig. 3 Optical diagram

where

s: the distance of the object at the focused position,
s_f: the largest defocus length of objects whose depth you want to accurately estimate,
Δd: desired depth resolution,
p: sensor pixel size,

(see Figure 3).

Derivation

Our guideline is based on the degree of parallax against defocus. It is expressed as

$$\epsilon(F,s,s_1) = \frac{1}{4}\frac{1}{(f+s)F}\left[f^2(1-\frac{s}{s_1})\right], \qquad (3)$$

where f is focal length, s is object position at focal point, s_1 is defocus from focal point, and F is the f-number. Eq.(1) derivated from basic geometric optics under paraxial approximation. Note that, principal light ray goes through center of each color filter apertures.

Eq.(1) shows that the parallax is changing significantly at small defocus, and

its changing rate is gradually small against defocus, and finally converge. Now, if the desirable depth resolution Δd, largest object position s_f and s are given, the smallest differential of parallax against depth can be expressed,

$$\Delta\epsilon_{min}(F,s,s_f) = \epsilon(F,s,s_f-\Delta d) - \epsilon(F,s,s_f) = \frac{1}{4}\frac{f^2}{(f+s)F}\left[\frac{s}{s_f}\frac{\Delta d}{(s_f-\Delta d)}\right]. \tag{4}$$

To distinguish all object depth from parallax, ϵ_{min} must be larger than image sensor pixel size p. Thus, first term for f-number selection is,

$$\Delta\epsilon_{min}(F,s,s_f) > p. \tag{5}$$

Eq.(2) and (3) show that the smallest f-number is most adequate for high depth resolution. However, image blur against defocus also must be considered. Fortunately, the degree of image blur against defocus can also express by eq.(3) under paraxial approximation, that is, the maximum image blur size is $\epsilon(F,s,s_1)$. Thus, if acceptable blur size B is given, second term for f-number selection is,

$$B > \epsilon(F,s,s_f). \tag{6}$$

Actually, it is difficult to determine actual B value because this value would depend on object texture, size and so on. Generally, MTF at 10/mm is the smallest criterion values for evaluate camera lens performance. Therefore, we decided that, at least resolution of 10/mm is necessary to capture object structure ($B = 0.1$-mm corresponding to $10p$in this paper).

3.2 Proposed Method 2: CFA with Multiple F-numbers

The second proposed method provides both of wider dynamic range and finer resolution, by using multiple images with different f-numbers.

Suppose we have n images, each of which is captured with a different f-number F_i for $i = 1,\ldots,n$. Let $I_r^{(i)}$, $I_g^{(i)}$, $I_r^{(i)}$ be the RGB components of the i-th image. We define the set of colors within a local window $w(x,y)$ by

$$S_{I^n}(x,y;d) = \{(I_r^{(i)}(x,y-\epsilon_i(d)), I_g^{(i)}(x-\epsilon_i(d),y), I_b^{(i)}(x+\epsilon_i(d),y) \mid (s,t)\in w(x,y), i=1,\ldots,n\}, \tag{7}$$

where $\{\epsilon_i\}$ denotes the parallaxes of the n images that depend on the *common* hypothesized distance d and optical conditions as follows:

$$\epsilon_i(d) = \frac{1}{4}\frac{1}{(f+s)F_i}\left[f^2(1-\frac{s}{d})\right]. \tag{8}$$

We minimize the following *color alignment measure* over d:

$$L(x,y;d) = \prod_{i=1}^{n}\frac{\lambda_0^{(i)}\lambda_1^{(i)}\lambda_2^{(i)}}{\sigma_r^{(i)2}\sigma_g^{(i)2}\sigma_b^{(i)2}}, \tag{9}$$

which is a simple extension of Eq.(1) to multiple f-numbers.

4. Experimental Results

In this section, we show experimental results that prove the advantage of our methods. The experiment was performed in following setting. Reflex camera of Nikon D-5000 and single short focus lens of $f = 50$-mm, F/1.8D were used. We can select the f-number among f/1.8, f/2.8, f/3.2, f/4, f/5.6, f/7.1, f/8, f/9, f/11 and f/16. We installed a rhomboid-shape colored filter just before the circular-aperture at the pupil plane. For the R, G, and B color filters we used Fujifilter SC-58, BPB-53, and BPB-45, respectively. All images were captured by RAW-mode to avoid JPEG artifacts. We applied no post-processing, e.g., the graph-cut technique to smooth the resulting depth map, in all the experiment.

4.1 Performance Gain by Proposed Method 1 (F-number Optimization)

Figure 4 shows the captured images (in the left column), and the resulting depth map (in the right column). Three rows correspond to different f-numbers. The average distance of objects from the camera is $s \approx 600$ (close-up setting), and $s_f = 1350$-mm, and each object are placed every $\Delta d = 250$-mm. Substituting

them into Eq.(2), we obtain the approperiate range for the f-number as

$$f/4.0 < F < f/7.1. \tag{10}$$

As our method suggests, the image with the f-number $f/11.0$, which is in the range (10), shows the best result among the shown three f-numbers,.

Images with $f/2.0$ and $f/5.6$ do not perform well, because the textural information of objects placed far from the focal distance is blurred too much due to short depth-of-field. A relatively large f-number is suitable for depth estimation in such a close-up capturing.

On the other hand, Figure 5 shows the result (in the same form as Figure 4) with another scene. The average distance of objects from the camera is $s \approx 1400$ (close-up setting), and $s_f = 2150$-mm, and each object are placed every $\Delta d = 250$-mm.

In this case, the image with $f/9.0$ does not perform well, because of the poor resolution of depth estimation. The image with $f/1.8$ also does not perform well for the same reason as in the first example (Figure.4).

4.2 Performance Gain by Proposed Method 2 (Multiple F-numbers)

The performance with the optimal f-number can be further improved by using multiple images with different f-numbers. In Figure.6, we show the captured images with $f/3.2$ and $f/5.6$ ((a) and (b)), and the corresponding depth map ((c) and (d)).

For single image processing, $f/3.2$ and $f/5.6$, which is recommend by the proposed method 1, performs well but has some estimation errors.

Figure.5(e) shows the resulting depth map,which is obtained from the two images (a) and (b) via the procedure proposed in Section 3.2. This proves that our multiple f-numbers approach improves the performance.

5. Conclusion

We showed that the performance of the color-filtered aperture (CFA) approach can be enhanced by controlling the f-number. Since the distances of objects could be measured by auto-focus sensor and the f-number could be adjusted automatically, our methods could be installed in a camera with fully automatic f-number control, which will make CFA approach more handy.

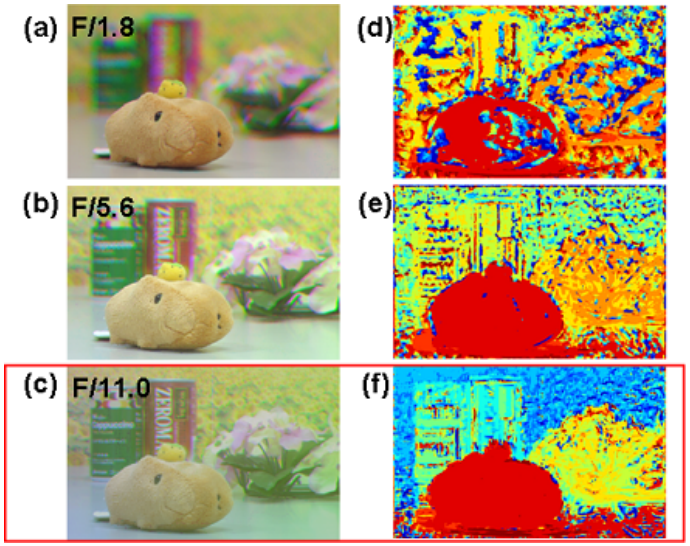

Fig. 4 (a)-(c): Captured image at (a) $f/1.8$, (b) $f/5.6$, and (c)$f/11.0$. (d)-(f): Estimated depth-map of (a), (b), and (c), respectively.

We also showed that multiple images with different f-numbers improves the accuracy. However, this approach has a potential problem; the camera as well as objects can move during the multiple acquisition. As future work, we would like to cope with this problem.

References

1) Y. Bando, C. Bing-Yu, and N. Tomoyuki, gExtracting Depth and Matte using a Color-Filtered Apertureh, SIGGRAPH ASIA 2008.
2) A. Levin, R. Fergus, F. Durang, and W. T. Freeman. : gImage and depth from a conventional camera with a coded apertureh, ACM Trans. Gr. 26, 3, 70:1?70:9 (2007).
3) A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, gDappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusingh, ACM Trans. Gr. 26, 3, 69:1?69:12 (2007).
4) R. Ng, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan, gLight field photography with a hand-held plenoptic camerah, Tech. Rep. CSTR 2005-02, Stanford Computer Science (2005).

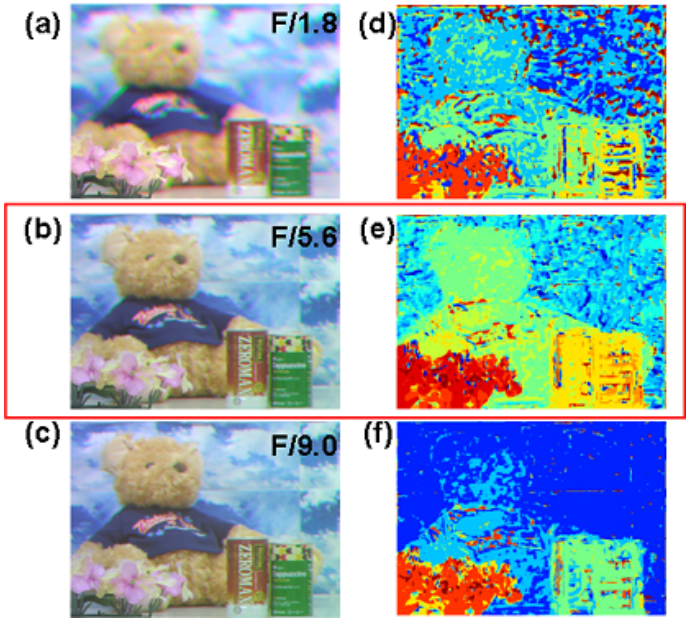

Fig. 5 (a)-(c): Captured image at (a) $f/1.8$, (b) $f/5.6$, and (c)$f/9.0$. (d)-(f): Estimated depth-map of (a), (b), and (c), respectively.

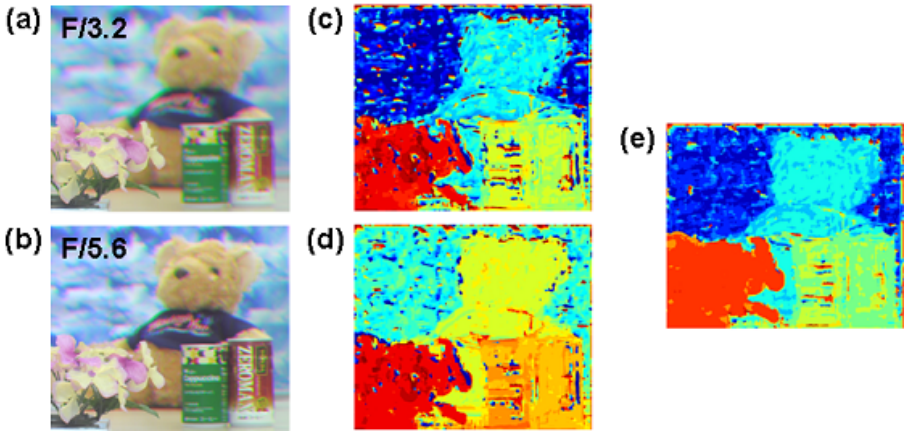

Fig. 6 (a)-(b): Captured image at (a) $f/3.2$ and (b) $f/5.6$. (c)-(d): Estimated depth-map of (a) and (b), respectively. (e) Estimated depth-map using (a) and (b).

5) P. Green, W. sun, W. Matusik and F. Durand, gMulti-aperture photograpyh, ACM Trans. Gr. 26, 3, 68:1?68:7.(2007).
6) C. K. Liang, T. H. Lin, B.-Y. Wong, C. Liu and H. H. Chen, gProgrammable aperture photography: multiplexed light field acquisitionh, ACM Trans. Gr. 27, 3, 55:1–55:10.(2008).
7) C. Zhou, S. Lin and S. Nayar, gCoded Aperture Pairs for Depth from Defocush, ICCV 2009.
8) C. Zhou and S. Nayar, gWhat are good apertures for defocus deblurring?h, In International Conference of Computational Photography, 2009.
9) H. Farid and E. Simoncelli, gRange estimation by optical differentiationh, Journal of the Optical Society of America A, 15(7):1777–1786, (1998).
10) S. Hiura and T. Matsuyama, gDepth measurement by the multi-focus camerah In CVPR, pages 953–959, (1998).
11) I. Omer, and M. Werman, gColor lines: image specific color representationh In Proc.CVPR, vol. 2, 946–953.
12) D. Scharstein and R. Szeliski, gA taxonomy and evaluation of dense two-frame stereo correspondence algorithmsh. IJCV (2002).
13) S. Chaudhuri and A. Rajagopalan, gDepth from defocus: A real aperture imaging approachh, Springer-Verlag, New York (1999).
14) E. Dowski and W. Cathey, gSingle-lens single-image incoherent passive-ranging systemsh, App Opt (1994).
15) A. Levin, S. Hasinoff, P. Green, F. Durand and W. Freeman, g4D frequency analysis of computationalcameras for depth of field extensionh, SIGGRAPH 2009
16) Y. Boykov, O. Veksler, and R. Zabih.; gFast approximate energy minimization via graph cutsh, PAMI 23, 1222-1239.
17) Y. Boykov, and G. Funka-Lea, gGraph Cuts and Efficient N-D Image Segmentationh, IJCV, vol. 70, no. 2, pp. 109-131 (2006).
18) C. Rother, V. Kolmogorov and A. Blake, "GrabCut: Interactive Foreground Extraction using Iterated Graph Cuts", SIGGRAPH 2004
19) A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr, gInteractive image segmentation using an adaptive GMMRF modelh, Proc. Eur. Conf. on Computer Vision, ECCV (2004).
20) D.Scharstein and R.Szeliski, gHigh-Accuracy Stereo Depth Maps Using Structured Lighth, CVPR 2003.