A New Method for Creating a Depth Map for Camera Auto Focus Using an All in Focus Picture and 2D Scale Space Matching

Earl Wong
Sony Electronics
Earl.Wong@am.sony.com

ABSTRACT

This paper describes a novel algorithm for creating a depth map from an arbitrary scene. The method utilizes the following two concepts: 1) The generation of an all in focus picture (infinite depth of field image) and 2) scale space theory. By combining ideas 1) and 2) with an intensity based block matching algorithm, a depth map is computed. Experimental results demonstrate the efficacy of the approach.

1. INTRODUCTION

Depth maps can be generated using a variety of monocular and binocular techniques. Here, we will focus on the monocular techniques. In Depth from Focus (DFF) methods, the scene sharpness is analyzed using a series of pictures captured at various camera focus positions [1],[2],[9]. For high accuracy, minimum depth of field pictures are captured. A depth map is then produced by using the knowledge of the camera parameters and the information contained in the captured pictures. In contrast, Depth from Defocus (DFD) approaches utilize one or two captured pictures [3],[9],[10]. These methods attempt to determine an associated blur quantity. If the blur quantity can be computed accurately, these methods provide a nice alternative to the DFF approach. However, accurate estimates of the blur quantity are difficult to obtain. In a future paper, we will present results from a DFD approach.

In this paper, we present a new technique for generating a depth map. Our method requires fewer captured pictures than the DFF approach and accurately computes the amount of blur present. This results in high level of depth accuracy. We accomplish our objective by fusing information captured in a sequence of maximum depth of field pictures to generate an all in focus picture. Scale space theory is then applied to the all in focus picture to generate a complete multi-scale representation of the scene. (To the best of our knowledge, this is the first time that scale space theory has been to the depth map generation problem.) Next, block matching is used to determine the quantity of blur present at various locations in the captured scene. Once the blur is known, a depth map can be easily computed.

2. BACKGROUND

The imaging process [1] can be described using the thin lens geometric optics model shown in Figure 1.

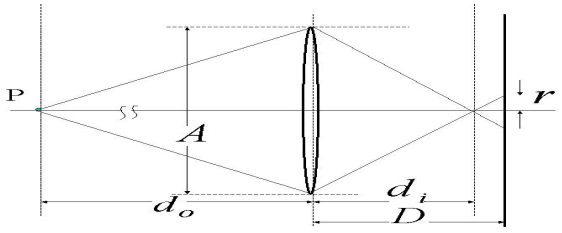

Figure 1: Blur formation process. P denotes the location of an object in the scene.

Using triangle equalities, we can write the following relationship:

$$\frac{(A/2)}{d_i} = \frac{r}{D - d_i} \rightarrow d_i = \frac{(A/2)D}{(A/2) + r}$$

Using the lens law equation, we can write the additional relationship:

$$\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f} \rightarrow d_o = \frac{fd_i}{d_i - f}$$

f denotes the camera focal length. Substituting d_i into the second equation, we obtain:

$$d_o = \frac{fAD}{AD - fA - 2fr} \qquad \textbf{Eqn (1)}$$

With the exception of blur term r, every term on the right hand side is known. Hence, once r is determined, the object depth d_o can be computed.

3. ALL IN FOCUS PICTURE GENERATION

Conceptually, an all in focus picture (infinite depth of field picture) can be captured using a pinhole camera. Since this is impractical, we generate the all in focus picture in the following manner. First, a sequence of pictures $\{I_1, I_2, ...I_M\}$ is captured using a single

camera that takes multiple pictures at different focus locations in the scene. Associated with each captured picture is a depth range where the scene information is in focus. For example, all objects lying within a to b feet from the camera may be in focus in picture I_1. All objects lying within b to c feet from the camera may be in focus in picture I_2, etc. Camera settings are chosen to maximize the in focus range for each captured picture. Focus locations are set to ensure that the in focus ranges are non-overlapping. (The latter two criteria minimize the number of captured pictures needed to generate the all in focus picture.) These ideas are illustrated in Figure 2.

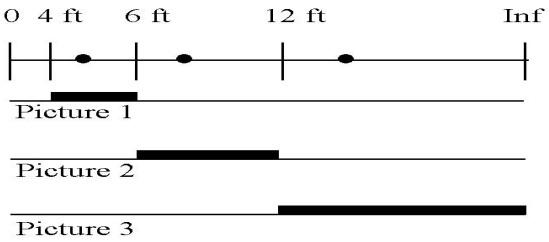

Figure 2. Black rectangles indicate the in focus regions of each captured picture. Pictures have been taken with the camera set for maximum depth of field exposure. In this example, the information from 3 pictures are needed to cover the entire depth range.

Next, the all in focus picture is generated by fusing the in focus information contained in the captured pictures. (Note: We assume that camera magnification effects are negligible and that the scene is static.) Many excellent image fusion techniques have been presented in the literature [5], [6]. Here, we perform image fusion by comparing the variance of the intensity at identical block locations in our captured pictures $\{I_1, I_2, ...I_M\}$ and selecting the intensity information associated with the block containing the largest variance. This process is repeated for all non-overlapping block locations in the scene. Our approach can be expressed succinctly using the following mathematical equations: **Eqn(2)**

$$K = \max_i \{\mathrm{var}(I_i(mN:(m+1)N, nN:(n+1)N))\}$$

$$AIF\_Pict(mN:(m+1)N, nN:(n+1)N) = I_{i=K}(mN:(m+1)N, nN:(n+1)N)$$

These equations are applied for all permissible values of m and n for block size N. We illustrate the results of our image fusion algorithm in Figure 3. In Picture 1, the foreground object is in focus. In Picture 2, no areas of the picture are in focus. In Picture 3, the background is in focus. By applying our algorithm, we obtain Picture 4, the all in focus picture.

Figure 3. Left to right and top to bottom: Picture 1, Picture 2, Picture 3 and Picture 4. Picture 4 (bottom right) = the fused image/all in focus picture.

4. SCALE SPACE REPRESENTATION

Scale space theory provides a concise mathematical framework for producing a multi-scale description of a signal or image [7]. In Witkin's seminal paper, the scale space representation of a 1D signal is given as the convolution of the signal with a family of gaussian blur kernels:

$$g(x,t) = \int_{-\infty}^{+\infty} g_0(x-\tau)h(\tau,t)d\tau \quad \textbf{Eqn(3)}$$

g_0 denotes the fine scale information contained in the original signal, h denotes the gaussian blur kernel with scale parameter t and $g(x,t)$ denotes the family of derived signals. Equivalently, for the gaussian case, the family of derived signals can be obtained as the solution of the linear heat-diffusion equation:

$$g_t = \Delta g = g_{xx}$$

Scale space descriptions have also been associated with non-linear differential equations such as the non-linear heat-diffusion equation (anisotropic diffusion) equation [8]:

$$g_t = \nabla \bullet (c(x,t)g_x)$$

In this paper, we will focus our attention on the 2D convolution representation. Two 2D blur kernels of interest include the previously mentioned gaussian blur kernel

$$h(x, y, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e - (\frac{x^2 + y^2}{2\sigma^2})$$

and the pillbox blur kernel

$$h(r) = \frac{1}{\pi r^2}$$

The family of derived images at various scales is created by applying $t = \sigma$ or $t = r$ for a range of values in the convolution relation. By construction, the gaussian blur kernel produces a gaussian blur family while the pillbox blur kernel produces a pillbox blur family. Figure 4, illustrates the result of applying the discrete version (time and space) of the 2D convolution equation using the pillbox blur kernel. The fine scale information contained in the original signal/picture (all in focus picture) is given in the first image. Pictures with increasing pillbox blur (members of the family of derived signals/pictures) are also shown. Note: Only four members of the pillbox blur family are shown in Figure 4. Many additional members exist.

Figure 4. Original picture and increasingly (pillbox) blurred pictures. $t = 0, r_1, r_2 \,\&\, r_3$. ($r_1 < r_2 < r_3$)

5. ALGORITHM

In Section 3, we introduced the idea of an all in focus picture and provided an algorithm for creating such a picture. In Section 4, we introduced the concept of creating a scale space description of a scene from the fine scale information contained in the original signal/picture (all in focus picture). Now, we use the scale space description to create a depth map.

From Section 2, we know that objects at different distances from the imaging device will undergo different amounts of blur. But how much blur is present at "region xy" in our captured picture? In order to answer this question, we turn to our family of derived pictures. By comparing the intensity value at "region xy" in our captured picture with the intensity values at "region xy" in our family of derived pictures, we can determine the best match. Once the best match is known, the blur quantity is known. Since the blur quantity in the derived image is approximately the same as the blur in the captured picture at "region xy", we now know the approximate depth at "region xy" in the captured picture.

Our complete algorithm consists of the following steps:

1) Generate the all in focus picture I_0 (Section 3, **Eqn(2)**).
2) Generate the scale space representation (family of derived images) of image I_0 from the all in focus picture. (Section 4, 2D signal version of **Eqn(3)**).
3) Select one of the captured pictures from $\{I_1, I_2, ...I_M\}$ used to generate the all in focus picture. For example, let $I_{select} = I_2$. (Or, take a new picture of the same scene with camera settings that "maximize" the blur content in the scene.)
4) For a specific location in the captured picture, find the best match between the $I_{select.}$ intensity data and the intensity data generated by the scale space representation (family of derived images) at the same location.

Example: Let I_{select} = Picture 1 (Figure 3). Let the specific location in the picture be denoted by the white box. This region is then compared with the white box regions in the family of derived images (three family members are shown in Figure 4) to determine the best match.

5) Determine the amount of blur applied to the picture in the family of derived pictures that produces the best match. Store this value.
6) Repeat this procedure for all non-overlapping locations in I_{select}, thereby producing a map containing the amount of blur present at all of the picture locations.
7) Apply **Eqn(1)** derived in Section 2 to obtain the corresponding depth. (In the case of gaussian blur, r should be scaled by a constant coefficient k).

6. EXPERIMENTAL RESULTS

We now provide experimental results to demonstrate the merits of our algorithm. For the sake of clarity, we will only present the blur map result. The depth map can be easily computed using Eqn. 1 in Section 2. Figures 5 and 7 show two different captured

pictures used in step 4 of the algorithm. Figures 6 and 8 show the results from step 6 of the algorithm. In both examples, the foreground and background blur have been accurately determined for the majority of areas in the picture. Note: Errors exist in the lower left hand corner of the picture and the sky region. This is because there is no image structure present in these regions. Our algorithm assumes a static scene. The accuracy of our algorithm is primarily limited by step 4 (matching step) in section 5. Matching routines are frequently employed in image processing due to their robust characteristics and performance.

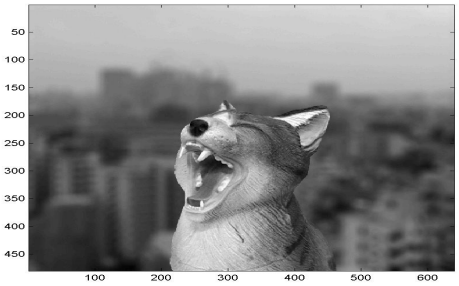

Figure 5: Foreground blur=0, Background blur=10

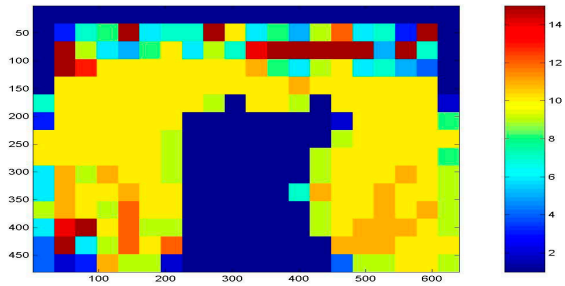

Figure 6: Resulting blur map.

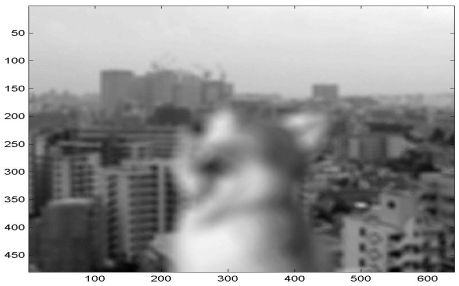

Figure 7: Foreground blur=15, Background Blur=5

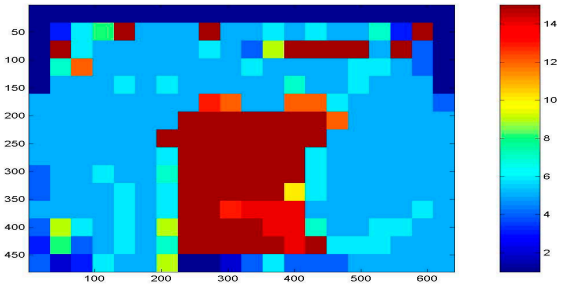

Figure 8: Resulting blur map.

7. CONCLUSIONS

We have presented a new method to compute a depth map for an arbitrary scene. The two main components of our approach are 1) the generation of the all in focus picture using data fusion and 2) the scale space representation of the all in focus picture (the family of derived signals with increasingly reduced detail). Block matching was then employed to determine the blur present at different locations in the captured picture. Experimental results were then provided to demonstrate the efficacy of the approach. To the best of our knowledge, this is the first time in the literature that scale space theory has been used to compute blur/depth maps.

The author would like to thank Makibi Nakamura for introducing him to this research topic. The author would like to recognize the valuable comments, suggestions and insights provided by Mr. Nakamura. The author would also like to acknowledge the assistance of Hidenori Kushida for developing the virtual camera simulation software.

REFERENCES

[1] T. Darrell and K. Wohn, "Pyramid Based Depth from Focus", Proceedings CVPR, pp. 504-09, June 1988.

[2] M. Subbarao and J. Tyan, "Selecting the Optimal Focus Meausre for Auto focusing and Depth From Focus", IEEE Trans. PAMI, Vol. 20, No. 8, pp. 864-70, August 1988.

[3] A.P. Pentland, "A New Sense for Depth of Field", IEEE Trans. PAMI, Vol. 9, No. 4, pp. 523-531, July 1987.

[4] Eugene Hecht, Optics, Addison Wesley, Reading, Mass., 1987.

[5] K. Kodama, K. Aizawa and M. Hatori, "Iterative Reconstruction of an All Focused Image by Using Multiple Differently Focused Images", ICIP Vol. 3, pp. 551-54, 1996.

[6] H. Li, B.S. Manjunath and S.K. Mitra, "Multi-Sensor Image Fusion Using the Wavelet Transform", ICIP, Vol. 1, pp. 51-55, 1994.

[7] T. Lindeberg, *Scale Space Theory in Computer Vision*, Kluwer Academic, Boston, Mass., 1994.

[8] P. Perona and J. Malik, "Scale Space and Edge Detection Using Anisotropic Diffusion", IEEE Trans. PAMI, Vol. 12, No. 7, pp. 629-39, June 1990.

[9] Personal correspondence with Makibi Nakamura.

[10] A.N. Rajagopalan and S. Chaudhuri, "An MRF model based approach to simultaneous recovery of depth and restoration from defocused images", PAMI, Vol. 21, No. 7, pp. 577-89, July 1999.