Depth Recovery from Focus-Defocus Cue by Entropy of DCT Coefficient

Huadong Sun[1,2(✉)], Zhijie Zhao[1,2], Xiaowei Han[1,2], and Lizhi Zhang[1,2]

[1] School of Computer and Information Engineering, Harbin University of Commerce, Harbin 150028, China
kof97_sun@163.com
[2] Heilongjiang Provincial Key Laboratory of Electronic Commerce and Information Processing, Harbin University of Commerce, Harbin 150028, China

Abstract. Depth recovery for single image is very important to 2D-3D image conversion, which is a challenging problem in computer vision. The focus-defocus as an effective pictorial cue, has been paid more and more attention. In this paper, we reveal the relationship between entropy of DCT coefficient and scale parameter of PSF. Then, a new method to depth recovery for single images using focus-defocus cue is proposed, in which the entropy of DCT coefficient is regarded as the measure of blur, and linear operation mapping the level of blur to depth is adopted. The proposed method, which can generate pixel-level depth map, is unnecessary to select threshold. The experimental results indicate that the new method is reliable and effective.

Keywords: Depth · Entropy of DCT coefficient · Focus-defocus cue · Measure of blur

1 Introduction

Three-Dimension display is an important form of expression to image information in future. Compared with traditional media, it provides outstanding intuitive and real scene feeling, diversified and comprehensive media interaction ability for the audience. Recently, Depth-image-based- rendering (DIBR) technique is applied to some advanced 3D TV system, in which a new 3D data representation, including tradition 2D image and its associated depth map, is adopted [1], which is efficient for rendering, transmission and coding [2]. But how to realize the depth information extraction from 2D image is a key problem.

In traditional image-forming system, 3D scene lost its depth information after projected by 2D image sensors. The efficient solution is to utilize signal processing or computer vision techniques to extract the lost depth information through employing various cues, such as linear perspective, motion parallax, texture gradient, atmospheric scattering, etc. [3]. In recent years, as an important pictorial cue, focus-defocus has been paid more and more attention.

In 1994, Gokstorp presented multi-resolution local frequency algorithm which needs two images of a scene obtained from the same view-point but using different

M. Jia et al. (Eds.): WiSATS 2019, LNICST 281, pp. 167–176, 2019.
https://doi.org/10.1007/978-3-030-19156-6_16

aperture settings, where the difference in defocused blur between two images can be calculated by local frequency representations of the two images and the sub-sampled scale-space pyramids [4]. In 2001, polynomial system identification method was presented by Rayala, which modeling the underlying phenomenon of defocusing as a linear system and a two-dimensional equation error algorithm is developed to calculate the coefficient of parametric transfer function [5]. In 2009, Mendapara introduced the exponentially decaying algorithm based on SUSAN operator, where a sequence of images acquired at varying focus is necessary [6]. In 2012, a formulation of unscented Kalman filter for depth estimation was designed, which is suitable to both motion and defocusing blur without constraining the PSF as Gaussian function [7]. In 2016, an iterative feedback method is presented which is for the simultaneous estimation of depth by joint spatiotemporal optimization, which needs all-in-focus videos from a defocused video pair [8]. In the same year, Xiao presented a method of multi-focus image fusion based on the depth recover, where the optical imaging of two multi-focus images can be simulated as the heat equations of positive regions, and the scene depth information is calculated by inhomogeneous diffusion equations [9]. Although providing good depth estimation, it is too restrictive that these algorithms require several images or videos to the same or similar scene screened by the different optical parameters.

Because of the multi-scale and multi-resolution characteristics, wavelet transform is used to estimate the level of blur in 2003, in which wavelet decomposition of macro-blocks within an image was carried on to analyze the high frequency information of that macro-block and the number of high-value wavelet coefficients was defined as the measure of blur [10]. However, how to select the threshold of wavelet coefficients was not mentioned in this method.

In this paper, a novel method is proposed, where the entropy of DCT coefficient is adopted as the measure of blur. In contrast to count number of high-value wavelet coefficients as a measure of blur in [10], entropy of DCT coefficient is unnecessary to select threshold and more effective.

2 Background

Focus-defocus is a significant cue to extract depth information for the single picture. Normally, the defocus phenomenon will happen when the object is not in the focal plane of the scene. The longer the distance between the focal plane and objects is, the worse the blurring is, and more smoothed objects' texture is. So, the level of blur is correlative to the depth of the objects.

The blur diameter can be referred as the intuitional measure to the level of blur. Suppose the background object defocus while the foreground object focus, the optical image-forming model of camera can be illustrated as Fig. 1. The parameters are defined as follows: f is the focus length, L is the optical lens aperture, p is the distance between focal plane and lens, q is the distance between imaging plane and lens, z is the distance between object and lens which is equivalent to the true depth value, and v is the image distance of object. The point object located at p will focus as one point in the imaging

plane. But the point object which located at z will defocus as a blur circle in the image-forming plane. Apparently, larger diameter d of blur circle is, more seriously blur phenomenon happens.

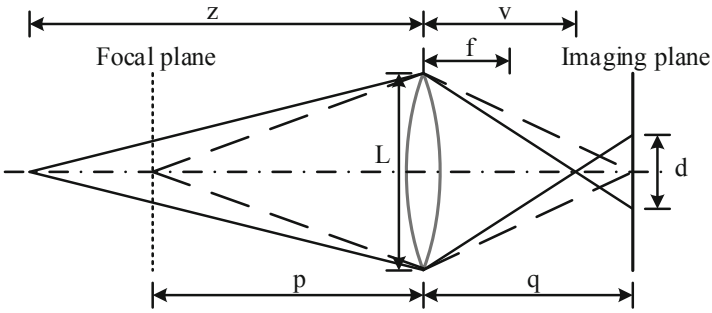


Fig. 1. The optical imaging model of camera.

The following equations can be obtained from the theory of optical lens image-forming and geometry relationship:

$$\frac{1}{p}+\frac{1}{q}=\frac{1}{f},\frac{1}{z}+\frac{1}{v}=\frac{1}{f},\frac{d}{L}=\frac{q-v}{v} \tag{1}$$

Through (1), the relationship between blur diameter (blur circle's diameter d) and true depth value z, can be calculated as

$$d=\frac{Lpf}{p-f}(\frac{1}{p}-\frac{1}{z}) \tag{2}$$

Obviously, when the true depth z increases, the blur diameter d will augment which indicates the blur phenomenon becomes worse.

Generally speaking, when the picture is not calibrated, the true depth can't be got because parameters $l, p.f$ are unknown. But the relative location between object and lens can be recovered, which means the relative depth. The relative depth is also enough to 2D-to-3D image conversion.

The mean square deviation of point spread function (PSF) can be regarded as the other measure to the level of blur. A 2D image point of a given object point can be regarded as the sum of the contributions of each point belonging to the surface in the neighbour of that 3D point which reflects light to the camera. 2D convolution operation can model above process. The observed defocused image $I'(x,y)$ can be described as the 2D convolution between PSF $g(x,y)$ and the ideal focused image $I(x,y)$.

$$I'(x,y)=I(x,y)\otimes g(x,y) \tag{3}$$

The PSF, which is determined by camera parameters and blur diameter, can be modeled as a 2D Gaussian function approximately,

$$g(x,y) = \frac{1}{2\pi\sigma_s^2}\exp(-\frac{x^2+y^2}{2\sigma_s^2}) \tag{4}$$

where the scale parameter σ_s of Gaussian function satisfies $\sigma_s = kd$, d is blur diameter, and k is the constant depended on the camera parameters. Thus, the larger value of σ_s indicates the more serious defocusing blur and the farther distance between lens and the object which can be regarded as relative depth.

3 Measuring the Blur by Entropy of DCT Coefficient

To an image, low frequency information is defined as small or slow gray value variation, such as the plain area of images. On the contrary, high frequency information is considered as a transient section which contains sharp or fast amplitude variation, such as edge and texture of images. Intuitively, to the same texture or edge, the focused image will have more high frequency component and less low frequency component than the defocused image. So, the more serious blur indicates less high frequency information and more low frequency information. It suggests that the frequency spectrum can be adopted to measure the level of blur.

3.1 Entropy of DCT Coefficient

Discrete Cosine Transform, especially type-II DCT, is widely used in signal processing and image compression. It profits from the strong characteristic, which most energy of natural signals (including sound and image) is concentrated at the low-frequency coefficients of DCT.

Let $f(m,n)$ denotes a pixel's gray value of the image, it's size is $N \times N$, then the type-II DCT of image can be described as.

$$A(k,l) = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} f(m,n)[a(m)cos\frac{\pi(2m+1)k}{2N}][a(n)cos\frac{\pi(2n+1)l}{2N}] \tag{5}$$

where

$$a(n) = \begin{cases} \sqrt{\frac{1}{N}}, & n = 0 \\ \sqrt{\frac{2}{N}}, & 1 \le n \le N-1 \end{cases}.$$

The frequency of corresponding cosine kernel will rise with the increase of k, l, and the DCT coefficient $A(k,l)$ can be considered as the map of image signal to the cosine kernel whose frequency increases. Therefore, the DCT coefficient can reflect the spectrum of image from low frequency to high frequency which can be used to measure the level of blur.

On the other hand, we can observe that $A(k,l)$ is the cosine weighted summation of $f(m,n)$ from (5). Suppose all $f(m,n)$ submit the identical distribution in different m, n, then as the weighted sum of random variables with the identical distribution, $A(k,l)$ can approximately submit to Gaussian distribution according to the central limit theorem. Although each $f(m,n)$ are spatial correlated, the central limit theorem can apply as long as the magnitude of correlation is less than 1. The correlation of typical image, is not too large and there are enough pixels to obtain a good approximation to Gaussian distribution. Moreover, because of the unitary nature of DCT, the mean of Gaussian distribution is zero. So, the DCT coefficient $A(k,l)$ submits a zero-mean Gaussian as follows.

$$p(A) = \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{A^2}{2\sigma^2}) \tag{6}$$

where σ^2 is variance, and A is $A(k,l)$ for short. Some examples about the distribution of DCT coefficient are shown in Fig. 2.

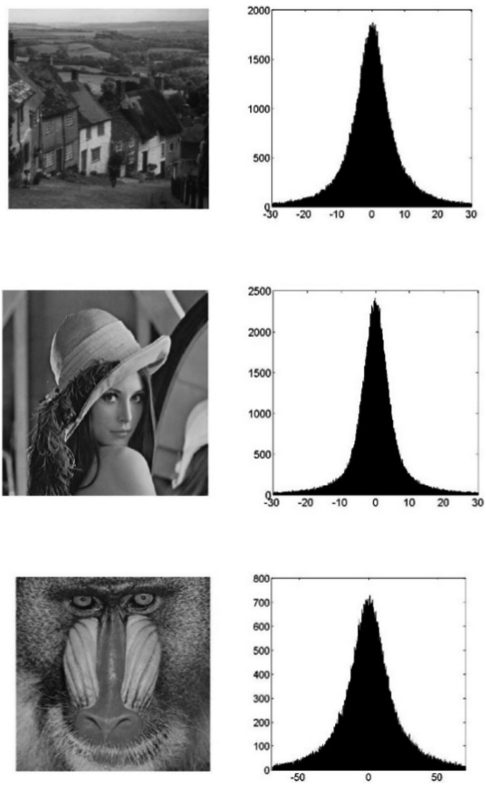


Fig. 2. Some standard images and their DCT coefficient histograms.

Thus, entropy of DCT coefficient can be obtained by

$$\begin{aligned} H &= -\int_{-\infty}^{+\infty} p(A) ln\, p(A) dA = -\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{A^2}{2\sigma^2}) \cdot [-ln(\sqrt{2\pi}\sigma) - \frac{A^2}{2\sigma^2}] dA \\ &= [\frac{1}{2} + ln(\sqrt{2\pi}\sigma)] \cdot \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{A^2}{2\sigma^2}) dA + \frac{1}{\sqrt{2\pi}\sigma} \cdot \frac{A}{2} \cdot exp(-\frac{A^2}{2\sigma^2})|_{-\infty}^{+\infty} \\ &= ln(\sqrt{2\pi e} \cdot \sigma) \end{aligned} \tag{7}$$

It can be seen obviously that the entropy of DCT coefficient is decided by σ only. The following proof will indicate the decreasing relationship between mean square deviation σ of DCT coefficient and the scale parameter σ_s of Gaussian PSF.

3.2 Relationship Between Entropy of DCT Coefficient and Level of Blur

The relationship between the level of blur and entropy of DCT coefficient can also be verified by the following simulations. Figure 3(a) is the original focused picture. After defocusing the original picture according to (3) and (4), Fig. 3(b) is defocused result of $\sigma_s = 1.5$, and Fig. 3(c) is that of $\sigma_s = 3$. Compared Fig. 3(b) with (c), it shows that the more value of σ_s is, the more serious blur happens. Figure 4 shows the decreasing characteristics between the mean square deviation of DCT coefficient and the scale parameter of PSF. Figure 5 illustrates the relationship between entropy of DCT coefficient and scale parameter of PSF, from which it can be shown obviously that entropy of DCT coefficient will decrease when more serious blur phenomenon happens.

(a) original picture

(b) result of σ_s=1.5

(c) result of σ_s=3

Fig. 3. Picture 'Lena' and its defocused results.

As the data discrete, another worth to mention is the calculation of DCT coefficient's entropy. Denote c_{max} and c_{min} as the maximum and the minimum of DCT coefficients, and the number of coefficients is K. Firstly, dividing the whole interval $[c_{min}, c_{max}]$ into several sub-interval with step length 0.1, and denoting all sub-intervals

with order number i. Secondly, calculating the number K_i of coefficients whose values locate at the ith sub-interval. Thus, the entropy of DCT coefficient can be obtained by

$$H = -\sum_i P_i \cdot log\, P_i = -\sum_i \frac{K_i}{K} \cdot log(\frac{K_i}{K}) \tag{8}$$

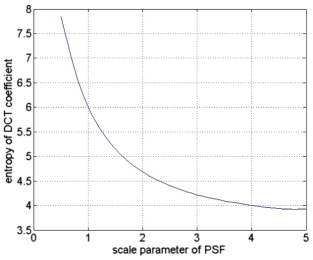


Fig. 4. Relationship between mean square deviation of DCT coefficient and scale parameter of PSF.

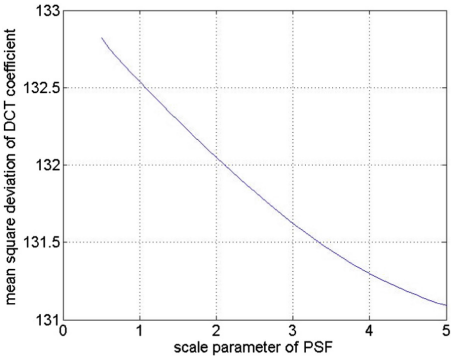


Fig. 5. Relationship between entropy of DCT coefficient and scale parameter of PSF.

4 Proposed Method for Depth Recovery

In this section, the proposed method will employ the entropy of DCT coefficient as the measure of blur, and adopt the linear operation mapping the blur into the depth. The algorithm flowing and linear mapping will be discussed in detail.

4.1 Outline of Algorithm

The entropy of DCT coefficient is adopted as the measure of blur in the presented method, whose flow is as follows.

Step 1: To each pixel (i,j), select the local window with size $N \times N$ from its neighbourhood, then carry on 2D DCT to this local window whose center is pixel (i,j). Generally, N is odd number.
Step 2: Calculate the entropy h of DCT coefficient according to (8) as the measure of blur in the pixel (i,j).
Step 3: Repeat Step 1 and Step 2, till every pixel of the image is traversed. Then the pixel-level depth map can be recovered after mapping the entropy h into depth with linear operation.

4.2 Linear Mapping

In our method, the large value of h denotes the front focused objects while the small value represents the defocused background. In order to generate the depth map with 256 gray-level, linear operation is necessary, which can be described as

$$D(i,j) = 255 \cdot \frac{h(i,j) - X_{min}}{X_{max} - X_{min}} \tag{9}$$

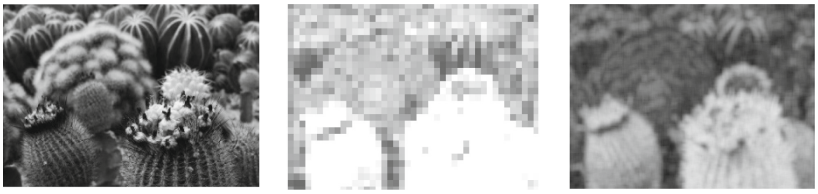
where $X_{max} = max\{h(i,j)\}$, $X_{min} = min\{h(i,j)\}$, and $D(i,j)$ is the depth of pixel (i,j).

5 Experimental Results and Analysis

In the Windows 7 system, Matlab2016 can be used to make the following experience. In our experiments, we select the local window's size as $N = 17$. Comparison of the depth maps between proposed method and algorithm in reference [10] is carried on. Figure 6(a) is the original image. Figure 6(b) is the depth map generated by the method in reference [10], in which white blocks are in front while black blocks are behind. Because the algorithm in reference [10] is the block-divided wavelet method, its initial depth map is blocky. Figure 6(c) is that of proposed method. We can observe distinctly that the depth map adopting the proposed pixel-level algorithm preserves more details.

(a) original picture (b) depth map of [10] (c) depth map of our method

Fig. 6. Picture 'leaf' and its depth map.



(a) original picture (b) depth map of [10] (c) depth map of our method

Fig. 7. Picture 'cactuses' and its depth map.

Another comparison is shown in Fig. 7. From Fig. 7(b), it can be seen that the cactuses can hardly be identified because of the series blocky effect.

Figure 8 illustrates other images and there depth maps generated by proposed algorithm in this paper. The left column shows the original images while the right column shows the corresponding depth maps, which demonstrate that the method presented here is reliable and robust.

Fig. 8. Other pictures and their depth maps.

6 Conclusion

In this paper, we propose a depth recovery algorithm for single images based on focus-defocus cue, which is enough for 3D rendering and stereo applications. Because of the decreasing relationship between entropy of DCT coefficient and scale parameter of PSF, this proposed algorithm employs the entropy of DCT coefficient as the amount of blur, and adopts the linear process mapping the blur into depth. The experimental result shows that this reliable method does a good job in depth recovery. The further work is to improve the accuracy of depth recovery combined with other monocular cues for single image.

Acknowledgement. This work is supported by the Harbin Science and Technology Bureau outstanding subject leader fund project (2017RAXXJ055), and Nature Science Foundation of Heilongjiang Province (F2018020).

References

1. Fehn, C.: A 3D-TV approach using depth-image-based rendering (DIBR). In: Image Process, pp. 482–487 (2003)
2. Liu, Y., Wang, J., Zhang, H.: Depth image-based temporal error concealment for 3-D video transmission. Circuits Syst. Video Technol. **20**, 600–604 (2010)
3. Zhang, L., Knorr, S.: 3D-TV content creation: automatic 2D-to-3D video conversion. Broadcasting 1–12 (2011)
4. Gokstorp, M.: Computing depth from out-of-focus blur using a local frequency representation. In: Pattern Recognition-Conference A: Computer Vision & Image Processing, vol. 1, pp. 153–158 (1994)
5. Rayala, J., Gupta, S., Mullick, S.K.: Estimation of depth from defocus as polynomial system identification. Image Signal Process. **148**, 356–362 (2001)
6. Mendapara, P., Minhas, R., Wu, Q.M.J.: Depth map estimation using exponentially decaying focus measure based on SUSAN operator. In: Systems, Man and Cybernetics, pp. 3705–3708 (2009)
7. Paramanand, C., Rajagopalan, A.N.: Depth from motion and optical blur with an unscented Kalman filter. IEEE Trans. Image Process. **21**, 2798–2811 (2012)
8. Lin, X., Suo, J., Dai, Q.: Extracting depth and radiance from a defocused video pair. IEEE J. Mag. **25**(4), 557–569 (2016)
9. Xiao, J., Liu, T., Zhang, Y., Zou, B., Lei, J., Li, Q.: Multi-focus image fusion based on depth extraction with inhomogeneous diffusion equation. Signal Process. **14**(1), 1–30 (2016)
10. Valenecia, S.A., Rodriguez-Dagnino, R.M.: Synthesizing stereo 3D views from focus cues in monoscopic 2D images. In: Proceedings of SPIE-IS&T Electronic Imaging, vol. 5006, pp. 377–388 (2003)