Depth Estimation from a Single Image Based on Gradient and Wavelet Analysis

Lixin He*[1], Zhi Cheng [1], Jing Yang [1], and Bin Kong [2]

[1] *School of Artificial Intelligence and Big Data, Hefei University, Hefei Anhui 230601, China*
[2] *Hefei Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei Anhui 230031, China*
hlxiniim@mail.ustc.edu.cn, cz_ganen108@126.com, jyang@iim.ac.cn, bkong@iim.ac.cn

Abstract
Recovering the depth information from a single image is a fundamental problem in computer vision field and has broad application prospects. To solve the problem that the accuracy of the depth recovered from a single image is not high enough, especially when there are several edges very close or intersecting, or when the edge is weak, a novel method to depth measurement is proposed in this article. Four steps are included in our method. Firstly, we obtain depth value of object edge point indirectly by measuring the defocusing degree of the object edge point. The gradient information of the 8 directions of edge point are employed during the process of measuring. Secondly, wavelet analysis is used to judge whether the measured depth value needs to be corrected or not. If necessary, it is corrected according to our formula. A sparse depth map is got when the depth values of all of edge points are measured and are corrected if necessary. Thirdly, the joint bilateral filter is employed to refine the sparse depth map. Lastly, the sparse depth map is extended to a dense depth map by the method of Matting Laplacian. The results of experiments show our method is effective.

Keywords: *Depth estimation, Dense depth map, Sparse depth map, Wavelet analysis.*

1 INTRODUCTION

When we take a scene with a camera, we only get a two-dimension image, and the depth information of the scene is lost. It causes machines could not perceive the information of distance and size of the object, and the speed of the moving objects in the scene, and so on. Therefore, recovering depth information from two-dimensional images is a fundamental problem in the field of computer vision. And it has become one of the research hotspots. [1] [3] [4] [15]

At present, many scholars use deep learning methods to estimate image depth information. There are two main categories: supervised learning and unsupervised learning. Supervised learning methods [1] [4] [15] require a lot of ground truth for deep data during training, and these ground truths are often difficult to obtain and costly. Unsupervised learning methods [2] [6] do not require ground truth for depth data, but they need to take at least one pair of images using two cameras which are placed on the same horizontal plane. And disparity d is found from the image pairs. Then d is combined with deep learning method to find the depth information from the monocular input image. But the requirements of the two cameras limit its application in many cases.

Motion parallax, linear perspective, occlusion, texture, shadow, and defocus, etc. clues can be used to obtain depth information [8] [10]. One or more of the first five clues only exist in the images of special scenes: there are relative motion that produces motion parallax, and there are vanishing line or vanishing point can be detected, and there are occlusive relation, rich texture and light and dark change in image, etc. However, defocus clues are ubiquitous in every image. Therefore, recovering depth based on defocus clue has a wider application field.

Pentland et al. [11] proposed a method for recovering depth information from planar images using defocus cues. Some researchers [5] [13] [14] subsequently improved the results, but all of those methods need to get two or more images of the same scene with different camera internal parameters. Zhuo et al. [17] proposed a method for obtaining relative depth information from a single defocus image. This method has a broader application prospect. Because all of the input data of this method is just an image.

B. Fox et al. (Eds.): IC-ICAIE 2022, AHCS 9, pp. 849-857, 2023.
https://doi.org/10.2991/978-94-6463-040-4_129

Therefore, the Zhuo's algorithm [17] has attracted many researchers to join the ranks of recovering depth information from a single image. For example, the literatures [9] [16] propose different methods to improve the accuracy of depth information, and the better results have been got. But there is still room for further improvement, especially the measurement accuracy at the weak edge points of the objects in image needs to be further improved. In view of this, we propose a novel method to improve measuring accuracy.

The rest of this article is organized as following: Section II introduces the basic principle on the measurement of the defocusing radius (or depth). Section III A, we describe the overview of our method. Section III B, the algorithms of the measurement of defocus radius R of edge points is presented. In the algorithm, the gradient information of the 8 directions of the object edge point are employed during the process of measuring in order to minimize the measurement error caused by image noise, etc. Section III C, we focus on how to judge the measurement value of the R need to be corrected, and how to correct it. Section III D, how to acquire a dense depth map is described; Section IV and V is Experiments and conclusion, respectively.

2 BASIC PRINCIPLE

The basic principle of the convex lens imaging shows that a clear image is formed when the point light source has a focal point that falls on the imaging plane after a convex lens, otherwise the point light source in the imaging plane is a defocus circle with a radius of R. As shown in Fig.1. The relationship between the u (object distance, namely depth) and the R can be expressed as follows:

$$\frac{1}{u} = \frac{1}{f} - \frac{1}{v_0} - \frac{1}{rv_0}R \tag{1}$$

where f is the focal length, and v_0 is the distance between the lens center and the imaging plane, and r is the aperture radius. The formula (1) can be derived according to the Fig.1., basic principles of optical imaging and knowledge of triangular geometry.

For clarity, the camera parameters f, r and v_0 are given, and we plot the relationship between R and u, as shown in Fig.2. Obviously, the R is a non-linear monotonically increasing function of u. Therefore, it is reasonable that we substitute defocus radius R for depth u.

Obviously, if the camera parameters are given, then we can get the value of depth (namely absolute depth) by using the R. If they are not given, we can distinguish between near and far (namely relative depth) by using the R.

The defocus edge point $d(x)$ in the image can be regarded as the convolution of the focused edge point $f(x)$ and the Gauss point spread function $g(x, \sigma)$, that is:

$$d(x) = f(x) \otimes g(x, \sigma) \tag{2}$$

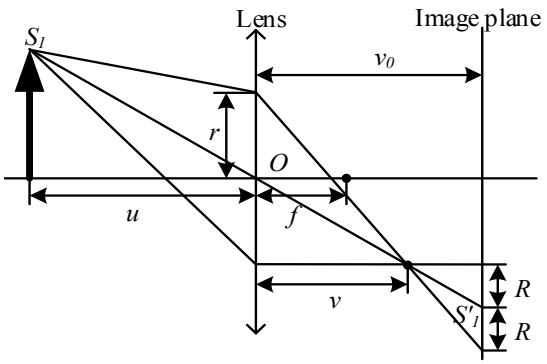


Figure 1: A diagrammatic sketch of convex lens imaging.

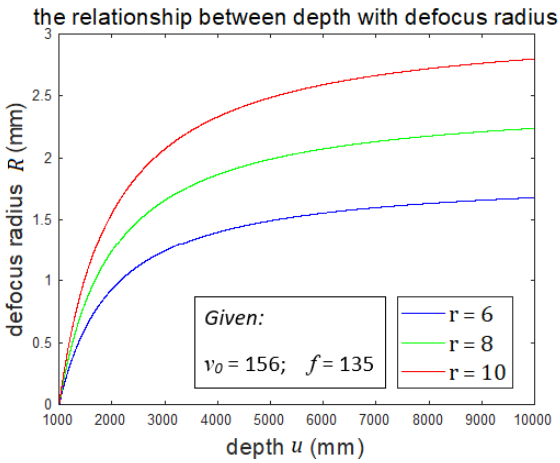
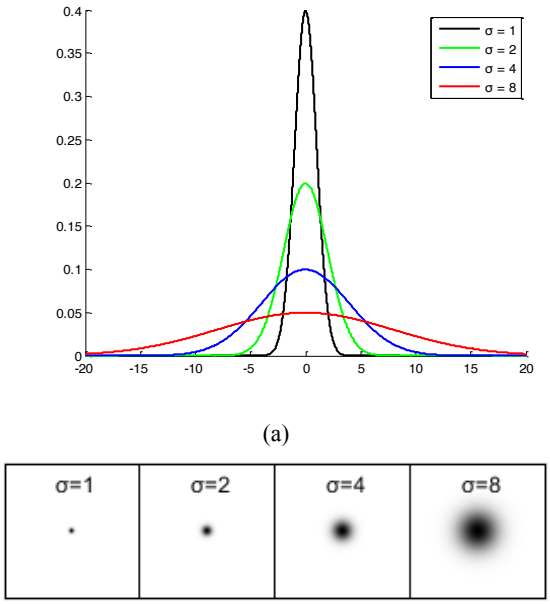


Figure 2: The relation between depth u and defocus radius R.

The relationship between σ and R is as follows: $\sigma = kR$, k is a real number [5] [8]. The Fig.3. (a) shows four Gaussian distribution function curves with different variances σ. Fig.3.(b) is a computer simulation images of the defocusing image of point light sources corresponding to different σ.

It can be seen from (b) that the larger σ is, the larger the defocus radius R is. That is to say, the grayscale of the image changes more slowly from the center of the point light source to each direction, and the corresponding gradient value also is smaller. Therefore, the defocusing degree can be measured by the gradient information at the edge point of the object in the image. However, if there are several edges crossing or very close to each other, or when the edges are weak, the defocusing radius value measured by this method is smaller than that of the actual value, so the measured value R should be corrected.



(a)

(b)

Figure 3: (a) Four Gaussian distribution function curves with different σ; (b) is a computer simulation of the defocusing image of point light sources corresponding to different σ.

3 OUR METHOD

3.1 Overview of Our Method

The According to the above basic principles, the idea of this article is as follows:

First, the original image is transformed into a gray image, then the edges of the objects are detected from the gray image by using the Canny method, and the gradient image of the gray image is obtained, and wavelet transformation is performed on the gray image.

Secondly, the defocus radius of the edge points R is calculated by using its gradient information.

Thirdly, wavelet analysis is used to determine whether the R needs to be corrected or not. If necessary, it is corrected by our formula. A sparse depth map is got when the R of all of the edge points are measured.

Fourthly, the sparse defocus image is refined by the JBF (Joint Bilateral Filtering) [7] [17].

Finally, the refined sparse depth map is extended to the dense depth map by using the Matting Laplacian method [12] [17].

The overview of our method is shown in Fig.4.

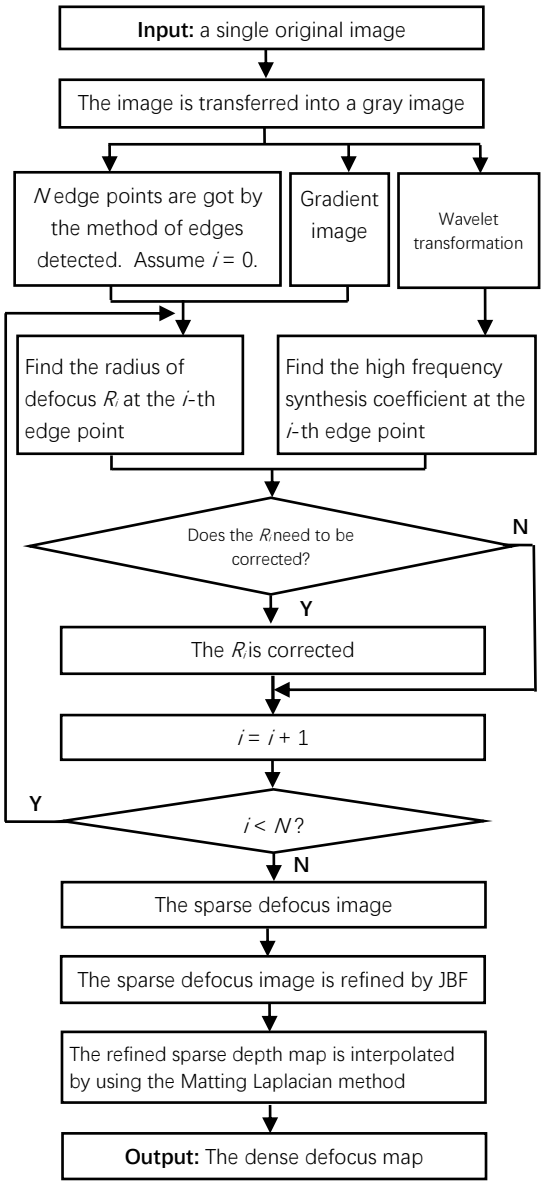


Figure 4: The overview of our method.

3.2 Measuring Defocus Radius

We assume that (x_i, y_i) is the i-th edge point p. In theory, its defocus radius R_i can be measured along any direction of the edge point. In order to minimize the measurement error caused by image noise, etc. In this article, firstly, we measure the defocus radii along the eight directions d_j ($j = 1, 2, \ldots, 8$), respectively, as shown in Fig.5. And we get the 8 values of defocus radius: $R_{i,1}$, $R_{i,2}$, $R_{i,3}$, ……, $R_{i,8}$,. Then the median value of $R_{i,j}$ ($j = 1,2,\ldots,8$) is taken as the measurement value of defocus radius of the i-th edge point p.

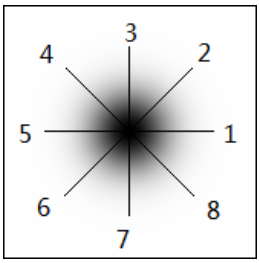


Figure 5: The 8 directions for measuring the defocus radius.

Assume that T_L is the lower threshold coefficient, T_H is the upper threshold coefficient, and G_p is the gradient value at the i-th edge point p, and the p' is the nearest point of p along the direction d_j ($j = 1, 2, ...,8$). ***Algorithm 1*** shows how to get the R_i of the i-th edge point p. In general, the value of R_i can be accurately measured using *Algorithm 1*.

Algorithm 1: Algorithm for measuring defocus radius

Input: an edge point $p(x_i , y_i)$, and a gradient image, ($i = 1, 2,..., N$).

Output: the defocus radius R_i of the i-th edge point $p(x_i , y_i)$

```
% Defocus radii are measured along the 8 directions in turn.
for j = 1 : 8
    p0 = p;   % Save the p point coordinates.
    Find the nearest point p' along the direction dj.;
    while(p' is in the image range) && ( p' is not the
          one of edge points) && (TLGp≤Gp'≤THGp)
        Ri,j = Ri,j + 1;
        p = p';
        Find the nearest point p' of p along the dj.;
    end
% Regain p coordinates to measure Ri,j+1 along the dj+1
    p = p0;
end
% Assign intermediate values to Ri.
Ri = median(Ri,1 , Ri,2 ,..., Ri,8);
```

The *Algorithm 1* shows that it may be lead to the measured value $R_{i,j}$ is less than its actual value when one of the three following cases occurred.

Case 1: The p is very close to one of the image four boundaries. This case may be lead to the coordinates of p' exceed the image range. It means that the value of (p' is in the image range) is *false.*

Case 2: There are several edges very close or intersecting. In this case, the p' may be an edge point. It means that the value of **(**p' is not the one of edge points**)** is *false*.

Case 3: The edge is a weak edge. In this case, the difference between the object and its background color is small. It may be lead to the $G_{p'}$ is smaller than $T_L G_p$. It means that the value of **(**$T_L G_p \le G_{p'} \le T_H G_p$**)** is *false*.

When any of the above three cases occur, the specified case of the while statement is false. Namely, **(**p' is in the image range**) && (**p' is not the one of edge points**) && (**$T_L G_p \le G_{p'} \le T_H G_p$**)** is *false*. Then the programing will be jump out of the while loop, and the defocus radius along the next direction d_{j+1} is measured immediately. It means that the $R_{i,j}$ defocus radius along d_j is truncated. Therefore, the R_i is less than its actual value.

The method that median value of $R_{i,j}$ ($j = 1,2,...,8$) is taken as the measurement value make *Algorithm 1* still work well if the case 1 and 2 does not occur more than three times. But the method does not work to the case 3.

From the above analysis, we can see: Although the measurement error can be avoided to some extent by the method that the intermediate value of defocus radii along 8 directions is taken as the R_i, the method doesn't work when the above case 1 and 2 are occurred in many directions at the same time or the case 3 occurred. It will seriously affect the measurement accuracy, and the so it must be corrected.

3.3 Correction

When the above adverse cases occurred, the R_i is less than or even far less than its actual value. We have found that the measured value R_i is very small ($R_i \le 3$) when the adverse cases occurred by many times experiments. Therefore, when the R_i is very small, we must judge whether it should be corrected or not.

If the R_i is very small and it is much smaller than its actual value, the image around the point p will be blurred, correspondingly, the high frequency signal at point p is less. On the contrary, if R_i is very small, but it is the reflection of the actual situation, the image around the point p will be clear, correspondingly, the high frequency signal at p is more. Therefore, we can transform the problem of the spatial domain into the frequency domain to analyze and solve.

Wavelet transform is a very useful tool for signal process in frequency domain. So we can use it to solve our problem. The wavelet transform coefficients can be used to judge whether the R_i needs to be corrected or not.

In this article, only one-scale wavelet transform is need, because the high frequency signal is mainly concentrated in the coefficients of the one-scale wavelet transform. The horizontal, vertical and diagonal wavelet transform coefficients those are obtained by one-scale wavelet decomposition are represented by $W_H(x_i , y_i)$, $W_V(x_i , y_i)$, $W_D(x_i , y_i)$, respectively. Then, $W(x_i , y_i)$ which is called high frequency synthesis wavelet coefficients is defined as:

$$W(x_i, y_i) = (W_H^2(x_i, y_i) + W_V^2(x_i, y_i) + W_D^2(x_i, y_i))^{1/2} \quad (3)$$

R_t is the threshold. If formula (4) is satisfied, R_i should be corrected according to the degree of blur at the point

p. In other words, we correct it according to the high frequency synthesis wavelet coefficients $W(x_i , y_i)$. The higher the degree of blur, the bigger the correction coefficient K_W. W_{tmax} is an upper threshold. When $W(x_i , y_i) \geq W_{tmax}$, we think that the R_i need not to be corrected because the point (x_i , y_i) is clear enough. W_{tmin} is an lower threshold. When $W(x_i , y_i) \leq W_{tmin}$, The R_i need to be corrected by the maximum correction coefficient K_w.

We can suppose that K_w is inversely proportional to $W(x_i , y_i)$ when $W_{tmin} \leq W(x_i , y_i) \leq W_{tmax}$. Therefore, K_w can be calculated by Formula (5). Then, the precise measurement value r_i can be obtained by formula (6).

$$R_i \leq R_t \tag{4}$$

$$k_W = \begin{cases} 0 & W(x_i, y_i) \geq W_{tmax} \\ 1 & W(x_i, y_i) \leq W_{tmin} \\ (W(x_i, y_i) - W_{tmax})/(W_{tmin} - W_{tmax}) & else \end{cases} \tag{5}$$

$$r_i = (1 + k_C k_W) R_i \tag{6}$$

where, both W_{tmax} and W_{tmin} are threshold, and k_C is the coefficient for adjustment, and k_C is a real number, and $k_C > 0$.

A more accurate sparse depth map can be obtained by calculating the defocus radius of each edge point one by one and judging and correcting the measured value according to formula (4), (5) and (6). The specific process is shown in *Algorithm 2*.

Algorithm 2: Algorithm for acquiring sparse defocus/depth image

Input: an original image I_o.

Output: a sparse defocus/depth image I_s.

Step1 The original image I_o is transferred into a gray image I_G.

Step2 N edge points are detected from I_G, besides, the gradient image and the first layer wavelet transform coefficients are got, respectively.

Step3 Assume $i = 1$.

Step4 The R_i is obtained using the **Algorithm 1**.

Step5 if $R_i \leq R_t$

if $W(x_i, y_i) \geq W_{tmax}$

$k_W = 0$;

elseif $W(x_i, y_i) \leq W_{tmin}$

$k_W = 1$;

else

$k_W = (W(x_i, y_i) - W_{tmax})/(W_{tmin} - W_{tmax})$;

end

end

Step6 $r_i = (1 + k_C k_W) R_i$, go to Step8.

Step7 $r_i = R_i$

Step8 $i = i + 1$; if $i <= N$, then go to Step4, else go to Step9.

Step9 A sparse defocus/depth image I_s is obtained, and over.

3.4 Dense Depth Map

Noise may lead to inaccurate measurement of defocus radius, so the JBF [7] [17] is employed to refine the sparse defocus/depth image. Finally, the sparse defocus image is extended to a dense depth map by the method of Matting Laplacian [12] [17]. This method converts the problem of expanding sparse defocus image to dense depth map into a problem of minimum cost function. It can be defined as

$$E(S) = S^T L S + \lambda (S - \hat{S})^T D (S - \hat{S}) \tag{7}$$

where D is a diagonal matrix. D_{ii} is its elements. If the pixel i is an edge point, $D_{ii} = 1$, else $D_{ii} = 0$. And λ is the smooth coefficient. And L is the Matting Laplacian matrix. It can be defined as

$$L(i,j) = \sum_{k|(i,j)\in\omega_k} (\delta_{ij} - \frac{1}{|\omega_k|}(1 + (I_i - \mu_k)^T (\sigma_k + \frac{\varepsilon}{|\omega_k|} U)^{-1} (I_j - \mu_k))) \tag{8}$$

where ω_k is the slip window whose center point is the pixel k. $|\omega_k|$ is the size of the window ω_k, namely, the number of pixels in the window. δ_{ij} is the Kronecker delta. I is the input image, and the I_i is the colors at the pixel i. And ε is a regularization parameter. μ_k is the mean matrix of the colors in ω_k, and σ_k is the covariance matrix of the colors in ω_k. U is 3×3 identity matrix.

4 EXPERIMENTS

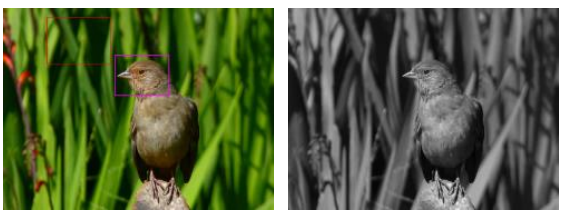
To verify the effectiveness of our method we carry out experiments and comparison. Fig.6. shows the result of the final experiment and the intermediate results of each step of our method. Fig.6.(a) is an original image which is captured by a common camera. The original image is transformed into a gray image that is shown in (b). (c) is an edge image which is obtained by processing the gray image with Canny algorithm. (d) is the gradient magnitude image of (b), and (e) is the results of one-scale wavelet transform of (b). (f) is the high frequency synthesis wavelet coefficients image that is defined by the formula (3). In order to clearer, (f) is shown by a color map instead of grayscale.

The red points in Fig.6.(g) indicate that theirs defocus measurement value need to be corrected by the *Algorithm 2*. For the convenience of observation, we select two representative regions from (g) to enlarge and show in (h) and (j), respectively. And the (i) and (k) is the corresponding regions from the original image (a).

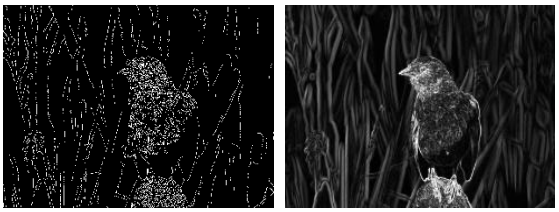
From Fig.6. (h) and (i), we can see that: the head of bird has a clear image, and the edge points on it are dense, close to each other or intersect, and the foreground and background color of the edge are very similar, but the defocus measurement values of most edge points measured by *Algorithm 1* do not need to be corrected. Because the area has a clear imaging, high signal frequency and large synthetic wavelet coefficients, which are larger than or far larger than W_{tmax}, and the correction coefficient k_W is 0 according to formula (5). In other words, no correction is needed. Certainly, it cannot excluded that very few edge points are labeled red due to noise and other factors, but these isolated points do not affect the measurement accuracy, because the adverse effects of these noise points will be eliminated by the subsequent joint bilateral filter in the filtering process.

Fig.6. (j) and (k) are the magnified image of the blue box in (g) and its corresponding magnified image in the region of the original image (a), respectively. From the (j) and (k), it can be seen that the image is blurred. When the foreground color is similar to the background color, the probability that defocus measurement values need to be corrected increase significantly. The yellow and pink boxes in (j) and (k) are the cases. However, if the color difference between the foreground and background is obvious, there is few defocus measurement values need to be corrected, and the region with green box in (j) and (k) is the case. In order to see clearly whether the defocus measurement value of each edge point need to be corrected, the region images of the yellow, pink and green boxes are enlarged and displayed in (l), (m) and (n), respectively.

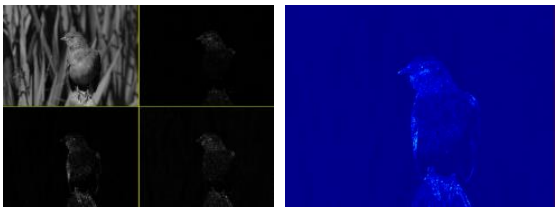
The following three points can be seen from the experiment. The first, when the image is blurred, the closer the object color is to its background color, the higher the probability that the defocus measurement value of the edge point needs to be corrected. The second, some of defocus measurement value of edge points need to be corrected when they are at the intersection of multiple edges or at the four bounds of the whole image. The third, when the image of the adjacent area of the edge point is very clear, few defocus measurement value of the edge point need to be corrected. All these are consistent with the previous theoretical analysis.



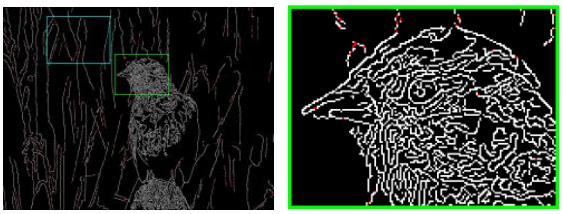
(a) (b)



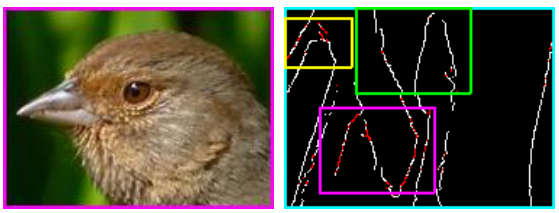
(c) (d)



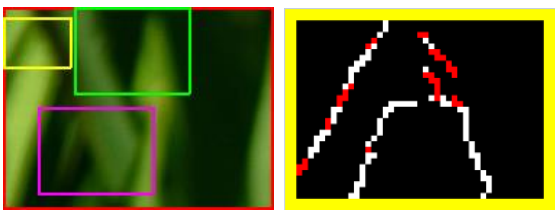
(e) (f)



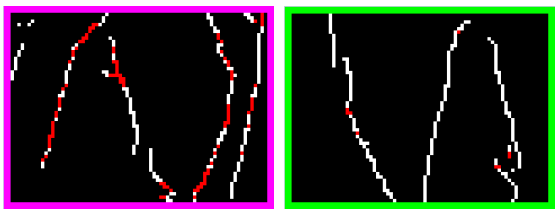
(g) (h)



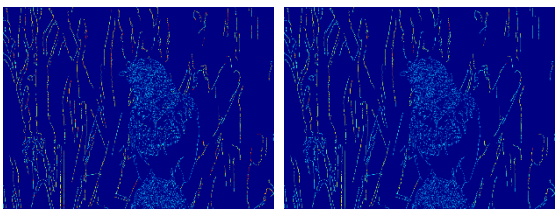
(i) (j)



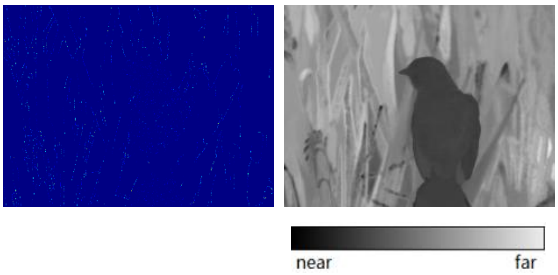
(k) (l)



(m) (n)



(o) (p)



(q) (r)

Figure 6: The result of the final experiment and the intermediate results of each step of our method. (a) An original image; (b) The gray image; (c) Result of edge detect; (d) The gradient image; (e) An one-scale wavelet transform; (f) The high frequency synthesis wavelet coefficients image; (g) Edge points marked red indicate that their measurement values need to be corrected; (h) The enlarged display of the image in the green box in (g); (i) The corresponding region image of (h) in (a); (j) The enlarged display of the image in the blue box in (g); (k) The corresponding region image of (j) in (a); (l) The enlarged display of the image in the green box in (j); (m) The enlarged display of the image in the yellow box in (g); (n) The enlarged display of the image in the pink box in (g); (o) The sparse depth/defocus image is obtained by the Algorithm 2; (p) The refined sparse depth/defocus image by JBF; (q) Difference image of (o) and (p); (r) The dense depth map is obtained by our method. The lower gray bar indicates the relative depth, and the smaller the gray value, the smaller the depth value.

The sparse depth/defocus image is obtained by the *Algorithm 2*, and it is shown as Fig.6.(o). Because of the influence of noise and other unfavorable factors, there are some measurement errors. In order to reduce these errors, the JBF is employed to filter the image (o), and the (p) is obtained. It can be seen that some noise has been filtered out in (p), which is much smoother than before filtering. The difference image is obtained by subtracting (o) and (p), as shown in (q). In order to clearer, we use the color maps instead of grayscale.

Fig.6. (r) is a dense depth map obtained by extending the sparse depth map (p) by Matting Laplacian method. The gray level of the image is used to represent the relative distance between the object in the image and the lens center of the camera. The smaller the gray value, the closer the pixel point is to the lens center of the camera. On the contrary, the larger the gray value, the farther is. It can be seen that the (r) accurately reflects the distance between the object in the image and the lens center of the camera.

In order to further verify the effectiveness of our method, several images with different real scenes are selected for our experiments. And all parameter are fixed in all of experiment: $W_{tmax} = 2.0$, $W_{tmin} = 0.5$, $R_t = 3$, $k_C = 2$, $T_L = 0.28$, $T_H = 1.2$. The results of our algorithm are compared with those of Zhuo's algorithms [17], as shown in Fig.7. Due to space limitations, only two scenes are shown. There are 2 columns in Fig.7. The (a) are the input images. (b) and (c) are the dense depth map obtained by Zhuo's algorithm and by our method, respectively. From Fig.7., we can see that the results of our algorithm are better than those of Zhuo's. Detailed description is as follows:

In the first column, the depth of the bird's head region and chest region are basically same, respectively. The result of our method well reflects the actual depth. But in the result of Zhuo's method, the depth of the bird's eyes are quite different from the surrounding parts. It is not consistent with the actual situation. The same problem occurs at the chest of birds in the result of Zhuo's method. In addition, there are serious measurement errors in the red box area in the result of Zhuo's method.

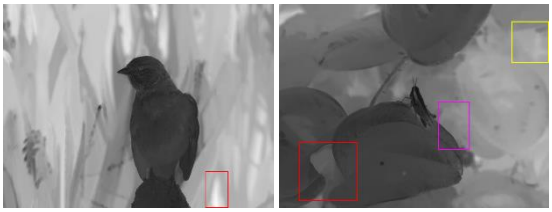
In the second column, the results of the two methods all reflect the actual depth well on the whole, but our method has better performance in dealing with some details. For example, our method better reflects the three different depths in the red-box labeled area, that is, the upper green leaves are the closest, and the lower green leave is a little farther, and the ground is the farthest. But the results of Zhuo's method is less precise than ours. Similarly, our method has higher measurement accuracy than Zhuo's method in the pink box and yellow box area.

By the comparison experiments and analysis, it can be seen that the method proposed in this article has higher depth measurement accuracy and higher practical value.

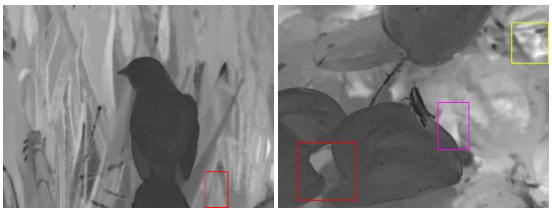
5 CONCLUSIONS

In conclusion, a novel depth measurement method is proposed in this article. Firstly, the defocus radius at the edge point of an object in a single defocusing image is measured by gradient information. We use the 8 directions gradient information and median value method in order to minimize the measurement error caused by image noise, image four boundary and several edges intersecting or very closing, etc. Then, wavelet analysis is used to judge whether the measured depth value needs to be corrected or not. If necessary, it is corrected according to our formula. Next, the JBF is used to obtain a higher accuracy sparse depth map. Finally, the sparse depth map is extended to dense depth map by interpolation with Matting Laplacian method. Our main contribution is that we proposed a novel effective method which use the gradient information and wavelet analysis to get a better sparse defocus depth map. The experiments show that the method proposed in this article has high measurement accuracy and practical value.

(a) Input images



(b) The results of Zhuo's method



(c) The results of our method

Figure 7: Comparison with Zhuo's method.

ACKNOWLEDGEMENTS

This work was supported in part by the University Outstanding Young Talent Visiting and Research Foundation of Anhui Province under Grant gxgnfx2019060, in part by the Talent Foundation of Hefei University under Grant 18-19RC28 and 18-19RC31, in part by the University Natural Sciences Research Project of Anhui Province under Grant KJ2017A541 and KJ2020A0659, in part by the Natural Science Foundation of Anhui Province under Grant 2008085QF290, in part by the National Natural Science Foundation of China under Grant 61304122 and 61672204, in part by the Outstanding Youth Talent Foundation of Hefei University under Grant 16YQ06RC.

REFERENCES

[1] Cao, Y. (2018). Estimating Depth From Monocular Images as Classification Using Deep Fully Convolutional Residual Networks, IEEE T CIRC SYST VID, vol. 28, no. 11, 3174-3182.

[2] Garg R., (2016). Unsupervised CNN for Single View Depth Estimation: Geometry to the Rescue, In Proc. ECCV, Netherlands.740-756.

[3] He, L., (2017). An Automatic Measurement Method for Absolute Depth of Objects in Two Monocular Images Based on SIFT Feature. Appl. Sci., vol. 7, no. 6, 517.

[4] Laina, C, (2016). Deeper Depth Prediction with Fully Convolutional Residual Networks, in Proc. 3DV, Stanford, CA, 239-248.

[5] Favaro P., (2008). Shape from defocus via diffusion, IEEE T PATTERN ANAL., vol. 30, no. 3, 518-531.

[6] Godard C., (2017). Unsupervised monocular depth estimation with left-right consistency, In Proc. CVPR. USA, 270-279.

[7] Levin, A., (2008). A closed-form solution to natural image matting," IEEE T PATTERN ANAL., vol. 30, no. 2, 228–242.

[8] Li C, (2015). Bayesian Depth-from-Defocus with Shading Constraints, IEEE T IMAGE PROCESS, vol. 25, no. 2,589-600.

[9] Liu S, (2016). Defocus Map Estimation From a Single Image Based on Two-Parameter Defocus Model, IEEE T IMAGE PROCESS, vol. 25, no. 12, 5943-5956.

[10] Palou G., (2013). Monocular depth ordering using T-junctions and convexity occlusion cues, IEEE T IMAGE PROCESS, vol. 22, no. 5,1926-1939.

[11] Pentland A. P. (1987). A new sense for depth of field, IEEE T PATTERN ANAL., vol. 9, no. 4, 523-531.

[12] Petschnig, G., (2004). Digital photography with flash and no-flash image pairs, ACM T GRAPHIC, vol. 23, no. 3, 664–672.

[13] Rajagopalan A N, (2004). Depth estimation and image restoration using defocused stereo pairs, IEEE T PATTERN ANAL., vol. 26, no. 11, 1521-1525.

[14] Subbarao, M., (1988). Depth Recovery from Blurred Edges, In Proc. CVPR, Ann Arbor, MI, USA, 498–503.

[15] Tao, T., (2018). Active depth estimation from defocus using a camera array, Appl. Opt., vol. 57, no. 18, 4960-4967.

[16] Zhang X, (2016). Spatially variant defocus blur map estimation and deblurring from a single image, J VIS COMMUN IMAGE R, vol. 35, no. 1, 257-264.

[17] Zhuo S, (2011). Defocus map estimation from a single image, PATTERN RECOGN, vol. 44, no. 9, 1852-1858.