Monocular defocus and texture cue based depth map estimation of 2d image

Vaibhav B Waman
M.E.(Signal Processing)
Jaihind College of Engineering,Pune

Dr.Rahul M Mulajkar
Assistant Professor
Jaihind College of Engineering,Pune

Abstract

Projecting 2D image and converting it into 3D scene is architecture of imaging. While with the conventional capturing of 2D images using camera has not covered depth information and to include depth information it is complex system with multiple cameras is necessary. Another approach to get this depth information is possible using monocular cues available into images. These cues could be possible to process and extract depth information from that. Here we are presenting methodology to get depth information using monocular cues such as defocus and texture. Depth information from defocus cue involves blurring of imaging more the blur more spatial extent of depth can be represented. Then we use another approach to get depth from texture information present in image. In this combined approach first we will calculate defocus using Gaussian model implementation. This mentioned as Sparse Defocus map. Using this sparse defocus output we create full defocus map. but this generation of defocus map has some ambiguity like hole in some regions. This issue we will further take into texture cues information and from that we generate again better defocus map. Integration of this two methods targets modifying errors present into some regions of defocus map from defocus cue. This earlier defocus map corrected using rules created during texture cues. Using region wise propagation method will increase accuracy and better depth map of an image.

Index Terms—*Defocus Estimation, Blur, Monocular Cues, Edge Detection, Superpixels*

I. INTRODUCTION

Perceiving 3D technology images and video is possible from monocular cues information extracted from 2D data. This cues which has important roles in the conversion of 2D to 3D having two major types which are monocular cues and binocular cues, in which binocular cues provides information using two eyes or two cameras. People worked on different approaches like Bae, Soonmin, Durand [15] and Zhuo, Shaojie and Sim [1] worked on monocular blur estimation for depth map, Cha Zhang, Zhaozheng Yin, Florencio [3] used motion parallax approach to get the 3D data from 2D images. Creation of 3D data using binocular cues or stereographic system which involves minimum of two cameras is complex and expensive system hence this cannot be affordable to all. Using supervised learning approach, Ashutosh saxena and Chung, Sung H and Ng, Andrew Y[3] created 3D images using

monocular cues like texture and haze. Oruc, Ipek and Maloney, Laurence T and Landy, Michael S [19] used combination of weighted depth maps from two different cues. With the help of multiple cues information depth map can be make more and more accurate for reconstruction of images. However learning of information requires complex architecture and time consuming algorithm so here we more focus more towards speed of algorithm using monocular cue defocus and texture information.

II. PROPOSED MODEL

We proposed method that involves using of defocus and texture cues to create defocus map. In first stage of operation we have create defocus map using blurring of images, here we have used defocus blur estimation using Gaussian model and then to propagate this defocus blur into entire image using integration method. In defocus blur estimation we have first edge detected into images from different edges of images blur is calculated. Then segmentation of image is done using available superpixel generation algorithm called as SLIC. Next stage of operation involves creation of texture energy using texture cues and to correct earlier created defocus map as earlier defocus map has some ambiguity present such as holes in depth map. This texture energy rules is corrected in this stage of operation. In second stage of operation involves optimization of earlier created depth map. The possible outcome of depth map from entire processing is shown in below example


Figure 1(a) Figure 1(b)

Figure 1(a) shows input image selected for processing and Figure 1(b) show relative color depth map of an image.

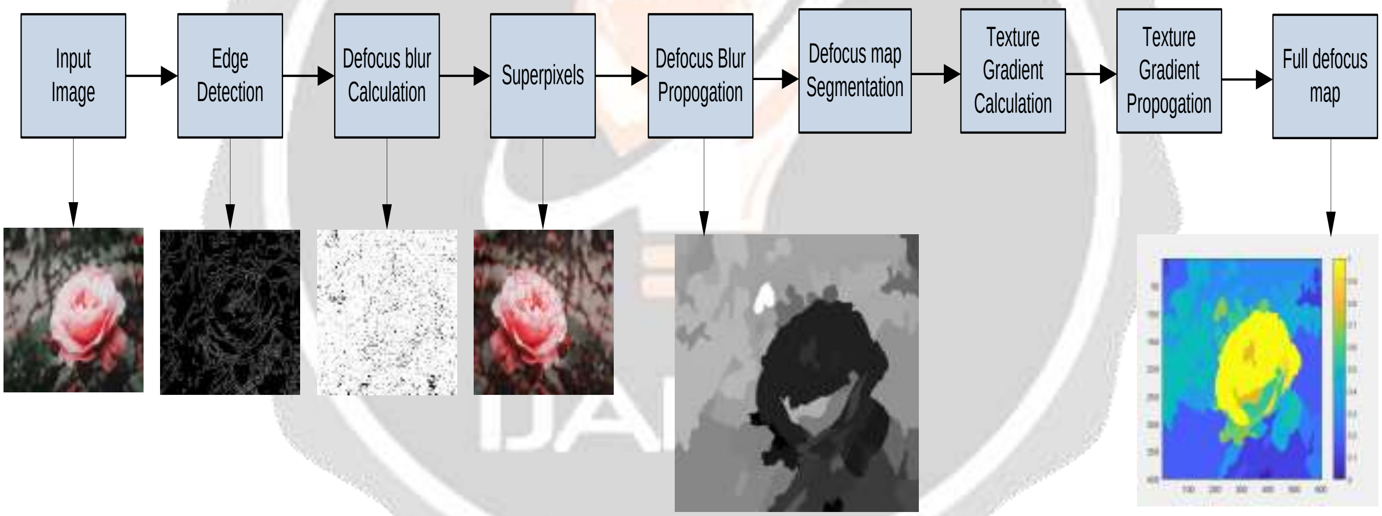

Figure 2 : Proposed depth map estimation model

Algorithm for Our Model

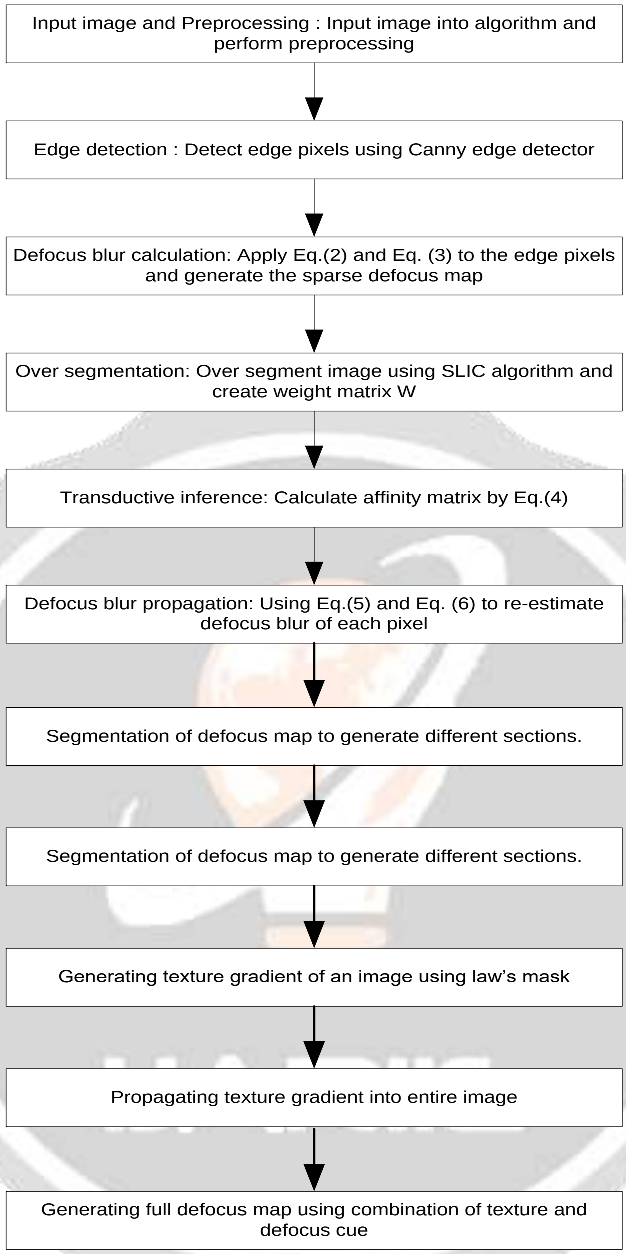

Figure 3 : Algorithm for our model

III. STAGE I : DEFOCUS CUE ESTIMATION

I. Defocus blur estimation

A. Edge Detection

As defocus blur is generated from image edge locations, here we applied edge detector on the image and result into set of E of edge pixels. We applied Canny Edge detection technique [9] and structure edge detector estimation method [10] for better performance and efficiency. Outputs of this edge detector have shown in below result section.

B. Defocus Blur Estimation

We have an edge function $e(x) = \alpha h(x) + \beta$, where α and β denote the amplitude and the offset of the edge respectively, h(x) denotes the step function, and x is a pixel location. The defocus blur can be modeled as a convolution of an edge pixels x with a Gaussian kernel $g(x, \sigma)$, where in standard deviation value σ gained from proportional value of circle diameter c.

Also blurred edge is hence defined by $b(x) = e(x) \otimes g(x, \sigma)$. In this case value of unknown standard deviation σ defined by blurriness of an edge pixel and could be used as to represent level of defocus blur which is on that edge pixel.

If we again re-blur given edge pixels using different Gaussian kernel, then in this case gradient of that re-blurred edge represented as

$$\nabla (b(x) \otimes g(x, \sigma_r)) = \nabla (e(x) \otimes g(x, \sigma) \otimes g(x, \sigma_r))$$

$$= \frac{\alpha}{\sqrt{2\pi(\sigma^2+\sigma^2 r)}} \exp\left(- \frac{x^2}{2(\sigma^2+\sigma^2 r)}\right) \quad \text{…... (1)}$$

Where σ_r denotes as standard deviation of re-blur Gaussian kernel. Zhuo and Sim [2] observed that gradient magnitude ratio R given in original blurred edge with the reblurred edge having maximum value at edge locations. Hence this, maximum value represented by

$$R = \frac{|\nabla b(x)|}{|\nabla b(x) \otimes g(x,\sigma r)|} = \sqrt{\frac{\sigma^2+\sigma^2 r}{\sigma^2}} \quad \text{……. (2)}$$

Thus, we calculate this unknown blur σ using gradient magnitude ratio defined as R at edge location with below equation by

$$\sigma = \frac{\sigma r}{\sqrt{R^2-1}} \quad \text{……. (3)}$$

Where σ_r already known value and R can be calculated from gradient of magnitude. Here that Eq. (3) is applicable for the edge locations only. Therefore, intermediate value outcome at current stage is a sparse defocus map on given edge pixels, this is shown in defocus blur model block in Fig. 1.

II. Defocus Blur Propagation

A. Superpixels Generation

Target from this step is of first creating the basic units called as (superpixels), and next to define similarity between neighborhood superpixels. Given an image, we first use the SLIC algorithm to oversegment the image into a superpixel set $S = \{ s1, s2, …, sN \}$. As per the superpixel set S, we state a weighted connections of graph $\mathcal{G} = (S,\varepsilon,\omega)$, in which the set of vertex is the superpixel set S and edge set ε containing pairs in adjacent every two superpixels. That, each of vertex S_i given the one single superpixel S, and also each edge $e_{ij} \in \varepsilon$ denotes adjacency connection between superpixels s_i and s_j. The weight function $\omega : \varepsilon \rightarrow [0,1]$ defines the corresponding weight ω_{ij} in each edge e_{ij} , expressed as terms of feature similarities. We thus define this weight matrix as $W = [\omega_{pq}]\ N \times N$.

B. Integration of Blur

The mentioned weight matrix *W* stated the similarity in between the any of two neighbor superpixels. As per the transductive inference proposed method by Zhou et al. [2], we can obtain an N-by-N affinity matrix A to defined transductive similar effect between any of two superpixels, However no matter if they are adjacent or not. The affinity matrix *A* can be stated by

$$A = (D - \gamma W)^{-1} I \quad \text{…………….(4)}$$

Where *D* is the diagonal matrix along with each diagonal entry equal to the row sum of *W*, γ is a parameter in (0, 1), and *I* is the N-by-N identity matrix. As the affinity matrix is encodes transductive similarity in between any of two superpixels, it is able to adjust defocus blur of any of superpixel combination pair using their affinity *A*.

IV. STAGE II : TEXTURE CUE ESTIMATION

A. *Extraction of different regions*

In this stage initial defocus map is processed, this defocus map segmented into multiple regions using segmentation and then for different segments histogram threshold is calculated. Based on that threshold level it is decided that same segment should have same amount of threshold level. Here different regions stated as *s1, s2, …, sk* and having corresponding threshold level of *t1, t2 ,…….., tk-1.*

B. *Calculation of texture gradient*

Image intensity present in the image is the most important channel to get texture information within image. Here we have applied Law's mask to this image intensity channel to calculate texture gradient value present in the image. As we have seen earlier we have calculated this energy for different regions of image. In this texture haze is displayed in the lower frequency of channel of color images. We capture this by using averaging filter (Law's mask) to this color channels. Then to calculate texture gradient that is robust to image noise, we convolve this intensity channel using six edge oriented filters. Multiple models could be used to calculate this texture energy of image.

C. *Correction in defocus map using texture rules*

In first stage our blur propagation by the defocus map interpolation given by equation (4).again we will take or texture energy into account and recalculate this formula to get final depth estimation with correction. Here in this formula our weight matrix W is multiplied by the factor γ. Therefore spreading the defocus map region wise into different regions is more effective results produced. This factor γ mainly depends on the class of depth in the image. We can make the choice of this factor by for different values. We have considered texture of each region along with value of γ. Here in this case higher the sharpness of the region blur should be lower so γ must be high. Also lower the edge sharpness of region higher value of defocus blur and hence γ chosen is low. For some regions having moderate edge sharpness as well as defocus blur γ should be chosen moderate.

D. *Propagation and filling ambiguity in defocus map*

In this stage we have propagate defocus map using texture gradient calculated in earlier stage.here in this stage for each of the superpixel we can have texture gradient calculated from [3 x 3] Law's mask and different filters. Earlier we have created affinity information consideration then we propagate this texture gradient H of each superpixel through entire image.

$$fsi = H [fs1 ; fs2 ; : : : ; fsN]T \qquad \text{……….(5)}$$

Where *H* denotes as an modified affinity matrix with texture gradient calculated. Each superpixel is convoluted with texture gradient to get texture cue depth map. This texture gradient utilizes multiple filters to utilize texture cue information present in the input image. Convolution filters has been used to create texture energy and texture gradient.

V. EXPERIMENTAL RESULTS

This algorithm has performed using Microsoft Windows based tool called as Matlab Ver. 2015B.The machine utilize for this having Intel Core I5-2500 CPU running at 2.4GHz and having RAM of 4GB.

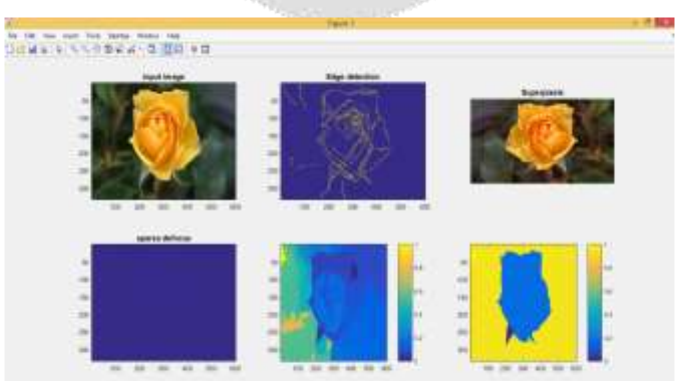

Figure 4 : Output at different stage of operation

Comparison of results with old and new algorithm:

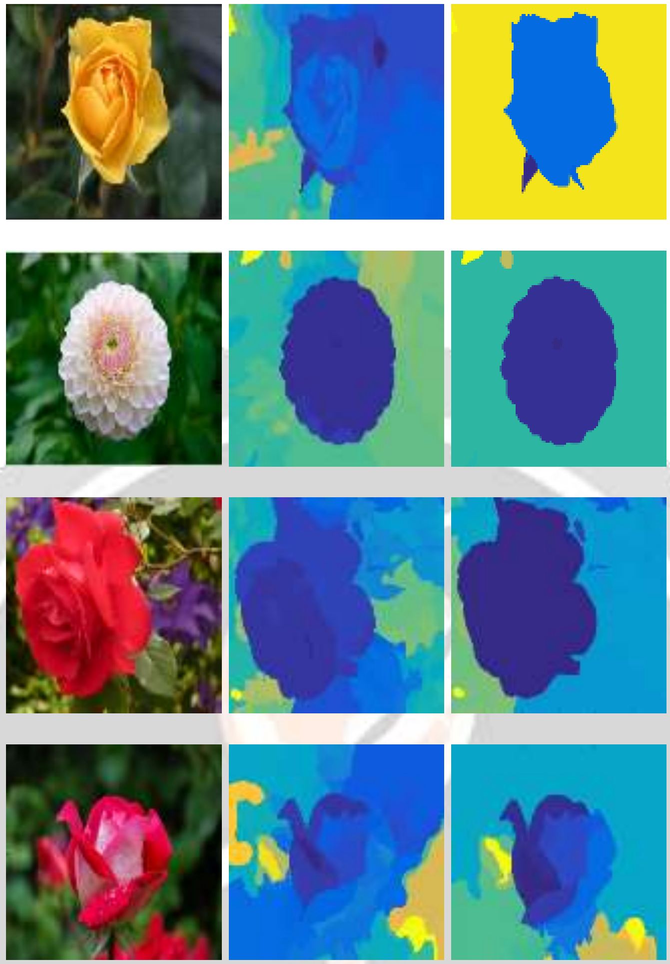

Figure 5(a) Figure 5(b) Figure 5(c)

Above example shows output at different stages. Here Figure 5(a) shows input image, Figure 5(b) shows depth map with old algorithm, Figure 5(c), shows depth map with new algorithm

VI. CONCLUSION

We have implemented estimation of depth map from single image using defocus blurring and texture cue. This algorithm has more accuracy than the earlier depth map which is generated from only single defocus cue. Also we have this method has increased speed of the propagation of blur estimation and texture gradient calculation into entire image. Also experimental results show method which we used having well visualization output. This method can be further extending with neural networks to create more visually better results.

VII. REFERENCES

[1] S. Zhuo and T. Sim, "*Defocus map estimation from a single image", Pattern Recognition*, vol. 44, no. 9, pp. 1852–1858, 2011.

[2] A. Saxena, S. H. Chung, and A. Y. Ng, "*Learning depth from single monocular images*," in Advances in Neural Information Processing Systems, pp. 1161–1168, 2005

[3] A. Saxena, M. Sun, and A. Y. Ng, "*Make3d: Learning 3d scene structure from a single still image*," Pattern Analysis and Machine Intelligence,IEEE Transactions on, vol. 31, no. 5, pp. 824–840, 2009.

[4] Davies, E. R. (1997). Laws' texture energy in TEXTURE. In *Machine vision: theory, algorithms, practicalities* (2nd ed.). San Diego: Academic Press.

[5] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, "*Digital image processing using MATLAB"*. Pearson Education India, 2004

[6] K. I. Laws, "*Textured image segmentation*", tech. rep., DTIC Document, 1980.

[7] C. Zhang, Z. Yin, and D. Florencio, "*Improving depth perception with motion parallax and its application in teleconferencing*," in Multimedia Signal Processing, 2009. MMSP'09. IEEE International Workshop on, pp. 1–6, IEEE, 2009.

[8] J. Shi, L. Xu, and J. Jia. "*Discriminative blur detection features"*. In 2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014, pages 2965–2972, 2014.

[9] Y. Tai and M. S. Brown. "*Single image defocus map estimation using local contrast prior"*. In Proceedings of the International Conference on Image Processing, ICIP 2009, 7-10 November 2009, Cairo, Egypt, pages 1797–1800, 2009.

[10] J. Canny. "*A computational approach to edge detection*". IEEE Trans. Pattern Anal. Mach. Intell., 8(6):679–698, 1986.

[11] P. Doll´ar and C. L. Zitnick. "*Fast edge detection using structured forests"*. IEEE Trans. Pattern Anal. Mach. Intell., 37(8):1558–1570, 2015.

[12] J. Lin, X. Ji, W. Xu, and Q. Dai. "*Absolute depth estimation from a single defocused image*". IEEE Transactions on Image Processing, 22(11):4545–4550, 2013.

[13] K. He, J. Sun, X. Tang,"*Single image haze removal using dark channel prior*", IEEE Trans. Pattern Anal. Mach. Intell. 99 (2010).

[14] Saxena, A., Schulte, J., & Ng, A. Y. (2007). "*Depth estimation using monocular and stereo cues*". In International joint conference on artificial intelligence (IJCAI).

[15] S. Bae and F. Durand, "*Defocus magnification*," in Computer Graphics Forum, vol. 26, pp. 571–579, Wiley Online Library, 2007.

[16] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. S¨usstrunk. "*SLIC superpixels compared to state-of-the-art superpixel methods"*. IEEE Trans. Pattern Anal.Mach. Intell., 34(11):2274–2282, 2012.

[17] J. H. Elder and S. W. Zucker. "*Local scale control for edge detection and blur estimation*". IEEE Trans. Pattern Anal. Mach. Intell., 20(7):699–716, 1998.

[18] X. Zhu, S. Cohen, S. Schiller, and P. Milanfar. "*Estimating spatially varying defocus blur from A single image"*.IEEE Transactions on Image Processing, 22(12):4879–4891, 2013.

[19] I. Oruc¸, L. T. Maloney, and M. S. Landy, "*Weighted linear cue combination with possibly correlated error*," Vision research, vol. 43, no. 23, pp. 2451–2468, 2003.

[20] S. Bae and F. Durand. "*Defocus magnification*". In Computer Graphics Forum, volume 26, pages 571–579. Wiley Online Library, 2007.