Research Article

A Robust Approach for Blur and Sharp Regions' Detection Using Multisequential Deviated Patterns

Awais Khan ⓘ,[1] Ali Javed ⓘ,[1] Aun Irtaza ⓘ,[1] and Muhammad Tariq Mahmood ⓘ[2]

[1]*Department of Computer Science, University of Engineering and Technology Taxila, Taxila, Pakistan*
[2]*Future Convergence Engineering, School of Computer Science and Engineering, Korea University of Technology and Education, Cheonan, Republic of Korea*

Correspondence should be addressed to Muhammad Tariq Mahmood; tariq@koreatech.ac.kr

Blur detection (BD) is an important and challenging task in digital imaging and computer vision applications. Accurate segmentation of homogenous smooth and blur regions, low-contrast focal regions, missing patches, and background clutter, without having any prior information about the blur, are the fundamental challenges of BD. Previous work on BD has emphasized much effort on designing local sharpness metric maps from the images. However, the smooth/blurred regions having the same patterns as sharp regions make them problematic. This paper presents a robust novel method to extract the local metric map for blurred and nonblurred regions based on multisequential deviated patterns (MSDPs). Unlike the preceding, MSDP extracts the local sharpness metric map on the images at multiple scales using different adaptive thresholds to overcome the problems of smooth/blur regions and missing patches. By using the integral values of the image along with image masking and Otsu thresholding, highly accurate segmented regions of the images are acquired. We argue/hypothesize that the local sharpness map extraction by using direct integral information of the image is highly affected by the threshold selected for distinction between the regions, whereas MSDP feature extraction overcomes the limitations substantially by using automatic threshold computation over multiple scales of the images. Moreover, the proposed method extracts the relatively accurate sharp regions from the high-dense blur and noisy images. Experiments are conducted on two commonly used SHI and DUT datasets for blur and sharp region classifications. The results indicate the effectiveness of the proposed method in terms of sharp segmented regions. Experimental results of qualitative and quantitative comparisons of the proposed method with ten comparative methods demonstrate the superiority of our method. Moreover, the proposed method is also computationally efficient over state-of-the-art methods.

1. Introduction

With the exponential growth of digital image capturing devices, i.e., DSLR cameras, cellphone cameras, wearable cameras, etc., we have witnessed a massive collection of digital photos captured and uploaded on social media on a daily basis. A good quality photo must be sharp and not contain any degradation such as noise and blurred regions. As for many applications, we need to highlight the target object from the images. To make the object highlighted, many techniques are being used nowadays, e.g., using high-definition camera sensors, adding blurriness to the background objects, etc. However, the high-definition image sensors impose blurriness in the background of the images to make the foreground objects more prominent. Consequently, blurriness is being used as an editing effect that is added purposely in modern-day image capturing.

Image blurriness can be categorized into motion blur and defocus blur. The motion blur can occur due to two potential reasons: (a) when you try to capture the moving objects and (b) camera motion either intentionally or unintentionally, whereas the defocus blur usually occurs due to special effects used by the photographers to highlight the focus and out-of-focus regions in the image. It is a visual effect added by the photographer using highly sophisticated techniques to make the target object sharp and the rest of the image blur. An image contains useful information that can be used in various computer vision and image processing

applications, i.e., background tracing, text retrieval, image retrieval, person authentication, etc. However, blur affects the contrast and sharpness details of the image that made the retrieval of information challenging. Similarly, photos are used as the key evidence in a criminal investigation, where it can be very challenging to extract the immersed information about the target object(s) in the presence of high-density blurred regions. For this purpose, we need to classify the image into blur and nonblur regions initially. The information of the objects lying in the nonblur regions is more reliable in comparison with those in blurred regions, as the information can be distorted in blur regions like an increase of edge thickness. Thus, blur detection and sharpening are required for the accurate extraction of the information from the images.

Defocus blur detection (DBD) is the classification problem of intentionally added blur by the highly sophisticated modern-day digital cameras. This classification has been paid substantial attention due to their significant potential applications, e.g., object detection [1], image segmentation [2], object augmentation [3], etc. Defocus blur affects the information and the sharpness details of the image making it more challenging for objects/regions detection. The DBD without having any prior information about the blur densities, blur type, or sensor settings of the camera is a challenging task.

Existing blur image detection approaches are divided into two categories: single image detection and multi-image detection. For multi-image detection, the knowledge of the blur densities, type, sensor information of the cameras, and other additional information is required [4]. In contrast, a single image can be split into sharp and blur regions without having any prior information about the blur and the device used to capture that image [5, 6]. Moreover, the existing approaches presented for DBD can be categorized into frequency-based [7–13], depth-based [14–17], or local sharpness metric map-based for segmentation of blur and nonblur regions [6, 18–21]. In [22], Zhu et al. used the local coherent map generated by the evaluation of gradient fields of the local spectrum. However, the use of flat areas information and color edges is not enough for accurate DBD detection. In [23], Chakrabarti et al. proposed a Point Spread Function (PSF) using the local frequency analysis to obtain the segmented map for DBD. This method has the limitation of generating erroneously labeled regions of the image. Su et al. [24] presented a method called singular value decomposition (SVD) based on single thresholding on image features to detect the blurred and nonblurred regions. Similarly, Xiao et al. extended the single threshold SVD into multiscale SVD in [8]. The fusion-based method is used to overcome the smooth/blur region problems from the images. In [25], Golestaneh estimated the level of blurriness at each location using a method called high-frequency multiscale Fusion and Sort Transform (HiFST) based on gradient magnitudes.

Depth-based methods [9, 14–17] also proved to be effective in defocus blur detection using the information about the blur densities and blurry edges. In [9], Liu et al. presented different local features, i.e., association congruence,

saturation, gradient histogram, and power bands to specify the type of blur from the images. In [26], a cross ensemble network is used along with a smaller defocus detector for diversity enhancement. However, this approach is computationally expensive and unable to differentiate the nonblur regions accurately in the presence of smooth regions. Furthermore, DBD measurements such as local variance, higher-order statistics, and variance of wavelength coefficient are also used in DBD with images containing a narrow depth of field (DOF) [27].

Most of the algorithms [6, 18, 19, 21] used the local sharpness metric approach for the detection of blur and nonblur regions. The local sharpness metric is like a filtering method such as energy function estimating the results based on the responses of blur energy from images. The low energy indicates the blur region, whereas the high energy represents the sharp region. In [12], Shi et al. introduced peculiar sharpness features, gradient histogram, and kurtosis span for DBD of local image regions. This method is unreliable and causes problems in the accurate detection of blur and sharp regions due to smooth homogenous regions. Zhu et al. analyze the blur from the images using PSF [22]. This method is unable to perform well in lightly blurred regions. In [21], Local Binary Patterns (LBP) are taken into consideration for defocus blur detection. The local metric map generated using the LBP segments the blur and nonblur regions but is unable to perform well in the presence of noisy images even in sharp regions. In our prior method [18], we proposed the Local Directional Mean Patterns (LDMP) to overcome the limitation of the sharp metric map in noisy situations. However, our prior method [18] is unable to detect simultaneous smooth blur and low-contrast regions.

Recently, deep learning methods have been heavily employed in various computer vision and image processing applications, i.e., saliency detection [1], semantic segmentation [2], automatic shadow detection [28] airplane detection using remote sensing images [29], ship detection from real-time images [30], vehicle detection [31], etc. The significance of deep learning algorithms is proven to be effective for defocus blur detection and segmentation, however, at the expense of increased computational cost. In [32], Kim introduced a deep learning method based on a convolution neural network (CNN) for the detection of sharp and blur regions of the image. The multiscale reconstruction loss function was used for the segmentation of blur regions. In [33], Park et al. encounter the DBD problem with patch level detection based on CNN. Unfortunately, the patch level DBD led to suppression in low-contrast regions. Tang et al. [34] proposed a Deep Neural Network (DNN) based technique Diffusion Network (DNet) that fused the refined features extracted by the networks to obtain the segmented blur and sharp regions. In [35], the author introduced global context-guided hierarchically residual feature refinement network "HRFRNet." The hierarchical features are used to enhance the final outcomes. Furthermore, a deep-guided fusion module is used for the refining process.

Accurate DBD has initiated extensive research interest from the last few years. However, it is still a significant yet

challenging computer vision problem. Although the afore-mentioned techniques can detect the defocus blur and nonblur regions, however, these approaches fail in certain cases, i.e., the presence of smooth blurry regions, missing sharp patterns, low contrast, etc. The DNN methods per-formed well for DBD, but all of these methods are com-putationally more complex and required high computational resources, i.e., GPUs, memory, etc. We aim to develop a robust method for DBD that can effectively extract the sharp targeted regions of the image in the presence of noise, smooth/blur, and low-contrast images. To address the aforementioned problems, we propose an efficient and robust Multisequential Deviated Patterns (MSDP) for accurate sharp region extraction from images at multiple scales. The extracted multiscale sharpness maps are further fused to get the refined map of the image. We used Otsu thresholding to segment the extracted sharp region into comparable binary representation. The pro-posed method is efficient due to using the local integral values of the images directly instead of using time-con-suming matting approaches used by the preceding methods to segment the binary images. The major contributions of the proposed work are as follows:

(i) We propose efficient and robust multisequential deviated patterns for accurate blur detection from high-density blur and noisy images

(ii) For feature computation, we extract the sharpness metric using adaptive thresholding on multiple image scales to overcome the smooth blur and missing sharp region problem of manual thresholding

(iii) For image segmentation, we fused the multiscale sharpness maps extracted from MSDP along with the image masking

(iv) Rigorous experiments were performed against several state-of-the-art methods over the latest DUT and SHI datasets to prove the effectiveness of the system

The rest of the paper is organized as follows. Section 2 presents the details of the proposed method. Section 3 provides the discussion on results of different experiments conducted to evaluate the performance of our method. Finally, Section 4 concludes our work.

2. Methods

This paper presents a novel method based on the integral use of the image, which detects the blur and sharp regions from high-dense blurry and noisy images. The proposed method used the local window of different sizes based on the input image scale to extract the sharp regions in the image. Firstly, the image is divided into three different scales S_1, S_2 and S_3. Secondly, MSDP is used to extract a sharpness map from each scaled image using an adaptive threshold. Lastly, image masking is used over the extracted sharpness maps, and fusion is performed to produce a more accurate single sharpness map. Next, Otsu thresholding is used to retrieve

the accurate binary images for comparison with state-of-the-art methods. The flow of the proposed method is shown in Figure 1.

2.1. Preprocessing and Image Scaling. The extraction of features using integral values of the image is challenging due to the influence of different factors, i.e., color, camera, object detail, etc., on the values. For accurate extraction of sharp regions, we used the RGB input image I_b consisting of sharp and blur regions. First, we convert the RGB color image into grayscale and apply the two-dimensional median filtering to reduce the noise.

$$
\begin{aligned}
I_b &= P_S + P_B, \\
I_g &= rgb2gray\,(I_b), \\
\mathrm{Im}_F &= medfilt2\,(I_g),
\end{aligned}
\tag{1}
$$

where P_S and P_B denote the sharp and blur pixel values of the image, and I_g represents the grayscale image. Im_F represents the 2D median filtered image with reduced noise obtained after applying the 2D median filtering function ($medfilt2$). Next, we represent the image into three different scales (S_1, S_2 and S_3) after employing the $medfilt2$.

2.2. Multisequential Deviated Patterns (MSDPs). Selecting an appropriate threshold for integral extraction is a compli-cated task. For example, the use of a high threshold in local feature extraction leads to exclusion of low/lesser sharp regions, whereas the selection of low threshold value causes the inclusion of additional useless details in the features, i.e., background object and noise, etc. The relationship of high and low thresholds in integral feature extraction is shown in Figure 2. For MSDP maps, we computed the upper and lower patterns of the image. In integral extraction of fea-tures, the three-level thresholding is proven effective [18] as compared to the two-level thresholding [21]. Moreover, two-level thresholding is not effective in the presence of a high density of noise in the images. Therefore, we employ 3-level thresholding for the extraction of sharp integral upper and lower features of the images. We extract the upper and lower features as follows:

$$
ULP = \sum_{P=1}^{8} 3^P s\,(im_p - im_c),
\tag{2}
$$

$$
\delta = \sqrt{\frac{\sum (q_i - \mu)^2}{N}},
$$

$$
P(im_p, im_c, \delta) = \begin{cases} 1 \text{ if } & im_p \geq im_c + \delta, \\ -1 \text{ if } & im_p \leq im_c - \delta, \\ 0, \text{ if } & im_p > im_c - \delta \,\&\, im_p < im_c + \delta, \end{cases}
\tag{3}
$$

where 3^P and q_i represent the 3-level thresholds and integral values of the pixels, N and μ denote the total number of

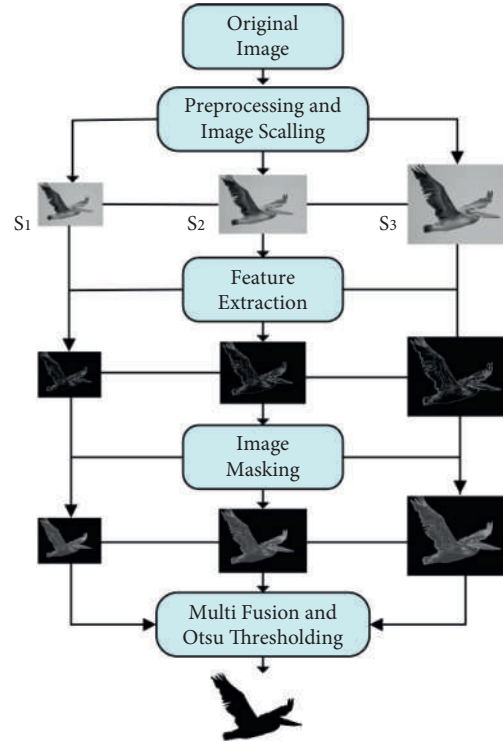FIGURE 1: Flow diagram of the proposed method.

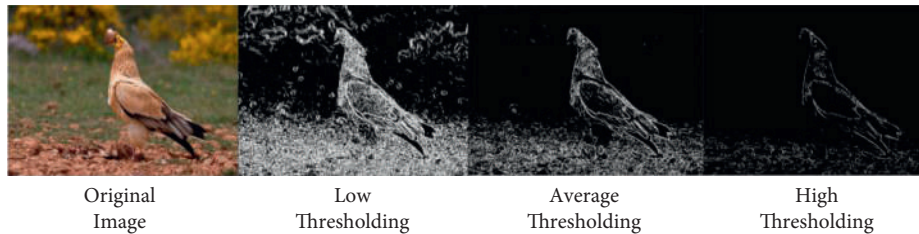| Original Image | Low Thresholding | Average Thresholding | High Thresholding |

FIGURE 2: Impact of the threshold in local integral pattern extraction.

integral and mean values, and im_p and im_c are the neighboring and center pixels of the image windows, while δ represents the adaptive threshold, which is calculated by the sequential deviation of the existing window as shown in equation (2). Instead of adopting the threshold selection approach of the existing methods, we computed an adaptive threshold for each region of the image by rotating the extraction window over the image. The adaptive threshold is computed automatically by adding and subtracting the center pixel value im_c of the window along with the standard deviation δ as shown in equation (3). An adaptive threshold based on δ and im_c is responsible for the extraction of sharp pixels from the images; i.e., high threshold value leads to highly sharp regions, and low threshold value leads to the inclusion of noise and other unwanted content. In contrast, the preceding methods mostly used a hard-coded threshold value in their algorithms, which makes these methods unable to effectively extract the low-dense sharp regions locally [18, 21]. Consequently, we applied three-level thresholding with an adaptive threshold for the extraction of three values including 1, −1, and 0 from the image. For instance, as shown in Figure 3, if a 3×3 window is used over the image integral values having a central pixel value (im_c) of 23 with the deviation of the neighboring pixel (δ) of 10.2, then the range of extraction lies between 13.2 and 33.2, which is shown as $im_p > im_c − \delta \& im_p < im_c + \delta$ in equation (3). For three-level value extraction from the image, the neighboring pixel in the window lies between 13.2 and 33.2 and is converted into "0," whereas the integral value of 1 is assigned to values greater than the threshold $im_c + \delta$ (33.2 in Figure 3), and −1 is assigned to the values below the threshold $im_c − \delta$ (13.2 in Figure 3). After replacing the integral values of the images, we obtained $P(im_p, im_c, \delta)$, which contains three-level values comprising 1, 0, and −1. The overall extraction of three-level values is shown in Figure 3. For instance, the image with Qn window from the integral values of pixels where n represents the 3×3 window having 9 values is shown in equation (4).
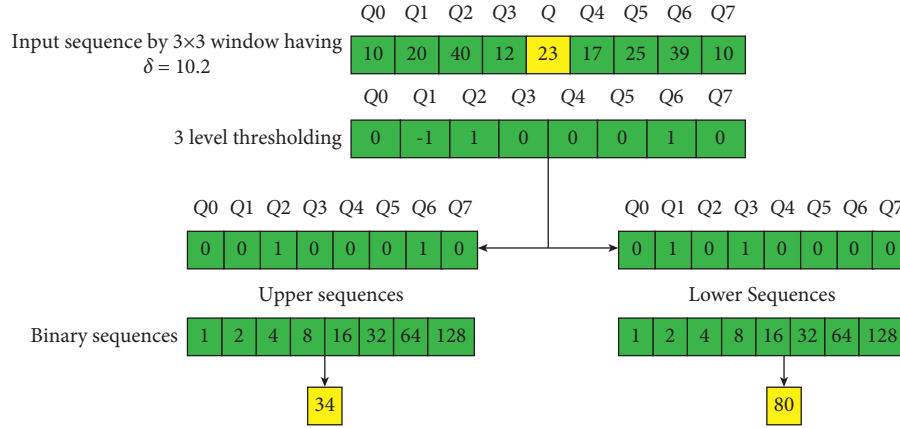
FIGURE 3: Extraction of upper and lower binary sequences.

$$\text{Im}_F = \sum_{i=1}^{n} Qn_{e \times e},$$
$$Q_{3 \times 3} = \{Q0, Q1, Q2 \ldots Q7\}, \tag{4}$$

where e denotes the dimension of the rotating window. In order to compute the sharp region, the upper and lower features of the image have to be computed separately. However, the proposed $P(im_p, im_c, \delta)$ extracts the combined features of the image. Therefore, to reduce the noise in the images, the three-level extracted patterns are further converted into two levels, i.e., upper and lower image patterns. For two levels of extraction, we need to replace the negative values from the obtained $P(im_p, im_c, \delta)$. For extracting the upper patterns of the image, we converted all −1 values from $P(im_p, im_c, \delta)$ into 0. And for lower patterns, 1 and −1 are replaced with 0 and 1 as shown in Figure 3. Ultimately, the resultant upper and lower patterns of the image are converted into binary bit streams using equation (5) that are further represented into their equivalent decimal values.

$$Upp_F = \left\{0 \Leftrightarrow \text{if} \quad P(im_p, im_c, \delta) = -1\right\},$$
$$Low_F = \begin{cases} 1 \Leftrightarrow \text{if} \quad P(im_p, im_c, \delta) = -1, \\ 0 \Leftrightarrow \text{if} \quad P(im_c, im_c, \delta) = 1, \end{cases} \tag{5}$$

where Upp_F denotes the upper features of the image, and Low_F represents the lower features. The $Z \times Z$ window is used to obtain the decimal values of upper and lower patterns using equation (6), where Z denotes the size of the window varying for each scale of the image. At last, we need to pick the sharp and blur patterns smartly by retrieving only the sharp pixels and neglecting the blur once. The proposed method uses the deviation of the pixels twice to observe the change in the patterns of blur and sharp regions. We computed the standard deviation of two-level patterns consisting of the upper and lower patterns of the image obtained from the last step. For this purpose, first, we have to convert the binary two-level patterns into their equivalent decimal numbers by using the window $Z \times Z$. All the neighboring pixels values in the window covert into their equivalent decimal number as follows:

$$\text{Im}_{upp} = \sum_{k=1}^{8} Upp_F(S_P) \times 2^{k-1},$$
$$\text{Im}_{low} = \sum_{k=1}^{8} Low_F(S_P) \times 2^{k-1}, \tag{6}$$

Im_{upp} and Im_{low} are the upper and lower patterns of equivalent decimal numbers from the extracted two-level binary patterns from equation (5). After that, we computed the deviation of the neighboring pixels again with decimal values of the upper and lower patterns computed from equation (6). The values higher than the deviation (δ) are considered as the sharp region values and retained, while the rest are neglected for being the blurry region. The extraction of Multilayered Sequential Pattern is shown in equations (7) and (8).

$$\delta_{upp} = \sqrt{\sum (qi - \mu)^2 / N},$$
$$\delta_{low} = \sqrt{\sum (qi - \mu)^2 / N}, \tag{7}$$

$$MSDP = \left\{\text{Im}_{upp} > \delta_{upp} \oplus \text{Im}_{low} > \delta_{low}\right\}, \tag{8}$$

where δ_{upp} and δ_{low} represent the deviation value of upper and lower patterns of the image, respectively. Finally, the sharpness map of sharp regions is extracted as $MSDP$ shown in equation (8). Moreover, we extracted three $MSDP$ sharpness maps for each scale of the image (S_1, S_2 and S_3) as shown in Figure 4.

2.3. Image Masking. The extracted features $MSDP$ contain only the sharp regions of the images at three scales S_1, S_2 and S_3. However, there are some regional variations among the images at all scales caused by the adaptive thresholding and different dimensions. The main reason for this variation is missing regions among the extracted maps. The empty integral values in the patterns usually occur due to the noise or background objects that ultimately lead to missing regions in the image. We applied a morphological operation on this image to fill the holes and gaps between the pixels. More
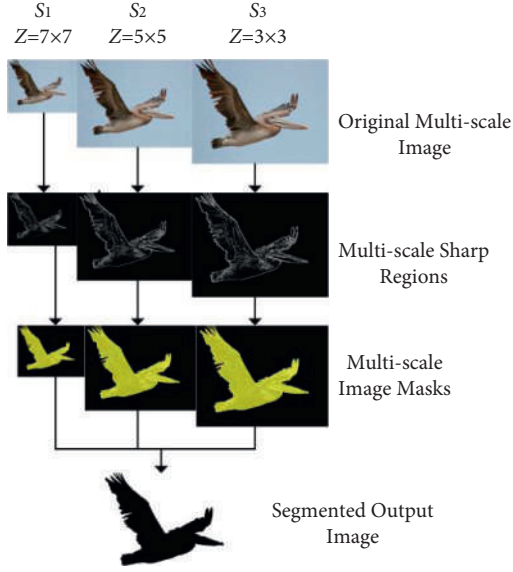
FIGURE 4: Multiscale image masking and Otsu thresholding.

specifically, we employed a binary filling operation to fill the gaps and holes using equation (9). Similarly, this filling process is applied at every scale (S_1, S_2 and S_3) of the image. In addition, we created an image mask containing sharp regions of the image before applying the Otsu thresholding for segmentation. For this purpose, we applied the Otsu global thresholding to select the sharp regions from the obtained image.

$$Trimap_{IM} = BF(MSDP). \tag{9}$$

2.4. Multifusion and Otsu Thresholding. In the last phase, we fused the scales (S_1, S_2 and S_3) into a single sharpness map and employed the Otsu thresholding [36] for binarization of the images. The masked image obtained from the previous phase is converted into a segmented binary image of the sharp and blur regions. The variance-based thresholding is used along with linear discriminant principles to segment the target object (foreground) from the heterogeneous and diverse background regions. The threshold for segmentation in Otsu thresholding is based on the variation of the integral values. The extracted patterns in $Trimap_{IM}$ are divided into two classes, i.e., sharp region and the background blur region. The global threshold is computed according to the variance of the classes. The regions with values higher than the threshold value are selected, whereas the regions below the threshold value are ignored. Finally, the highly segmented binary image with the sharp object as foreground and the black background is obtained as follows:

$$B_{img} = c_0(\varepsilon)v_0^2(\varepsilon) + c_1(\varepsilon)v_1^2(\varepsilon),$$
$$c_0(\varepsilon) = \sum_{j=1}^{rw-1} i(img),$$
$$c_1(\varepsilon) = \sum_{j=\varepsilon}^{cl-1} i(img), \tag{10}$$

where c_0 and c_1 are the integral values of sharp and blur region classes separated by a threshold ε, whereas v_0^2 and v_1^2 denote the variance of the classes. Overall, we extracted the MSDP from images at multiple scales (S_1, S_2 and S_3) using $Z \times Z$ windows of multiple sizes (i.e. 3×3, 5×5, 7×7). The reason behind the extraction of the same $MS\ DP$ at different scales (S_1, S_2, S_3) is to overcome the missing region problem from the extracted sharp regions. The sharpness map is extracted from single $MSDP$ image containing some missing areas inside the sharp regions, whereas the extraction of sharpness map over different scales overcomes this problem to some extent as discussed in Section 4. The intensity of the blur varies in each region of the image; i.e., some regions are highly affected by the blur, whereas regions far away from the sharp objects are less affected. Therefore, the extraction of the integral value is highly affected by the selected threshold value for classification of pixels as sharp and blur. We employed an adaptive threshold calculated from the deviated ratio of the patches along with the central pixel values of the regions. The combination of local central pixel values of the regions and the overall deviation between the pixels for thresholding make our method robust in the extraction of the sharp regions from the blurry images. The overall computation process at multiple scales of the image is shown in Figure 4.

3. Experimental Results

This section provides a discussion on the results of different experiments performed to measure the performance of the proposed method. We have provided a detailed comparison of qualitative and quantitative results along with the analysis of the computational complexity of our method. The details of the datasets and evaluation metrics are also presented in this section.

3.1. Datasets. The performance of our method is evaluated on two standard and commonly used datasets (Shi and DUT) for blur detection. Shi dataset [12] is the first public dataset collected and evaluated for blur detection [12]. This dataset [12] is used by almost all state-of-the-art descriptors to show the effectiveness of the methods. Shi dataset [12] contains 1000 blur images, where 704 images are partially blurred using defocus, and the rest of the images contain the motion blur. Additionally, the manually annotated ground truth images are also available along with the blur images. Most of the images in the Shi dataset are of 640×427 resolution.

DUT dataset [37] is the second commonly used publicly available dataset consisting of 500 images with defocus blur. Similarly, this dataset is also provided with manually annotated ground truth images. In comparison with the Shi dataset [12], the DUT dataset [37] is more challenging for blur detection and segmentation for various reasons; i.e., many of the images contain homogeneous smooth blur regions, have cluttered background, and are low in contrast images.

3.2. Evaluation Metrics.

Three standard and commonly used metrics including precision, recall, and $F1$-score are used to evaluate the performance of the proposed method. We selected these metrics as adopted by the comparative methods for performance evaluation. We calculated the precision and recall as

$$pre = \frac{\text{Im}_E \cap \text{Im}_G}{\text{Im}_E},$$
$$rec = \frac{\text{Im}_E \cap \text{Im}_G}{\text{Im}_G}, \tag{11}$$

where Im_E represents the pixels within the detected sharp areas, and Im_G corresponds to the pixels in the manually annotated ground truth image. Similarly, $F1$-score is computed as

$$F1 = 2 \times \frac{pre * rec}{pre + rec}, \tag{12}$$

where pre and rec denote the precision and recall of the proposed method.

3.3. Performance Evaluation of the Proposed Method.

The objective of this experiment is to evaluate the effectiveness of the proposed method for blur detection on two diverse datasets. For this purpose, we computed the results of our method on the images of Shi [12] and DUT [37] datasets separately and reported the results in Table 1. In the case of Shi [12] dataset, the proposed method dominates in every comparison, i.e., quantitative analysis, qualitative analysis, and PR curve. However, in DUT [37] dataset, the precision of the proposed system slightly deteriorates. The DUT [37] dataset is more challenging than Shi dataset [12] due to exceeding homogenous and low cluttered regions. The homogenous regions are always difficult to locate in the local extraction of the regions. Although the precision of the proposed system is a little low, however, the other quantitative results (recall, $F1$-score, and computational cost) prove the effectiveness of the method.

3.4. Performance Comparison of the Proposed and Existing Methods.

The objective of this experiment is to measure the robustness of the proposed method for blur detection over state-of-the-art methods. For this purpose, we have provided both the qualitative and quantitative analyses of the proposed and comparative approaches.

3.4.1. Qualitative Comparative Analysis.

This experiment is designed to show the qualitative analysis of the proposed and comparative methods on Shi [12] and DUT [37] datasets. The visual quality of processed images is presented in Figures 5 and 6 to show the effectiveness of the proposed method over state-of-the-art methods. From the images depicted in Figures 5 and 6, we can observe that the proposed method can detect highly accurate results from Shi [12] and the DUT [37] datasets. Specifically, in Figure 6, we can see that images of DUT dataset [37] contain more homogenous smooth regions and low local cluttered regions that are effectively classified using our proposed method.

3.4.2. Quantitative Comparative Analysis.

This experiment is designed to evaluate the performance of our method in terms of quantitative analysis. For this purpose, we used three standard metrics, i.e., precision, recall, and $F1$-score, to measure the performance of our system against the comparative methods. The precision-recall (PR) curve is calculated on the results of the proposed method from Shi [12] and DUT [37] datasets and presented in Figures 7 and 8, respectively. A separate PR curve comparison is shown for Shi [12] and DUT [37] datasets. The PR curves demonstrate that our method consistently outperforms all the comparative methods. On the other hand, our method effectively addresses the problems of homogenous smooth and low local cluttered regions in the images of a more challenging DUT dataset [37]. The PR curve of the proposed method on DUT [37] dataset continuously dominates throughout the period as shown in Figure 7.

Additionally, the $F1$-score is measured and compared for both the Shi [12] and DUT [37] datasets as shown in Figures 9 and 10. Although DNet [34] achieves almost comparable results as of the proposed method, however, our method is computationally very efficient over the DNet method [34] as shown in Table 2. This comparative analysis illustrates the superiority of the proposed method for defocus blur detection over the comparative approaches.

3.4.3. Computational Cost Analysis.

Although DFD methods must be effective in terms of detecting blurred regions from the images, however, producing such accurate results in a minimum time is also crucial, especially in real-time applications. This experiment is designed to evaluate the efficiency of the proposed and comparative approaches for defocus blur detection. The results of this comparative analysis of time complexity for both the detection and segmentation are provided in Table 2. The proposed method not only dominates in terms of accurate blur detection over the state-of-the-art methods, but also executes exceptionally fast and computationally very efficient over comparative approaches. From the results, we can clearly observe that the proposed method has the 2nd lowest computational cost after LBP [21] as compared to all comparative methods. More precisely, [21] performs the best by achieving 5 seconds to segment an image into blur and sharp regions, whereas our method performed second best and achieved the time complexity of 7 sec. On the contrary, [22] performed the worst by taking the highest computational cost of 12 minutes. The main reason behind achieving such low time complexity of our method is the direct extraction of the local sharpness metric and removal of the time-consuming matting procedures used by several competitive methods. Moreover, image masking and Otsu thresholding used to produce the segmented maps are also very fast in the execution. In our method, the multiscale inference phase consumes the majority of the execution times.

TABLE 1: Precision, recall, and F1-score of the proposed method.

	Precision	Recall	F1-score
Shi dataset	0.91	0.90	0.92
DUT dataset	0.89	0.91	0.88

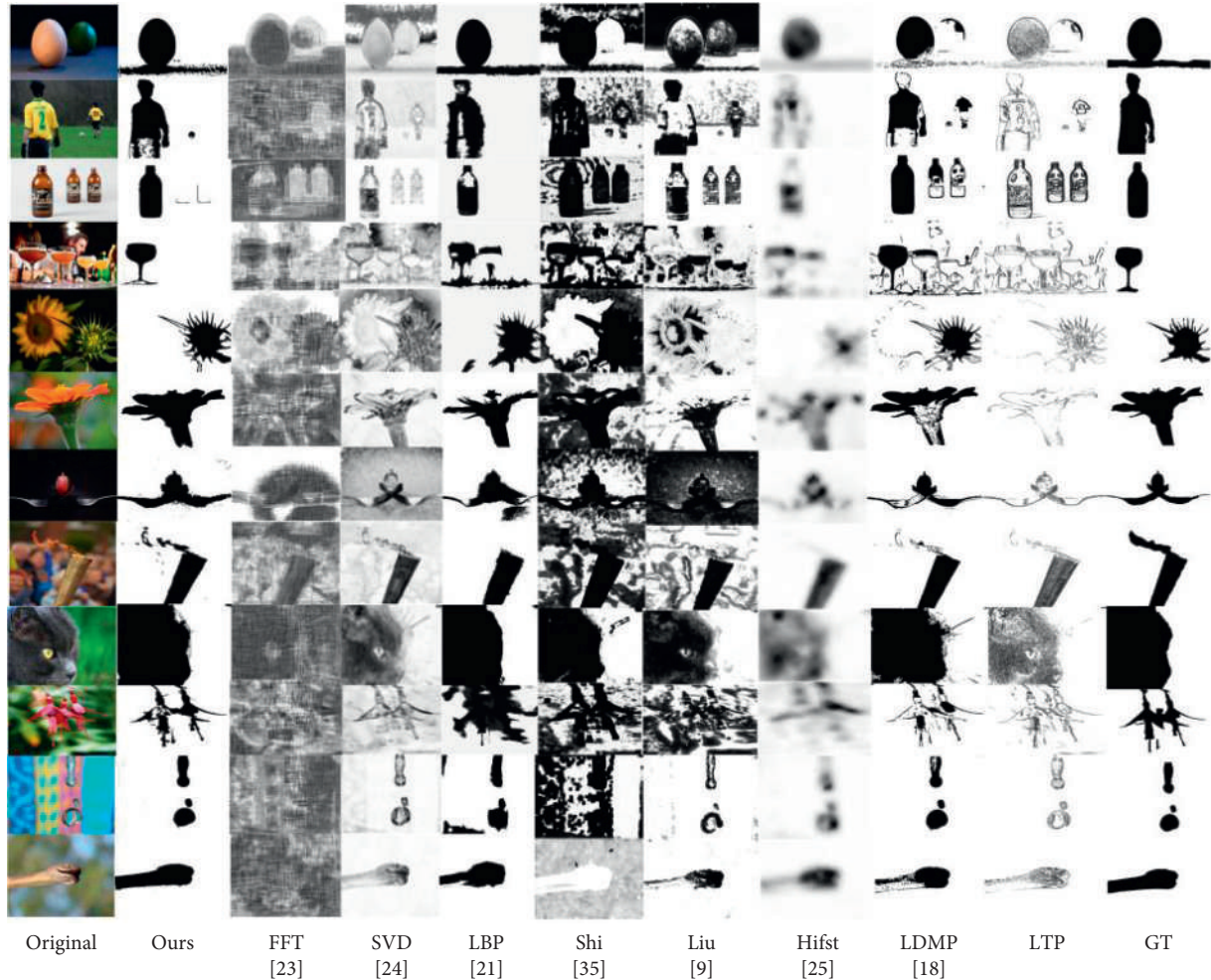| Original | Ours | FFT [23] | SVD [24] | LBP [21] | Shi [35] | Liu [9] | Hifst [25] | LDMP [18] | LTP | GT |

FIGURE 5: Visual comparison of the proposed method against comparative methods on the Shi dataset.

For this comparative analysis, we used the codes of comparative methods that are publicly available along with our own implementation. In the proposed method, the original image was scaled into 256×256, 128×128 and 64×64 dimensions for S_1, S_2 and S_3. Additionally, the window size $Z \times Z$ is selected as 3×3, 5×5 and 7×7 for image scale S_1, S_2 and S_3 respectively. The final values of window size Z for the specific scale Sn are selected after the detailed observations and experiments. Moreover, automatic global thresholding is used in image masking and Otsu thresholding for binarization of the images. We have implemented the proposed and comparative methods on Intel(R) Core (TM) m3-7Y30 CPU @ 1.00 GHz, 1.61 GHz with 8 GB memory system.

4. Discussion

The present study analyzes the findings about the selection of adaptive threshold and the impact of neighboring pixels in local extraction of the image regions. Experimental results demonstrate two facts. First, the global or hard-coded fix threshold value is not reliable for all types of images. Second, the neighboring pixels used to differentiate the integral values have a major impact on the extracted patterns. This is an important finding in the understanding of the local patterns and direct integral extraction from the images. Some sample results from LBP [21] and LDMP [18] are shown in Figure 11 to defend this fact. Figure 11 clearly demonstrates that the

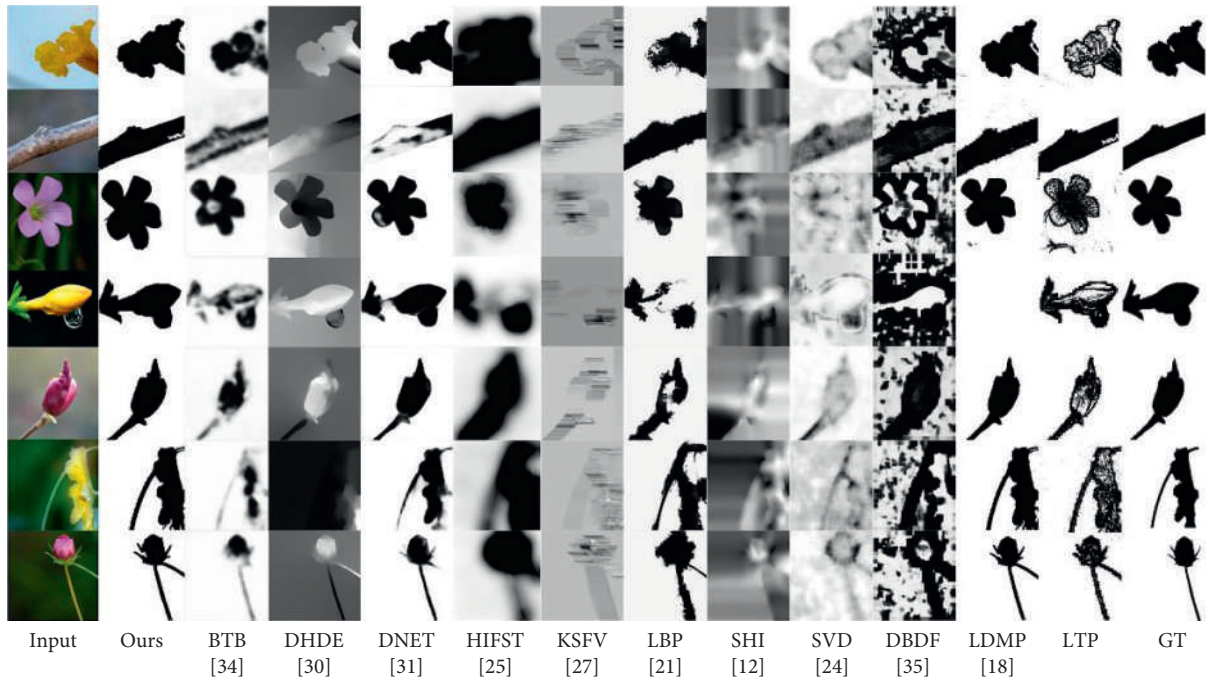| Input | Ours | BTB [34] | DHDE [30] | DNET [31] | HIFST [25] | KSFV [27] | LBP [21] | SHI [12] | SVD [24] | DBDF [35] | LDMP [18] | LTP | GT |

Figure 6: Visual comparison of the proposed method against comparative methods on the DUT dataset.
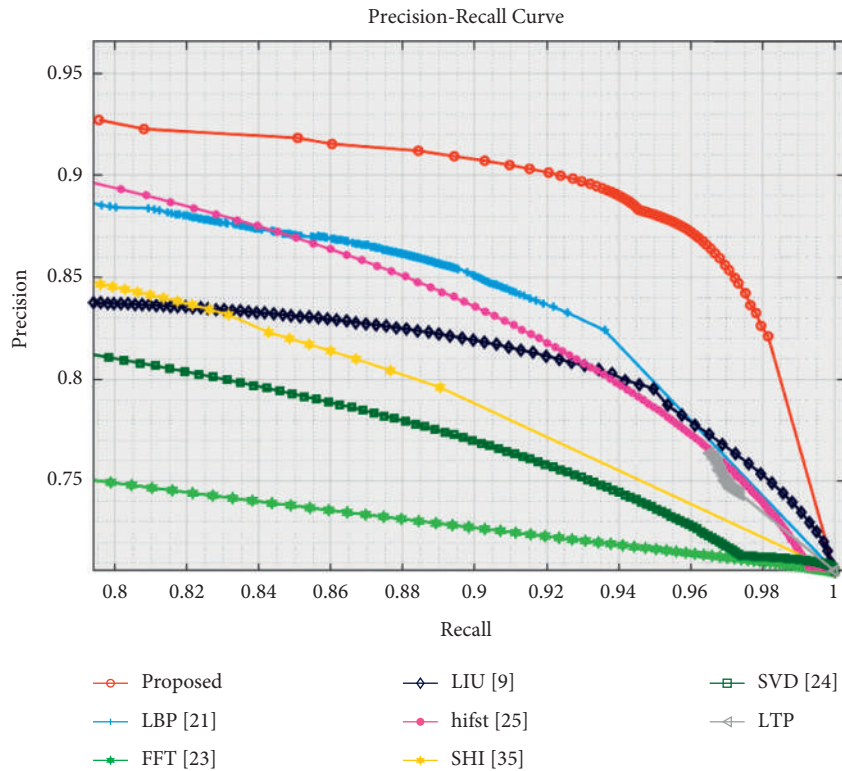

Figure 7: PR curve comparison on the Shi dataset.

comparative methods LBP [21] and LDMP [18] are unable to perform well for many images. Two methods [18, 21] used a similar approach of extraction from local integral

values of the images but used either a fixed static threshold or from a specified range. This hard-coded threshold scheme results in performance degradation, as shown in
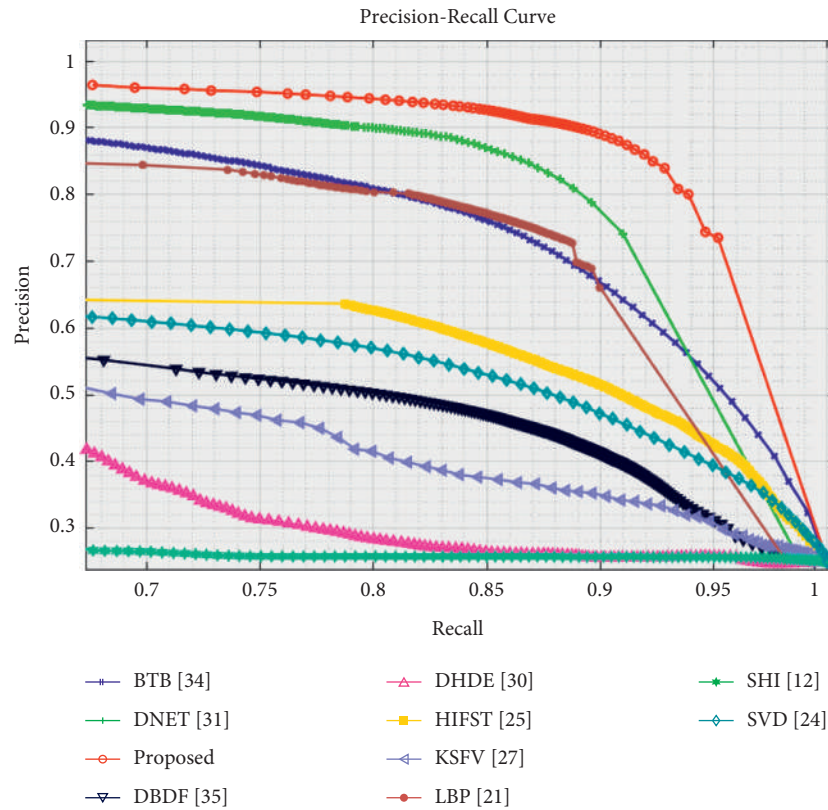
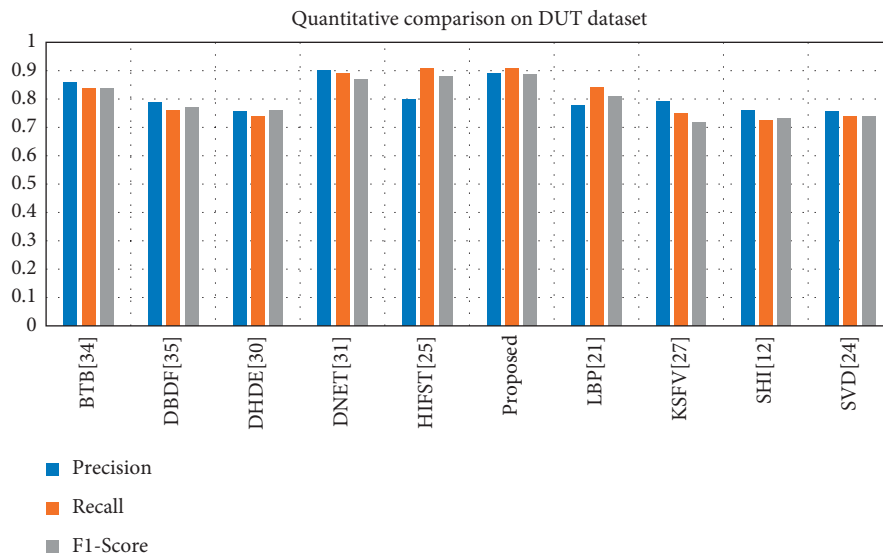Figure 8: PR curve comparison on the DUT dataset.

Figure 9: Quantitative comparison on the DUT dataset.

Figure 11 where [21] is unable to detect the regions in the image, whereas the proposed method used an adaptive threshold computed using the deviation between the neighboring pixels and performed very well. Our experiments further reveal that the results of [18, 21] indicate the gaps between the extracted sharp regions, whereas the proposed method produced the filled regions approximately.

Extraction of the pixels using a small number of neighboring pixels can cause the overall selection of the region. We have used varying numbers of adjacent neighbors ($3 \times 3, 5 \times 5, 7 \times 7$) to extract the regions based on the deviation between the neighbors. This provides significantly better results due to the large deviation of the region. Our results indicate that the pixel selection based on a small number of adjacent pixels affects the

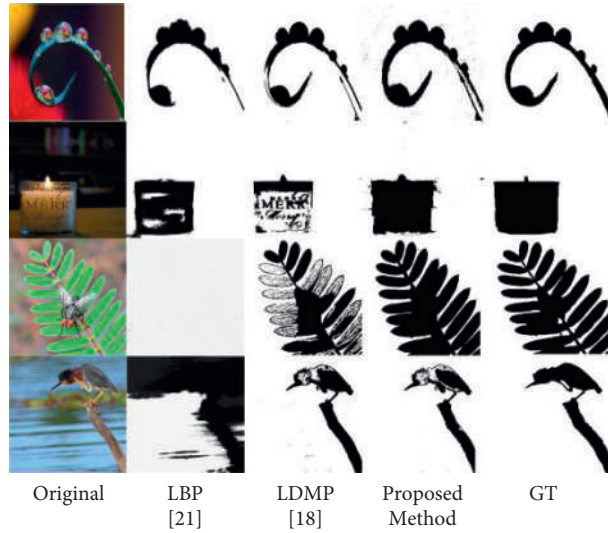| Original | LBP [21] | LDMP [18] | Proposed Method | GT |

Figure 10: Comparison of local pattern-based approaches.

Table 2: Computational cost comparison of the proposed method against state-of-the-art methods.

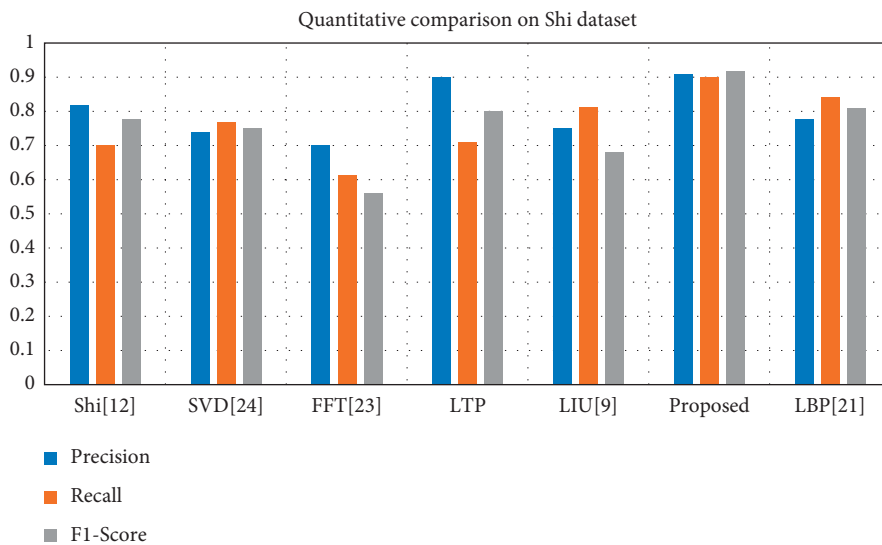| DBD methods	Avg. computational cost (detection, segmentation)
DBDF [38]	7 m, 13 m
LBP [21]	5 s, 58 ms
Vu [13]	1 m 20 s, 35 s
SVD [24]	50 s, 55 s
Shi [12]	55 s
Zhou [28]	45 s
Zhu [22]	12 min
Proposed method	7 s

Figure 11: Quantitative analysis on the Shi dataset.

results, whereas using a large number of adjacent pixels to determine the pixel selection leads to good results. Existing defocus blur detection methods, based on using the integral values directly, fail to operate well for the motion blur. High variation in the integral values makes these methods unable to perform better on the motion blur. We aim to develop a unified

method in the future that can effectively detect both the motion and focus blur.

5. Conclusion

We have proposed an effective and efficient method for defocus blur detection problem without having the prior information about blur, camera configuration, pixels densities, etc. The local sharpness metric map is extracted directly from the images at different scales along with different patch sizes of the images. The local deviations between the neighboring pixels are used for the extraction of sharp regions. The automatic image masking and Otsu thresholding provide highly accurate and optimal segmented results over state-of-the-art methods. Our experimental results demonstrated that the proposed method performs far better than many hard-coded threshold-based algorithms. Additionally, the proposed method has a significant speed advantage over several comparative segmentation algorithms, i.e., alpha matting, KNN matting, global matting by GPU implementation, etc.

Data Availability

The datasets used and analyzed in this paper are publicly available.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

References

[1] W. Wang, "Salient object detection in the deep learning era: an in-depth survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, In press, 2021.

[2] C. Li, W. Chen, and Y. Tan, "Point-sampling method based on 3D U-net architecture to reduce the influence of false positive and solve boundary blur problem in 3D CT image segmentation," *Applied Sciences*, vol. 10, no. 19, p. 6838, 2020.

[3] X. Chen, "Generative adversarial U-net for domain-free medical image augmentation," 2021, https://arxiv.org/abs/2101.04793.

[4] A. Garnica-Carrillo, "Multi-focus image fusion for multiple images using adaptable size windows and parallel programming," *Signal, Image and Video Processing*, vol. 14, no. 7, pp. 1293–1300, 2020.

[5] A. Shen, H. Dong, K. Wang, Y. Kong, J. Wu, and H. Shu, "Automatic extraction of blur regions on a single image based on semantic segmentation," *IEEE Access*, vol. 8, pp. 44867–44878, 2020.

[6] M. T. Mahmood, U. Ali, and Y. K. Choi, "Single image defocus blur segmentation using Local Ternary Pattern," *ICT Express*, vol. 6, no. 2, pp. 113–116, 2020.

[7] X. Chen, "Motion blur detection based on lowest directional high-frequency energy," in *Proceedings of the 2010 IEEE International Conference on Image Processing*, September 2010.

[8] H. Xiao, W. Lu, R. Li et al., "Defocus blur detection based on multiscale SVD fusion in gradient domain," *Journal of Visual Communication and Image Representation*, vol. 59, pp. 52–61, 2019.

[9] R. Liu, Z. Li, and J. Jia, "Image partial blur detection and classification," in *Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition*, June 2008.

[10] E. Kalalembang, K. Usman, and I. P. Gunawan, "DCT-based local motion blur detection," in *Proceedings of the International Conference on Instrumentation, Communication, Information Technology, and Biomedical Engineering*, November 2009.

[11] E. Mavridaki and V. Mezaris, "No-reference blur assessment in natural images using fourier transform and spatial pyramids," in *Proceedings of the 2014 IEEE International Conference on Image Processing (ICIP)*, October 2014.

[12] J. Shi, L. Xu, and J. Jia, "Just noticeable defocus blur detection and estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, June 2015.

[13] C. T. Vu, T. D. Phan, and D. M. Chandler, "${\bf S} _ {3} $: a spectral and spatial measure of local perceived sharpness in natural images," *IEEE Transactions on Image Processing*, vol. 21, no. 3, pp. 934–945, 2011.

[14] N. D. Narvekar and L. J. Karam, "A no-reference image blur metric based on the cumulative probability of blur detection (CPBD)," *IEEE Transactions on Image Processing*, vol. 20, no. 9, pp. 2678–2683, 2011.

[15] R. Huang, W. Feng, M. Fan, L. Wan, and J. Sun, "Multiscale blur detection by learning discriminative deep features," *Neurocomputing*, vol. 285, pp. 154–166, 2018.

[16] L. D'Andrès, J. Salvador, A. Kochale, and S. Susstrunk, "Nonparametric blur map regression for depth of field extension," *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1660–1673, 2016.

[17] S. Gur and L. Wolf, "Single image depth estimation trained via depth from defocus cues," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, June 2019.

[18] A. Khan, "Defocus blur detection using novel local directional mean patterns (LDMP) and segmentation via KNN matting," *Frontiers of Computer Science*, 2020.

[19] A. Khan, "Segmentation of defocus blur using local triplicate Co-occurrence patterns (LTCoP)," in *Proceedings of the 2019 13th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS)*, December 2019.

[20] M. Ma, W. Lu, and W. Lyu, "Defocus blur detection via edge pixel DCT feature of local patches," *Signal Processing*, vol. 176, Article ID 107670, 2020.

[21] X. Yi and M. Eramian, "LBP-based segmentation of defocus blur," *IEEE Transactions on Image Processing*, vol. 25, no. 4, pp. 1626–1638, 2016.

[22] X. Zhu, S. Cohen, S. Schiller, and P. Milanfar, "Estimating spatially varying defocus blur from a single image," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4879–4891, 2013.

[23] A. Chakrabarti, T. Zickler, and W. T. Freeman, "Analyzing spatially-varying blur," in *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2010.

[24] B. Su, S. Lu, and C. L. Tan, "Blurred image region detection and classification," in *Proceedings of the 19th ACM international Conference on Multimedia*, Singapore, 2011.

[25] S. A. Golestaneh and L. J. Karam, "Spatially-varying blur detection based on multiscale fused and sorted transform coefficients of gradient magnitudes," in *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, July 2017.

[26] W. Zhao, "Enhancing diversity of defocus blur detectors via cross-ensemble network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, June 2019.

[27] Y. Pang, H. Zhu, X. Li, and X. Li, "Classifying discriminative features for blur detection," *IEEE Transactions on Cybernetics*, vol. 46, no. 10, pp. 2220–2227, 2015.

[28] T. Zhou, H. Fu, C. Sun, and S. Wang, "Shadow detection and compensation from remote sensing images under complex urban conditions," *Remote Sensing*, vol. 13, no. 4, p. 699, 2021.

[29] Y. Zhong, Z. Zheng, A. Ma, X. Lu, and L. Zhang, "COLOR: cycling, offline learning, and online representation framework for airport and airplane detection using GF-2 satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 58, no. 12, pp. 8438–8449, 2020.

[30] R. Liu, W. Yuan, X. Chen, and Y. Lu, "An enhanced CNN-enabled learning method for promoting ship detection in maritime surveillance system," *Ocean Engineering*, vol. 235, Article ID 109435, 2021.

[31] S. Arabi, A. Haghighat, and A. Sharma, "A deep-learning-based computer vision solution for construction vehicle detection," *Computer-Aided Civil and Infrastructure Engineering*, vol. 35, no. 7, pp. 753–767, 2020.

[32] B. Kim, "Defocus and motion blur detection with deep contextual features," in *Computer Graphics Forum* Wiley Online Library, Hoboken, NJ, USA, 2018.

[33] J. Park, Y.-W. Tai, D. Cho, and I. S. Kweon, "A unified approach of multi-scale deep and hand-crafted features for defocus estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, July 2017.

[34] C. Tang, X. Zhu, X. Liu, L. Wang, and A. Zomaya, "Defusionnet: defocus blur detection via recurrently fusing and refining multi-scale deep features," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, USA, June 2019.

[35] Y. Zhai, J. Wang, J. Deng, G. Yue, W. Zhang, and C. Tang, "Global context guided hierarchically residual feature refinement network for defocus blur detection," *Signal Processing*, vol. 183, Article ID 107996, 2021.

[36] X. Yang, X. Shen, J. Long, and H. Chen, "An improved median-based otsu image thresholding algorithm," *AASRI Procedia*, vol. 3, pp. 468–473, 2012.

[37] W. Zhao, "Defocus blur detection via multi-stream bottom-top-bottom fully convolutional network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, June 2018.

[38] J. Shi, L. Xu, and J. Jia, "Discriminative blur detection features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA, June 2014.