Enhancing Noise Robustness in Focus Measure Using Tight Framelet Features

Yan-Ran Li, Junwei Liu, Zhangtao Ye, Lixin Shen, Xiaosheng Zhuang

***Abstract*—Focus measures are widely used to assess image clarity in various fields, such as photography and computer vision. However, many existing focus measures face challenges in balancing noise robustness and measurement capability. In this letter, a novel focus measure called Variance of Tight Framelet Feature (VTFF) is proposed to address this challenge. VTFF leverages the advantages of tight framelet features and variance information in feature maps to provide a robust and accurate assessment of image focus. Experimental results on both synthetic and real-world data demonstrate its superior performance compared to recent focus measures in measurement capability, noise robustness, and real-time performance.**

***Index Terms*—Focus measure, tight framelet features, noise robustness**

I. Introduction

FOCUS measures calculate a value that quantifies the sharpness of an image based on variations in image intensity, gradient information, and other factors. This critical value is widely used in technologies such as autofocus [1]–[3], 3D shape estimation [4]–[6], depth-from-focus [7], [8], and multi-focus image fusion [9]–[11]. For instance, in autofocus systems, the imaging device measures the focus values at different lens positions and adjusts the lens position accordingly to achieve a sharply focused image of the target.

Existing focus measures are generally categorized into four main categories: derivative-based methods [1], [4], [10]–[12]; statistical-based methods [13]–[16]; transformation-based methods [17]–[23]; and other methods not belonging to the above three categories [7], [24], [25].

Among these four types of methods, derivative-based methods are the most typical ones and have been studied for the longest time. These methods extract gradient information from the image using operators such as the Sobel operator [1] and the Laplacian operator [4], [11], and then calculate the focus measure value by summing or squaring the absolute values of the gradient information. Multi-scale Weighted Modified Laplacian (MSWML) [10] and Reduced Tenengrad (RT) [12] are examples of derivative-based methods.

As an image transitions from blurry to sharp, the intensity distribution of its pixels shifts from dispersed to concentrated, which is reflected in the image's statistical properties. This relationship has led to the development of methods based on image statistics, such as variance [13], absolute central moment [14], and eigenvalues [15]. A focus measure leveraging improved Tsallis entropy [26] was also proposed to enhance the robustness and accuracy of focus assessment [16].

Transformation-based methods use various transformation techniques to convert the image from the spatial domain to the frequency domain and then calculate the focus measure value by directly quantifying the high-frequency information contained in the frequency domain. Commonly used transformation methods include wavelet transform [17], curvelet transform [18], and Hadamard transform [19]. Among these, methods based on Discrete Cosine Transform (DCT) have been extensively studied [20], [21], [22], [23]. For instance, a focus measure based on improved DCT coefficients was proposed by analyzing the energy distribution of DCT coefficients [22]. Furthermore, the Ratio of Higher and Lower DCT coefficients (RHLD), which examines the ratio of higher-order to lower-order DCT coefficients, was introduced in [23].

In addition to these categories, several other innovative focus measures have been proposed in recent years. For example, a method based on feature information extracted using a ring difference filter is introduced in [8]. Another method applies mean and median filters to correct feature information extracted by a max-min filter for improved feature measurement [24]. The Difference of Gaussian (DoG) focus measure uses the DoG operator to extract image features and calculates the focus measure value using the Minkowski distance [25].

Focus measures can also be broadly categorized as those that directly or indirectly analyze the high-frequency information in an image. In practical applications, the presence of noise in images inevitably impacts the accuracy and reliability of these focus measures [27]. Noise encountered in images captured by imaging devices is often modeled as Gaussian noise or speckle noise [25]. Since noise also resides in the high-frequency domain, current focus measures are highly sensitive to noise, often resulting in poor noise robustness. This sensitivity can lead to inaccurate focus assessments, as noise may be misinterpreted as sharp details. For example, though the Difference of Gaussian (DoG) method [25] demonstrates

The work of Y.-R. Li was supported in part by the National Natural Science Foundation of China (NSFC 12471400) and the Shenzhen Science and Technology Program (JCYJ20230808105610021). The work of L. Shen was supported in part by the National Science Foundation under grants DMS-1913039 and DMS-2208385, and Syracuse CUSE grant. The work of X. Zhuang was supported in part by the Research Grants Council of Hong Kong (Project no. CityU 11309122, CityU 11302023, and CityU 11301224) and a grant from the Innovation and Technology Commission of Hong Kong (Project no. MHP/054/22).

Y.-R. Li, J. Liu, and Z. Ye are with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, 518060, China (e-mails: lyran@szu.edu.cn, 2200271023@email.szu.edu.cn, 1085723418@qq.com).

L. Shen is with the Department of Mathematics, Syracuse University, Syracuse, NY 13244 (e-mail: lshen03@syr.edu).

X. Zhuang is with the Department of Mathematics, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong, Hong Kong SAR, China (e-mail: xzhuang7@cityu.edu.hk).

some level of noise resilience, its feature extraction process tends to cause detail loss and introduce blocky artifacts.

To address these challenges, we propose in this letter a novel focus measure that leverages tight framelet features. This approach's motivation is to extract image features effectively while mitigating noise interference, ensuring both noise robustness and reliable focus measurement, even in noisy environments.

The remainder of this letter is organized as follows: Section II presents in detail the proposed focus measure based on tight framelet features; Section III evaluates the performance of the proposed method using multiple metrics and compares it against recently developed methods on both synthetically generated and real-world datasets; and Section IV concludes the letter with a summary and contributions.

II. Variance of Tight Framelet Feature (VTFF)

In this section, we propose a focus measure called the Variance of Tight Framelet Features (VTFF), designed to achieve both noise robustness and accurate measurement capability.

Framelet systems are mathematical representation systems used in signal and image processing that build upon the framework of multiresolution (multiscale) analysis (MRA). The framelet systems provide a structured approach to decomposing signals into components at different scales or resolutions, which is particularly useful for detecting localized features and analyzing images. The framelet systems are, in general, redundant systems. The redundancy of framelet systems makes them robust against noise and errors in data. Their effectiveness has been demonstrated in various applications [28]–[34].

In this work, we employ the Two-level Non-stationary Tight Framelet (TNTF) system from our previous work [35] to extract multi-scale and multi-order features from the underlying images as well as to suppress noise. The TNTF system combines the directional Haar framelet (DHF) in [32] at the first level and the discrete cosine transform (DCT) tight framelet in [30] at the second level. Specifically, the low-pass subband of the DHF system produces smoothed or blurred versions of the original image, significantly mitigating the effects of noise. This preprocessing step enhances the DCT system's ability to accurately extract second-order features from the images, even under noisy conditions.

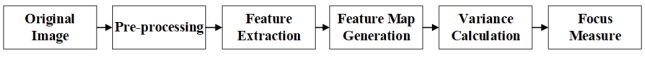

Fig. 1. Framework of Variance of Tight Framelet Features (VTFF).

The proposed VTFF framework for focus measurement, illustrated in Fig. 1, consists of four key steps:

1) **Preprocessing:** To meet the real-time requirements of practical applications, color images are converted into grayscale during preprocessing. This ensures computational efficiency without compromising accuracy.
2) **Feature Extraction:** The TNTF system is applied to the grayscale images from the preprocessing step to extract feature information. This process yields one low-frequency subband and twelve high-frequency subband features. Through extensive testing with 2000 images from the Kadis-700K database [36] under Gaussian and speckle noise conditions (detailed in Appendix 1[1]), we identify and select specific subband features that optimize real-time performance and noise robustness for focus measurement. The experimental results demonstrate the use of the low-pass filter $\tau_0 = \frac{1}{4}\begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}$ from the DHF tight framelet to obtain a coarse approximation (a smoothed image) of the input image. Useful information is then extracted from the smoothed image using the high-pass filters $\kappa_1 = \frac{\sqrt{6}}{18}\begin{bmatrix}1 & 0 & -1\\ 1 & 0 & -1\\ 1 & 0 & -1\end{bmatrix}$ and $\kappa_2 = \frac{\sqrt{6}}{18}\begin{bmatrix}1 & 1 & 1\\ 0 & 0 & 0\\ -1 & -1 & -1\end{bmatrix}$ from the DCT tight framelet. This approach leverages the TNTF system to focus on key subband features, striking a balance between computational efficiency and robust feature extraction.
3) **Feature Map Generation:** The high-pass subband features obtained from κ_1 and κ_2 are combined into a single feature map by summing the squares of their coefficients. To extract statistical information from the feature map, we adopt the block variance statistics method [25], known for its effectiveness in suppressing noise and providing reliable results. The feature map is divided into N non-overlapping image blocks of a specified size. For each block, say the n-th block, we calculate S_n, the sum of all features within that block.
4) **Variance Calculation:** The final focus measure value, FM, is computed as the variance of the sums of features from all image blocks: $\mathrm{FM} = \frac{1}{N}\sum_{n=1}^{N}(S_n - \overline{S})^2$, where $\overline{S}$ represents the mean of the sums of features across all blocks. The performance is affected by the choice of block sizes. Experimental results (2000 images from the Kadis-700K dataset) in Appendix 1 indicate that a block size of 16×16 achieves optimal performance.

To ensure the VTFF measure performs effectively, extensive experimental analysis was conducted to determine the optimal combination of high-pass subbands and the appropriate block size for generating the global feature map. Detailed experimental procedures and results are provided in Appendix 1.

III. Numerical Experiments

In this section, we will quantitatively analyze the noise robustness, measurement capability, and real-time performance of the VTFF measure using synthetic and real-world data. Additionally, we will conduct performance comparisons with several state-of-the-art and representative focus measures: MMAM [24], MSWML [10], RHLD [23], RT [12], and DoG [25], to evaluate the effectiveness of our VTFF measure.

A. Experimental Data

The experimental process utilizes 81 images from the Kadid-10k database [36] and 25 images from the TID2013 database [37] as original data, following the experimental data generation methods referenced in [22] and [25]. Gaussian (kernel) functions with standard deviations ranging from 0 to

[1]Code and appendices are available at: http://staffweb1.cityu.edu.hk/xzhuang7/softs/#VTFF

3.75 (step size 0.25) are convolved with the original images to simulate a defocusing process and generate sequences of blurred images.

To evaluate the robustness of the focus measures against noise, these blurred image sequences are further corrupted with image noise, resulting in sequences of noisy blurred images. This process is illustrated in Fig. 2. In image processing, noise in images acquired by imaging devices is typically modeled as Gaussian or speckle noise [25]. Therefore, in each image frame, the added noise in this letter we use are the Gaussian and the speckle noise (both with a mean of 0 and a variance of 0.02). The experimental results based on the Kadid-10K database are presented in Table I and II, while the results from the TID2013 database are provided in Appendix 1. Although Salt-and-Pepper noise is not common in real-world scenarios in this case, we have still included it (with a noise density 0.02) in our experiments to demonstrate the effectiveness of our method. The experimental results involving Salt-and-Pepper noise are also presented in Appendix 1.

B. Evaluation Metrics

The evaluation metrics (See Appendix 2[2] for detailed definitions) for assessing the performance of focus measures can be categorized into the following three aspects.

1) **Measurement Capability.** The Effective Range (ER) [25] metric and Sensitivity Detection Ability (SDA) [22] metric are utilized to evaluate the performance of focus measures. The ER metric evaluates the range of variation in the focus measure curves across different blur levels and the SDA metric evaluates the variation between adjacent focus measure values near the correct focus position. A higher ER value indicates that the focus measure is capable of detecting a wider range of focus levels effectively, while a higher SDA value signifies sensitivity in detecting the sharpest focus.
2) **Noise Robustness.** To assess noise robustness, we use three metrics: Difference of Curve (DoC), Difference of ER (DoER), and Difference of SDA (DoSDA). The DoC measures the stability of the trend of the focus measure between blurred image sequences with and without noise. The DoER and DoSDA Quantify the degradation in ER and SDA metrics due to noise. Smaller values of these measures reflect better robustness to noise.
3) **Real-time Performance.** Real-time performance is evaluated based on the average processing time per image, referred to as Run Time. Experiments are conducted on a system equipped with an Intel i7-7700 3.60GHz CPU and 16GB of memory to ensure consistent and fair evaluation.

C. Comparison on Synthetic Data

This subsection presents a comparative analysis of the VTFF measure against five state-of-the-art focus measures according to their measurement capability, noise robustness, and real-time performance in the output experimental results in Tables I and II.

[2]Code and appendices are available at:http://staffweb1.cityu.edu.hk/xzhuang7/softs/#VTFF

a) **Measurement Capability.** From Table I, for the blurred image sequence, the MSWML method and the proposed VTFF method achieve the best results in the ER and SDA metrics, respectively. However, when Gaussian noise or speckle noise is introduced, the ER and SDA metrics of the MMAM, RHLD, MSWML, and RT methods degrade significantly. In contrast, the DoG and VTFF methods are less affected by noise, with the VTFF outperforming the DoG method in both metrics. This highlights the VTFF's superior measurement capability compared to the other methods.
b) **Noise Robustness.** The noise robustness comparison of the six methods is shown in Table II. The proposed VTFF method achieves the lowest values for the DoC, DoER, and DoER, showing that the VTFF method consistently outperforms the other five methods under both Gaussian and speckle noise. This highlights the VTFF's superior noise robustness compared to the other methods.
c) **Real-time Performance.** In terms of Run Time, the RT method achieves the shortest execution time. Although the VTFF method takes slightly longer, it still operates at the millisecond level, demonstrating strong real-time performance.

D. Comparison on Real-World Data

To further validate the robustness of the VTFF method against real-world noise, we conduct additional experiments using real-world data. The experimental data comprises 104 sets of image sequences captured in real-world scenarios, including both day- and night-time scenes. Each sequence consists of 50 frames, documenting the process of objects within the scene transitioning from blurred to clear and back to blurred again. For instance, in a noisy nighttime scene shown on the left side of Fig. 3, we selected a target region indicated by a red square. By extracting this region across all 50 frames, we generated the image sequence displayed on the right side of Fig. 3. Additionally, the experimental results for two more sets of nighttime noise data are provided in Appendix 1.

For this image sequence, values of the focus measure for each of the six methods are calculated for each frame resulting in six different lines in Fig. 4 (Left). Each focus measure (sample points in the figure) indicates the clarity of each frame in the image sequences. The frame with the highest focus measure value is identified as the clearest and best target, indicating the point of optimal focus within the sequence.

From Fig. 4 (Left) (and their numerical values), we see that the MSWML and DoG methods achieve their maximal values at Frame-47 and Frame-11, respectively, but the other four methods all achieve their maximal values at Frame-14. We display these three frames in Fig. 5. We can further see that Frame-11 and Frame-14 are generally similar, but the zoom-in areas, shown in Fig. 5(d) and (e) reveal that the target in Frame-14 is sharper and more clear at the edges compared to Frame-11. Therefore, we can conclude that Frame-14 is the correct focus position.

Despite peaking at Frame-14, the MMAM, RHLD, and RT, exhibit limited variation ranges and local maxima in their focus

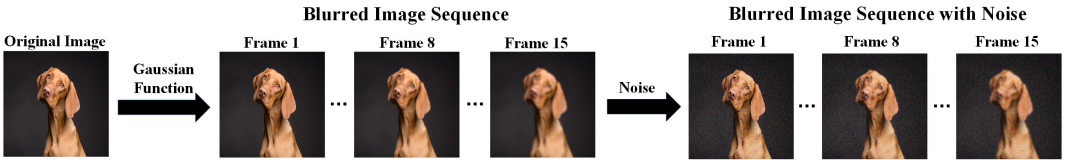

Fig. 2. The process of generating experimental data.

TABLE I
MEASUREMENT CAPABILITY AND REAL-TIME PERFORMANCE OF DIFFERENT FOCUS MEASURES

Focus Measure	Blurred Image Sequence		Blurred Image Sequence with Gaussian Noise		Blurred Image Sequence with Speckle Noise		Run Time(Seconds)↓
	$\overline{ER}\uparrow$	$\overline{SDA}\uparrow$	$\overline{ER}\uparrow$	$\overline{SDA}\uparrow$	$\overline{ER}\uparrow$	$\overline{SDA}\uparrow$	
MMAM [24]	1.0761	0.3403	0.0825	0.1268	0.2872	0.2358	1.7510
RHLD [23]	1.2108	0.3604	0.1225	0.1911	0.5888	0.3034	0.1029
MSWML [10]	**1.3662**	0.3476	0.0176	0.0752	0.0791	0.0924	0.0120
RT [12]	1.0092	0.3587	0.2040	0.2217	0.5106	0.3040	**0.0067**
DoG [25]	0.5414	0.3564	0.4595	0.3436	0.5168	0.3531	0.0136
VTFF	0.9324	**0.3692**	**0.8918**	**0.3686**	**0.9147**	**0.3688**	0.0188

TABLE II
ANTI-NOISE PERFORMANCE OF DIFFERENT FOCUS MEASURES

Focus Measure	$DoC\downarrow$	$DoER\downarrow$	$DoSDA\downarrow$
	Gaussian Noise		
MMAM [24]	2.2892	9.2533	1.9626
RHLD [23]	2.1946	9.9341	1.5892
MSWML [10]	3.0262	12.2469	2.4583
RT [12]	1.6287	7.3707	1.3012
DoG [25]	0.2044	0.8101	0.1358
VTFF	**0.0996**	**0.0439**	**0.0024**
	Speckle Noise		
MMAM [24]	1.3857	7.3638	1.0546
RHLD [23]	0.7748	6.0044	0.6657
MSWML [10]	2.5128	11.6819	2.3242
RT [12]	0.7361	4.7580	0.5950
DoG [25]	0.0616	0.2749	0.0410
VTFF	**0.0471**	**0.0210**	**0.0013**

Fig. 3. The real-world scene used for evaluation and the target image sequence.

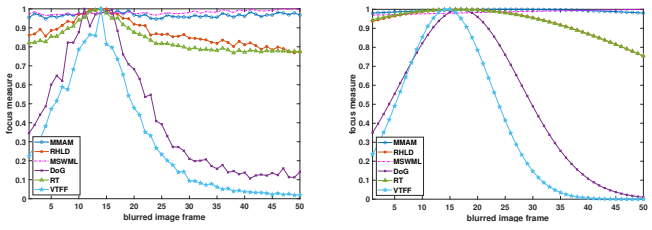

Fig. 4. Left: Measurement results of image sequences by different focus measure methods. Right: Curve fitting results of four focus measures.

(a) (b) (c) (d) (e)

Fig. 5. The comparisons of frame 11, 14, 47. (a) Frame 11, (b) Frame 14 , (c) Frame 47, (d) and (e) corresponding zoom-in part of (a) and (b).

measure curves as indicated in Fig. 4 (Left). These issues may prevent subsequent steps from accurately locating the correct focus position. In applications like autofocus and 3D shape estimation, the hill-climbing method [38] and curve fitting method [39] are commonly used to address such issues. The curve fitting method randomly samples points on the focus measure values that fit these points using a Gaussian curve model and eventually determines the focus position based on the maximum value of the fitted curve.

We performed such a curve-fitting method as well. The fitted curves from data in Fig. 4 (Left) for the six methods are shown in Fig. 4 (Right). From the fitted curves, we can find the frame with the maximum focus measure for each method as follows: MMAM (Frame-26), RHLD (Frame-17), MSWML (Frame-50), RT (Frame-17), DoG (Frame-16), and VTFF (Frame-14). We can see that, except for our VTFF, the fitted curves of the other five focus measure methods show a shift in the correct focus position. This shift can propagate errors to subsequent processes, potentially leading to inaccuracies in the final results.

IV. Conclusion

In this letter, we propose a novel focus measure VTFF method, which leverages the advantages of TNTF features and variance information in feature maps to provide a robust and accurate assessment of image focus. Experiments on both synthetic and real-world data scenarios demonstrate that VTFF measure outperforms state-of-the-art focus measures in terms of measurement capability, noise robustness, and real-time performance.

REFERENCES

[1] E. Krotkov, "Focusing," *International Journal of Computer Vision*, vol. 1, no. 3, pp. 223–237, 1988.

[2] M. Subbarao, T.-S. Choi, and A. Nikzad, "Focusing techniques," *Optical Engineering*, vol. 32, no. 11, pp. 2824–2836, 1993.

[3] Y. Zhang, L. Liu, W. Gong, H. Yu, W. Wang, C. Zhao, P. Wang, and T. Ueda, "Autofocus system and evaluation methodologies: A literature review." *Sensors & Materials*, vol. 30, 2018.

[4] S. K. Nayar and Y. Nakagawa, "Shape from focus," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 16, no. 8, pp. 824–831, 1994.

[5] B. Ahmad, H. Mutahira, M. Li, and M. S. Muhammad, "Measuring focus quality in color space," in *2019 2nd International Conference on Communication, Computing and Digital systems (C-CODE)*. IEEE, 2019, pp. 115–119.

[6] M. T. Mahmood and U. Ali, "Measuring focus quality in vector valued images for shape from focus," in *2022 26th International Conference on Pattern Recognition (ICPR)*. IEEE, 2022, pp. 3399–3405.

[7] M. Subbarao and J.-K. Tyan, "Selecting the optimal focus measure for autofocusing and depth-from-focus," *IEEE transactions on pattern analysis and machine intelligence*, vol. 20, no. 8, pp. 864–870, 1998.

[8] H.-G. Jeon, J. Surh, S. Im, and I. S. Kweon, "Ring difference filter for fast and noise robust depth from focus," *IEEE Transactions on Image Processing*, vol. 29, pp. 1045–1060, 2019.

[9] W. Huang and Z. Jing, "Evaluation of focus measures in multi-focus image fusion," *Pattern recognition letters*, vol. 28, no. 4, pp. 493–500, 2007.

[10] Z. Hu, W. Liang, D. Ding, and G. Wei, "An improved multi-focus image fusion algorithm based on multi-scale weighted focus measure," *Applied Intelligence*, vol. 51, no. 7, pp. 4453–4469, 2021.

[11] J. Wang, H. Qu, Y. Wei, M. Xie, J. Xu, and Z. Zhang, "Multi-focus image fusion based on quad-tree decomposition and edge-weighted focus measure," *Signal Processing*, vol. 198, p. 108590, 2022.

[12] I. Helmy and W. Choi, "Reduced tenegrad focus measure for performance improvement of astronomical images," in *2022 International Conference on Electronics, Information, and Communication (ICEIC)*. IEEE, 2022, pp. 1–4.

[13] L. Firestone, K. Cook, K. Culp, N. Talsania, and K. Preston Jr, "Comparison of autofocus methods for automated microscopy," *Cytometry: The Journal of the International Society for Analytical Cytology*, vol. 12, no. 3, pp. 195–206, 1991.

[14] M. V. Shirvaikar, "An optimal measure for camera focus and exposure," in *Thirty-Sixth Southeastern Symposium on System Theory, 2004. Proceedings of the IEEE*. IEEE, 2004, pp. 472–475.

[15] C.-Y. Wee and R. Paramesran, "Measure of image sharpness using eigenvalues," *Information Sciences*, vol. 177, no. 12, pp. 2533–2552, 2007.

[16] J. Rajevenceltha and V. H. Gaidhane, "A novel approach for image focus measure," *Signal, Image and Video Processing*, vol. 15, no. 3, pp. 547–555, 2021.

[17] H. Xie, W. Rong, and L. Sun, "Wavelet-based focus measure and 3-d surface reconstruction method for microscopy images," in *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2006, pp. 229–234.

[18] R. Minhas, A. A. Mohammed, and Q. J. Wu, "Shape from focus using fast discrete curvelet transform," *Pattern Recognition*, vol. 44, no. 4, pp. 839–853, 2011.

[19] L. Guo, X. Cao, and L. Liu, "A novel autofocus measure based on weighted walsh-hadamard transform," *IEEE Access*, vol. 7, pp. 22 107–22 117, 2019.

[20] S.-Y. Lee, Y. Kumar, J.-M. Cho, S.-W. Lee, and S.-W. Kim, "Enhanced autofocus algorithm using robust focus measure and fuzzy reasoning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 9, pp. 1237–1246, 2008.

[21] S.-Y. Lee, J.-T. Yoo, Y. Kumar, and S.-W. Kim, "Reduced energy-ratio measure for robust autofocusing in digital camera," *IEEE Signal Processing Letters*, vol. 16, no. 2, pp. 133–136, 2009.

[22] Z. Zhang, Y. Liu, Z. Xiong, J. Li, and M. Zhang, "Focus and blurriness measure using reorganized dct coefficients for an autofocus application," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 1, pp. 15–30, 2016.

[23] X. Nie, B. Xiao, X. Bi, W. Li, and X. Gao, "A focus measure in discrete cosine transform domain for multi-focus image fast fusion," *Neurocomputing*, vol. 465, pp. 93–102, 2021.

[24] S. Liu, Y. Lu, J. Wang, S. Hu, J. Zhao, and Z. Zhu, "A new focus evaluation operator based on max–min filter and its application in high quality multi-focus image fusion," *Multidimensional Systems and Signal Processing*, vol. 31, no. 2, pp. 569–590, 2020.

[25] L. Guo and L. Liu, "A perceptual-based robust measure of image focus," *IEEE Signal Processing Letters*, vol. 29, pp. 2717–2721, 2022.

[26] M.-S. Kang, J.-H. Bae, S.-H. Lee, and K.-T. Kim, "Efficient isar autofocus via minimization of tsallis entropy," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 6, pp. 2950–2960, 2016.

[27] V. A. Pimpalkhute, R. Page, A. Kothari, K. M. Bhurchandi, and V. M. Kamble, "Digital image noise estimation using dwt coefficients," *ieee transactions on image processing*, vol. 30, pp. 1962–1972, 2021.

[28] R. H. Chan, T. F. Chan, L. Shen, and Z. Shen, "Wavelet algorithms for high-resolution image reconstruction," *SIAM Journal on Scientific Computing*, vol. 24, no. 4, pp. 1408–1432, 2003.

[29] R. H. Chan, S. D. Riemenschneider, L. Shen, and Z. Shen, "Tight frame: an efficient way for high-resolution image reconstruction," *Applied and Computational Harmonic Analysis*, vol. 17, no. 1, pp. 91–115, 2004.

[30] Y.-R. Li, L. Shen, D.-Q. Dai, and B. W. Suter, "Framelet algorithms for de-blurring images corrupted by impulse plus gaussian noise," *IEEE Transactions on Image Processing*, vol. 20, no. 7, pp. 1822–1837, 2011.

[31] L. Shen, I. Kakadiaris, M. Papadakis, I. Konstantinidis, D. Kouri, and D. Hoffman, "Image denoising using a tight frame," *IEEE Transactions on Image Processing*, vol. 15, no. 5, pp. 1254–1263, 2006.

[32] Y.-R. Li, R. H. Chan, L. Shen, Y.-C. Hsu, and W.-Y. Isaac Tseng, "An adaptive directional haar framelet-based reconstruction algorithm for parallel magnetic resonance imaging," *SIAM Journal on Imaging Sciences*, vol. 9, no. 2, pp. 794–821, 2016.

[33] Y.-R. Li, L. Shen, and X. Zhuang, "A tailor-made 3-dimensional directional haar semi-tight framelet for pmri reconstruction," *Applied and Computational Harmonic Analysis*, vol. 60, pp. 446–470, 2022.

[34] Y. Li, R. H. Chan, L. Shen, X. Zhuang, R. Wu, Y. Huang, and J. Liu, "Exploring structural sparsity of coil images from 3-dimensional directional tight framelets for sense reconstruction," *SIAM Journal on Imaging Sciences*, vol. 17, no. 2, pp. 888–916, 2024.

[35] Y.-R. Li, R. H. Chan, L. Shen, and X. Zhuang, "Regularization with multilevel non-stationary tight framelets for image restoration," *Applied and Computational Harmonic Analysis*, vol. 53, pp. 332–348, 2021.

[36] H. Lin, V. Hosu, and D. Saupe, "Kadid-10k: A large-scale artificially distorted iqa database," in *2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX)*. IEEE, 2019, pp. 1–3.

[37] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti *et al.*, "Image database tid2013: Peculiarities, results and perspectives," *Signal processing: Image communication*, vol. 30, pp. 57–77, 2015.

[38] C.-M. Chen, C.-M. Hong, and H.-C. Chuang, "Efficient auto-focus algorithm utilizing discrete difference equation prediction model for digital still cameras," *IEEE Transactions on Consumer Electronics*, vol. 52, no. 4, pp. 1135–1143, 2006.

[39] H. Mir, P. Xu, and P. Van Beek, "An extensive empirical evaluation of focus measures for digital photography," in *Digital Photography X*, vol. 9023. SPIE, 2014, pp. 167–177.

Appendix 1
Selection of High-Pass Subband Features and Block Size of Feature Map

This appendix provides details on (I) how images are processed through the Two-level Non-stationary Tight Framelet (TNTF) system to extract features, (II) the selection of high-pass subband combinations of the TNTF system for best performance, (III) the determination of feature map block size, and (IV) further experiments on synthetic and real-world data.

I. Features Extraction via the Two-Level Non-Stationary Tight Framelet System

The TNTF system consists of the DHF tight framelet system and the DCT tight framelet system. First, the input (grayscale) image undergoes feature extraction using the DHF tight framelet system. This process is accomplished through the filter bank $\{\tau_0, \tau_1, \ldots, \tau_6\}$ corresponding to the DHF tight framelet, as shown in Eqs. (1). Here, τ_0 is the low-pass filter, and the other filters are high-pass filters designed to capture directional information in the image. Specifically, τ_1 , τ_2, τ_3 and τ_4 are used to extract feature information in the 45° and 135°, horizontal and vertical directions, respectively. The roles of τ_5 and τ_6 are the same as τ_3 and τ_4. To reduce redundancy, only $\tau_1, \tau_2, \tau_3, \tau_4$ are used for feature extraction in practice. Then, the low-pass subband feature (i.e., through τ_0) extracted by the DHF is further processed by the DCT tight framelet system using the filter bank $\{\kappa_0, \kappa_1, \ldots, \kappa_8\}$ shown in Eqs.(2). Here, κ_0 is the low-pass filter. κ_1 and κ_3 are used to extract first-order features in the horizontal and vertical directions, respectively. κ_2 and κ_6 are used to extract second-order features in the horizontal and vertical directions. The remaining filters are used to extract higher-order image features.

$$\begin{aligned} \tau_0 = \tfrac{1}{4}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \tau_1 = \tfrac{1}{4}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \tau_2 = \tfrac{1}{4}\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad \tau_3 = \tfrac{1}{4}\begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}, \\ \tau_4 = \tfrac{1}{4}\begin{bmatrix} 1 & 0 \\ -1 & 0 \end{bmatrix}, \quad \tau_5 = \tfrac{1}{4}\begin{bmatrix} 0 & 0 \\ 1 & -1 \end{bmatrix}, \quad \tau_6 = \tfrac{1}{4}\begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix}. \end{aligned} \tag{1}$$

$$\begin{aligned} \kappa_0 = \tfrac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \kappa_1 = \tfrac{\sqrt{6}}{18}\begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix}, \quad \kappa_2 = \tfrac{\sqrt{6}}{18}\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}, \\ \kappa_3 = \tfrac{\sqrt{2}}{18}\begin{bmatrix} 1 & -2 & 1 \\ 1 & -2 & 1 \\ 1 & -2 & 1 \end{bmatrix}, \quad \kappa_4 = \tfrac{1}{18}\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}, \quad \kappa_5 = \tfrac{\sqrt{2}}{18}\begin{bmatrix} 1 & 1 & 1 \\ -2 & -2 & -2 \\ 1 & 1 & 1 \end{bmatrix}, \\ \kappa_6 = \tfrac{\sqrt{3}}{18}\begin{bmatrix} 1 & -2 & 1 \\ 0 & 0 & 0 \\ -1 & 2 & -1 \end{bmatrix}, \quad \kappa_7 = \tfrac{\sqrt{3}}{18}\begin{bmatrix} 1 & 0 & -1 \\ -2 & 0 & 2 \\ 1 & 0 & -1 \end{bmatrix}, \quad \kappa_8 = \tfrac{1}{6}\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}. \end{aligned} \tag{2}$$

To illustrate the above TNTF feature extraction process, We use the color image in Fig. 1 as an example. After converting the color image to grayscale, it is first processed by the DHF tight frame, which extracts one low-pass subband and four high-pass subband features, as shown in Fig. 1. Specifically, Fig. 1(c) and (d) contain first-order features in the 45° and 135° directions, while Fig. 1(e) and (f) contain first-order features in the horizontal and vertical directions. Then, the low-pass subband feature (Fig. 1(b)) extracted by the DHF is further processed by the DCT, extracting additional features. This yields one low-pass

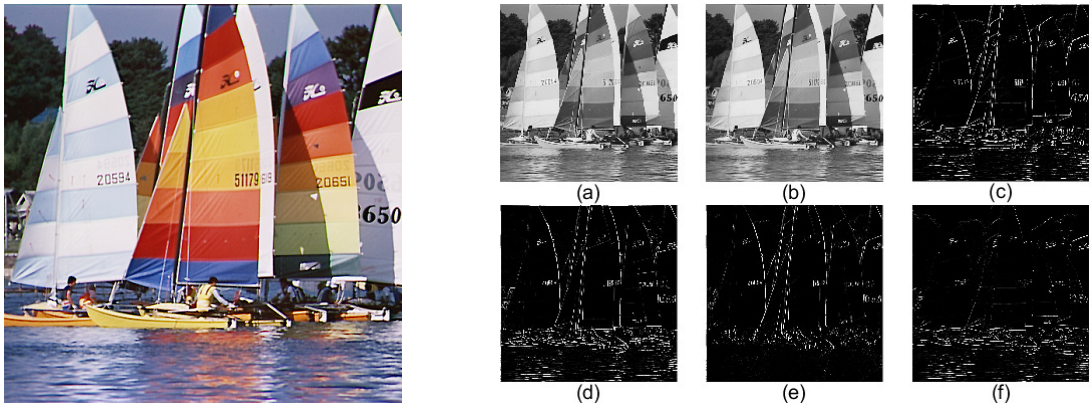

Fig. 1. A color sample image and its DHF tight framelet features. (a) Input grayscale image, converted from the color sample image. (b) Low-pass subband extracted by the DHF system. (c)-(f) High-pass subbands extracted by the DHF system.

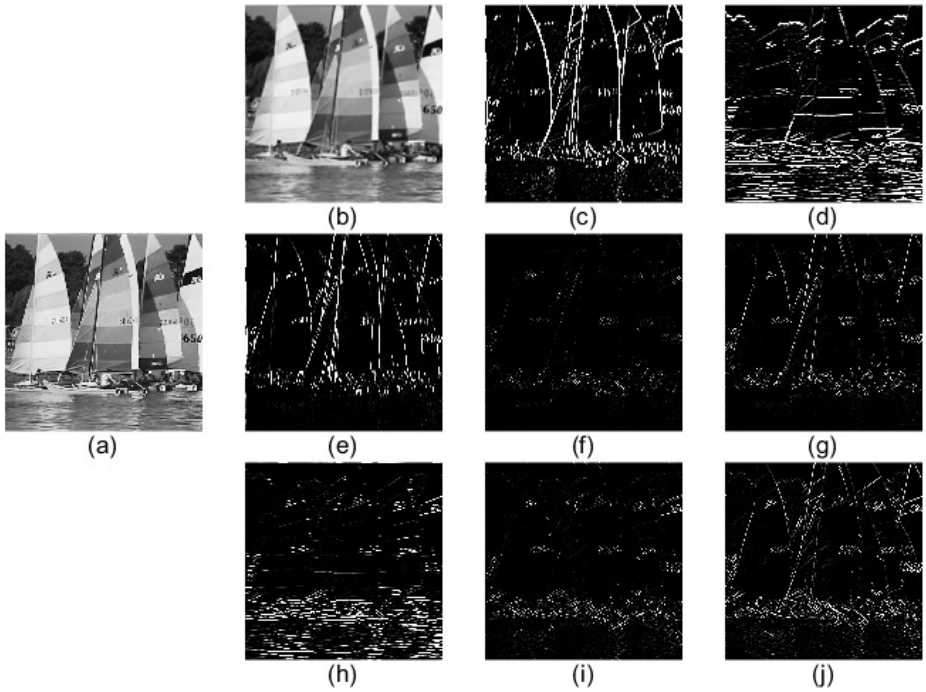

Fig. 2. High-pass features of the low-pass subband in Fig. 1(a) by the DCT tight framelet. (a) Low-pass subband extracted by the DHF system in Fig. 1(a). (b) Low-pass subband extracted by the DCT system. (c)-(j) High-pass subbands extracted by the DCT system.

subband and eight high-pass subband features, as shown in Fig. 2. Here, Fig. 2(c) and (d), (e) and (h) contain first-order and second-order features in the horizontal and vertical directions, respectively, while Fig. 2(f), (g), (i), and (j) contain higher-order features of the image.

It should be noted that the DCT uses the low-pass subband extracted by the DHF as input. According to the reference [1], this is because the DCT tight framelet involves the extraction of higher-order feature information, which is more susceptible to image noise. The low-pass subband results from the original image after smoothing, reducing the effect of noise. Therefore, using the low-pass subband as input makes the feature extraction process of the DCT more reliable.

The proposed VTFF focus measure uses high-pass subband combinations from the TNTF system to form the feature map, which is then divided into image blocks of specified sizes. The focus measure value is calculated by computing the variance of the sum of features within all image blocks. This framework involves two key aspects: the selection of high-pass subband combinations and the setting of the block size

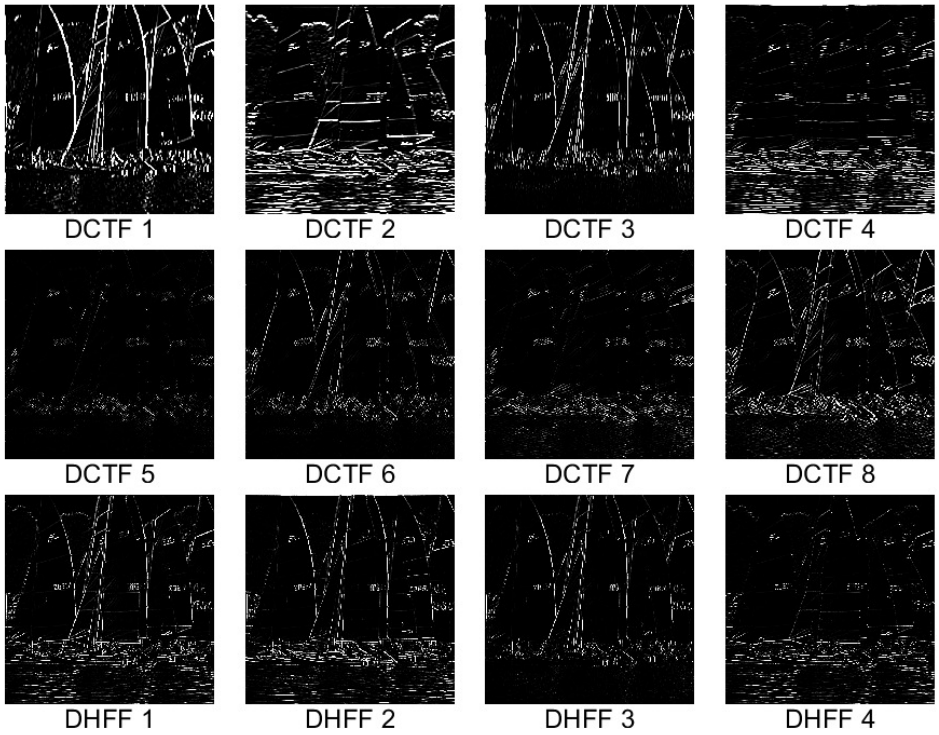

Fig. 3. The re-named DHF and DCT tight framelet high-pass subbands.

for the feature map. To this end, we next explore the performance of the VTFF with various combinations and block size using noise-free and noisy blurred image sequences, which are generated from 2000 images in the Kadis-700K database [2]. Through experimental comparative analysis, we aim to determine the optimal high-pass subband combinations and the appropriate block size for the feature map.

II. Selection of High-Pass Subband Combinations

For the 12 high-pass subband features extracted from the TNTF, this section will name the 8 high-pass subband features extracted by the DCT as 'DCTF1' to 'DCTF8' (w.r.t. $\kappa_1, \ldots, \kappa_8$), and the 4 high-pass subband features extracted by the DHF as 'DHFF1' to 'DHFF4' (w.r.t. $\tau_1, \ldots, \tau_4$) as shown in Fig. 3. These 12 high-pass subband features contain edge and texture details of the image.

Despite the denoising and smoothing operations during feature extraction by the TNTF system, some high-pass subband features remain sensitive to noise, potentially affecting the VTFF performance. Therefore, this experiment first analyzes the extent to which each subband is affected by noise and its sensitivity to noise. Based on this analysis, we select appropriate subband combinations and verify their noise robustness. Through experimental analysis, we aim to identify the optimal combination of high-pass subbands for achieving the best performance of the VTFF.

Firstly, Gaussian noise (or Speckle noise) with a mean of 0 and a variance of 0.02 (the parameters remain the same throughout the subsequent experiments), is added to the sample image shown on the left side of Fig. 1. The image features extracted by TNTF are shown in Fig. 4. Comparing these features with Fig. 3, it is evident that noise affects each high-pass subband to different extents. To evaluate this, Mean Square Error (MSE) is employed in this experiment to assess the extent to which the 12 high-pass subbands are affected by noise.

In this appendix, 2000 sets of experimental data are used, with each image sequence containing 15 frames. For the kth set of experimental data, I_{ktn} and $\hat{I}_{ktn}$ represent the nth high-pass subband extracted by the TNTF from the tth frame of the noise-free and nosiy blurred sequence, respectively. The resolution

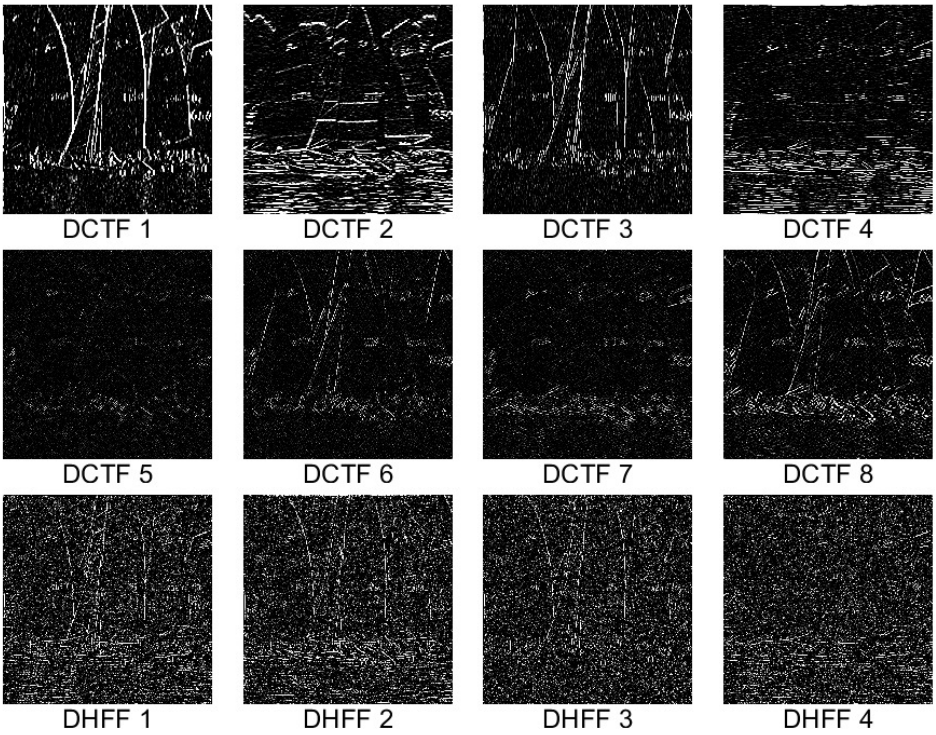

Fig. 4. The 12 high-pass subbands extracted from the color sample image in Fig. 1 after adding Gaussian noise.

of the high-pass subbands is $M \times N$. The MSE between the two high-pass subbands is calculated as shown below:

$$MSE_{ktn} = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \left(\hat{I}_{ktn}(i,j) - I_{ktn}(i,j) \right)^2 .$$

All MSE_{ktn} values can be used to further calculate the average MSE for each high-pass subband, denoted as $\overline{MSE}_n$, as shown below:

$$\overline{MSE}_n = \frac{1}{K \times T} \sum_{k=1}^{K} \sum_{t=1}^{T} MSE_{ktn},$$

where, K represents the number of experimental data sets, and T is the total number of frames in the image sequence. Specifically, $K = 2000$ and $T = 15$. A larger $\overline{MSE}_n$ value signifies a greater impact of noise on that high-pass subband. Depending on the type of noise, $\overline{MSE}_n$ values for high-pass subband features extracted by the TNTF system under Gaussian noise and speckle noise are shown in Table I and Table II, respectively.

TABLE I
$\overline{MSE}_n$ OF HIGH-PASS SUBBANDS EXTRACTED BY THE TNTF SYSTEM UNDER GAUSSIAN NOISE

Subband	DCTF1	DCTF2	DCTF3	DCTF4	DCTF5	DCTF6
$\overline{MSE}_n \downarrow$	14.20	14.12	13.81	13.65	13.93	13.64

Subband	DCTF7	DCTF8	DHFF1	DHFF2	DHFF3	DHFF4
$\overline{MSE}_n \downarrow$	13.73	13.87	62.30	62.35	62.11	62.45

The $\overline{MSE}_n$ index in Tables I and II indicates that the eight high-pass subband features extracted by the DCT are similarly affected by noise. The same observation applies to the four high-pass subband features

TABLE II
$\overline{MSE}_n$ OF HIGH-PASS SUBBANDS EXTRACTED BY THE TNTF SYSTEM UNDER SPECKLE NOISE

Subband	DCTF1	DCTF2	DCTF3	DCTF4	DCTF5	DCTF6
$\overline{MSE}_n \downarrow$	4.24	4.33	4.27	4.18	4.20	4.20

Subband	DCTF7	DCTF8	DHFF1	DHFF2	DHFF3	DHFF4
$\overline{MSE}_n \downarrow$	4.18	4.20	19.04	19.04	19.05	19.11

extracted by the DHF, with the former being less affected by noise than the latter. This is related to the feature extraction process of the TNTF system. As mentioned earlier, the DHF takes the original image as input, whereas the DCT uses the low-pass subband extracted by the DHF as input. The low-pass subband is the result of the original image being smoothed and denoised. Therefore, the DCT processes images with less noise, making its extracted high-pass subband features less susceptible to noise. If the image features were directly extracted by the DCT without processing by the DHF, the $\overline{MSE}_n$ of the high-pass subbands under different types of noise would be as shown in Tables III and IV.

TABLE III
$\overline{MSE}_n$ OF THE HIGH-PASS SUBBANDS EXTRACTED USING ONLY THE DCT SYSTEM UNDER GAUSSIAN NOISE

Subband	DCTF1	DCTF2	DCTF3	DCTF4	DCTF5	DCTF6	DCTF7	DCTF8
$\overline{MSE}_n \downarrow$	55.47	55.79	55.27	55.14	55.88	54.70	55.30	55.01

TABLE IV
$\overline{MSE}_n$ OF THE HIGH-PASS SUBBANDS EXTRACTED USING ONLY THE DCT SYSTEM UNDER SPECKLE NOISE

Subband	DCTF1	DCTF2	DCTF3	DCTF4	DCTF5	DCTF6	DCTF7	DCTF8
$\overline{MSE}_n \downarrow$	17.00	16.87	16.84	16.87	16.93	16.89	17.07	16.92

By comparing the $\overline{MSE}_n$ values for corresponding high-pass subband features in Tables I, II, III, and IV, it is evident that the extent to which high-pass subbands are affected by noise significantly increases when using only the DCT system to extract image features. This demonstrates that using the low-pass subband features extracted by the DHF as input makes the DCT extraction process more effective and reliable.

The data from the $\overline{MSE}_n$ metric suggest that the high-pass subband features extracted by the DCT system are relatively less affected by noise than those extracted by the DHF system, making them suitable candidates for application in VTFF. To further select the appropriate high-pass subbands from the eight candidates, this experiment uses the Noise Energy Ratio (NER) metric to analyze the sensitivity of the candidate high-pass subbands to noise. The specific definition of NER is as follows:

$$NER_{ktn} = \frac{\left|\hat{E}_{ktn} - E_{ktn}\right|}{E_{ktn}},$$

where, $\hat{E}_{ktn}$ represents the energy contained in the nth high-pass subband feature extracted by the DCT from the tth frame of the noisy blurred sequence in the kth set of experimental data, termed as noise signal energy, given by $\hat{E}_{ktn} = \sum_{i=1}^{M}\sum_{j=1}^{N}(\hat{I}_{ktn}(i,j))^2$. E_{ktn} is the original signal energy, which is the energy contained in the nth high-pass subband feature extracted by the DCT system from the tth frame of the noise-free blurred sequence, defined as $E_{ktn} = \sum_{i=1}^{M}\sum_{j=1}^{N}(I_{ktn}(i,j))^2$. The difference between these two values represents the energy produced by noise in the nth high-pass subband, known as the noise

energy. The NER_{ktn} values computed from all experimental data are used to obtain the average NER for each high-pass subband using Equation

$$\overline{NER}_n = \frac{1}{K \times T} \sum_{k=1}^{K} \sum_{t=1}^{T} NER_{ktn},$$

where K and T are the same as before. A larger $\overline{NER}_n$ indicates greater sensitivity of high-pass subband to noise. The $\overline{NER}_n$ values for the candidate high-pass subbands under Gaussian noise and speckle noise are shown in Tables V and VI, respectively.

TABLE V
$\overline{NER}_n$ OF THE HIGH-PASS SUBBANDS EXTRACTED BY THE DCT SYSTEM UNDER GAUSSIAN NOISE

Subband	DCTF1	DCTF2	DCTF3	DCTF4	DCTF5	DCTF6	DCTF7	DCTF8
$\overline{NER}_n \downarrow$	**0.87**	**0.73**	16.17	694.81	10.51	184.16	163.95	11.64

TABLE VI
$\overline{NER}_n$ OF THE HIGH-PASS SUBBANDS EXTRACTED BY THE DCT SYSTEM UNDER SPECKLE NOISE

Subband	DCTF1	DCTF2	DCTF3	DCTF4	DCTF5	DCTF6	DCTF7	DCTF8
$\overline{NER}_n \downarrow$	**0.27**	**0.24**	5.22	229.92	3.33	60.88	53.40	3.70

According to the data in the Tables V and VI, DCTF 1 and DCTF 2 are less sensitive to noise compared to other high-pass subband features. Therefore, combinations formed solely by these two subbands {DCTF1} and {DCTF2}, or the combination of the two {DCTF1, DCTF2}, or combinations formed by adding other high-pass subbands to these two {DCTF1, DCTF2, DCTF3}, {DCTF1, DCTF2, DCTF4}, {DCTF1, DCTF2, DCTF5}, {DCTF1, DCTF2, DCTF6}, {DCTF1, DCTF2, DCTF7}, {DCTF1, DCTF2, DCTF8}, {DCTF1, DCTF2, DHFF1}, {DCTF1, DCTF2, DHFF2}, {DCTF1, DCTF2, DHFF3}, and {DCTF1, DCTF2, DHFF4}, a total of 13 high-pass subband combinations, are all potential candidates for providing good noise robustness in the VTFF.

Therefore, this experiment evaluates the noise robustness of the VTFF using these 13 high-pass subband combinations, using DoC, DoER, and DoSDA as performance metrics. Since each DoC metric corresponds to a single set of experiments, the average value of the DoC metrics from multiple sets of experiments, $\overline{DoC}$, is used to ensure the validity of the experiment. To control variables, the experiment does not partition the total feature map, but instead calculates the variance of the feature map on a per-pixel basis. Based on the aforementioned experimental setup, the noise robustness of the VTFF using the 13 high-pass subband combinations under Gaussian noise and speckle noise is shown in Tables VII and VIII.

According to the data in Tables VII and VIII, when the feature maps are not partitioned into blocks, the VTFF using {DCTF1, DCTF2} as the high-pass subband combination achieves the smallest values for $\overline{DoC}$, $DoER$, and $DoSDA$ under the influence of Gaussian and speckle noise. This indicates that the method has the best anti-noise performance. Therefore, the {DCTF1, DCTF2} combination will be used as the high-pass subband combination for our VTFF.

III. FEATURE MAP BLOCK SIZE SETTING

After selecting the appropriate high-pass subband combination, this section further analyzes the impact of total feature map block size on the performance of the VTFF through experiments. The goal is to determine the optimal block size for the total feature map. The experiments measure the anti-noise performance of the VTFF using the DoC, DoER, and DoSDA metrics for seven common block size schemes: no blocking (block size of 1), and block sizes of 2, 4, 8, 16, 32, and 64. Similarly, the average

TABLE VII
ANTI-NOISE PERFORMANCE OF VTFF WITH DIFFERENT HIGH-PASS SUBBAND COMBINATIONS UNDER GAUSSIAN NOISE (WITHOUT BLOCK PARTITIONING OF FEATURE MAPS)

High-pass Subband Combinations	$\overline{DoC} \downarrow$	$DoER \downarrow$	$DoSDA \downarrow$
{DCTF1}	0.2895	5.6880	1.3666
{DCTF2}	0.2581	4.9144	1.1857
{DCTF1, DCTF2}	**0.1850**	**3.9204**	**0.8912**
{DCTF1, DCTF2, DCTF3}	0.2784	6.0418	1.2632
{DCTF1, DCTF2, DCTF4}	0.3518	6.6495	1.3808
{DCTF1, DCTF2, DCTF5}	0.2671	5.8424	1.2188
{DCTF1, DCTF2, DCTF6}	0.3290	6.6641	1.3682
{DCTF1, DCTF2, DCTF7}	0.3281	6.6826	1.3649
{DCTF1, DCTF2, DCTF8}	0.2595	5.8538	1.2178
{DCTF1, DCTF2, DHFF1}	0.4799	8.9803	2.1718
{DCTF1, DCTF2, DHFF2}	0.4820	8.9993	2.1795
{DCTF1, DCTF2, DHFF3}	0.5266	9.3527	2.2953
{DCTF1, DCTF2, DHFF4}	0.5189	9.2407	2.2718

TABLE VIII
ANTI-NOISE PERFORMANCE OF VTFF WITH DIFFERENT HIGH-PASS SUBBAND COMBINATIONS UNDER SPECKLE NOISE (WITHOUT BLOCK PARTITIONING OF FEATURE MAPS)

High-pass Subband Combinations	$\overline{DoC} \downarrow$	$DoER \downarrow$	$DoSDA \downarrow$
{DCTF1}	0.0832	2.2640	0.5259
{DCTF2}	0.0757	1.9202	0.4570
{DCTF1, DCTF2}	**0.0512**	**1.2927**	**0.2990**
{DCTF1, DCTF2, DCTF3}	0.0888	2.2963	0.4950
{DCTF1, DCTF2, DCTF4}	0.1319	2.8782	0.5685
{DCTF1, DCTF2, DCTF5}	0.0842	2.1558	0.4725
{DCTF1, DCTF2, DCTF6}	0.1142	2.7490	0.5575
{DCTF1, DCTF2, DCTF7}	0.1138	2.7575	0.5561
{DCTF1, DCTF2, DCTF8}	0.0795	2.1708	0.4643
{DCTF1, DCTF2, DHFF1}	0.1571	3.7477	0.8872
{DCTF1, DCTF2, DHFF2}	0.1580	3.7521	0.8894
{DCTF1, DCTF2, DHFF3}	0.1859	4.1789	0.9791
{DCTF1, DCTF2, DHFF4}	0.1813	4.0669	0.9610

TABLE IX
ANTI-NOISE PERFORMANCE OF VTFF WITH DIFFERENT BLOCK SIZES UNDER GAUSSIAN NOISE (USING THE HIGH-PASS SUBBAND COMBINATION {DCTF1,DCTF2})

Total Feature Map Block Size	$\overline{DoC} \downarrow$	$DoER \downarrow$	$DoSDA \downarrow$
No Blocking	0.1850	3.9204	0.8912
Block Size of 2	0.1276	4.3015	0.2429
Block Size of 4	0.1028	3.0699	0.1734
Block Size of 8	0.0826	2.2287	**0.1282**
Block Size of 16	**0.0715**	**1.8338**	**0.1108**
Block Size of 32	**0.0662**	**1.7345**	**0.1126**
Block Size of 64	**0.0653**	**1.8330**	0.1343

TABLE X
ANTI-NOISE PERFORMANCE OF VTFF WITH DIFFERENT BLOCK SIZES UNDER SPECKLE NOISE (USING THE HIGH-PASS SUBBAND COMBINATION {DCTF1,DCTF2})

Total Feature Map Block Size	$\overline{DoC}\downarrow$	$DoER\downarrow$	$DoSDA\downarrow$
No Blocking	0.0512	1.2927	0.2990
Block Size of 2	0.0442	1.5436	0.0729
Block Size of 4	0.0393	1.2401	**0.0571**
Block Size of 8	**0.0355**	**1.0576**	**0.0540**
Block Size of 16	**0.0349**	**1.0280**	**0.0633**
Block Size of 32	**0.0371**	**1.1428**	0.0913
Block Size of 64	0.0432	1.4805	0.1693

value $\overline{DoC}$ of the DoC metric results from multiple experiments is used in this section. Tables IX and X present the experimental results under the influence of Gaussian noise and speckle noise, respectively.

The experimental results of Tables IX and X indicate that under Gaussian noise, the VTFF demonstrates better anti-noise performance with block sizes of 16, 32, and 64. Under speckle noise, the VTFF performs better with block sizes of 8, 16, and 32. Therefore, when using the {DCTF1, DCTF2} combination as the high-pass subband combination, a block size of 16 is suitable for both Gaussian and speckle noise, showing good anti-noise performance. Consequently, setting the block size of the total feature map to 16 is most appropriate.

Based on the experimental analysis in these two aspects, the VTFF finally **adopts {DCTF1, DCTF2} as the high-pass subband combination** and **sets the block size of the total feature map to 16** to ensure optimal performance.

IV. THE EXPERIMENTS WITH SYNTHETIC DATA AND REAL-WORLD SCENES

A. Kadid-10K Dataset

In image processing, noise present in images acquired by imaging devices is typically modeled as Gaussian noise or speckle noise [7]. To further validate the robustness of the proposed algorithm, we assess the noise robustness, measurement capability, and real-time performance of the VTFF using synthetic data under salt-and-pepper noise (with noise density of 0.02). Through experiments, we find that the optimal subband combination {DCTF1, DCTF2} and block size 16 of the total feature map obtained from Gaussian noise and Speckle noise tests are also applicable under salt-and-pepper noise. We conduct testing on the Kadid-10K database, which contains 81 images [2].

TABLE XI
MEASUREMENT CAPABILITY AND REAL-TIME PERFORMANCE OF DIFFERENT FOCUS MEASURES ON THE KADID-10K DATABASE [2] UNDER SALT-AND-PEPPER NOISE

Focus Measure	Blurred Image Sequence with Salt-and-Pepper Noise		Run Time(Seconds)↓
	$\overline{ER}\uparrow$	$\overline{SDA}\uparrow$	
MMAM [3]	0.4451	0.2880	4.1834
RHLD [4]	0.2997	0.2659	0.3155
MSWML [5]	0.2328	0.2165	0.0310
RT [6]	0.4366	0.2920	**0.0100**
DoG [7]	0.5176	0.3528	0.0384
VTFF	**0.9314**	**0.3691**	0.0554

As shown in Table XI, VTFF outperforms the other methods in both ER and SDA metrics, indicating its superior measurement capability under Salt-and-Pepper noise. Although slower than RT in terms of Run time, VTFF still operates at the millisecond level. The performance of VTFF in terms of DoC, DoER,

TABLE XII
ANTI-NOISE PERFORMANCE OF DIFFERENT FOCUS MEASURES ON THE KADID-10K DATABASE [2] UNDER SALT-AND-PEPPER NOISE

Focus Measure	Salt-and-Pepper Noise		
	$DoC\downarrow$	$DoER\downarrow$	$DoSDA\downarrow$
MMAM [3]	0.8998	5.8617	0.5749
RHLD [4]	1.5016	8.3642	0.9967
MSWML [5]	1.7082	10.2983	1.2444
RT [6]	0.8860	5.3246	0.6867
DoG [7]	0.0512	0.2566	0.0409
VTFF	**0.0038**	**0.0344**	**0.0015**

and DoSDA indices under salt-and-pepper noise conditions, as shown in Table XII, further demonstrates its superior noise robustness, despite not testing the optimal subband combination and block size of the total feature map under these conditions.

TABLE XIII
MEASUREMENT CAPABILITY AND REAL-TIME PERFORMANCE OF DIFFERENT FOCUS MEASURES ON THE TID2013 DATABASE [8]

Focus Measure	Blurred Image Sequence		Blurred Image Sequence with Gaussian Noise		Blurred Image Sequence with Speckle Noise		Blurred Image Sequence with Salt-and-Pepper Noise		Run Time(Seconds)↓
	$\overline{ER}\uparrow$	$\overline{SDA}\uparrow$	$\overline{ER}\uparrow$	$\overline{SDA}\uparrow$	$\overline{ER}\uparrow$	$\overline{SDA}\uparrow$	$\overline{ER}\uparrow$	$\overline{SDA}\uparrow$	
MMAM [3]	1.2976	0.3254	0.0950	0.1233	0.3361	0.2364	0.5379	0.2904	4.9990
RHLD [4]	1.3876	0.3505	0.1211	0.1801	0.6098	0.2975	0.3064	0.2572	0.2898
MSWML [5]	**1.6342**	0.3261	0.0239	0.0702	0.1009	0.1002	0.3198	0.2323	0.0202
RT [6]	1.1247	0.3547	0.2109	0.2154	0.5642	0.3051	0.4674	0.2883	0.0274
DoG [7]	0.5407	0.3560	0.4450	0.3405	0.5153	0.3527	0.5111	0.3514	**0.0067**
VTFF	0.9494	**0.3681**	**0.9057**	**0.3674**	**0.9157**	**0.3673**	**0.9477**	**0.3680**	0.0364

TABLE XIV
ANTI-NOISE PERFORMANCE OF DIFFERENT FOCUS MEASURES ON THE TID2013 DATABASE [8]

Focus Measure	Gaussian Noise			Speckle Noise			Salt-and-Pepper Noise		
	$DoC\downarrow$	$DoER\downarrow$	$DoSDA\downarrow$	$DoC\downarrow$	$DoER\downarrow$	$DoSDA\downarrow$	$DoC\downarrow$	$DoER\downarrow$	$DoSDA\downarrow$
MMAM [3]	2.4122	6.0954	1.0361	1.3748	4.8958	0.4967	0.8240	3.8832	0.2083
RHLD [4]	2.3249	6.3665	0.8842	0.7908	4.0516	0.3066	1.5185	5.4606	0.5176
MSWML [5]	3.1035	8.0724	1.2846	2.4740	7.6993	1.1408	1.4403	6.6245	0.4797
RT [6]	1.7078	4.6159	0.7331	0.7061	2.9057	0.2867	0.9094	3.3587	0.3710
DoG [7]	0.2407	0.5076	0.0874	0.0719	**0.1400**	0.0193	0.0647	0.1654	0.0268
VTFF	**0.0865**	**0.2317**	**0.0069**	**0.0668**	0.1915	**0.0062**	**0.0046**	**0.0211**	**0.0013**

B. TID2023 Dataset

Next, we will perform experiments on the TID2013 dataset [8], which consists of 25 images. To simulate a defocusing process, Gaussian functions with standard deviations ranging from 0 to 3.75 in steps of 0.25 are convolved with the original images, resulting in sequences of blurred images. Additionally, we will assess the noise robustness, measurement capability, and real-time performance of the proposed method using three metrics, under salt-and-pepper noise with a variance of 0.02, Gaussian noise with a mean of 0 and variance of 0.02, and speckle noise with a variance of 0.02.

As shown in Table XIII, except for the ER metric in the Blurred Image Sequence and average Run time, our VTFF method achieves the highest metric values. In Table XIV, with the exception of the DoER metric under Speckle noise, our method also outperforms others in all other conditions. This demonstrates that the proposed method offers superior noise robustness, measurement capability, and real-time performance on the TID2013 dataset compared to other focus measures.

C. Real-world Data

We manually collected a total of 104 real-world image sequences, which include 48 indoor sequences, 48 outdoor daytime sequences, and 8 outdoor nighttime sequences. Since images captured in nighttime scenes generally contain more noise compared to those taken during the day, in addition to a set of comparative data mentioned in the letter, we will next provide two real-world nighttime scene examples to evaluate the noise robustness of the proposed method in comparison with other methods.

Fig. 5. The first real-world nighttime scene used for evaluation and the target image sequence.

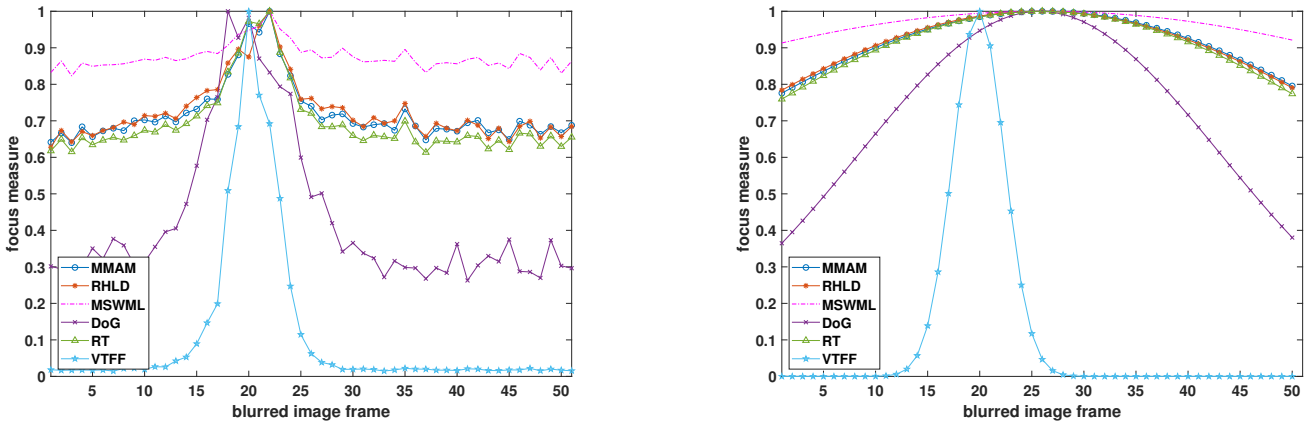

Fig. 6. The focus measure results and Gaussian curve fitting results for the scene in Fig. 5. Left: Measurement results of image sequences by different focus measure methods. Right: Curve fitting results of four focus measures.

The left side of both Fig. 5 and Fig. 8 shows a noisy nighttime scene, from which we have selected a target area using a red square. We then extracted 50 frames from the same region to generate the image sequences shown on the right side of Fig. 5 and Fig. 8.

The measurement results on the left side of Fig. 6 show that MMAM, RHLD, MSWML, and RT all achieve their maximum values at Frame-22, while DoG reaches its maximum at Frame-18. In contrast, our VTFF method achieves its maximum value at Frame-20. Fig. 7 includes images from Frame-18, Frame-20, and Frame-22, along with zoomed-in parts of the central regions. As indicated by the red arrow in the second-row images of Fig. 7, it can be observed that Fig. 7(f) exhibits more distinct edge details,

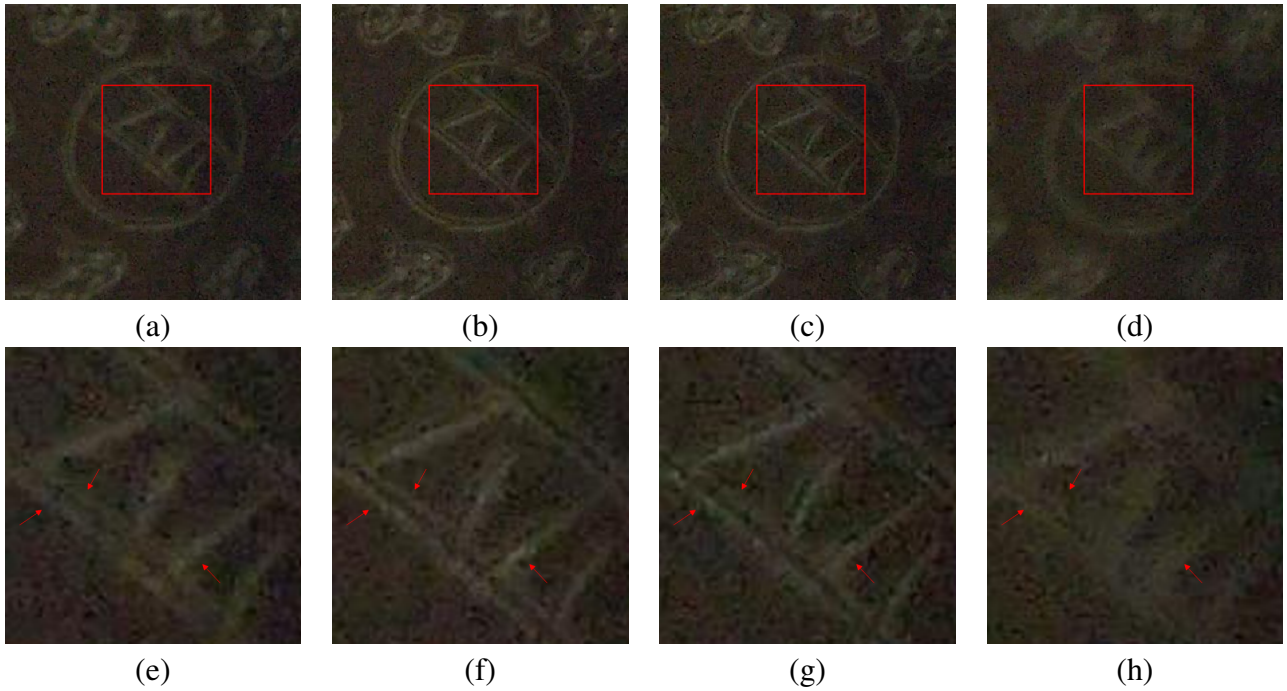

Fig. 7. The comparisons of frames 18, 20, 22, and 26 for the scene in Fig. 5. (a) Frame-18, (b) Frame-20, (c) Frame-22, (d) Frame-26. (e), (f), (g), and (h) are the zoomed-in part of (a), (b), (c), and (d), respectively.

confirming that Frame-20 is the correct focus position. The images on the right side ofFig. 6 show the Gaussian curve fitting results of the focus measures. Except for the VTFF method, where the fitted curve reaches its maximum at Frame-20, the Gaussian curves of the other methods all achieve their maximum at Frame-26. Fig. 7(d) and Fig. 7(h) show the image from Frame-26, along with zoomed-in portions of its central region. It can be observed that the details in this frame are blurry, confirming that it cannot be the correct focus position. Thus, only our VTFF method provides the correct results both in the measurement results and the Gaussian curve fitting results.

Fig. 8. The second real-world nighttime scene used for evaluation and the target image sequence.

Fig. 8 shows the metric results and Gaussian curve fitting results for the scene in Fig. 8. From the left

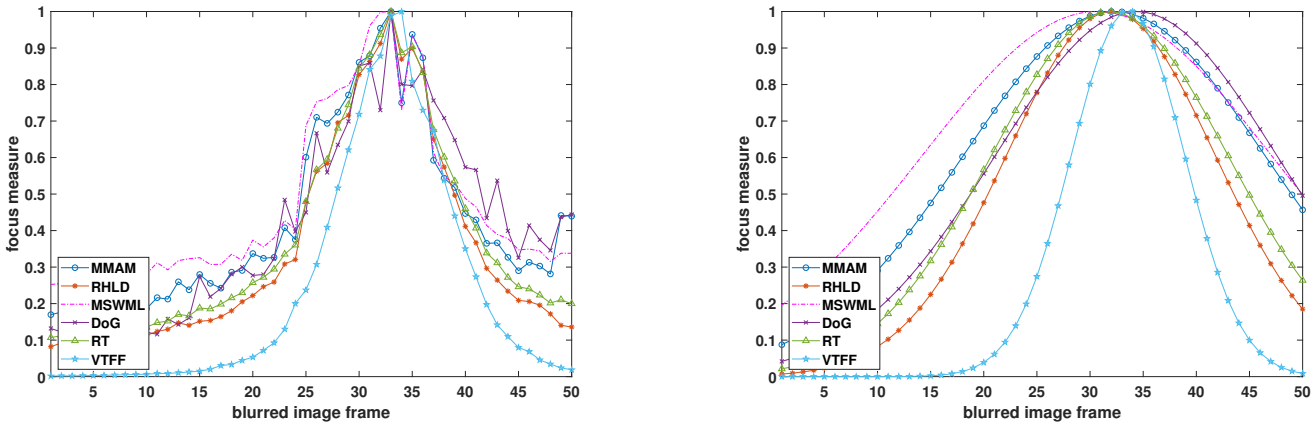

Fig. 9. The focus measure results and Gaussian curve fitting results for the scene in Fig. 8. Left: Measurement results of image sequences by different focus measure methods. Right: Curve fitting results of four focus measures.

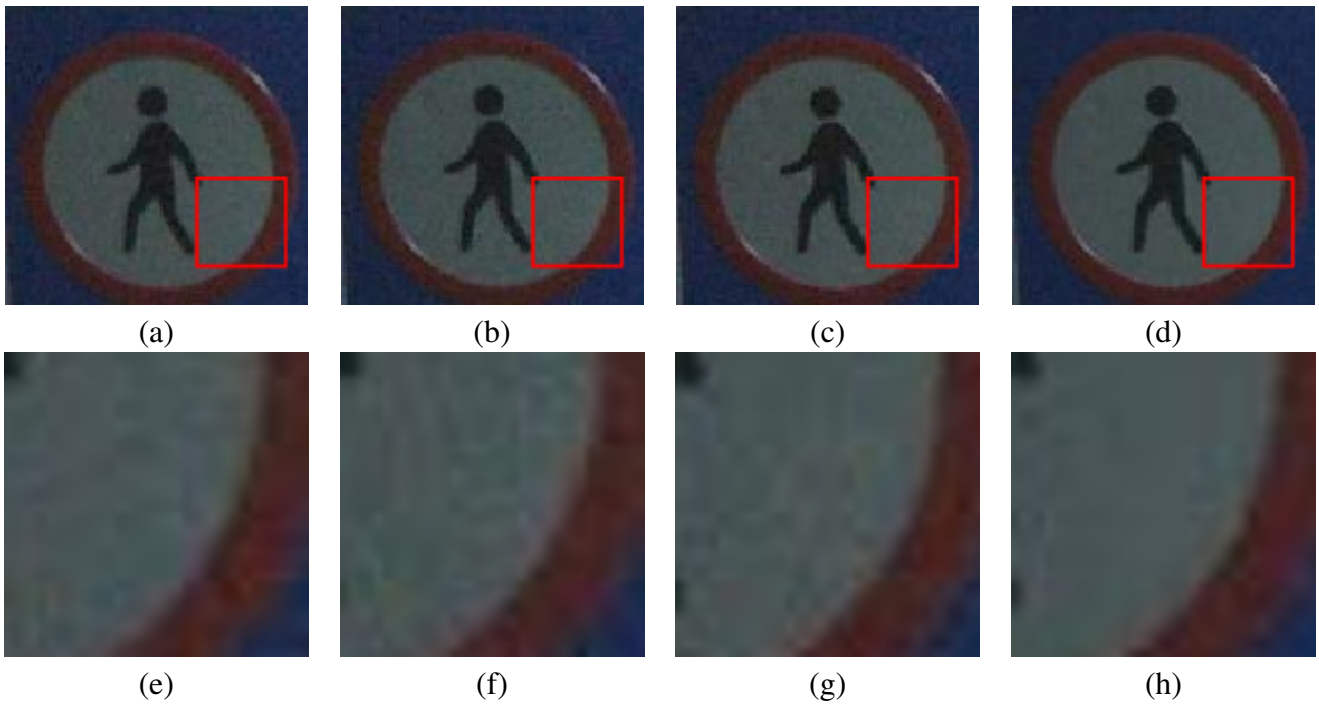

Fig. 10. The comparisons of frames 31, 32, 33, and 34 for the scene in Fig. 8. (a) Frame-31, (b) Frame-32, (c) Frame-33, (d) Frame-34. (e), (f), (g), and (h) are the zoomed-in part of (a), (b), (c), and (d), respectively.

side of Fig. 8, it can be seen that except for the VTFF method, which reaches its maximum at Frame-34, all other methods achieve their maximum at Frame-33. Fig. 10 includes images from frames 33 and 34, along with their corresponding zoomed-in images. As seen in Fig. 10(g) and (h), the image in (h) exhibits clearer details with less noise, confirming that Frame-34 is the correct focus position.

In the Gaussian fitting results on the right side of Fig. 9, MMAM, RHLD, and DoG achieve their maximum values at Frame-32, MSWML at Frame-31, and RT and VTFF both at Fram-34. Fig. 10(a) and (b) show the images from frames 31 and 32, whileFig. 10(e) and (f) present the corresponding zoomed-in regions. The details in (e) and (f) are not as clear as in (h), confirming that frames 31 and 32 are not the correct focus positions. Therefore, only VTFF correctly identifies the focus position, both in the measurement results and the Gaussian curve fitting results. In summary, our VTFF method ensures a significant change in the curve in noisy nighttime scenes, maintaining a monotonically increasing curve

to the left of the correct focus position and a monotonically decreasing curve to the right. This behavior supports a better Gaussian fitting curve when sampled randomly from the focus measure curve, benefiting subsequent processes. Moreover, our VTFF method exhibits excellent noise robustness, accurately identifying the correct focus position even in high-noise environments.

REFERENCES

[1] Y.-R. Li, R. H. F. Chan, L. Shen, and X. Zhuang, "Regularization with multilevel non-stationary tight framelets for image restoration," Applied and Computational Harmonic Analysis, vol. 53, pp. 332–348, Jul. 2021.

[2] H. Lin, V. Hosu, and D. Saupe, "KADID-10k: A Large-scale Artificially Distorted IQA Database," in 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), Berlin, Germany: IEEE, Jun. 2019, pp. 1–3.

[3] S. Liu, Y. Lu, J. Wang, S. Hu, J. Zhao, and Z. Zhu, "A new focus evaluation operator based on max–min filter and its application in high quality multi-focus image fusion," Multidim Syst Sign Process, vol. 31, no. 2, pp. 569–590, Apr. 2020.

[4] X. Nie, B. Xiao, X. Bi, W. Li, and X. Gao, "A focus measure in discrete cosine transform domain for multi-focus image fast fusion," Neurocomputing, vol. 465, pp. 93–102, Nov. 2021.

[5] Z. Hu, W. Liang, D. Ding, and G. Wei, "An improved multi-focus image fusion algorithm based on multi-scale weighted focus measure," Appl Intell, vol. 51, no. 7, pp. 4453–4469, Jul. 2021.

[6] I. Helmy and W. Choi, "Reduced Tenegrad Focus Measure for Performance Improvement of Astronomical Images," in 2022 International Conference on Electronics, Information, and Communication (ICEIC), Jeju, Korea, Republic of: IEEE, Feb. 2022.

[7] L. Guo and L. Liu, "A Perceptual-Based Robust Measure of Image Focus," IEEE Signal Process. Lett., vol. 29, pp. 2717–2721, 2022.

[8] N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti, and C.-C. Jay Kuo, "Image database TID2013: Peculiarities, results and perspectives," *Signal Processing: Image Communication*, vol. 30, pp. 57–77, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0923596514001490

Appendix 2
Detailed definitions of the measurement capability and noise robustness metrics

In this appendix, we outline the experimental setup used to evaluate the method's performance. Initially, the experimental data is generated by convolving the original images in the database with Gaussian functions of 15 different variances, starting from 0 and increasing by increments of 0.25 up to 3.75. This process simulates the defocusing process, yielding a sequence of blurred images. Subsequently, Gaussian or speckle noise with a mean of 0 and a variance of 0.02 is added to each frame of the blurred image sequence to generate the noisy blurred image sequences, as shown in Fig. 1.

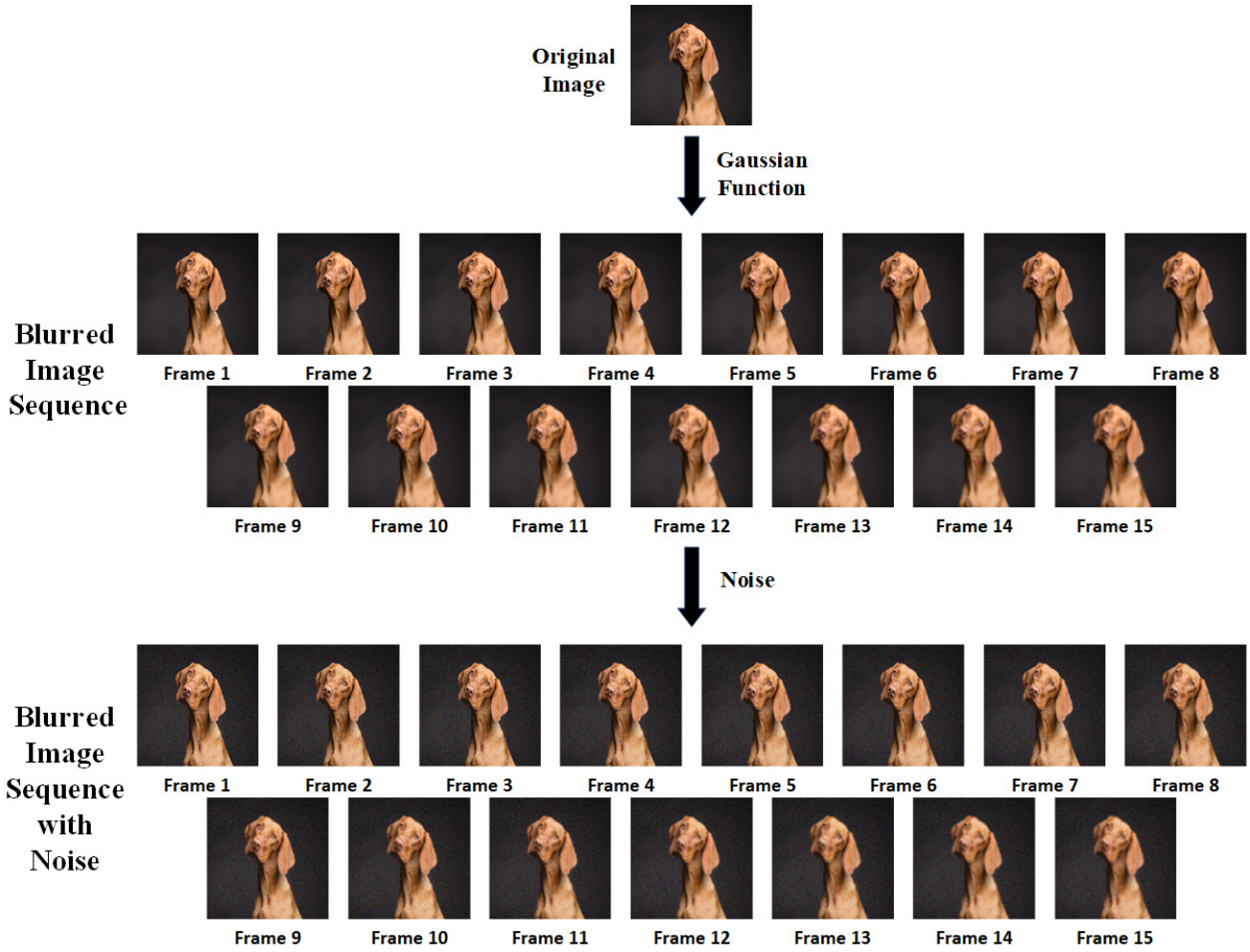

Fig. 1. The process of generating experimental data.

We apply one focus measure (RHLD [1]) to both the blurred image sequence and the noisy blurred image sequence in Figure 1, yielding two focus measure curves as shown in Figure 2. By analyzing these curves, we can quantitatively evaluate the method's performance based on various metrics, including the range of curve variation, differences between adjacent values, and other relevant information. The evaluation metrics used in the letter include measurement capability, noise robustness, and real-time performance.

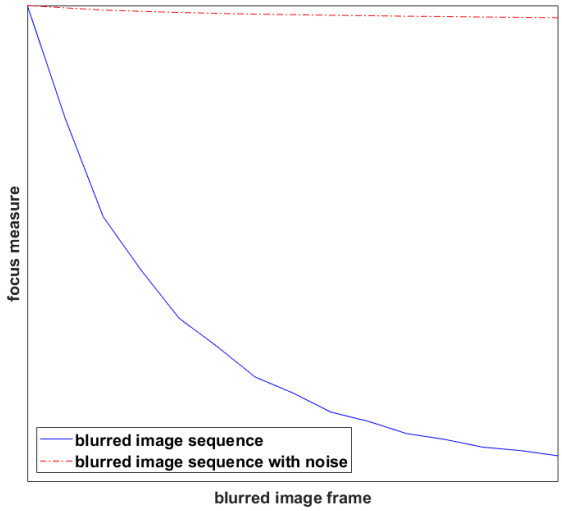

Fig. 2. Results of RHLD applied to the blurred image sequences with and without noise in Figure 1.

Firstly, the metrics for measurement capability and real-time performance are referenced from [2] and [3]. For real-time performance, the average runtime of the program is used as the evaluation criterion. For measurement capability, two metrics are employed: Sensitivity Detection Ability (SDA) [2] and Effective Range (ER) [3]. This appendix provides a detailed introduction to the SDA and ER metrics.

A. *Measurement Capability Metrics*

1) Sensitivity Detection Ability (SDA): The calculation formula for SDA is as follows:

$$SDA = \frac{1}{T-1}\sum_{t=1}^{T-1}\left(1 - e^{-\left|\frac{M_{t+1}-M_t}{\sigma_{t+1}-\sigma_t}\right|\frac{1}{\sigma_{t+1}}}\right),$$

where T is the total number of frames in the blurred image sequence, M_t is the focus measure value of the tth frame in the sequence, and σ_t is the standard deviation of the Gaussian function used in the tth frame. In the experimental data used in this study, the standard deviation of the Gaussian function increases by a constant value. In this case, a larger SDA indicates that the focus measure values between adjacent lens positions have greater differences, meaning the focus measure values can effectively distinguish different degrees of blur. Therefore, a larger SDA value indicates better measurement capability of the focus measure.

2) Effective Range (ER): The specific form of the ER metric is as follows:

$$ER = \frac{\sigma}{\mu},$$

where σ is the standard deviation of the focus measure curve and μ is the mean of the focus measure curve. When the focus measure has good measurement capability, the obtained focus measure curve is similar to the without noise curve in Fig. 2. The focus measure value decreases rapidly as the target becomes more blurred, exhibiting a large range of variation. In this case, the standard deviation of the focus measure curve is large, while the mean is small, resulting in a large ER. Conversely, the focus measure curve measured by the method is like the noise curves in Fig. 2, where the focus measure value shows a smaller range of variation, and the focus measure value no longer changes When the target reaches a certain level of blur. In this situation, the variation in the focus measure value cannot correctly reflect the changes in the target contrast. The standard deviation of the focus measure curve is small, and

the mean is large, resulting in a small ER. Therefore, the larger the ER, the greater the range of variation in the focus measure values measured by the method, which can effectively reflect changes in the degree of target blur and indicate better measurement capability.

B. Noise Robustness Metrics

The noise robustness of the focus measure is a key focus of the study in the letter. To evaluate this, we use existing experimental data and metrics are used to assess the stability of the curve trend, ER metric, and SDA metric under noise through three metrics: Difference of Curve, Difference of ER, and Difference of SDA. These metrics reflect the noise robustness of the method. In the following content, this appendix will provide a detailed introduction to the implementation principles and specific definitions of these three metrics, thereby demonstrating their rationality and effectiveness.

1) Difference of Curve (DoC): When measuring the same blurred image sequence with and without noise, the measurement results of a focus measure with poor noise robustness are shown in Fig. 2. The two curves will have significant differences in the overall trend. Conversely, when the method has good noise robustness (such as the DoG [3]), the overall trends of the two curves are basically the same, as shown in Fig. 3.

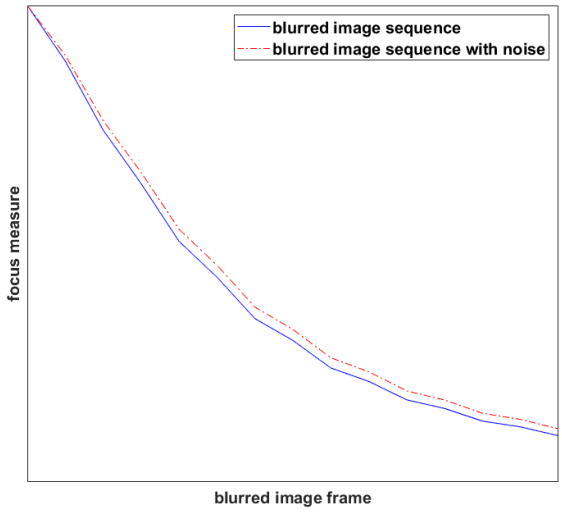

Fig. 3. Results of DoG applied to the blurred image sequences with and without noise in Figure 1.

Based on this point, the difference in trend between the measurement results of the focus measure with and without noise can serve as a basis for evaluating the noise robustness of the method. Accordingly, we calculates the difference between the focus measure curves measured on noisy and noise-free blurred sequences using the following formula:

$$DoC = \sqrt{\sum_{t=1}^{T}(M_{Nt} - M_{Ot})^2},$$

where T is the total number of frames in the blurred image sequence, M_{Ot} and M_{Nt} are the focus measure values obtained by the focus measure for the tth frame of the noise-free and noisy blurred sequences, respectively. A smaller DoC indicates that the two curves are more similar in overall trend, suggesting better noise robustness of the method.

2) Difference of ER (DoER) and Difference of SDA (DoSDA): Under the interference of noise, the measurement capability of the focus measure will also be affected. Similarly, taking the DoG [3] with good noise robustness and the RHLD [1] with poor noise robustness as examples, the corresponding ER and SDA metrics for the two methods when measuring the blurred image sequences with and without noise in Fig. 1 are shown in Table I.

TABLE I
ER AND SDA FOR DOG AND RHLD WHEN MEASURING BLURRED IMAGE SEQUENCES WITH AND WITHOUT NOISE IN FIG. 1

Focus Measure	Blurred Image Sequence		Blurred Image Sequence with Gaussian Noise	
	$ER\uparrow$	$SDA\uparrow$	$ER\uparrow$	$SDA\uparrow$
DoG	0.7279	0.3652	0.6842	0.3625
RHLD	0.9496	0.3713	0.0078	0.0842

From the data in the Table I, it can be seen that the ER and SDA metrics for the DoG are very close under both noisy and noise-free conditions. However, for the RHLD, both ER and SDA metrics show a significant decrease when affected by Gaussian noise. These results indicate that the variability in the measurement capability of focus measures under noise interference can reflect their noise robustness. Therefore, we will calculate the difference in ER metrics between corresponding noisy and noise-free blurred sequences, specifically examining how the method's ER metric changes due to noise interference, thereby reflecting its noise robustness. The formula is

$$DoER = \sqrt{\sum_{k=1}^{K} (ER_{Nk} - ER_{Ok})^2},$$

where, K represents the number of experimental groups. For the kth group of experimental data, ER_{Nk} and ER_{Ok} respectively denote the ER metrics of the focus measure when measuring the corresponding noisy and noise-free blurred sequences of that group. A smaller DoER indicates that the method maintains its measurement capability under noise influence, demonstrating better noise robustness. Similarly, this concept applies to the SDA metric. We evaluates the stability of the focus measure's SDA metric under noise influence, thereby reflecting its noise robustness. The specific formula is:

$$DoSDA = \sqrt{\sum_{k=1}^{K} (SDA_{Nk} - SDA_{Ok})^2}$$

where, SDA_{Nk} and SDA_{Ok} represent the SDA metric of the method when measuring the corresponding noisy and noise-free blurred sequences of the kth group of experimental data. A smaller value of DoSDA suggests that the method's measurement capability is less affected by noise, demonstrating better noise robustness.

REFERENCES

[1] X. Nie, B. Xiao, X. Bi, W. Li, and X. Gao, "A focus measure in discrete cosine transform domain for multi-focus image fast fusion," Neurocomputing, vol. 465, pp. 93–102, Nov. 2021.

[2] Z. Zhang, Y. Liu, Z. Xiong, J. Li, and M. Zhang, "Focus and Blurriness Measure Using Reorganized DCT Coefficients for an Autofocus Application," IEEE Trans. Circuits Syst. Video Technol., vol. 28, no. 1, pp. 15–30, Jan. 2018.

[3] L. Guo and L. Liu, "A Perceptual-Based Robust Measure of Image Focus," IEEE Signal Process. Lett., vol. 29, pp. 2717–2721, 2022.