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

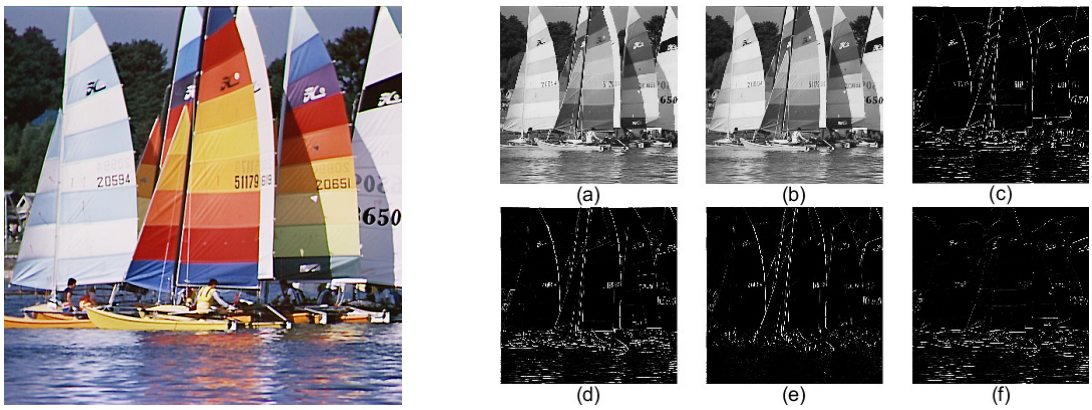

Fig. 1. A color sample image and its DHF tight framelet features. (a) Input grayscale image, converted from the color sample image. (b) Low-pass subband extracted by the DHF system. (c)-(f) High-pass subbands extracted by the DHF system.

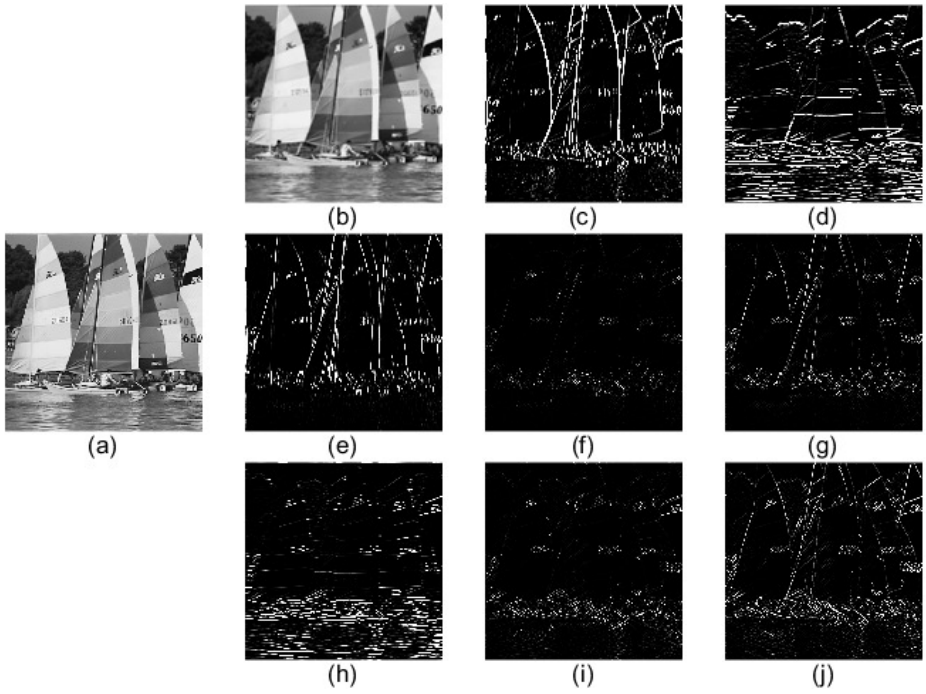

Fig. 2. High-pass features of the low-pass subband in Fig. 1(a) by the DCT tight framelet. (a) Low-pass subband extracted by the DHF system in Fig. 1(a). (b) Low-pass subband extracted by the DCT system. (c)-(j) High-pass subbands extracted by the DCT system.

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

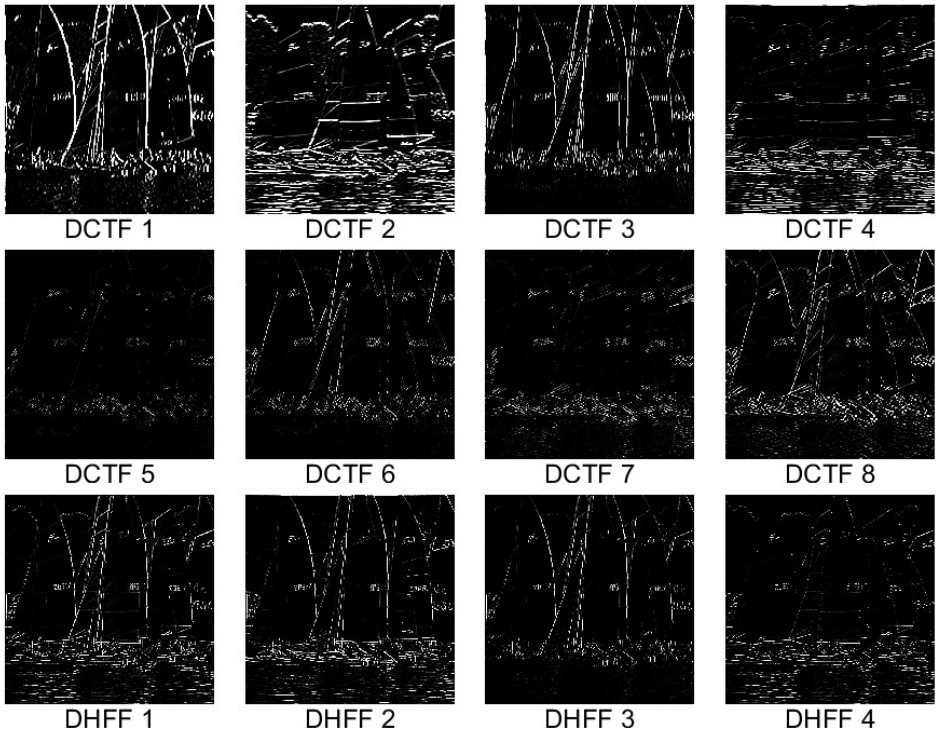

Fig. 3. The re-named DHF and DCT tight framelet high-pass subbands.

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

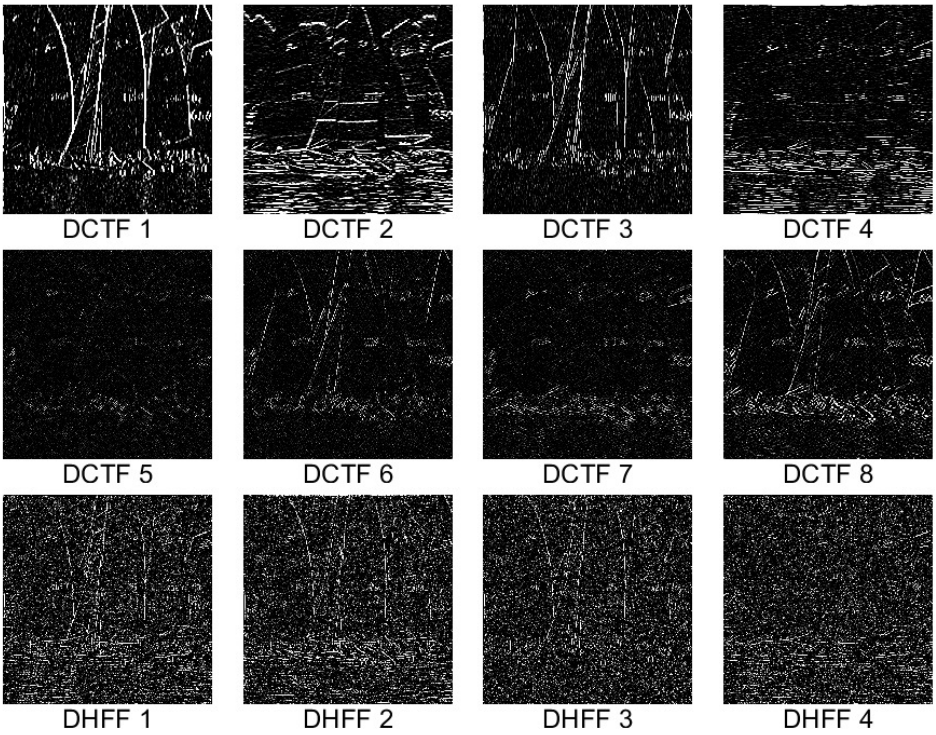

Fig. 4. The 12 high-pass subbands extracted from the color sample image in Fig. 1 after adding Gaussian noise.

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.

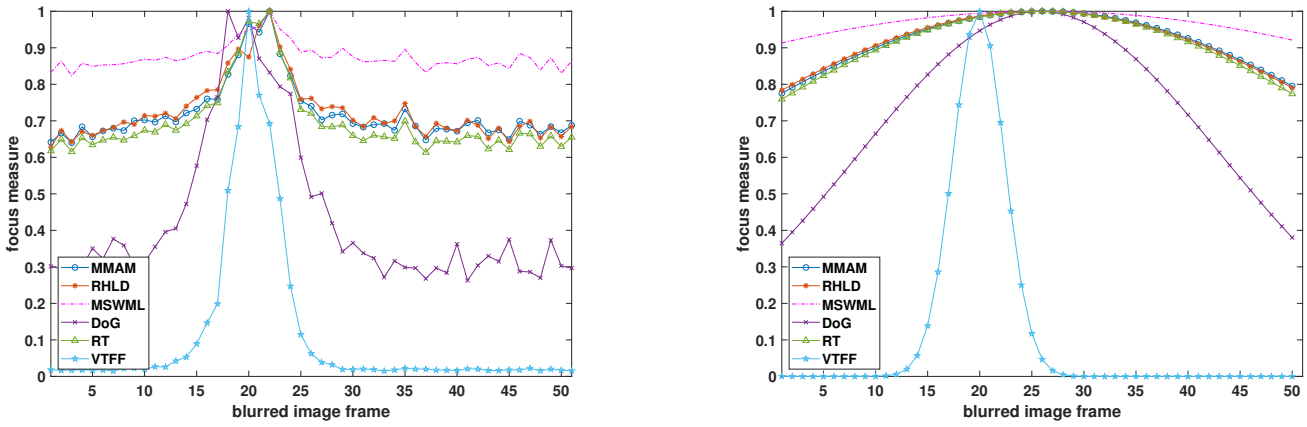

Fig. 6. The focus measure results and Gaussian curve fitting results for the scene in Fig. 5. Left: Measurement results of image sequences by different focus measure methods. Right: Curve fitting results of four focus measures.

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,

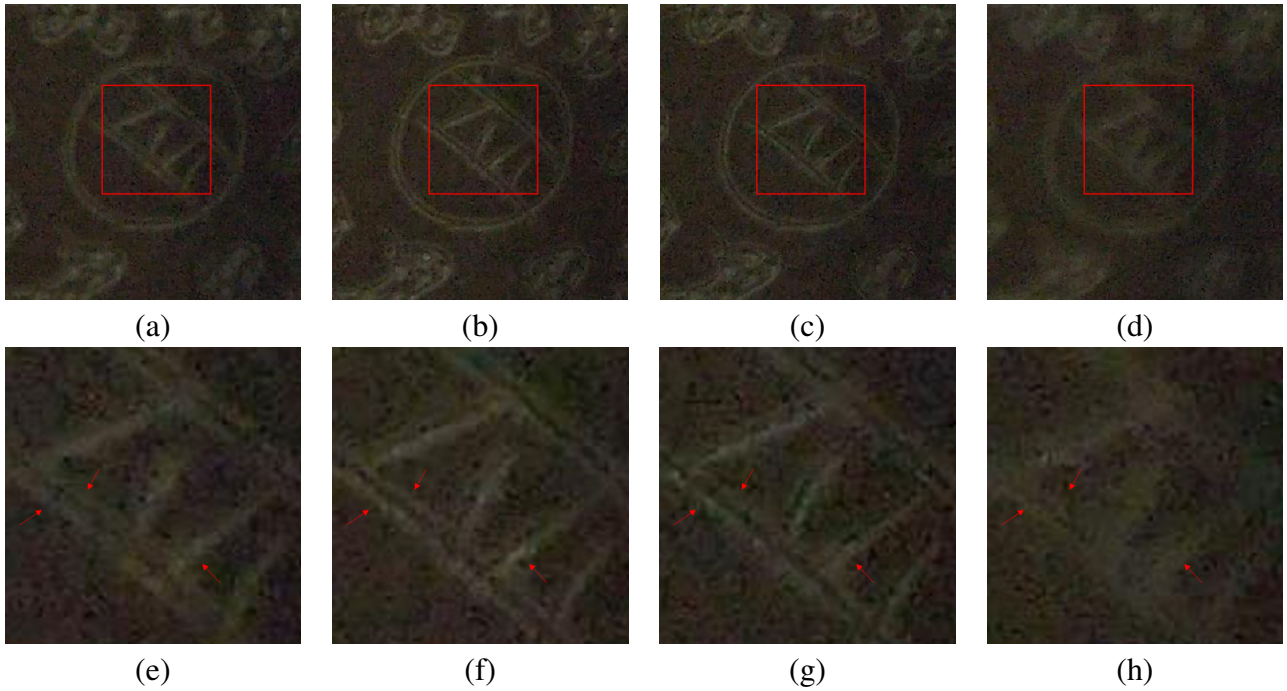

Fig. 7. The comparisons of frames 18, 20, 22, and 26 for the scene in Fig. 5. (a) Frame-18, (b) Frame-20, (c) Frame-22, (d) Frame-26. (e), (f), (g), and (h) are the zoomed-in part of (a), (b), (c), and (d), respectively.

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

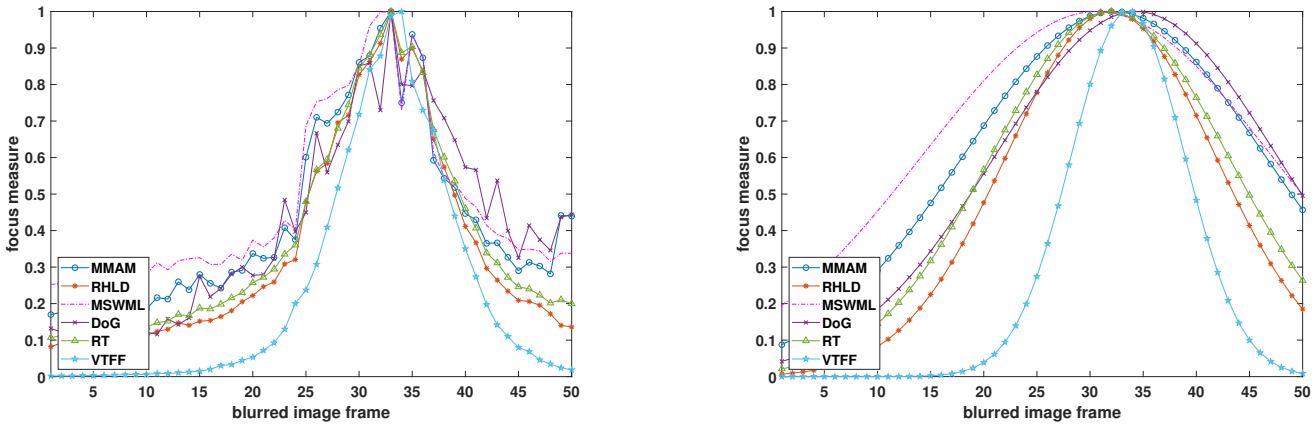

Fig. 9. The focus measure results and Gaussian curve fitting results for the scene in Fig. 8. Left: Measurement results of image sequences by different focus measure methods. Right: Curve fitting results of four focus measures.

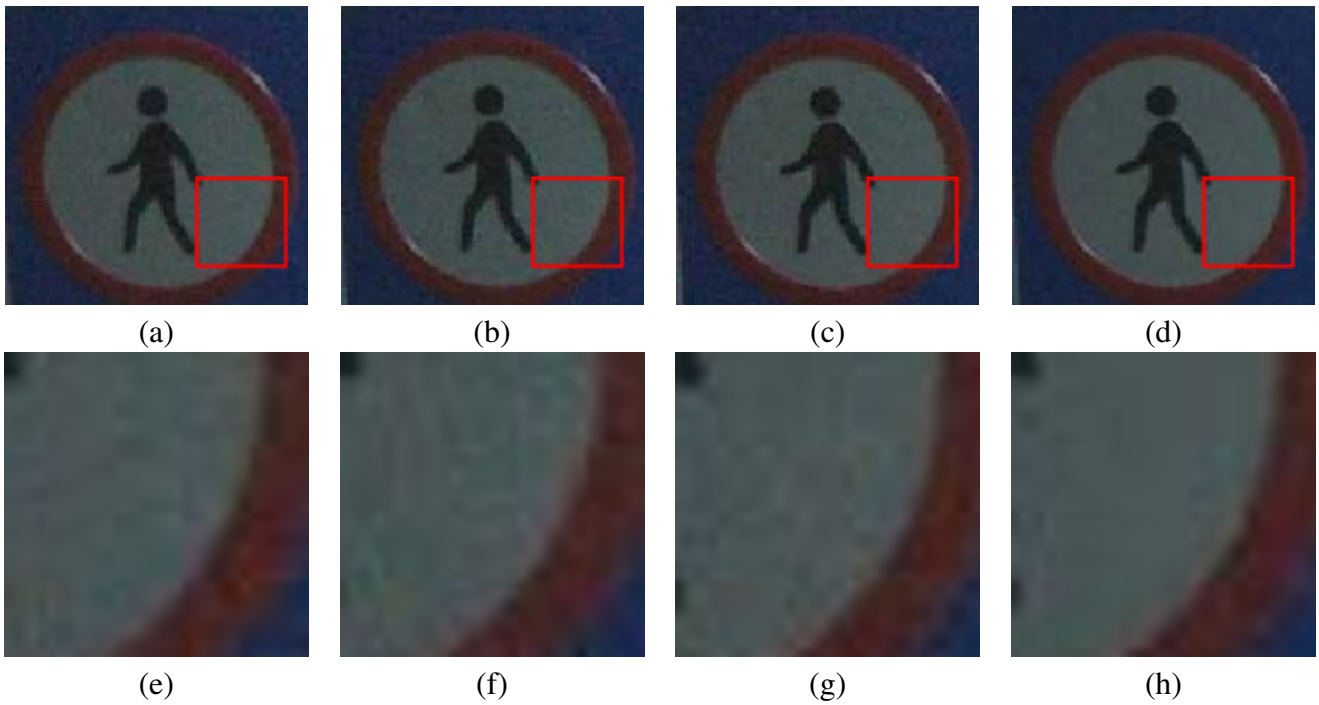

Fig. 10. The comparisons of frames 31, 32, 33, and 34 for the scene in Fig. 8. (a) Frame-31, (b) Frame-32, (c) Frame-33, (d) Frame-34. (e), (f), (g), and (h) are the zoomed-in part of (a), (b), (c), and (d), respectively.

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.

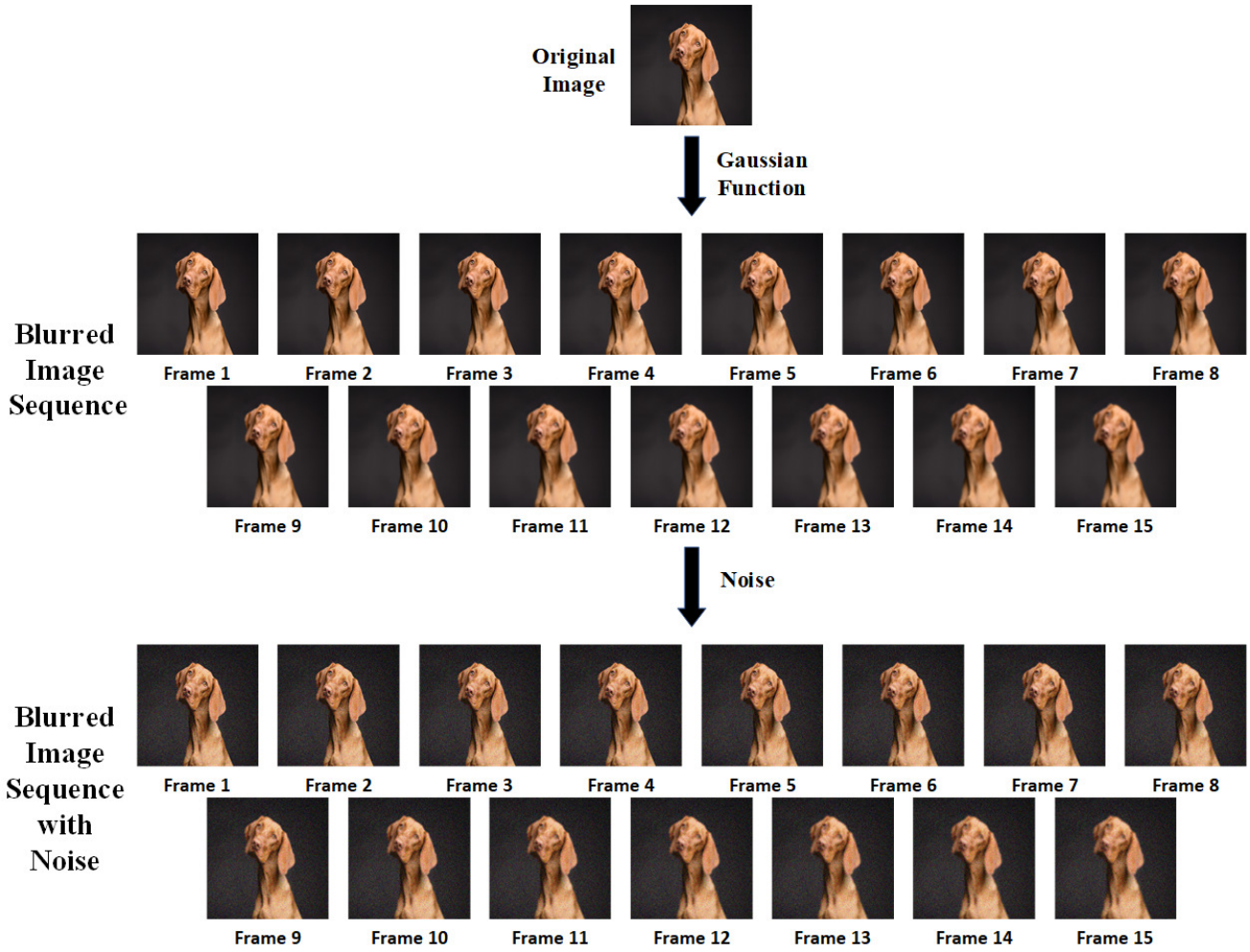

Fig. 1. The process of generating experimental data.

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.

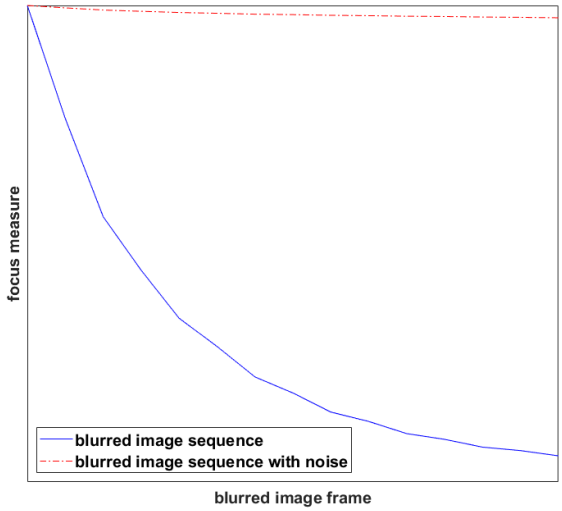

Fig. 2. Results of RHLD applied to the blurred image sequences with and without noise in Figure 1.

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.

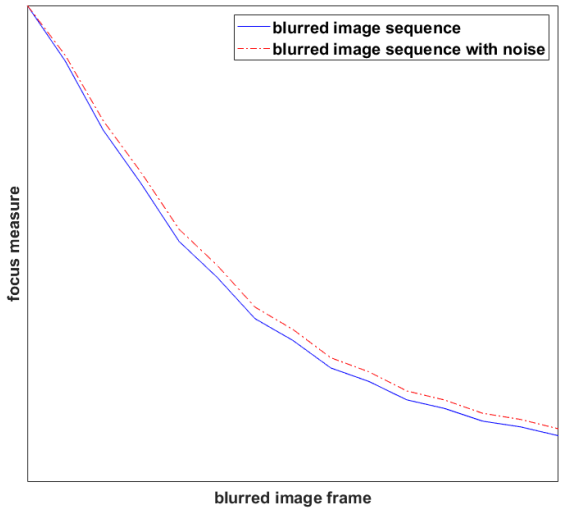

Fig. 3. Results of DoG applied to the blurred image sequences with and without noise in Figure 1.

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.