Fast Distribution Fitting for Parameter Estimation of Range-Weighted Neighborhood Filters

Chao-Tsung Huang, *Member, IEEE*

***Abstract*—The range variance of neighborhood filters is well estimated via distribution fitting of a chi scale mixtures model proposed in our previous work. However, it introduced computation overheads for deriving empirical distributions and performing iterative fitting. In this letter, we discuss how to greatly reduce the overheads for practical usage while maintaining denoising quality. For empirical distributions, a grid-subsampling strategy is adopted for acceleration. Regarding distribution fitting, two different methods are studied: equal-frequency merged distribution and L-moment fitting. The former reformulates the fitting process into entropy optimization for only few merged bins. It provides 6-13x speedup for model fitting with negligible quality loss and 9-20x speedup with $\leq$0.1 dB PSNR drop by using 20 and 10 bins respectively. The latter performs table lookup of L-moments, instead of conventional moments, for robust fitting of heavy-tailed distributions. The fitting time then becomes negligible with $\leq$0.2 dB drop in most cases, e.g. the overall run time for bilateral 9×9 filtering can be thus accelerated by around 6x. Experiments on bilateral and non-local means filters are also given to show the speedup, quality and robustness.**

***Index Terms*—Bilateral filter, non-local means, denoising, empirical Bayesian, L-moment, parameter estimation.**

I. Introduction

Range-weighted formulation provides simple and effective neighborhood filters for edge-preserved denoising, especially the bilateral filter [1] and non-local means (NLM) [2]. Many extensions were proposed for computer vision applications, and the reader is referred to [3] for more details. Its general form can be expressed using adaptive weighted averaging:

$$\hat{\mathbf{z}}_l = \frac{\sum_{i\in\Lambda_l} w_{l,i} d_{l,i} \cdot \mathbf{y}_i}{\sum_{i\in\Lambda_l} w_{l,i} d_{l,i}}, \tag{1}$$

where $\mathbf{y}$ and $\hat{\mathbf{z}}$ are the observed and filtered signals respectively, and Λ_l represents the neighborhood of the pixel at position l. The kernels $w_{l,i}$ and $d_{l,i}$ represent the adaptive weights for intensity range and pixel distance respectively. For the bilateral filter, the range-weighted $w_{l,i} = K_r(\frac{\|\mathbf{y}_l - \mathbf{y}_i\|_2^2}{2\sigma_r^2})$ adapts to pixel similarity for edge preservation. For the NLM, the pixel differences are replaced by patch differences for robustness. A conventional choice of the kernel $K_r(\cdot)$ is Gaussian function, and the performance is controlled by the range variance σ_r^2. The filtered result will be over blurred or still noisy if it is too large or too small. Therefore, how to estimate the best σ_r^2 is an important issue in practice.

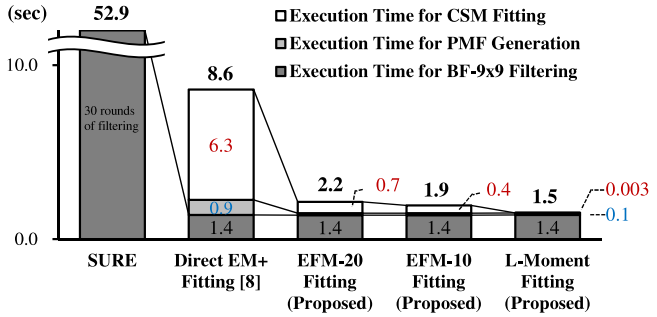

Fig. 1. Comparison of average execution time per image for the first-iteration bilateral 9×9 filter (**BF-9×9**). Twelve standard images are tested. Detailed results are listed in Table II in Section III. EFM-20/10 stands for the proposed equal-frequency merged distribution fitting with 20/10 bins. Grid-subsampled probability mass functions are used for all the proposed methods.

A general approach of parameter estimation is the Stein's unbiased risk estimate (SURE) [4], [5], [6], [7]. Due to theoretical limitations, it only works when the following assumptions hold: independent Gaussian noise and a weakly differentiable kernel. It provides unbiased estimation of the mean squared error (MSE) from noisy and filtered images. The optimal parameters can then be selected based on the filtered results using different parameter combinations. Though accurate, the complexity is quite high because each combination needs to filter the image separately.

In our previous work [8], we proposed a neighborhood noise model for specifically reasoning the range-weighted filters and also accurately estimating the range variance σ_r^2. An expectation-maximization-plus (EM+) algorithm was devised for parameter estimation by iteratively fitting the empirical distributions of pixel differences to a chi scale mixtures (CSM) model. Fig. 1 shows an example of the speed advantage over SURE. In addition, it can overcome the theoretical limitations of SURE and thus enables iterative filtering to improve denoising quality and also parameter estimation for non-differentiable kernels, e.g. NLM with motion estimation. However, the computation overhead introduced by the direct EM+ fitting is significant compared to the filtering itself. As shown in Fig. 1, deriving the probability mass function (PMF) and performing the iterative fitting may take 0.9 and 6.3 seconds respectively while the filtering only needs 1.4 seconds. In this letter, we aim to reduce the overhead to a sufficiently small extent for practical usage. In Section II, a grid-sampling strategy will be presented first for speeding up the PMF derivation, and then two different methods will be proposed

This work was supported by the Ministry of Science and Technology, Taiwan, R.O.C. under Grant No. MOST 103-2218-E-007-008-MY3.

C.-T. Huang is with the National Tsing Hua University, Hsinchu 30013, Taiwan (e-mail: chaotsung@ee.nthu.edu.tw).

for accelerating the distribution fitting. One merges the PMF into only few equal-frequency bins to speed up the EM+ fitting. And the other one uses moment fitting of L-moments [9], [10], [11] which are robust for heavy-tailed distributions. Experimental results on true-color image denoising will be given in Section III, and conclusion remarks in Section IV.

II. Fast CSM Fitting

In [8], we built an observable CSM model for the pixel difference $s_{l,i} \triangleq \| \mathbf{y_l} - \mathbf{y_i} \|_2$ with a marginal distribution

$$f_s(s;\sigma,\epsilon,\alpha) = \int_\epsilon^1 f_{s,w}(s,w;\sigma,\epsilon,\alpha)\,dw, \tag{2}$$

where $f_{s,w}$ is a chi distribution scaled by a hidden random variable w, and σ and (ϵ,α) are the scale and shape parameters respectively. The range variance σ_r^2 is equal to $\alpha\sigma^2$. The EM+ fitting was then developed to iteratively optimize the Kullback-Leibler divergence (KLD) between the empirical distribution $P(s)$ and the estimated $\tilde{P}(s;\sigma,\epsilon,\alpha)$. Its complexity issue results from that the marginal distribution $f_s(s)$ and its partial derivatives have no analytical expressions. Therefore, many numerical integrals are required in each iteration. In the following, we will discuss how to accelerate the fitting process.

A. Grid-subsampled empirical distribution

The naive way in [8] to derive $P(s)$ for the bilateral filter is accumulating the histogram of all $s_{l,i}$. The complexity is then of the same order as the filtering: $O(|\Lambda||\mathcal{I}|)$ where $\mathcal{I}$ is the pixel set of the whole image. In this letter, we apply a grid-subsampling strategy that only samples l at grid positions such that their neighborhood Λ_l are mutually exclusive and collectively exhaustive. Then the complexity becomes only $O(|\mathcal{I}|)$, and the collected $P(s)$ still includes information from all pixels. Similarly, for the NLM we sample l such that their center patches are exclusive and exhaustive.

B. Equal-frequency merging (EFM) for distribution fitting

The complexity of the EM+ fitting is independent of image resolution and, instead, mainly related to the histogram bin number. For the empirical $P(s_j)$, we can merge the bins s_j into T different and contiguous intervals $\mathbf{S}_0, \mathbf{S}_1, ..., \mathbf{S}_{T-1}$, and each one has the discrete probability $P_t = \sum_{s_j \in \mathbf{S}_t} P(s_j)$. Suppose a probability density function $p(s)$ is behind the PMF $P(s)$. Then we can find a representative s_t^* in each interval such that $p(s_t^*) = P_t/|\mathbf{S}_t|$ according to the mean value theorem.

Distribution fitting can be performed by minimizing the KLD between the merged P_t and an estimated $\tilde{P}_t = f_s(s_t^*)\cdot|\mathbf{S}_t|$. This is equivalent to minimizing $\sum_t -P_t \log f_s(s_t^*)$ which exactly can be optimized by the EM+ fitting. The bin merging accelerates the fitting process by reducing the number of numerical integrals proportionally.

To find the representative s_t^* for each interval, we approximated the $p(s)$ by polynomial fitting on $\log P(s)$. A high polynomial degree may be preferred to reduce fitting error, but overfitting should be avoided because the empirical $P(s)$ is noisy. On the other hand, a low degree could be sufficient since the distribution $f_s(s)$ is unimodal and varies smoothly. This tradeoff was verified by the experiments in Fig. 2. As a result, we chose the degree 5 for its balanced performance.

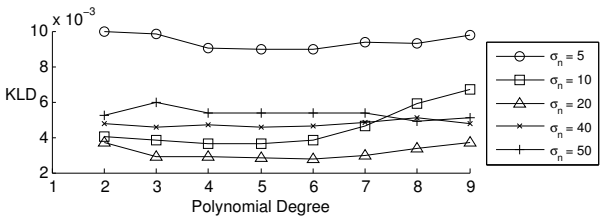

Fig. 2. Average KLD between $P(s)$ and $\tilde{P}(s)$ versus degree of polynomial fitting (EFM, $T = 10$) for the twelve tested images with different noise intensities σ_n.

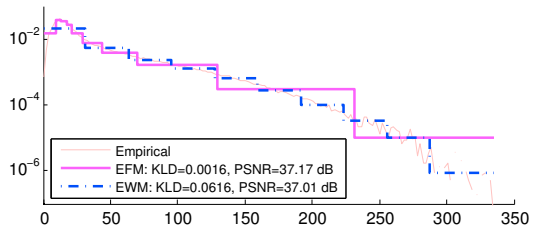

Fig. 3. Density functions of bin merging ($T = 10$) for **BF-9×9** on *Barbara* with noise intensity $\sigma_n = 5$.

The last issue is how to merge the bins. A conventional way is the equal-width merging (EWM), i.e. equal $|\mathbf{S}_t|$. Instead, we used the EFM approach, i.e. equal P_t, because the most information can be preserved in terms of discrete entropy. Fig. 3 shows an example of the advantage. The EFM keeps details near the density peak even when T is small, while the EWM loses too much information there. Therefore, given the same fitting quality the EFM can use a smaller T to provide more speedup. Note that if T is too small (e.g. $T = 5$), the proposed EFM will still fail for two reasons: the polynomial fitting cannot fit long intervals well, and too little information about $P(s)$ is preserved.

C. L-moment fitting

A much faster way for distribution fitting is to compare and match moments, but the common moments cannot perform well due to their instability for the heavy-tailed $f_s(s)$. In contrast, the L-moments remain viable for $f_s(s)$ and can be used for the fitting. They are formed by the expected values of order statistics, e.g. $\lambda_2 = \frac{1}{2}\mathbf{E}[\mathbf{X}_{2:2} - \mathbf{X}_{1:2}]$ where $\mathbf{X}_{j:n}$ denotes the jth smallest element of a random sample of size n drawn from $f_{\mathbf{X}}(x)$ [11]. And in a more general form, the trimmed L-moments discard the first a and the last b elements of the random sample, e.g. $\lambda_2^{(a,b)}$.

In this work, we use the L-moment ratios $\tau_r^{(0,b)} \triangleq \lambda_r^{(0,b)}/\lambda_2^{(0,b)}$ to fit the shape parameters (ϵ,α) and then the L-moment λ_2 for the scale parameter σ. For a 2-D grid of (ϵ,α), the $\tau_r^{(0,b)}$ and λ_2 of CSM distributions $f_s(s;\sigma=1,\epsilon,\alpha)$ are precalculated and stored in tables, e.g. $\tau_3(\epsilon,\alpha)$. Then the estimated $(\hat{\epsilon},\hat{\alpha})$ can be found by table lookup for the empirical L-moment ratios,

TABLE I
R^2 OF $y = x$ FITTING FOR EMPIRICAL AND CSM-FITTED L-MOMENTS

σ_n	λ_2	τ_3	τ_4	$\tau_3^{(0,1)}$	$\tau_3^{(0,2)}$
5	0.997	0.922	**0.931**	0.879	**0.929**
20	1.000	**0.998**	0.973	**0.989**	0.970
50	0.996	**0.991**	**0.977**	0.910	0.816

e.g. $\tau_3[P(s)]$, and the estimated $\hat{\sigma}$ is derived as the ratio of the empirical $\lambda_2[P(s)]$ to the estimated $\lambda_2(\hat{\epsilon}, \hat{\alpha})$.

The fitting accuracy depends on if the L-moments and ratios of $f_s(s; \sigma, \epsilon, \alpha)$ can match those of the empirical $P(s)$ well, i.e. how accurately a simple $y = x$ can fit them. Table I shows the fitness in terms of R^2 values for the first-iteration **BF-9×9** on tested noisy images. While the λ_2 shows great predictability, the performance of the ratios $\tau_r^{(0,b)}$ is varied in different cases of noise intensity σ_n. Also, we observed that for each $\tau_r^{(0,b)}(\epsilon, \alpha)$ the valley of the absolute gradient, or insensitive area, locates at different regions. Therefore, the fitting accuracy and robustness of each ratio are related to the latent $(\sigma, \epsilon, \alpha)$, which leads to a chicken-and-egg situation.

To address the accuracy and robustness issues, we adopted a confidence-based approach. Two possible sets of $(\hat{\epsilon}, \hat{\alpha})$ are first estimated using (τ_3, τ_4) and $(\tau_3^{(0,1)}, \tau_3^{(0,2)})$, respectively, and they are also accompanied with corresponding confidence values. The one with the highest confidence is then chosen and used to derive $\hat{\sigma}$ via the table of λ_2. In this work, we used paraboloid approximation to refine $(\hat{\epsilon}, \hat{\alpha})$ from the table grid and applied Gaussian curvature to measure the confidence.

For simplicity, only the case of using (τ_3, τ_4) is detailed in the following. The tables of τ_3 and τ_4 are discretely 2-D indexed, and so are the corresponding ϵ, α and λ_2. Given the empirical $\tau_3[P(s)]$ and $\tau_4[P(s)]$, the best index (m, n) can be chosen by minimizing the cost E in L^2-norm:

$$E_{m,n} = \| \, (\tau_3[P(s)], \tau_4[P(s)]) - (\tau_3, \tau_4)_{m,n} \, \|_2, \quad (3)$$

$$(m, n) = \arg\min_{m', n'} E_{m', n'}. \quad (4)$$

Then we use a paraboloid to fit the cost E locally at the (m, n) and have the fractional refinement

$$\triangle m = \frac{E_{m-1,n} - E_{m+1,n}}{2A}, \triangle n = \frac{E_{m,n-1} - E_{m,n+1}}{2B}, \quad (5)$$

where $A = E_{m-1,n} - 2E_{m,n} + E_{m+1,n}$ and $B = E_{m,n-1} - 2E_{m,n} + E_{m,n+1}$. The confidence can be derived as AB by the definition of Gaussian curvature. Finally, the refined $\hat{\epsilon}$, $\hat{\alpha}$ and $\lambda_2(\hat{\epsilon}, \hat{\alpha})$ are bilinearly interpolated from the 2-D tables.

III. EXPERIMENTAL RESULTS

The bilateral filter, **BF-9×9**, and modified NLM filter, **MNLM-DCT**, are tested using the same twelve standard true-color images in [8] with the Gaussian kernel $K_r(\cdot)$. The EFM fitting is tested in two conditions: EFM-20 and EFM-10 for bin numbers $T = 20$ and 10 respectively. The L-moments are calculated using the representation of quantile function in [11] and the piecewise linear cumulative density functions approximated by the PMFs. The tables of the L-moments and ratios of the CSM distributions were generated in a dense 2-D grid ($\alpha = [3 : 0.125 : 14], \epsilon = 10^{[-4:0.125:-1]}$). The test platform was MATLAB R2010b on a 3.4 GHz CPU with 8 MB cache, and the MATLAB code was optimized by multicore processing and vectorization.

TABLE II
COMPARISON OF DISTRIBUTION FITTING METHODS (**BF-9×9**, AWGN)

Bilateral 9x9	σ_n	1st Iteration							Recursive Filtering			
		Fitting Result			Execution Time				Fitting Result			
		KLD (x10^{-3})	ΔPSNR (dB)	$\lvert\Delta\sigma_r^2\rvert$ (rel.)	Total (s)	Filter (s)	PMF (s)	Fitting (s)	$\bar{m}$	ΔPSNR (dB)	PSNR (dB)	Δ_{EM+} (dB)
Direct EM+ Fitting	5	8.2	-0.2	38%	5.6	1.4	0.9	3.4	1.0	-0.2	37.5	-
	10	1.7	-0.1	18%	6.6	1.4	0.8	4.3	1.2	-0.1	33.5	-
	20	1.6	-0.3	30%	8.4	1.4	0.9	6.1	2.0	0.4	30.2	-
	40	1.9	-0.3	35%	10.9	1.4	0.9	8.6	2.2	0.7	26.7	-
	50	1.9	-0.2	35%	11.6	1.4	0.9	9.3	2.2	0.7	25.5	-
EFM-20 Fitting	5	8.6	-0.2	40%	2.0	1.4	0.1	0.5	1.0	-0.2	37.5	-0.03
	10	2.2	-0.1	21%	2.1	1.4	0.1	0.6	1.2	-0.1	33.5	-0.01
	20	2.9	-0.3	35%	2.1	1.4	0.1	0.6	2.0	0.4	30.2	-0.04
	40	1.4	-0.2	32%	2.3	1.4	0.1	0.8	2.2	0.7	26.6	-0.04
	50	1.6	-0.2	39%	2.2	1.4	0.1	0.7	2.2	0.6	25.4	-0.05
EFM-10 Fitting	5	8.9	-0.2	35%	1.9	1.4	0.1	0.4	1.1	-0.2	37.5	-0.06
	10	3.6	-0.2	22%	1.9	1.4	0.1	0.4	1.3	-0.2	33.5	-0.05
	20	2.8	-0.3	29%	2.0	1.4	0.1	0.5	2.0	0.4	30.2	-0.01
	40	4.6	-0.6	40%	2.0	1.4	0.1	0.5	2.2	0.6	26.6	-0.09
	50	5.4	-0.6	50%	2.0	1.4	0.1	0.5	2.5	0.7	25.5	-0.02
L-Moment Fitting	5	23.8	-0.3	48%	1.5	1.4	0.1	0.003	1.1	-0.4	37.3	-0.19
	10	3.5	-0.1	17%	1.5	1.4	0.1	0.003	1.3	-0.1	33.6	0.04
	20	2.1	-0.1	23%	1.5	1.4	0.1	0.003	2.0	0.4	30.2	-0.01
	40	1.5	-0.1	20%	1.5	1.4	0.1	0.003	2.3	0.6	26.6	-0.12
	50	1.0	-0.1	31%	1.5	1.4	0.1	0.003	2.3	0.6	25.4	-0.06

△PSNR is the difference from the best first-iteration PSNR;
$|\triangle\sigma_r^2|$ is in form of relative percentage of absolute difference;
$\bar{m}$ stands for the average number of filtering iterations;
$\triangle_{EM+}$ is the PSNR difference from the result using direct EM+ fitting.

A. Bilateral 9×9 filter

The averaged results for **BF-9×9** with additive white Gaussian noise (AWGN) are summarized in Table II, and the direct EM+ fitting, if used, dominates the total execution time. The proposed grid-subsampling strategy can save the computation time of PMF derivation from 0.9 to 0.1 s. The EFM-20 fitting provides nearly the same results as the direct EM+ fitting and accelerates the fitting time by 6-13x. The speedup can be increased to 9-20x by using the EFM-10 fitting, but a 0.1 dB PSNR drop is possible. The L-moment fitting can achieve a nearly negligible fitting time, 0.003 s, and the quality drop is about 0.2 dB for the small noise intensity $\sigma_n = 5$ and around 0.1 dB for larger σ_n. More experimental details and also the MATLAB code of the proposed fast fitting methods are available online[1].

To examine the quality issues for the L-moment and EFM-10 fitting, we show two examples of fitting and filtering in Fig. 4. For the small noise intensity $\sigma_n = 5$ on the left column, the L-moment fitting is not accurate. This is because it relies on the similarity between the empirical distribution and CSM, but the optimal KLD between them is high in this case. The situation is different for large noise intensities, e.g. $\sigma_n = 50$, for which the KLD is small but its sensitivity is low as discussed in [8]. The results show that the L-moment fitting tends to estimate larger range variances. On the contrary, the EFM-10 fitting often provides smaller ones because the long tails are not well

[1] http://www.ee.nthu.edu.tw/chaotsung/fast_fitting

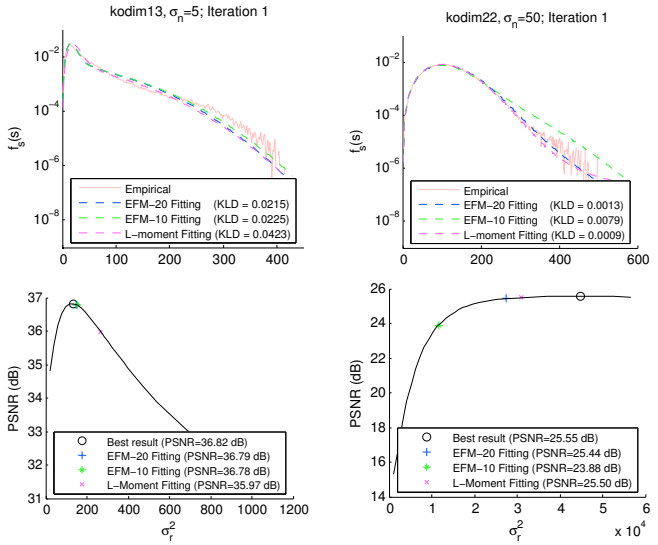

Fig. 4. Examples of **BF-9×9** results. The top row shows the fast model fitting of empirical distributions, and the bottom presents the filtering results.

TABLE III
SUMMARY FOR DIFFERENT NOISE TYPES (**BF-9×9**)

Bilateral 9x9	Noise Type	1st Iteration: Fitting Result KLD ($\times10^{-3}$)	ΔPSNR (dB)	$\lvert\Delta\sigma_r^2\rvert$ (rel.)	Recursive Filtering: Fitting Result $\bar{m}$	ΔPSNR (dB)	PSNR (dB)	Δ_{EM+} (dB)
Direct EM+ Fitting	U5	7.4	-0.2	42%	1.0	-0.2	37.5	-
	U20	9.1	-0.1	21%	2.0	0.4	30.1	-
	U50	13.6	-0.1	36%	2.2	0.5	25.3	-
	POS	6.9	-0.4	36%	1.2	-0.4	32.5	-
	SNP	86.1	-0.1	775%	1.0	-0.1	25.1	-
EFM-20 Fitting	U5	7.5	-0.2	36%	1.0	-0.2	37.6	0.04
	U20	8.8	-0.1	22%	2.0	0.4	30.2	0.02
	U50	13.3	-0.1	37%	2.1	0.5	25.2	-0.02
	POS	6.6	-0.4	34%	1.2	-0.4	32.6	0.06
	SNP	85.8	-0.1	857%	1.0	-0.1	25.1	0.00
EFM-10 Fitting	U5	8.2	-0.2	36%	1.0	-0.2	37.6	0.04
	U20	12.1	-0.2	28%	1.9	0.3	30.1	-0.08
	U50	18.4	-0.3	41%	2.3	0.5	25.3	0.04
	POS	7.6	-0.5	40%	1.4	-0.4	32.6	0.05
	SNP	89.0	-0.1	950%	1.0	-0.1	25.1	-0.01
L-Moment Fitting	U5	15.6	-0.3	48%	1.1	-0.3	37.4	-0.12
	U20	27.5	-0.1	30%	1.5	0.0	29.8	-0.39
	U50	27.1	-0.3	52%	2.2	0.5	25.2	-0.04
	POS	11.4	-0.3	29%	1.2	-0.2	32.7	0.17
	SNP	783.8	-0.1	542%	1.0	-0.1	25.1	0.00

U5/U20/U50: Uniform noise with standard deviation 5/20/50;
POS: Poisson noise with the pixel value as the number of photons;
SNP: Salt and pepper noise with 1% density.

captured using only 10 bins.

Other noise types were also tested and summarized in Table III. The EFM methods perform well and show robustness in practice. The L-moment fitting also has similar results to the AWGN case, except for the uniform noise of intensity 20. In that case, the resulting KLD is usually high, and the estimated range variance in the first iteration is usually too large to turn on the second iteration.

B. MNLM-DCT filter

Considering practical usage, we overlapped the patches by every other pixel and reduced the search window to 21×21 for an 8x speedup (with ≤ 0.1 dB drop) for the filtering process compared to [8]. The fast algorithm using an integral image in [12] was also implemented. The results are summarized in Table IV. The proposed grid-subsampling strategy reduces the PMF derivation time from 10.2 to 3.6 s. The EFM-20 fitting provides very similar results compared to the direct EM+ fitting, which again demonstrates that the grid-subsampling and 20 merged bins are sufficient for the CSM fitting. The EFM-10 fitting performs even better, about 0.1 dB gain, for its smaller range variances which are preferred for the DCT-Wiener filter. In contrast, the quality degradation for using the L-moment fitting can be up to 0.3 dB because larger range variances are often inferred.

TABLE IV
SUMMARY FOR THE **MNLM-DCT** FILTER

MNLM-DCT	σ_n	1st Iteration: Fitting Result KLD ($\times10^{-3}$)	ΔPSNR (dB)	$\lvert\Delta\sigma_r^2\rvert$ (rel.)	Execution Time: Total (s)	Filter (s)	PMF (s)	Fitting (s)	Recursive Filtering: Fitting Result $\bar{m}$	ΔPSNR (dB)	PSNR (dB)	Δ_{EM+} (dB)
Direct EM+ Fitting	5	5.2	-0.3	268%	27.3	13.2	9.9	4.2	1.0	-0.3	37.3	-
	10	1.3	-0.3	109%	28.5	13.3	10.0	5.2	1.0	-0.3	34.4	-
	20	1.9	-0.2	71%	29.2	13.2	10.1	5.9	1.0	-0.2	31.3	-
	40	1.6	-0.1	44%	30.5	13.2	10.4	7.0	2.0	0.6	28.5	-
	50	2.0	-0.1	61%	31.5	13.2	10.5	7.8	2.0	0.8	27.6	-
EFM-20 Fitting	5	5.5	-0.3	255%	17.4	13.2	3.5	0.6	1.0	-0.3	37.3	0.01
	10	1.6	-0.2	95%	17.5	13.2	3.5	0.7	1.0	-0.2	34.5	0.07
	20	1.9	-0.2	73%	17.5	13.2	3.5	0.7	1.0	-0.2	31.3	0.00
	40	2.6	-0.1	49%	17.7	13.3	3.6	0.8	2.0	0.6	28.5	0.00
	50	3.0	-0.1	61%	17.6	13.2	3.6	0.8	2.0	0.9	27.6	0.02
EFM-10 Fitting	5	8.7	-0.2	212%	17.3	13.3	3.6	0.3	1.0	-0.2	37.4	0.08
	10	5.7	-0.2	90%	17.4	13.4	3.6	0.5	1.0	-0.2	34.5	0.14
	20	10.4	-0.1	53%	17.4	13.3	3.6	0.5	1.0	-0.1	31.3	0.04
	40	5.7	-0.1	44%	17.5	13.4	3.6	0.5	2.0	0.6	28.6	0.05
	50	8.2	-0.1	61%	17.4	13.3	3.6	0.5	2.0	0.9	27.6	0.02
L-Moment Fitting	5	16.8	-0.3	258%	16.9	13.3	3.6	0.003	1.0	-0.3	37.3	0.01
	10	2.6	-0.4	126%	16.9	13.3	3.6	0.003	1.0	-0.4	34.4	-0.04
	20	1.8	-0.4	159%	16.9	13.3	3.6	0.003	1.0	-0.4	31.1	-0.19
	40	2.1	-0.2	108%	16.9	13.3	3.6	0.003	1.9	0.3	28.3	-0.28
	50	2.5	-0.1	98%	17.0	13.4	3.6	0.003	2.0	0.7	27.5	-0.15

IV. CONCLUSION

In this letter, we study how to accelerate the CSM fitting process while maintaining the denoising quality. The complexity overheads consist of PMF derivation and distribution fitting. The grid-subsampling strategy reduces the PMF derivation time greatly, e.g. from 0.9 to 0.1 s for the **BF-9×9** and from 10.2 to 3.6 s for the **MNLM-DCT**. Regarding the distribution fitting, the direct EM+ fitting could dominate the total execution time, and the proposed EFM fitting can achieve 6-12x and 9-20x speedup by using 20 and 10 bins respectively. An even faster approach is the confidence-based L-moment fitting via table lookup and paraboloid refinement. The fitting time becomes as negligibly small as 0.003 s with $\leq$0.2 dB drop in most cases.

This work not only enhances the applicability of the CSM fitting but also provides some heuristics on practical usage. For example, the EFM-10 fitting is especially suitable for the **MNLM-DCT**. And the L-moment fitting is useful when the computing resource is limited or only one single filter iteration is allowed. We also believe that the proposed approaches can be extended to fit other heavy-tailed distributions which have no analytical expressions in a fast and robust way.

REFERENCES

[1] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE ICCV*, 1998, pp. 839–846.

[2] A. Buades, B. Coll, and J.-M. Morel, "A review of image denoising algorithms, with a new one," *SIAM Journal on Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, 2005.

[3] S. Paris, P. Kornprobst, J. Tumblin, and F. Durand, "Bilateral filtering: Theory and applications," in *Foundations and Trends in Computer Graphics and Vision*, vol. 4, no. 1, 2008, pp. 1–73.

[4] C. M. Stein, "Estimation of the mean of a multivariate normal distribution," *The Annals of Statistics*, vol. 9, no. 6, pp. 1135–1151, 1981.

[5] H. Peng, R. Rao, and S. A. Dianat, "Multispectral image denoising with optimized vector bilateral filter," *IEEE Trans. Image Process.*, vol. 23, no. 1, pp. 264–273, Jan. 2014.

[6] D. V. D. Ville and M. Kocher, "SURE-based non-local means," *IEEE Signal Process. Lett.*, vol. 16, no. 11, pp. 973–976, Nov. 2009.

[7] Y. Chen and K. J. R. Liu, "Image denoising games," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 10, pp. 1704–1716, Oct. 2013.

[8] C.-T. Huang, "Bayesian inference for neighborhood filters with application in denoising," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 4299–4311, Nov. 2015.

[9] J. R. M. Hosking, "L-moments: Analysis and estimation of distributions using linear combinations of order statistics," *J. Roy. Statist. Soc. B*, vol. 52, pp. 105–124, 1990.

[10] E. A. Elamir and A. H. Seheult, "Trimmed L-moments," *Comput. Statist. Data Anal.*, vol. 43, pp. 299–314, 2003.

[11] J. R. M. Hosking, "Some theory and practical uses of trimmed L-moments," *J. Statist. Plann. Inference*, vol. 137, pp. 3024–3039, 2007.

[12] J. Darbon, A. Cunha, T. F. Chan, S. Osher, and G. J. Jensen, "Fast nonlocal filtering applied to electron cryomicroscopy," in *IEEE International Symposium on Biomedical Imaging*, 2008, pp. 1331–1334.