# Estimation of Gaussian, Poissonian-Gaussian, and Processed Visual Noise and its Level Function

Meisam Rakhshanfar, *Student Member, IEEE,* and Maria A. Amer, *Senior Member, IEEE*

***Abstract*—We propose a method for estimating image and video noise of different types: white Gaussian (signal-independent), mixed Poissonian-Gaussian (signal-dependent), or processed (non-white). Our method also estimates the noise level function (NLF) of these types. We do so by classifying image patches based on their intensity and variance in order to find homogeneous regions that best represent the noise. We assume the noise variance is a piecewise linear function of intensity in each intensity class. To find noise representative regions, noisy (signal-free) patches are first nominated in each intensity class. Next, clusters of connected patches are weighted, where the weights are calculated based on the degree of similarity to the noise model. The highest ranked cluster defines the peak noise variance, and other selected clusters are used to approximate the NLF. The more information we incorporate, such as temporal data and camera settings, the more reliable the estimation becomes. To account for processed noise, (i.e., remaining after in-camera processing), we consider the ratio of low- to high-frequency energies. We address noise variations along video signals using a temporal stabilization of the estimated noise. Objective and subjective simulations demonstrate that the proposed method outperforms other noise estimation techniques, both in accuracy and speed.**

***Index Terms*—Video, image, noise estimation, white Gaussian, Poissonian-Gaussian, frequency-dependent noise, signal-dependent noise, noise level function, intensity classification.**

## I. Introduction

NOISE measurement is required in many image and video processing techniques (e.g., enhancement and segmentation, as adapting their parameters to the noise level can significantly improve their accuracy). Noise is added to an image from different sources [1]–[3] such as sensor (fixed pattern noise, dark current noise, shot noise, and amplifier noise), post-filtering (processed noise), and compression (quantization noise). In digital cameras, due to the physical properties of sensors noise is signal-dependent, and due to post-capture filtering or Bayer interpolation noise becomes frequency-dependent. We classify image and video noise into: *additive white Gaussian noise* AWGN which is frequency- and signal-independent, *Poissonian-Gaussian noise* PGN which is frequency-independent but signal-dependent (i.e., AWGN for a fixed intensity), and *processed Poissonian-Gaussian noise* PPN which is frequency- and signal-dependent (i.e., non-white Gaussian for a fixed intensity).

Many noise estimation approaches assume the noise is AWGN, which is not accurate in practical video applications, where video noise is signal- (intensity) or frequency-dependent. Techniques that estimate signal-dependent noise, on the other hand, do not accurately handle white Gaussian noise. Furthermore, noise estimation approaches rely on the assumption that high-frequency (HF) components of the noise exist, which makes them fail in real-world non-white (processed) noise. This is even more problematic in approaches using small patches (e.g., $5 \times 5$ pixels) [4]–[9] because the probability of finding a small patch with a variance much less than the noise power is higher than in a large patch.

Our contributions in this paper are the following: 1) an automatic and fast estimation of the variance of AWGN, PGN, and PPN; 2) a non-parametric estimation of the noise level functions (NLF) of these noises; 3) relating the input signal and its downsampled version so to reject non-homogeneous patches; 4) ranking noise representative regions using patch statistics and connectivity; 5) weighting of patches based on intra-frame (spatial) and on inter-frame (temporal) features; 6) integrating both capture settings and user input, if available, to enhance the estimation.

This work is an extension of [10], but it a) estimates both the noise variance and the NLF, b) estimates both processed and unprocessed noise, and c) broadens the solution by adding many new features such as temporal data. In the following, section II discusses related methods, section III presents the noise model, section IV presents the proposed method, section V gives objective ad subjective results, and section VI concludes the paper.

## II. Related work

AWGN estimation techniques can be categorized into filter-based, transform-based, edge-based, and patch-based methods. Filter-based techniques [11], [12] first smooth the image using a spatial filter and then estimate the noise from the difference between the noisy and smoothed images. In such methods, spatial filters are designed based on parameters that represent the image noise. WT or DCT based methods [13]–[19] extract the noise from the diagonal band coefficients. [18] proposed a statistical approach to analyze the DCT filtered image and suggested that the change in kurtosis values results from the input noise. They proposed a model using this effect to estimate the noise level in real-world images. Although the global processing makes transform-based methods robust, their edge versus noise differentiation lead to inaccuracy under low-level noise or in high-structured images. [18] tries to solve this problem by applying a block-based transform. [19] uses self-similarity of image blocks, where similar blocks are represented in 3D form via a 3D DCT transform. The noise variance is estimated from HF components assuming image structure is concentrated in LF. Edge-based methods [20]–[22]

This work was supported jointly by TandemLaunch Inc., wrnch Inc., and Mitacs Canada. Some images are courtesy of wrnch Inc.

The authors are with the Electrical and Computer Engineering Department, Concordia University, Montreal, QC, Canada.

select homogeneous segments via edge-detection. In patch-based methods [6]–[9], noise estimation relies on identifying pure noise patches (usually blocks) and averaging the patch variances. Overall local methods that deal with subsets of images (i.e., homogeneous segments or patches) are more accurate, since they exclude image structures more efficiently. [6] utilizes local and global data to increase robustness. In [7], a threshold adaptive *Sobel* edge detection selects the target patches, then averages the convolutions over the selected blocks for accurate estimation. Based on principal component analysis, [8] first finds the smallest eigenvalue of the image block covariance matrix and then estimates the noise variance. Gradient covariance matrix is used in [9] to select "weak" textured patches through an iterative process to estimate the noise variance. Patch size is critical for patch-based methods. A smaller patch is better for low-level noise, while larger patches make the estimation more accurate under higher noise level. For all patch sizes, estimation is error prone under processed noise; however, by taking more LF components into account, larger patches are less erroneous. By adapting the patch size in these estimators to image resolution, it is more likely to find noisy (signal-free) patches, which consequently increases the performance. Logically finding image subsets with lower energy under AWGN conditions leads to accurate results. However, under PGN conditions underestimation normally occurs. Under AWGN, [7]–[9] outperform others; however, noise underestimation in PGN makes them impractical for real-world applications.

PGN estimation methods express the noise as a function of image brightness. The main focus of related work is to first simplify the variance-intensity function and second to estimate the function parameters using many candidates as fitting points. In [4], [23], the NLF is defined as a linear function of intensity $I$, $\sigma^2(I) = aI + b$, and the goal is to estimate the constants $a$ and $b$. WT [4] and DCT [23] analysis are used to localize the smooth regions. Based on the variance of selected regions, each point in the curve is considered to perform the maximum likelihood fitting. [24] estimates noise variation parameters using a maximum likelihood estimator. This iterative procedure brings up the initial value selection and convergence problems. The same idea is applied in [21] by using a piecewise smooth image model. After image segmentation, the estimated variance of each segment is considered as an overestimate of the noise level. Then, the lower envelope variance samples versus mean of each segment is computed and based on that, the noise level function by a curve fitting is calculated. In [25], particle filters are used as a structure analyzer to detect homogeneous blocks, which are grouped together to estimate noise levels for various image intensities with confidences. Then, the noise level function is estimated from the incomplete and noisy estimated samples by solving its sparse representation under a trained basis. The curve fitting, using many variance-intensity pairs, requires enormous computations, which is not practical for many applications especially when the curve estimation is needed to be presented as a single value. As a special case of PGN with zero dependency, AWGN cases are not examined in these NLF estimation methods. In [26], a variance stabilization transform (VST) converts the properties of the noise into AWGN. Instead of processing the Gaussianized image and inverting back to Poisson model, a Poisson denoising method is applied to avoid an inverted VST.

PPN is not yet an active research area and few estimation methods exist. In [27], first, candidate patches are selected using their gradient energy. Then, the 3D Fourier analysis of the current frame and other motion-compensated frames is used to estimate the amplitude of noise. A wider assumption is in [28] by considering both frequency and signal dependency. In this method and its extended version [29], the similarity between patches and neighborhood in DCT domain is used to differentiate the noise and image structure. Using the exhaustive search, candidate patches are selected and noise is estimated in each DCT coefficient and ultimately for the whole image.

## III. Noise modeling

### A. White noise

The input noisy video frame (or still image) $F$ can be modeled as $F = F_{org} + n_d + n_g + n_q$, where $F_{org}$, $n_d$, $n_g$, and, $n_q$ are the noise-free image, white signal-dependent noise, white signal-independent noise, and, quantization and amplification noise, respectively. With modern camera technology, $n_q$ can be ignored since it is very small compared to $n_o = n_d + n_g$. $n_d$ and $n_g$ are assumed zero-mean random variables with variance $\sigma_d^2(I)$ and $\sigma_g^2$, respectively. The NLF according to each intensity $I$ can be assumed

$$\sigma^2(I) = \sigma_d^2(I) + \sigma_g^2. \tag{1}$$

We define $\sigma_o^2 = \max(\sigma^2(I))$ as the *peak* of $\sigma^2(I)$. When a video application (e.g., motion detection) requires a single noise variance, the best descriptive value is the maximum level, since a boundary can be effectively designated to discriminate between signal and noise.

### B. Processed noise

Processing technologies such as Bayer pattern interpolation, noise removal, bit-rate reduction, and resolution enlargement, are increasingly embedded in digital cameras. For example, spatial filtering is used to decrease the bit-rate. Accurate data about in-camera processing is not available in many cameras. However, processing can be bypassed manually, which allows for an assessment of the statistical properties of noise before and after processing. Fig. 1 shows parts of two images taken under the same conditions in raw and processed image modes. This figure also shows the frequency spectrum of noise in both modes. We studied the noise using homogeneous image regions that we manually selected from 35 images taken by seven different cameras (Canon EOS 6D, Fujifilm x100, Nikon D700, Olympus E-5, Panasonic LX7, Samsung NX200, Sony RX100). As we can see, filtering changes the frequency spectrum of the noise and makes it processed (frequency-dependent). We have analyzed noise characteristics in these cameras and noticed that their spatial filters remove low-power HF components of the noise (compared to noise power). As a result, low-frequency (LF) and impulse shaped noise remains (see also [30]). The difference between two images capturing the same scene (with no motion) under the same conditions contains only noise. We used these differences to calculate the noise characteristics of an image. Let us assume the peak variance of noise after in-camera filtering is $\sigma_p^2$. Depending on the applications, both noise levels before $\sigma_o^2$ and after $\sigma_p^2$ in-camera processing can be useful. We aim to estimate both. In applications such as denoising, where LF noise should be removed, the noise level before in-camera filtering can better represent LF noise. This is because LF noise remains intact after processing. Such an estimation is challenging since some noise frequency components are removed and the calculation of the pre-processing (original) noise level by its current power (e.g., variance of homogeneous patches) is no longer accurate.

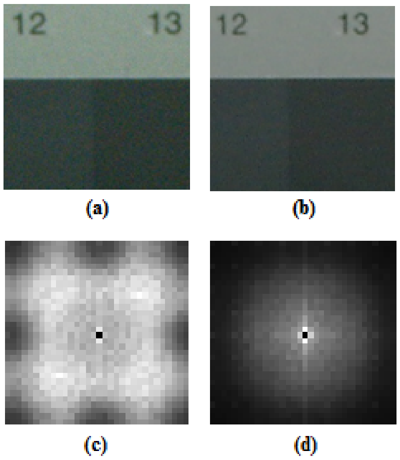

Fig. 1. Images captured with the same camera in raw (a) and processed modes (b). Average of noise frequency magnitude of 35 different images taken by seven cameras in raw (c) and processed modes (d).

When PGN becomes processed, we model the resulting image as $F_p = F_{org} + n_p$ with $n_p$ as the PPN with a peak variance $\sigma_p^2$. We model the before in-camera processing image $F$ as $F = F_p + n_f$ with $n_f$ as the filtering distortion with power $\sigma_f^2$. We thus differentiate between PGN $n_o$, PPN $n_p$, and filtering distortion $n_f$, where $n_o = n_p + n_f$. Let $1 \leq \gamma \leq \gamma_{max}$ be the degree (power) of processing on $\sigma_o^2$. We estimate

$$\sigma_o^2 = \gamma \cdot \sigma_p^2. \tag{2}$$

$\gamma = 1$ means the observed noise is PGN; $\gamma \leq \gamma_{max}$ means $F$ was *not heavily* processed (see Fig. 8). *Heavily processed* means the nature of PGN was heavily changed resulting in large $\sigma_\gamma^2$ compared to $\sigma_p^2$, i.e., $\sigma_\gamma^2 \gg \sigma_p^2$ since the mean absolute difference of $F$ and $F_p$ is large.

### *C. Noise level function*

A better adaptation of video processing applications to noise can be achieved by considering the NLF instead of a single value. However, as there is no guarantee that pure noise (signal-free) pixels are available for all intensities, NLF estimation is challenging. The NLF strongly depends on camera or capture settings and histogram modifications such as white balancing and gamma correction [21]. As shown in Fig. 2, many possibilities for the NLF shape exist and any assumption about the shape of NLF (such as a linear summation of Poisson and Gaussian distribution) cannot be taken. Thus, we consider a general shape for NLF, where the peak of noise can occur at any intensity.

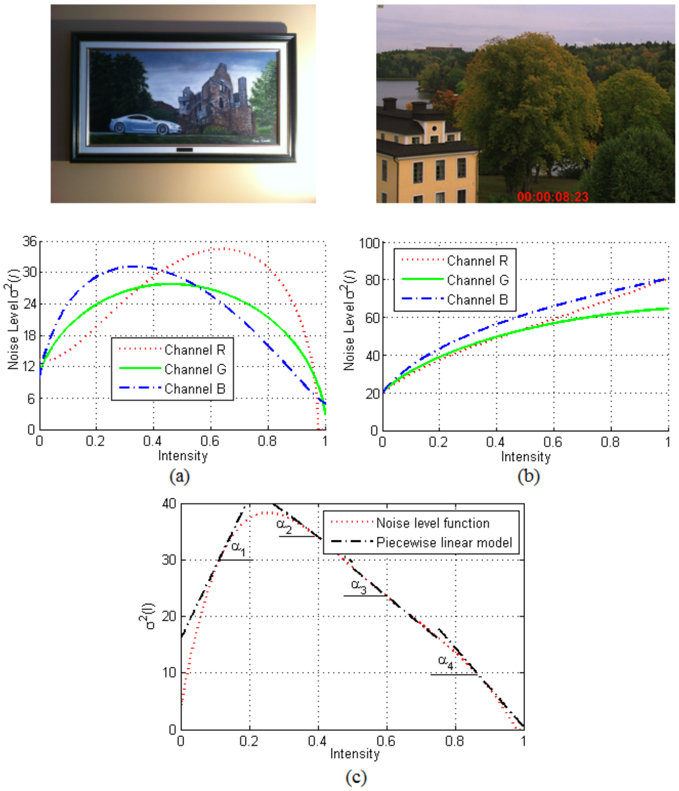

Fig. 2. NLF approximation: (a) and (b) show two sample images and their NLF in RGB channels. (c) shows piecewise linear modeling of the NLF.

Let the intensity range be divided into $M$ sub-intensity classes. A piecewise linear function, see Fig. 2(c), can approximate the NLF in intensity class $l$ as

$$\sigma_l^2(I) = \alpha_l \cdot \sigma_{rep_l}^2 \cdot (I - I_{rep_l}) + \sigma_{rep_l}^2, \tag{3}$$

where $l \in \{1, \ldots, M\}$, $I \in \{I_l^{\min}, I_l^{\max}\}$. $I_l^{\min}$ and $I_l^{\max}$ define the class boundaries, $\sigma_{rep_l}^2$ is a representative point of $\sigma_l^2(I)$ and $I_{rep_l}$ is its corresponding intensity. $\sigma_{rep_l}^2$ can be, for example, the median of $\sigma_l^2(I)$. $\alpha_l$ represents the slope of a line approximating the NLF in the class $l$ as illustrated in Fig. 2(c). If $M$ is appropriately selected, $|\alpha_l|$ does not exceed $\alpha_{max} \geq \max(|\alpha_l|)$, which we estimated experimentally in analyzing different images and cameras. With $\max(|I - I_{rep_l}|) = 1/M$ and $|\alpha_l| \leq \alpha_{max}$,

$$\sigma_l^2(I) \leq \sigma_{max_l}^2 = \alpha_{max} \cdot \sigma_{rep_l}^2 \cdot \max(|I - I_{rep_l}|) + \sigma_{rep_l}^2. \tag{4}$$

Image regions with variances greater than $\sigma_{max_l}^2$ do not fit in the NLF curve and should be rejected as non-homogeneous regions. This can thus be used to target homogeneous regions, as shown in section IV-B, where we use $\alpha_{max}$ to locate patches that fit into the linear approximation of NLF. In section IV-E, we propose an approximation of the NLF.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2016.2588320, IEEE Transactions on Image Processing

## IV. Proposed method

The proposed method is based on the classification of intensity (or color) variances of signal patches (blocks) in order to find homogeneous regions that best represent the noise. We assume that noise variance is linear, with limited slope, to the intensity of a class. To find homogeneous regions, the method works on the downsampled input image and divides it into patches. Each patch is assigned to an intensity class, whereas outlier patches are rejected. Clusters of connected patches in each class are formed and some weights are assigned to them. Then, the most homogeneous cluster is selected and the mean variance of patches of this cluster is considered as the noise variance peak of the input noisy signal. To account for processed noise, an adjustment procedure is proposed based on the ratio of LF to HF energies. To account for noise variations along video signals, a temporal stabilization of the estimated noise is proposed. The block diagram in Fig. 3 shows our noise estimator for one image or video frame without temporal considerations. Fig. 4 shows how the method is stabilized using temporal processing in video. The proposed noise estimation based on intensity-variance homogeneity classification (*IVHC*) can be summarized as in Algorithm 1. In the remainder of this section, section IV-A builds homogeneous patches; section IV-B classifies patches; section IV-C builds clusters of connected patches and estimates the noise peak variance; section IV-D estimates parameters of processed noise; section IV-E approximates the NLF; section IV-F temporally stabilizes the estimate; sections IV-G and IV-H compute intra-frame and inter-frame weights; section IV-I extends the method to camera settings and user input.

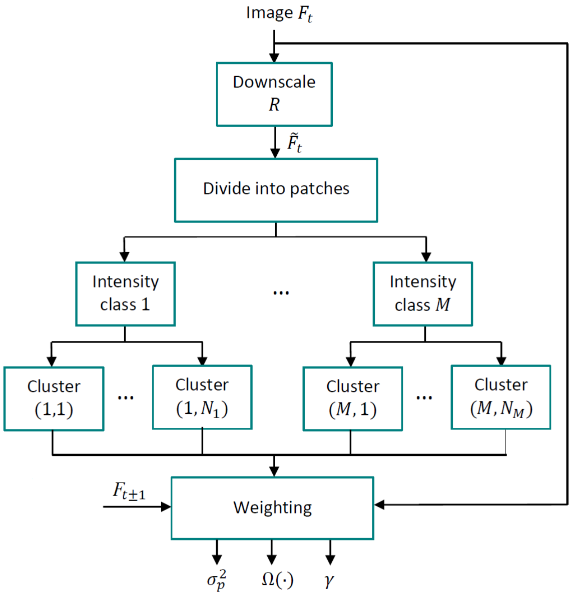

Fig. 3. Intra-frame block diagram of the proposed estimator operating spatially on a single image or video frame. $F_{t\pm1}$ is either preceding or subsequent frame (see section IV-H) and is used only for video signals.

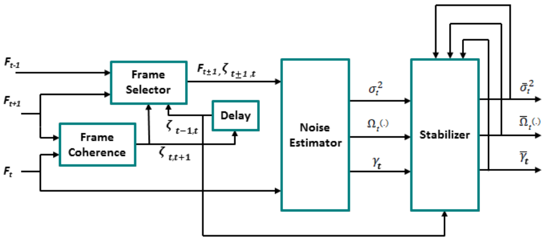

Fig. 4. Block diagram of the proposed estimator operating spatio-temporally in a video signal. The noise estimator block is shown in Fig. 3

**Algorithm 1** *IVHC* based noise estimation

```
Input: Noisy image $F$.
Output: NLF $\Omega(I)$, noise level peak $\sigma_p^2$ and the degree of processing $\gamma$.
Downscale the image $F$ to $\tilde{F}$ via (8) and divide $\tilde{F}$ into $W \times W$ patches via (5).
for each class $l = 1$ to $M$ do
    Find target connected clusters in class $l$.
    for each cluster $k = 1$ to number of clusters do
        Find the corresponding cluster $\ddot{\Phi}(l,k)$ in $F$ and remove outlier patches via (12).
        Calculate weights $\omega_1(l,k)$ to $\omega_{11}(l,k)$ for $k^{\text{th}}$ cluster.
    end for
end for
Find the cluster $\hat{\Phi}$ with the highest weight via (14).
Compute the noise variance $\sigma_p^2$ of selected cluster $\hat{\Phi}$ via (15).
Estimate the noise level function $\Omega(I)$ (18).
Estimate the in-camera processing degree $\gamma$ (17).
Stabilize the estimates $\sigma_p^2$, $\Omega(I)$, and $\gamma$ temporally via (19).
```

### A. Homogeneity guided patches

Homogeneous patches are image blocks $\tilde{B}_i$ of size $W \times W$,

$$\tilde{B}_i = \left\{ \tilde{F}(x,y) \,\middle|\, x,y \in P_i \right\}, \quad P_i = \{(x,y)| \quad \tfrac{i}{r} \le x \le \tfrac{i}{r} + W - 1, \; \text{mod}(i,r) \le y \le \text{mod}(i,r) + W - 1\} \tag{5}$$

where $\tilde{F}(x,y)$ is the downsampled version of the input noisy image at the spatial location $(x,y)$, $i$ is the patch number, mod() is the modulus after division, and $r$ is the image height (number of rows). After decomposing the image into non-overlapped patches, the noise $n_i$ of each patch can be described as $\tilde{B}_i = Z_i + n_i$, where $\tilde{B}_i$ is the observed patch corrupted by independent and identically-distributed (i.i.d) zero-mean white Gaussian noise $n_i$, and $Z_i$ is the original non-noisy image patch. The variance $\sigma^2(\tilde{B}_i)$ of a patch represents the level of homogeneity $\tilde{H}_i$ of $\tilde{B}_i$ as

$$\tilde{H}_i = \sigma^2(\tilde{B}_i) = \frac{\sum \left( \tilde{B}_i - \mu(\tilde{B}_i) \right)^2}{W^2 - 1}; \quad \mu(\tilde{B}_i) = \frac{\sum \tilde{B}_i}{W^2}. \tag{6}$$

A small $\tilde{H}_i$ expresses high patch homogeneity. Under PGN conditions, noise is i.i.d for each intensity level. If an image is classified into classes of patches with the same intensity level, the homogeneity model $\tilde{H}_i$ can be applied to each class. Assuming $M$ intensity classes and $\tilde{L}_l$ represents the patches of the $l^{\text{th}}$ intensity class,

$$\tilde{L}_l = \left\{ \tilde{B}_i \,\middle|\, I_l^{\text{min}} \le \mu(\tilde{B}_i) \le I_l^{\text{max}} \right\}, \quad l \in \{1:M\}. \tag{7}$$

### B. Adaptive patch classification

Images contain statistically more LF than HF. But small image patches show more HF than LF. Thus, small patches have the advantage of better signal-noise differentiation. Large image patches, on the other hand, are less likely to fall in the local minima, especially when noise is processed. To benefit from both, we propose image downscaling with rate $R$ with a coarse block-wise averaging filter as

$$\tilde{F}(x,y) = \frac{1}{R^2}\sum_{\hat{i},\hat{j}=0}^{R-1} F(xR+\hat{i}, yR+\hat{j}), \tag{8}$$

where $\hat{i}$ and $\hat{j}$ are indexes, and $F$ and $\tilde{F}$ are the observed and downsampled images. This gives small patches in $\tilde{F}$ and large patches in $F$. Furthermore, the processed noise converges to white in the downscaled image. Other desirable effects of downscaling are: 1) noise estimation parameters can be fixed for a lowest possible resolution of the images (note that $R$ varies depending on the input image resolution); and 2) since the down-scaled image contains more LF, the signal to noise ratio is higher. Assuming $\tilde{L}$ represents the set of patches in $\tilde{F}$; we binary classify the patches of the $l^{\text{th}}$ intensity class in $\tilde{F}$ into $\tilde{L}_l = \left\{\tilde{L}_l^0, \tilde{L}_l^1\right\}$, where $\tilde{L}_l^1$ are the target patches as

$$\tilde{L}_l^1 = \left\{\tilde{B}_i \;\middle|\; \tilde{H}_i \le \tilde{H}_{th}(l)\ ,\ \tilde{B}_i \in \tilde{L}_l\right\}. \tag{9}$$

(9) uses the homogeneity $\tilde{H}_i$ and a threshold $\tilde{H}_{th(l)}$ to perform a binary classification on $\tilde{L}_l$. Assuming the maximum value of the slopes $\alpha_l$ of the NLF in (3) is $\alpha_{\max}$. We define $\tilde{H}_{th}(l)$ as

$$\tilde{H}_{th}(l) = \alpha_{\max}\tilde{H}_{\text{med}}(l) + \beta, \tag{10}$$

where $\beta = 1$ and $\alpha_{\max} = 3$. To calculate $\tilde{H}_{\text{med}}(l)$, we first divide $\tilde{L}_l$ into three subclasses, then find the minimum $\tilde{H}_i$ in each subclass and finally we find the median of the three values. When class $l$ contains overexposed or underexposed patches, $\tilde{H}_{\text{med}}(l)$ becomes very small. Therefore, the offset $\beta$ is considered to include noisy patches. Fig. 5 shows sample target patches and their connectivity with $M = 4$. (9) may reject some homogeneous patches such as the observed holes in Fig. 5; however, other selected homogeneous patches are sufficient for an accurate estimation.

### C. Cluster selection and peak variance estimation

The classifier (9) simply uses the power criterion, which is only useful for rejecting part of non-homogeneous patches and not for accurately finding the homogeneous ones. Thus, we add a geometrical analysis to the patch selection process. We consider the connectivity of patches in both horizontal and vertical directions to form clusters of similar patches. For each cluster of connected patches in the downsampled image $\tilde{F}$, we firstly find the corresponding connected patches $B_i$ (with the size of $R\cdot W \times R\cdot W$) in the input noisy image $F$ to form the cluster $\ddot{\Phi}(l,k)$. Secondly, we eliminate certain patches of $\ddot{\Phi}(l,k)$ indicated as outliers. Finally, we assess each (outlier-removed) cluster based on the intra- and inter-frame weights $\omega_1$ to $\omega_{11}$, defined in the sections IV-G and IV-H. $\ddot{\Phi}(l,k)$ represents the $k^{\text{th}}$ cluster of connected patches in the class $l$ before outlier removal.

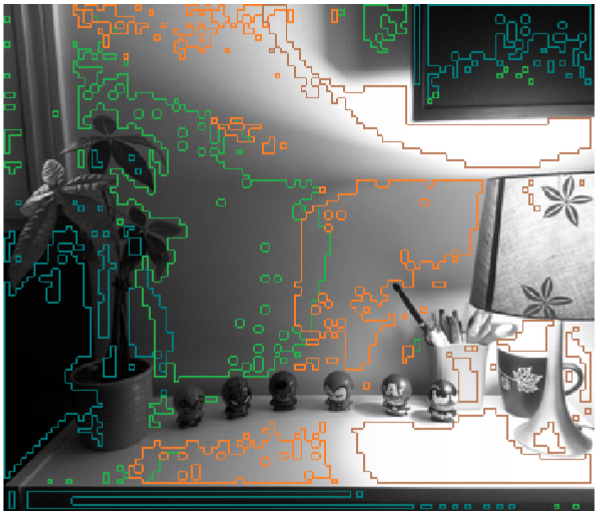

Fig. 5. Target patches: different intensity classes are shown with different colors; each class consists of several clusters of different sizes.

*1) Outlier removal:* The removal of outliers in each cluster is based on Euclidean distance of both the mean and the variance. For each cluster, the most homogeneous patch is defined as the reference patch. Patches beyond certain Euclidean distance are considered as outliers. Assuming $\ddot{\Phi}(l,k)$ represents the $k^{\text{th}}$ cluster of connected patches in the class $l$ before outlier removal, we define the reference value of variance and mean of each cluster as

$$\sigma^2_{ref}(l,k) = \min\{\sigma^2_{B_i}\}, \quad \mu_{ref}(l,k) = \text{mean}\left[B_{ref}(l,k)\right], \\ B_{ref}(l,k) = \arg\min_{B_i\in\ddot{\Phi}(l,k)}\{\sigma^2_{B_i}\}, \tag{11}$$

where $B_{ref}(l,k)$ is the patch with the minimum variance in $\ddot{\Phi}(l,k)$ and its variance $\sigma^2_{ref}(l,k)$ and mean $\mu_{ref}(l,k)$ are considered references. By defining two intervals using two thresholds, the cluster after outlier removal becomes

$$\Phi(l,k) = \big\{B_i \;\big|\; |\sigma^2_{B_i} - \sigma^2_{ref}(l,k)| \le t_\sigma(l,k) \wedge \\ |\mu_{B_i} - \mu_{ref}(l,k)| \le t_\mu(l,k) \wedge B_i \in \ddot{\Phi}(l,k)\big\} \tag{12}$$

where $t_\sigma(l,k)$ and $t_\mu(l,k)$ are the variance and the mean thresholds that are directly proportional to $\sigma^2_{ref}(l,k)$ as

$$t_\sigma(l,k) = C_\sigma \cdot \sigma^2_{ref}(l,k); \quad t_\mu(l,k) = C_\mu \cdot \frac{\sigma_{ref}(l,k)}{R\cdot W}, \tag{13}$$

where $C_\sigma = 3$ and $C_\mu = 4$. Since in the first step of cluster formation (9), the intensity is not taken into account, it is possible that $\ddot{\Phi}(l,k)$ contains a wide range of intensities. By applying (12), many homogeneous patches may be rejected. Thus, for the patches that meet the variance condition and not the mean condition, intensity range is divided into $t_\mu(l,k)$ intervals to form other clusters, which are then added to the pool of clusters for ranking.

*2) Cluster ranking:* For each outlier-reduced connected cluster $\Phi(l,k)$, we first compute the weights $\omega_j(l,k)$ and then select the final homogeneous cluster $\hat{\Phi}$ as

$$\hat{\Phi} = \arg\max_{\Phi(l,k)}\left(\sum_{j=1}^{11}\omega_j(l,k)\right). \tag{14}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2016.2588320, IEEE Transactions on Image Processing

Then, we define the peak noise level $\sigma_p^2$ in the input noisy image as the average of the patch variances in $\hat{\Phi}$ the cluster ranked highest (i.e., best represents random noise). Thus,

$$\sigma_p^2 = \frac{\sum \sigma_{B_i}^2}{N\{\hat{\Phi}\}}\ ,\ B_i \in \hat{\Phi}, \tag{15}$$

where $N\{\hat{\Phi}\}$ is the number of patches in $\hat{\Phi}$. $\sigma_p^2$ is the *peak* variance because we give higher weights to clusters with higher variances. Estimates of $\{0 \leq \omega_j(l,k) \leq 1\}$ are proposed in sections IV-G-IV-H. Fig. 6 shows the highest-ranked clusters in the different intensity classes of Fig. 5.

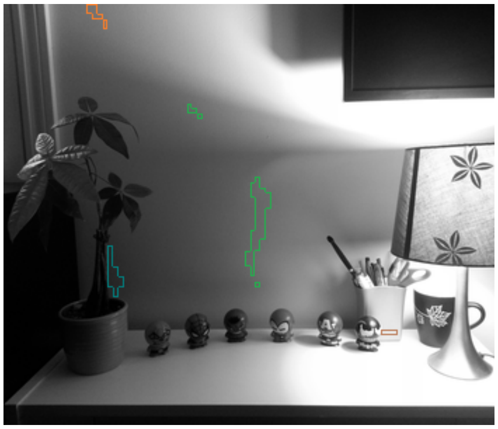

Fig. 6. Highest-ranked clusters in different intensity classes, $M = 4$. Since, each cluster represents noise of a single intensity, clusters that contain a wide range of intensities are divided into smaller clusters. We use the highest ranked cluster as a reference and other clusters to estimate the NLF.

### D. Processed noise estimation

In low-pass filtered images, the assumption that the noise is frequency-independent in each homogeneous cluster is incorrect. In such situations, $\sigma_p^2$ the variance of selected cluster (15) does not represent the true level of the noise in the unprocessed noisy image because some frequency components of the noise have been removed. In some applications such as enhancement, the level of the unprocessed (original) noise is required. To estimate this original noise, the relation between LF and HF components is necessary to trace the deviation from whiteness because we assume that the degree of noise removal in HF and LF is different. Let $E(L_f)$ represent the variance of low-pass filtered pixels of $\Phi(l,k)$ and $E(H_f)$ represent the median of the power of high-pass filtered pixels of $\Phi(l,k)$. We define

$$E_r = \frac{E(L_f)}{E(H_f)} = \frac{C_e \cdot \text{Var}\left\{h_{lp} * \Phi(l,k)\right\}}{\text{Median}\left\{|h_{hp} * \Phi(l,k)|^2\right\}} \tag{16}$$

where $*$ is convolution, $h_{lp}$ is a 3×3 moving average filter, and $h_{hp} = \mathbf{1} - h_{lp}$ a high-pass filter. $\mathbf{1}$ has zero elements except one at the center. With the given low-pass filter, according to the median of Chi-squared distribution $C_e = 8(1-\frac{2}{9})^3 = 3.7$. The ratio $E_r$ increases with stronger low-pass filtering. Since HF noise after filtering is not uniformly removed, especially when the filter is edge-stopping, we use the median operation for $E(H_f)$. In many cameras, the filtering process is optional, which allowed us to study the effect of this filtering on processed noise. Fig. 7 shows the low-to-high ratio of homogeneous regions in different raw and processed images. The more noise deviates from whiteness, the higher $E_r$ is.

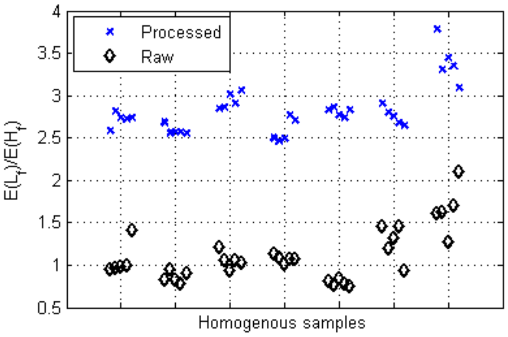

Fig. 7. Low-to-High frequency power ratio of homogeneous regions in raw and processed images taken by seven different cameras (Canon EOS 6D, Fujifilm x100, Nikon D700, Olympus E-5, Panasonic LX7, Samsung NX200, Sony RX100). Homogeneous regions were manually selected.

To approximate the processing degree $\gamma$ of (2), we studied the effect of applying anisotropic diffusion [31] and bilateral filters [32] on synthetic AWGN. Fig. 8 shows the relation between $E(L_f)$ and $E(H_f)$ and how $E_r$ relates to $\gamma$. We propose a linear approximation of $\gamma$ as

$$\gamma = 1.4E_r. \tag{17}$$

We temporally stabilized $\gamma$ as in section IV-F. As shown in Fig. 8(b) at $\gamma \approx 3.5$, the approximation becomes less accurate.

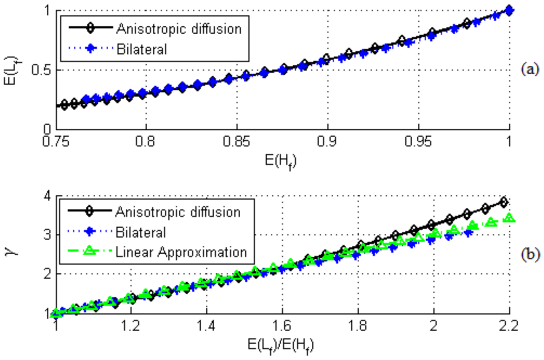

Fig. 8. Relation between the filter strength and low-to-high average frequency power ratio (a). Linear approximating $\gamma$ using the low-to-high ratio (b).

### E. Noise level function approximation

We estimate the NLF based on the peak noise variance $\sigma_p^2$ of the selected cluster $\hat{\Phi}$ defined in (15) and employ other outlier-removed clusters $\Phi(l,k)$ to approximate the NLF. First, we initialize NLF points $\hat{\Omega}(I)$ to $\sigma_p^2$, which means the noise level is identical in all intensities (white Gaussian). Then, we update $\hat{\Omega}(I)$ based on $N\{\Phi(l,k)\}$ the size (i.e., number of patches) and on $\sigma^2(l,k)$ the average of the variances of cluster $\Phi(l,k)$. We assign a weight (confidence) $\lambda(l,k)$ to $\sigma^2(l,k)$: the larger $N\{\Phi(l,k)\}$ is, the better $\sigma^2(l,k)$ represents the noise at intensity $\mu(l,k)$, meaning the closer $\lambda(l,k)$ should be to 1. The point-wise NLF $\hat{\Omega}(I)$ becomes

$$\hat{\Omega}\left(\mu\left(l,k\right)\right) = \min\left(\sigma_p^2, \frac{1}{\lambda(l,k)} \cdot \sigma^2(l,k)\right). \tag{18}$$

$\lambda(l,k) = 1 - \exp(-\frac{N\{\Phi(l,k)\}}{C_\lambda})$, meaning clusters with a smaller number of patches, are less reliable. $C_\lambda = 5$ was calculated numerically as follows: let the large clusters with 15 (or more) patches be completely reliable (i.e., $\lambda(l,k) = 1$), then from the $3\sigma$ rule $C_\lambda = 5$. Finally, the continuous NLF $\Omega(I)$ can be approximated from $\hat{\Omega}(I)$ by applying a regression analysis (e.g., curve fitting as illustrated in Fig. 9 using *polyfit* of *Matlab*). Under AWGN, $\hat{\Omega}(\mu(l,k))$ is a constant equal to $\sigma_p^2$. Under PPN, $\hat{\Omega}(\mu(l,k))$ is reduced by factor $\gamma$, but the normalized NLF shape is not altered. Thus, with $\sigma_o^2 = \gamma \cdot \sigma_p^2$ in (2) the proposed method can estimate the NLF whether the noise is processed or white.

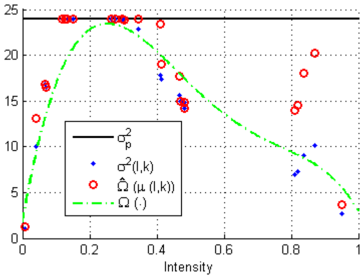

Fig. 9. Illustration of NLF approximation.

### F. Estimate temporal stabilization

In many video applications, instability of noise level is intolerable, unless the temporal coherence between frame is very small (e.g., a scene change). Let $\zeta_{t-1,t}$ represent the similarity between the current $F_t$ and previous frame $F_{t-1}$; $0 \leq \zeta_{t-1,t} \leq 1$. $\zeta$ determines how the statistical properties of new observations (i.e., image) are related to previous observations. Consider a process (such as median) $\mathbf{O}(\sigma^2_{t-N^t}, ..., \sigma^2_{t-1}, \sigma^2_t)$ to filter out outliers from the set of current $\sigma_t^2$ and previous $N^t$ estimates $\sigma^2_{t-N^t}$ to $\sigma^2_{t-1}$. When $\zeta_{t-N^t,t} = 1$, the accurate estimate should be $\mathbf{O}(\sigma^2_{t-N^t}, ..., \sigma^2_{t-1}, \sigma^2_t)$; when $\zeta_{t-1,t} = 0$, the accurate estimate is $\sigma_t^2$ itself. Thus, we propose the following linear stabilization:

$$\bar{\sigma}_t^2 = \mathbf{O}(\sigma^2_{t-N^t}, ..., \sigma^2_{t-1}, \sigma^2_t) \cdot \zeta_{t-1,t} + (1 - \zeta_{t-1,t}) \cdot \sigma_t^2, \quad (19)$$

where, $\bar{\sigma}_t^2$ is the stabilized final noise variance in $F_t$. Note $\sigma_t^2$ in (19) is $\sigma_p^2$ in (15) at time $t$. This stabilization process can be performed on both $\gamma$ and the NLF to get $\bar{\gamma}_t^2$ and $\bar{\Omega}_t(I)$.

### G. Intra-frame weighting

*1) Noise in low frequencies:* Image signal is more concentrated in LF; however, noise is equally distributed. Downsampled versus input images can be exploited to analyze noise in the LF components. The variance of finite Gaussian samples follows a *scaled chi-squared* distribution. But here we utilize an approximation benefiting the normalized Euclidean distance:

$$\omega_1(l,k) = \exp(-C_1 \cdot \frac{(\sigma^2(l,k) - R^2 \cdot \tilde{\sigma}^2(l,k))^2}{(\sigma^2(l,k))^2}), \quad (20)$$

where $\exp(\cdot)$ symbolizes the exponential function, $\sigma^2(l,k)$ and $\tilde{\sigma}^2(l,k)$ are the average of variances of the input and downsampled patches in the cluster after outlier removal $\Phi(l,k)$. The positive constant $C_1$ (e.g., 0.4) varies depending on the $R$ and the $W$. Low values of $\omega_1(l,k)$ account for image structure, where the signal is concentrated in LF.

*2) Noise in high frequencies:* The dependency of neighboring pixels is another criterion to extract image structure. The median absolute deviation (MAD) in the horizontal, vertical, and diagonal directions expresses this dependency as

$$\begin{aligned} \tau_i = \; & \text{median}\,\{|B_i(m,n+1) - B_i(m,n)|, \\ & |B_i(m+1,n) - B_i(m,n)|, \\ & |B_i(m+1,n+1) - B_i(m,n)|\}\,, \end{aligned} \quad (21)$$

where $0 \leq m,n \leq R \cdot W - 2$. According to half-normal distribution $\sigma^2_{B_i} = 2\text{erf}^{-1}(0.5) \cdot \tau_i^2 = 1.1\tau_i^2$, where $\text{erf}^{-1}$ is the inverse error function. We profit from this property to extract the likelihood function of neighborhood dependency. Assuming for each $\Phi(l,k)$, $\tau(l,k)$ is the average of $\tau_i$ of the patches in the $\Phi(l,k)$. Under AWGN, we define the following likelihood function:

$$\omega_2(l,k) = \exp(-C_2 \cdot R^2 \frac{(\sigma^2(l,k) - 1.1\tau^2(l,k))^2}{(\sigma^2(l,k))^2}). \quad (22)$$

For a Gaussian random variable, $C_2$ can be computed by numerical analysis; however, we considered a more relaxed value $C_2 = 0.2$ to handle both unprocessed and processed noise. Low values of $\omega_2(l,k)$ mean a strong neighboring dependency, which is a hint of image structure. In case of white noise, we analyze the MAD versus variance to specify whether or not the patch contains structure. Thus, in the final estimation step, we use $1.1\tau^2(l,k)$ instead of $\sigma^2(l,k)$ for patches with structure.

*3) Size of the cluster:* A large cluster has a high probability of being a homogeneous region. This is because we make sure all the connected patches of a cluster have the same statistics. Such a large cluster can provide sufficient information about the noise. However, continuing to increase the cluster size does not lead to significant accuracy improvement. In other words, in a competition between two large clusters, size is less important compared to other decision criteria. Thus, a linear relationship between the size and the corresponding weight is not efficient. We propose the following nonlinear weight for the size of the cluster:

$$\omega_3(l,k) = 1 - \exp(-C_3 \cdot \frac{N\{\Phi(l,k)\}}{N\{F\}}), \quad (23)$$

where $N\{\Phi(l,k)\}$ and $N\{F\}$ are the number of patches in $\Phi(l,k)$ and the input image, respectively. We compute $C_3$ numerically: assuming we divide the image by 5 in each dimension, each section containing 4% of the image, is large enough to give $\omega_3(l,k) = 1$; with the $3\sigma$ rule, $C_3 = \frac{3}{0.04} = 75$.

*4) Variance of means and variance of variances:* In a homogeneous cluster with a relatively large number of pixels in each patch (here $R \cdot W \times R \cdot W$), the normalized value of the variance of variances $\nu(l,k)$ and variance of means $\epsilon(l,k)$ of $\{B_i \in \Phi(l,k)\}$, should be small. So we propose:

$$\omega_4(l,k) = \omega_3(l,k)\exp(-\frac{\nu(l,k)}{\sigma^4(l,k)}), \quad (24)$$

$$\omega_5(l,k) = \omega_3(l,k)\exp(-\frac{\epsilon(l,k)}{\sigma^2(l,k)}), \tag{25}$$

where

$$\nu(l,k) = \frac{\sum\left(\sigma^2_{B_i} - \sigma^2(l,k)\right)^2}{(N\{\Phi(l,k)\})^2 - 1}, \epsilon(l,k) = \frac{\sum\left(\mu_{B_i} - \mu(l,k)\right)^2}{(N\{\Phi(l,k)\})^2 - 1}.$$

In equations (24) and (25) $\omega_4(l,k)$ and $\omega_5(l,k)$ are directly proportional to $\omega_3(l,k)$. Without this, it is probable to assign high values to $\omega_4(l,k)$ and $\omega_5(l,k)$ when the cluster has a small number of patches even though it is not homogeneous. Uniformity of mean and variance describes cluster homogeneity and leads to a high value for $\omega_4(l,k)$ and $\omega_5(l,k)$.

*5) Intensity margins:* Excluding the intensity margins from the estimation procedure can be problematic when the signal margins are informative. For instance, the elimination of dark intensities in an underexposed image leads to the removal of the majority of the data and, consequently, inaccurate estimation. We propose negative weights to the margins thus:

$$\omega_6(l,k) = -(\frac{\max(\mu(l,k) - I_H, 0)}{1 - I_H} + \frac{\max(I_L - \mu(l,k), 0)}{I_L}). \tag{26}$$

When the average intensity of a cluster has an extremely low value ($\mu(l,k) < I_L$) or high value ($\mu(l,k) > I_H$), the cluster is not informative. In those cases, we assign negative value to $\omega_6(l,k)$. Otherwise, when $I_L \leq \mu(l,k) \leq I_H$, $\omega_6(l,k) = 0$. We set $I_H = 0.9$ and $I_L = 0.06$.

*6) Variance margins:* There are cases where underexposed or overexposed image parts with very low variances are not observed in the intensity margins. On the other hand, extremely high variances signify image structure. For consumer electronic related applications, the PSNR usually is not below a certain value (e.g., 22dB). Thus, similar to intensity margins, variance margins also affect the homogeneity characterization. We thus propose the following weight:

$$\omega_7(l,k) = -\exp\left(-\frac{\sigma^2(l,k)}{\sigma^2_{min}}\right) - \exp\left(-\frac{\delta(l,k)}{\sigma^2_{max}}\right), \tag{27}$$

where $\delta(l,k) = \max(\sigma^2(l,k) - \sigma^2_{max}, 0)$, $\sigma^2_{min} = 5$ and $\sigma^2_{max} = 200$ are variance margins.

*7) Maximum noise level:* Under PGN, the maximum noise level distinguishes the signal and noise boundary. Hence, the maximum noise level and the corresponding intensity can be used to estimate the NLF. As a result, the $\Phi(l,k)$ with the maximum noise level should be ranked higher. However, some consideration should be given in order to exclude clusters containing image structures for this weighting procedure. The basic assumption that the noise variance slope is limited helps to restrict the maximum level of noise in each intensity class. So,

$$\sigma^2_p(l) = \min\left\{\alpha_{\max}\cdot\text{median}\left[\sigma^2(l,k)\right],\ \max\left[\sigma^2(l,k)\right]\right\}, \tag{28}$$

where $\sigma^2_p(l)$ is the expected peak of noise in the class $l$. By considering a valid noise variance interval, the weight can be defined as follows:

$$\omega_8(l,k) = \exp\left(-\frac{[\sigma^2_p(l) - \sigma^2(l,k)]^2}{\sigma^4(l,k)}\right). \tag{29}$$

*8) Clipping factor:* Due to bit-depth limitations, the intensity values of the input images are often clipped in low and high margins. We propose a weight according to $3\sigma$ bound:

$$\begin{aligned} \omega_9(l,k) &= \exp(-\tfrac{\mu^2_{clip}}{2\sigma^2(l,k)}) - 1; \\ \mu_{clip} = &\ \max\left[\mu(l,k) + 3\sigma(l,k) - 1, 0\right] + \\ &\ \max\left[\mu(l,k) - 3\sigma(l,k), 0\right], \end{aligned} \tag{30}$$

where 1 and 0 are maximum and minimum possible intensities. If all pixels are in the $3\sigma$ bound, $\mu_{clip} = 0$.

### H. Inter-frame weighting

Utilizing only spatial data in video signals may lead to estimation uncertainty, especially in processed noise, where the relation between LF and HF components deviates from AWGN, which in turn makes structure and noise differentiation more challenging. Another issue to consider in video is robust estimation over time, especially in joint video noise estimation and enhancement applications.

*1) Temporal variation:* Assume $B_{(i,t)}$ is $i^{\text{th}}$ patch in the noisy frame $F_t$ at time $t$, and $B_{(i,t\pm1)}$ is the corresponding patch in the adjacent noisy frame at time $t \pm 1$. Based on which adjacent frame (previous or following) has less temporal variation for the whole frame, we select $-1$ or $+1$. Assuming the noise level does not change over time, the matching (or temporal consistency) factor can be defined as

$$\omega_{10}(l,k) = \sum_{i=0}^{N\{\Phi(l,k)\}-1} \exp\left(-\frac{\left[\sigma_{(B_i,t)} - \sigma_{(B_i,t\pm1)}\right]^2}{\sigma^2_{(B_i,t)}}\right), \tag{31}$$

where $B_{(i,t)} \in \Phi_t(l,k)$, and $\Phi_t(l,k)$ is the $k^{\text{th}}$ connected cluster of class $l$ in $F_t$. Since the homogeneity detection is applied on the input noisy image, there is no guarantee that the temporal $B_{(i,t\pm1)}$ is also homogeneous. Therefore, a high temporal difference of few patches should not significantly affect $\omega_{10}(l,k)$. For this, we analyze each patch separately and aggregate all matching degrees. This is more reliable than assessing the aggregated temporal variances.

*2) Previous estimates:* In video applications, noise estimation should be stable over time and coarse noise level jumps are only acceptable when there is a scene (or lighting) change. Therefore, the cluster with the variance closer to the previous observation is more likely to be the target cluster. Assuming $\sigma^2_{t-1}$ is the estimated peak noise for previous frame, we define the following to increase temporal consistency:

$$\omega_{11}(l,k) = \zeta_{t-1,t}\cdot\exp\left(-\frac{[(\sigma_{t-1} - \sigma(l,k)]^2}{\sigma^2_{t-1}}\right), \tag{32}$$

where $0 \leq \zeta_{t-1,t} \leq 1$ measures the scene change estimated at patch level. Assuming the temporally matched patches have the mean error less than $2\sigma^2_{max}/W^2$, the ratio of temporally matched patches to the whole patches defines the $\zeta_{t-1,t}$. Note that (32) guides the estimator to find the most similar homogeneous region in $F_{t-1}$.

### *I. Camera settings and user input*

The noise type and level can be desirably modeled using camera parameters such as ISO, shutter speed, aperture, and flash on/off. However, creating a model for each camera requires excessive data processing. Also, such metadata can be lost, for example, due to format conversion and image transfer. Thus, we cannot only rely on the camera or capturing properties to estimate the noise; however, these data, if available, can support the selection of homogeneous regions and thereby increase estimation robustness. Assuming the camera settings give a probable range of noise level. Patch selection threshold $\tilde{H}_{th}(l)$ in (10) can be modified according to this range. We can also use variance margin weights in (27) to reject out of range values as we show in section V-E.

In some video applications, such as post-production, users require manual intervention to adjust the noise level for their specific needs. Assuming user knowledge about the noise level can define the valid noise range, the variance margin used in (27) can be used to reject the out of range clusters.

## V. Experimental results

The patch size $R{\cdot}W$, the number $M$, and interval $[I_l^{\min} I_l^{\max}]$ of intensity classes are the key parameters of proposed *IVHC*. All other constant parameters used in the proposed weights are given directly after their respective equations, and we have used the same set of values in all results in this paper. The patch size is a trade-off between the number of homogeneous patches and the variance of estimation error. As the patch size decreases, more homogeneous patches can be found. On the other hand, according to the sample variance rule, the variance of estimation error increases. Resolution of the image also has a key impact on the existence of homogeneous patches. As the resolution increases, the probability of larger homogeneous patches increases. For instance, a 5×5 patch in a CIF-resolution (352×288) image corresponds to a 27×18 patch in a HD (1920×1080) image. Thus, we made $R$ and hence patch size $R \cdot W$, a function of image resolution: $R = 1$ for QCIF, $R = 2$ for 720p, and $R = 3$ for HD. We have set the downsampled patch size $W$ to 5, which is an efficient trade-off between the number of found homogeneous patches and the variance of error (see [33] and [6]). Higher class number $M$ makes our assumption of limited slope of NLF more accurate. It also leads to a better NLF approximation due to availability of more samples per intensity. As $M$ increases, the number of homogeneous patches in each intensity decreases, which leads to invalid statistics such as $\sigma^2_{rep_l}$. To find $M$, we have manually extracted homogeneous regions of 30 noisy images captured by different cameras and estimated their NLF. Then, we found the class intervals that meet the limited slope criteria and lead to minimum $M$. Accordingly, we found the upper bound of intensity class $I_l^{\max} = \{0.2, 0.45, 0.85, 1\}$ and we set the overlap between classes to 0.03, which led to $I_l^{\min} = \{0, 0.17, 0.42, 0.82\}$. The reason the intervals are not equal is due to Gamma correction and white balancing, which cause the slope of NLF to be higher in lower intensities.

The proposed homogeneous cluster selection can be performed either on one channel of a color space or on each channel separately. Normally, the Y channel is less manipulated in capturing process and, therefore, noise property assumptions are more realistic. Our observation confirms that adapting the estimation to Y channel leads to better video denoising. We, therefore, use estimated target cluster in Y channel as a guide to select corresponding patches in chroma. Utilizing these patches, we calculate the properties of chroma noise (i.e., $\sigma_p^2$ and $\gamma$). Due to space constraint, simulation results here are given for the Y channel.

Target patches in (9) can be recalculated in a second iteration by adapting the $\tilde{H}_{\min}(l)$ to $\sigma_p^2$ (estimated in first iteration). A finer estimation can be performed by limiting the slope, meaning smaller value for $\alpha_{\max}$. The rest of the method is the same as in the first iteration. Since patch statistics are already computed, the complexity of a second iteration is minor. However, our tests show that a second iteration improves the estimation results slightly, not justifying iterative estimation.

In the remainder of this section, we evaluate the proposed estimation of AWGN, PGN, PPN, and NLF separately, and we show how camera settings and user input improve the estimation. We also discuss implementation issues.

### *A. Additive white Gaussian noise (AWGN)*

We have selected six state-of-the-art approaches [5]–[9], [18] and evaluated their performance on 14 test images as in Fig.10. We generated noisy images by adding a zero-mean AWGN to the ground-truth, with 4 levels of standard deviation, from 4 to 16 with the step of 4 and we run 10 Monte-Carlo experiments for each noise level. Table I demonstrates the better performance in mean of absolute errors of the proposed method compared to related methods. The average variance of the error for our method compared to related methods is similar and is not given here. Method [8] and [9] give the closest results. Fig.11 shows examples of selected homogeneous clusters.

Fig. 10. Test images for AWGN experiment: *Lena*, *Barbara*, *Boat*, *Peppers*, and ten images from the *Kodak* database.

TABLE I
AWGN: Average of absolute errors using test images in Fig. 10.

| $\sigma$ | Ref [7] | Ref [8] | Ref [9] | Ref [18] | Ref [6] | Ref [5] | Ours |
|---|---|---|---|---|---|---|---|
| **4** | 0.69 | 0.25 | 0.23 | 0.80 | 0.82 | 0.51 | **0.22** |
| **8** | 0.46 | 0.17 | **0.15** | 0.72 | 0.50 | 0.33 | **0.15** |
| **12** | 0.31 | 0.15 | 0.15 | 0.93 | 0.73 | 0.33 | **0.14** |
| **16** | 0.22 | 0.16 | 0.24 | 1.21 | 0.78 | 0.42 | **0.15** |

We also tested the proposed method for video signals. Fig. 12 shows the average result with and without temporal aid for the first 100 frames of two sequences. Collaboration of inter-frame weighting (31), (32) and temporal stabilization (19) improves the estimation. In this figure, we also compare our results to [9] (i.e., closest related work from Table I).

TABLE II
METRICQ AND NIQE COMPARISON OF PGN REMOVAL.

| | MetricQ | | | | | | | NIQE | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Image | Ref [7] | Ref [8] | Ref [9] | Ref [18] | Ref [6] | Ref [5] | Ours | Ref [7] | Ref [8] | Ref [9] | Ref [18] | Ref [6] | Ref [5] | Ours |
| *Church* | 10.35 | 7.90 | 10.10 | 8.41 | 10.69 | 10.70 | **11.08** | 2.91 | 3.78 | 3.01 | 3.02 | 3.50 | 3.19 | **4.27** |
| *Intotree* | 9.34 | 7.71 | 7.24 | 8.98 | 10.56 | 10.06 | **11.49** | 3.74 | 5.03 | 3.84 | 3.79 | 3.77 | 3.72 | **5.07** |
| *Painting1* | 22.48 | 17.19 | 20.37 | 25.20 | 22.26 | 21.57 | **25.27** | 3.61 | 5.66 | 4.22 | 4.09 | 3.90 | 3.66 | **6.03** |
| *Painting2* | 19.58 | 15.62 | 16.86 | 20.14 | 20.67 | 20.11 | **21.83** | 2.97 | 3.25 | 2.86 | 2.66 | 3.14 | 3.15 | **3.49** |
| *Office* | 12.08 | 10.01 | 10.18 | 11.93 | 11.60 | 10.61 | **13.10** | 2.80 | 3.12 | 3.32 | 2.75 | 3.42 | 2.99 | **3.90** |
| *Room* | 11.06 | 9.56 | 10.31 | 11.18 | 10.84 | 10.01 | **12.49** | 3.97 | 3.74 | 3.64 | 3.91 | 4.21 | 3.70 | **5.26** |
| *Tears* | 12.05 | 11.09 | 10.89 | 11.22 | 12.23 | 12.02 | **14.14** | 3.55 | 3.60 | 3.67 | 3.74 | **4.67** | 3.55 | 4.55 |
| *Average* | 13.85 | 11.30 | 12.28 | 13.87 | 14.12 | 13.58 | **15.63** | 3.36 | 4.03 | 3.51 | 3.42 | 3.80 | 3.42 | **4.65** |

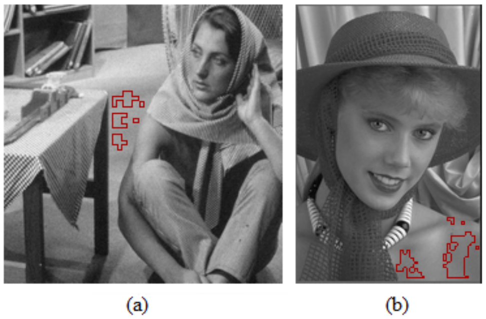

Fig. 11. The highest ranked cluster under AWGN $\sigma = 8$ (a) and $\sigma = 4$ (b). Other lower ranked homogeneous clusters are not shown here.

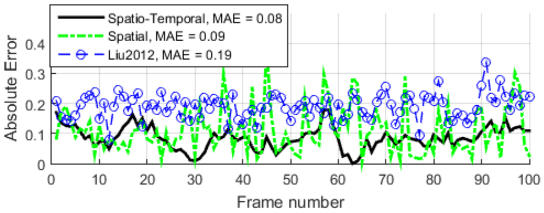

Fig. 12. Stability of the proposed method in video signals under AWGN $\sigma = 8$ with and without temporal weights. We give the mean of absolute error (MAE) over 100 frames of the *Stefan* and *Flower* sequences. Both inter-frame weighting (31), (32) and estimate stabilization (19) led to better estimate.

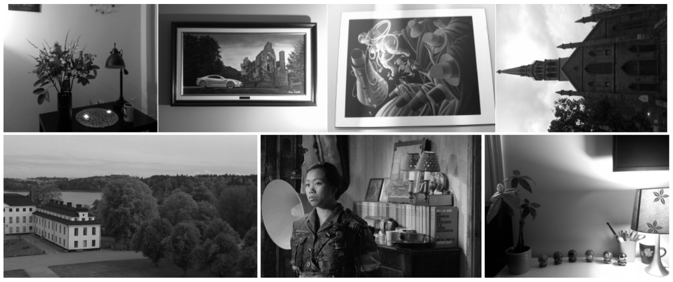

Fig. 13. Real-world PGN images: *room* (1296×968), *painting1* (1296×968), *painting2* (1296×968), *church* (1296×968), *intotree* (1920×1080), *tears* (1600×1080) and *office* (1400×1080).

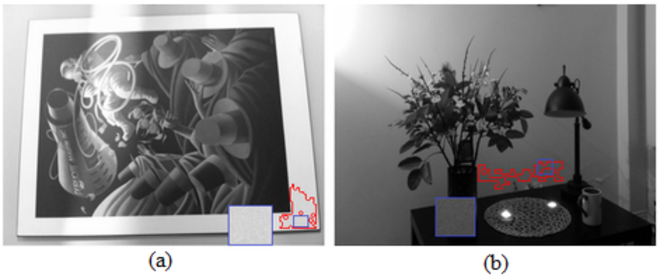

Fig. 14. Examples of highest ranked homogeneous clusters for real PGN.

### B. Poissonian-Gaussian noise (PGN)

To evaluate the performance of the proposed estimation of PGN, we tested six state-of-the-art approaches [5]–[9], [18] on seven real-world test images see Fig.13, *intotree* from *SVT HD Test Set*, *tears* from *Mango Blender* and five other real-world noisy images that were taken in raw mode, where noise is visibly signal-dependent. To objectively evaluate the PNG estimator without a reference frame, we combine the denoising method *BM3D* [34] with noise levels provided from ours and related estimators. To verify the performance, we tested different no-reference quality assessment methods such as MetricQ [35] and NIQE [36]. Table II compares MetricQ and NIQE of denoised images with a higher value indicating better quality. The proposed method yields higher quality than related methods. *IVHC* avoids underestimation by selecting the cluster with higher variance. Fig.14 shows examples of selected homogeneous clusters and Fig.15 shows visual comparison of noisy and noise-reduced image parts. As we can see, by using *IVHC* noise is better removed.

We have also evaluated our PGN estimator to denoise video signals using *BM3D*. Fig. 16 confirms the better quality of our method compared to closest related methods (from Table II) for 150 frames of the *intotree* sequence.

### C. Processed Poissonian-Gaussian noise (PPN)

If the observed noise is PPN, downscaling has the effect of converging it to white. This in turn leads to a better patch selection under processed noise. Moreover, since our method uses a large patch size $R \cdot W$, it includes more LF (i.e., a more realistic estimation). Fig. 17 shows the better performance of the proposed method with adjustment in (2), and compared to [9] (i.e., the closest to our method from Table I).

To evaluate our method under real-world processed noise, we chose 6 real images (4 from *iPhone 5* and 2 from *iPhone 6*) and applied *BM3D* [34] using noise levels provided by [8], [9], and proposed *IVHC*. Table III and Fig. 18 show that objectively and subjectively noise is better removed based on *IVHC*. We used no-reference quality assessment methods, MetricQ [35] and BIQI [37] to quantify the image quality.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TIP.2016.2588320, IEEE Transactions on Image Processing

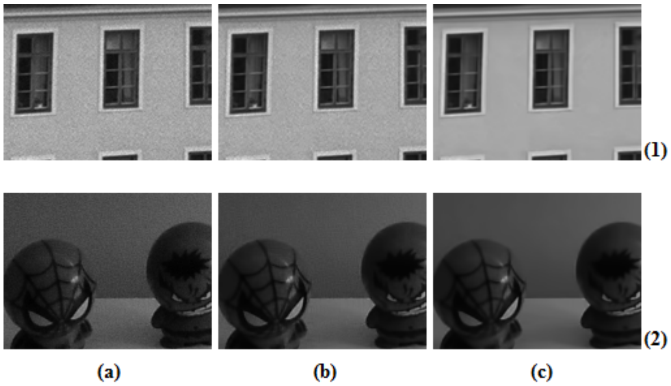

Fig. 15. Real-noise removal examples using *BM3D*. (a) original. (b) noise estimated using [7]. (c) noise estimated using *IVHC*. Noise is left in (b) while it is efficiently removed in (c).

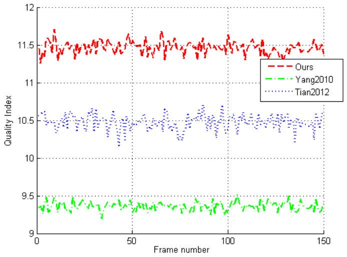

Fig. 16. PNG: MetricQ of real noise removal using different noise estimators for *Intotree* sequence.

We have compared the performance of our method with the PPN estimator [29]. We considered 14 test images of Fig.10 as the ground-truth. We added two types of frequency-dependent (processed) noise to them and estimated the noise using both estimators. To generate PPN, we processed synthetic AWGN with different standard deviations ($\sigma_a = \{10, 15\}$) using isotropic and anisotropic approaches. For isotropic processing, we used 3×3 Gaussian blur with different sigmas $\sigma_{GB} = \{0.45, 0.5, 0.55\}$. These different parameters change the standard deviation of noise to $\{0.74\sigma_a, 0.64\sigma_a, 0.56\sigma_a\}$. For anisotropic processing, we used the bilateral filter with the radius of 1 and different filtering power $\sigma^2_{BL} = \{0.5, 1, 2\}\sigma^2_a$. These different parameters change the standard deviation of noise to $\{0.85\sigma_a, 0.73\sigma_a, 0.59\sigma_a\}$. Table IV compares the mean of absolute errors for our method and for [29]. Our proposed method outperforms in both isotropic and anisotropic processing.

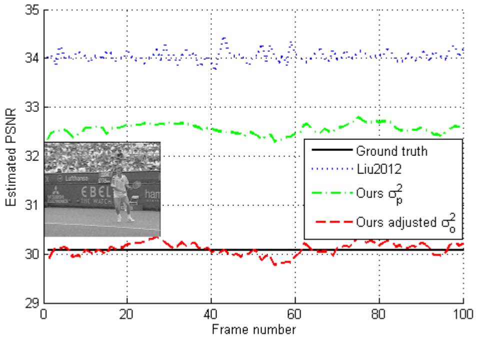

Fig. 17. Processed synthetic noise in *Stefan* video: $\sigma^2_p$ and $\sigma^2_o = \gamma\sigma^2_p$ in PSNR (original AWGN $\sigma = 8$ then filtered by bilateral filter [32]).

TABLE III
REAL-WORLD PROCESSED NOISE REMOVAL USING *BM3D* FOR 6 IMAGES CAPTURED BY SMARTPHONES.

| Method: | Ref [8] | Ref [9] | Ours |
|---|---|---|---|
| Average MetricQ: | 13.95 | 15.34 | 18.77 |

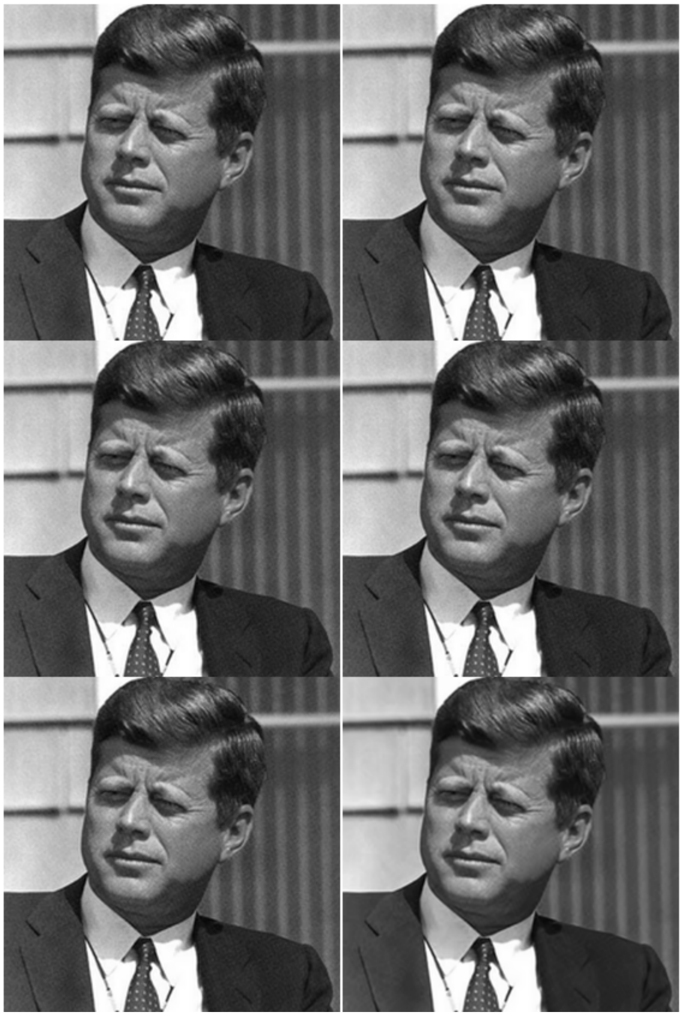

Fig. 18. Real-world processed noise removal using *BM3D* and noise estimation: from left to right and top to bottom, original image (MetricQ=36.1, BIQI=73.6), noise estimated by Ref [8] (MetricQ=38.7, BIQI=75.7), Ref [9] (MetricQ=36.6, BIQI=72.5), Ref [18] (MetricQ=36.3, BIQI=73.1), Ref [29] (MetricQ=37.2, BIQI=72.8), and proposed (MetricQ=42.7, BIQI=78.1).

### *D. Noise level function*

We applied the proposed NLF estimation on images with synthetic and real PGN. The ground-truth for real PGN images has been extracted manually (i.e., subjectively extracted homogeneous regions). Two state-of-the-art methods [21] and [4] were selected for comparison. Fig. 19 shows NLF results and Table V shows the root mean squared error (RMSE) and the maximum error comparison. The proposed *IVHC* has a better performance of finding the noise level peak, especially

TABLE IV
PROCESSED NOISE: AVERAGE OF ABSOLUTE ERROR USING TEST IMAGES IN FIG. 10.

| | Isotropic processed noise using Gaussian Blur | | | | | |
|---|---|---|---|---|---|---|
| | $\sigma_a = 10$ | | | $\sigma_a = 15$ | | |
| | $\sigma_{GB} =$ 0.45 | $\sigma_{GB} =$ 0.5 | $\sigma_{GB} =$ 0.6 | $\sigma_{GB} =$ 0.45 | $\sigma_{GB} =$ 0.5 | $\sigma_{GB} =$ 0.6 |
| Ref [29] | 0.96 | 0.98 | 0.99 | 0.68 | 0.73 | 0.70 |
| Ours | **0.18** | **0.19** | **0.23** | **0.24** | **0.27** | **0.33** |
| | Anisotropic processed noise using bilateral filter | | | | | |
| | $\sigma_a = 10$ | | | $\sigma_a = 15$ | | |
| | $\sigma^2_{BL} = 0.5\sigma^2_a$ | $\sigma^2_{BL} = \sigma^2_a$ | $\sigma^2_{BL} = 2\sigma^2_a$ | $\sigma^2_{BL} = 0.5\sigma^2_a$ | $\sigma^2_{BL} = \sigma^2_a$ | $\sigma^2_{BL} = 2\sigma^2_a$ |
| Ref [29] | 0.60 | 0.65 | 0.52 | 1.34 | 1.38 | 0.95 |
| Ours | **0.24** | **0.28** | **0.38** | **0.27** | **0.32** | **0.44** |

when the level is greater in higher intensities (e.g., *Intotree* signal).

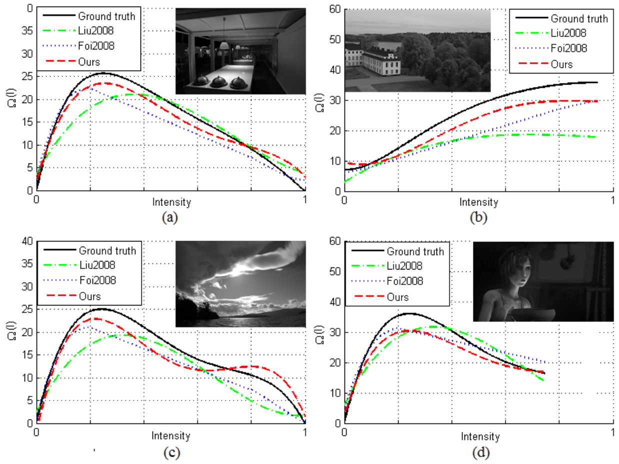

Fig. 19. Estimated NLF for SRx100II (a), Intotree (b), Salpha77 (c) and Sintel (d). Noise in (a) and (b) is real and in (c) and (d) synthetically added.

TABLE V
RMSE AND MAXIMUM OF ERROR OF NLF IN NOISY IMAGES *SRx100II* (REAL), *Intotree* (REAL), *Salpha77* (SYNTHETIC) AND *Sintel* (SYNTHETIC).

| Image | Ref [21]<br>Liu2008 | | Ref [4]<br>Foi2008 | | Ours | |
|---|---|---|---|---|---|---|
| | RMSE | MAX | RMSE | MAX | RMSE | MAX |
| *SRx100II* | 3.41 | 7.29 | 3.30 | 5.47 | **1.99** | **3.35** |
| *Intotree* | 11.17 | 17.87 | 7.31 | 9.95 | **4.10** | **6.04** |
| *Salpha77* | 4.40 | 7.45 | 3.38 | 5.63 | **2.52** | **3.89** |
| *Sintel* | 3.88 | 7.44 | 3.49 | 6.03 | **3.55** | **5.59** |
| *Average* | 5.71 | 10.01 | 4.37 | 6.77 | **3.04** | **4.72** |

### *E. Camera settings and user input*

The more image information is provided, the more reliable the estimation can be. Capturing properties, if available as a metadata, can be useful for guiding the cluster selection procedure. To test this, we selected 10 highly-textured images taken by a mobile camera (*Samsung S5*) in the burst mode without motion. First, we manually found the ground-truth peak of the noise by analyzing the homogeneous patches and temporal difference of burst mode captured images. Second, we applied our noise estimator using only Intra-frame weights and the estimated PSNR when compared the ground-truth showed an average estimation error of 1.2 dB. In the last step, we adapted both the patch selection threshold $\tilde{H}_{th}(l)$ in (10) and the variance margin weight $\omega_7(l,k)$ in (27) to the meta-data brightness value and ISO. This led to more reliable estimations with an average error of 0.34dB in PSNR.

Performance of image and video processing methods improves if expertise of their users can be integrated. Our method easily allows such an integration. For example, if the user of an offline application can define a possible noise range, the proposed variance margin (27) can be used to reject the out of range clusters.

### *F. Implementation issues*

The source codes of [7]–[9], [18], and [29] were obtained from the authors' websites. We implemented [6], [5], and our method using *Matlab*. We measured the processing time of related methods using a 3.07 GHz, i7 CPU. Table VI shows the results. The proposed method is significantly faster than the related methods. This is mainly because our method rejects most of non-homogeneous patches at the first step. We placed our *Matlab* software and other supplementary materials on our project website http://users.encs.concordia.ca/~amer/NEstIVHC/.

TABLE VI
AVERAGE OF ELAPSED TIME IN SECONDS TO PROCESS 10 HD (1920×1080) FRAMES FROM *intotree* SEQUENCE.

| Ref [7] | Ref [8] | Ref [9] | Ref [18] | Ref [6] | Ref [5] | Ref [29] | Ours |
|---|---|---|---|---|---|---|---|
| 1.2 | 17.7 | 7.6 | 9.7 | 21.0 | 10.1 | 41.4 | 0.1 |

## VI. CONCLUSION

Noise estimation methods typically assume image or video noise is white Gaussian. This paper bridges the gap between the well studied white Gaussian noise and the more complicated white signal-dependent and non-white processed types. We proposed a noise estimation method that widens noise assumptions based on the classification of intensities and on the extraction of weights using statistical noise properties and homogeneous regions in the input image. The use of connected clusters of homogeneous patches allowed us to approximate the noise level function with satisfactory results. We estimated the degree of processed noise versus white noise as a ratio of low to high frequency energies in the image. To better differentiate between noise and image structure, we used both the input noisy image and its downscaled version. We showed that the proposed method robustly handles different types of visual noise: white Gaussian, white Poissonian-Gaussian, and processed (non-white) Gaussian noise. Subjective and objective (with/without reference) results showed the higher performance of our method both in accuracy and speed. Simulation results in this paper are given for the gray-level format of noisy signals. However, we have tested our method on color sequences and it also outperforms related work.

## REFERENCES

[1] R. Szeliski, *Computer vision: algorithms and applications*, Springer, 2010.

[2] Y. Tsin, V. Ramesh, and T. Kanade, "Statistical calibration of CCD imaging process," in *Computer Vision ICCV, IEEE Int. Conf. on*. IEEE, 2001, vol. 1, pp. 480–487.

[3] G.E. Healey and R. Kondepudy, "Radiometric CCD camera calibration and noise estimation," *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, vol. 16, no. 3, pp. 267–276, Mar 1994.

[4] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical Poissonian-Gaussian noise modeling and fitting for single-image raw-data," *Image Processing, IEEE Trans. on*, vol. 17, no. 10, pp. 1737–1754, 2008.

[5] M. Ghazal and A. Amer, "Homogeneity localization using particle filters with application to noise estimation," *Image Processing, IEEE Trans. on*, vol. 20, no. 7, pp. 1788–1796, 2011.

[6] J. Tian and Li Chen, "Image noise estimation using a variation-adaptive evolutionary approach," *Signal Processing Letters, IEEE*, vol. 19, no. 7, pp. 395–398, 2012.

[7] Sh.-M. Yang and Sh.-Ch. Tai, "Fast and reliable image-noise estimation using a hybrid approach," *Journal of Electronic Imaging*, vol. 19, no. 3, pp. 033007–033007, 2010.

[8] S. Pyatykh, J. Hesser, and Lei Zheng, "Image noise level estimation by principal component analysis," *Image Processing, IEEE Trans. on*, vol. 22, no. 2, pp. 687–699, 2013.

[9] X. Liu, M. Tanaka, and M. Okutomi, "Noise level estimation using weak textured patches of a single noisy image," in *Image Processing (ICIP), IEEE Int. Conf. on*, 2012, pp. 665–668.

[10] M. Rakhshanfar and A. Amer, "Homogeneity classification for signal-dependent noise estimation in images," in *Image Processing (ICIP), IEEE Int. Conf. on*, Oct 2014, pp. 4271–4275.

[11] T.-A. Nguyen and M.-Ch. Hong, "Filtering-based noise estimation for denoising the image degraded by Gaussian noise," in *Advances in Image and Video Technology*, pp. 157–167. Springer, 2012.

[12] D.-H. Shin, R.-H. Park, S. Yang, and J.-H. Jung, "Block-based noise estimation using adaptive Gaussian filtering," *Consumer Electronics, IEEE Trans. on*, vol. 51, no. 1, pp. 218–226, 2005.

[13] D.L. Donoho, , and J.M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[14] E.J. Balster, Y.F. Zheng, and R.L. Ewing, "Combined spatial and temporal domain wavelet shrinkage algorithm for video denoising," *Circuits and Systems for Video Technology, IEEE Trans. on*, vol. 16, no. 2, pp. 220–230, 2006.

[15] J. Yang, Y. Wang, W. Xu, and Q. Dai, "Image and video denoising using adaptive dual-tree discrete wavelet packets," *Circuits and Systems for Video Technology, IEEE Trans. on*, vol. 19, no. 5, pp. 642–655, 2009.

[16] M. Hashemi and S. Beheshti, "Adaptive noise variance estimation in Bayes-Shrink," *Signal Processing Letters, IEEE*, vol. 17, no. 1, pp. 12–15, 2010.

[17] H.H. Khalil, R.O.K. Rahmat, and W.A. Mahmoud, "Chapter 15: Estimation of noise in gray-scale and colored images using median absolute deviation (MAD)," in *Geometric Modeling and Imaging, 3rd Int. Conf. on*, July 2008, pp. 92–97.

[18] D. Zoran and Y. Weiss, "Scale invariance and noise in natural images," in *Computer Vision, IEEE 12th Int. Conf. on*, Sept 2009, pp. 2209–2216.

[19] A. Danielyan and A. Foi, "Noise variance estimation in nonlocal transform domain," in *Local and Non-Local Approximation in Image Processing LNLA, Int. Workshop on*. IEEE, 2009, pp. 41–45.

[20] Sh.-Ch. Tai and Sh.-M. Yang, "A fast method for image noise estimation using Laplacian operator and adaptive edge detection," in *Communications, Control and Signal Processing ISCCSP, Int. Symposium on*, 2008, pp. 1077–1081.

[21] Ce Liu, R. Szeliski, S.B. Kang, C.L. Zitnick, and W.T. Freeman, "Automatic estimation and removal of noise from a single image," *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, vol. 30, no. 2, pp. 299–314, 2008.

[22] P. Fu, Q.S. Sun, Z.X. Ji, and Q. Chen, "A new method for noise estimation in single-band remote sensing images," in *Fuzzy Systems and Knowledge Discovery, Int. Conf. on*, May 2012, pp. 1664–1668.

[23] A. Foi, "Practical denoising of clipped or overexposed noisy images," in *EUSIPCO, 16th European Signal Processing Conf.*, 2008, pp. 1–5.

[24] A Jezierska, C. Chaux, J.-C. Pesquet, H. Talbot, and G. Engler, "An EM approach for time-variant Poisson-Gaussian model parameter estimation," *Signal Processing, IEEE Trans. on*, vol. 62, no. 1, pp. 17–30, Jan 2014.

[25] J. Yang, Zh. Wu, and Ch. Hou, "Estimation of signal-dependent sensor noise via sparse representation of noise level functions," in *Image Processing (ICIP), 19th IEEE Int. Conf. on*, Sept 2012, pp. 673–676.

[26] X. Jin, Zh. Xu, and K. Hirakawa, "Noise parameter estimation for Poisson corrupted images using variance stabilization transforms," *Image Processing, IEEE Trans. on*, vol. 23, no. 3, pp. 1329–1339, 2014.

[27] A. Kokaram, D. Kelly, H. Denman, and A. Crawford, "Measuring noise correlation for improved video denoising," in *Image Processing (ICIP), 19th IEEE Int. Conf. on*, Sept 2012, pp. 1201–1204.

[28] M. Colom, M. Lebrun, A. Buades, and J.M. Morel, "A non-parametric approach for the estimation of intensity-frequency dependent noise," in *Image Processing (ICIP), 21th IEEE Int. Conf. on*, Oct 2014.

[29] M. Colom, M. Lebrun, A. Buades, and J. M. Morel, "Nonparametric multiscale blind estimation of intensity-frequency-dependent noise," *IEEE Trans. on Image Processing*, vol. 24, no. 10, pp. 3162–3175, Oct 2015.

[30] Ce Liu and William T Freeman, "A high-quality video denoising algorithm based on reliable motion estimation," in *Computer Vision–ECCV 2010*, pp. 706–719. Springer, 2010.

[31] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, vol. 12, no. 7, pp. 629–639, 1990.

[32] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Computer Vision, Int. Conf. on*, Jan 1998, pp. 839–846.

[33] A. Amer and E. Dubois, "Fast and reliable structure-oriented video noise estimation," *Circuits and Systems for Video Technology, IEEE Trans. on*, vol. 15, no. 1, pp. 113–118, 2005.

[34] K.. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *Image Processing, IEEE Trans. on*, vol. 16, no. 8, pp. 2080–2095, 2007.

[35] X. Zhu and P. Milanfar, "Automatic parameter selection for denoising algorithms using a no-reference measure of image content," *Image Processing, IEEE Trans. on*, vol. 19, no. 12, pp. 3116–3132, 2010.

[36] A. Mittal, R. Soundararajan, and A. C. Bovik, "Making a "completely blind" image quality analyzer," *Signal Processing Letters, IEEE*, vol. 20, no. 3, pp. 209–212, March 2013.

[37] A.K. Moorthy and A.C. Bovik, "A two-step framework for constructing blind image quality indices," *Signal Processing Letters, IEEE*, vol. 17, no. 5, pp. 513–516, May 2010.

**Meisam Rakhshanfar** received the B.Sc. and M.Sc. degree in electrical engineering from Polytechnic University of Tehran, Iran, in 2002 and 2005 (with highest honors), and is currently pursuing the Ph.D. degree in the Electrical and Computer Engineering Department, Concordia University, Montreal, QC, Canada. He has over 10 publications including three patents. His research interests include video quality enhancement (noise estimation and reduction), video compression, motion estimation, real-time signal processing. He received 3 internships from Mitacs Canada and he won a leadership award from TandemLaunch inc.

**Maria A. Amer** received her Diploma degree in Computer Engineering from Universität Dortmund, Germany, in 1994 and her Ph.D. degree in Telecommunications from the INRS, Université du Québec, QC, Canada, in 2001. Dr. Amer is currently an associate professor with the Department of Electrical and Computer Engineering, Concordia University, Montréal, QC, Canada. From 2011 to 2014, she was the CTO of a research spin-off (in video enhancement technologies), co-founded with TandemLaunch Inc. From 1995 to 1997, she was with Siemens-AG/Munich and University of Dortmund as a Research and Development Associate. Dr. Amer holds 8 patents and has published over 60 papers. One of her noise reduction algorithms has been implemented in Siemens-AG TV chip sets; her recent technologies for noise estimation, reduction, and quality assessment have been adopted by wrnch Inc. Her research interests are in video processing such as enhancement, quality assessment, segmentation, and object tracking. Dr. Amer is an Associate Editor for the EURASIP Journal on Image and Video Processing.