Nonlinearly Weighted First-order Regression for Denoising Monte Carlo Renderings

Benedikt Bitterli\textsuperscript{1} Fabrice Rousselle\textsuperscript{1} Bochang Moon\textsuperscript{1} José A. Iglesias-Guitián\textsuperscript{1} David Adler\textsuperscript{2} Kenny Mitchell\textsuperscript{1,3} Wojciech Jarosz\textsuperscript{4} Jan Novák\textsuperscript{1}

\textsuperscript{1}Disney Research \textsuperscript{2}Walt Disney Animation Studios \textsuperscript{3}Edinburgh Napier University \textsuperscript{4}Dartmouth College

Abstract
We address the problem of denoising Monte Carlo renderings by studying existing approaches and proposing a new algorithm that yields state-of-the-art performance on a wide range of scenes. We analyze existing approaches from a theoretical and empirical point of view, relating the strengths and limitations of their corresponding components with an emphasis on production requirements. The observations of our analysis instruct the design of our new filter that offers high-quality results and stable performance. A key observation of our analysis is that using auxiliary buffers (normal, albedo, etc.) to compute the regression weights greatly improves the robustness of zero-order models, but can be detrimental to first-order models. Consequently, our filter performs a first-order regression leveraging a rich set of auxiliary buffers only when fitting the data, and, unlike recent works, considers the pixel color alone when computing the regression weights. We further improve the quality of our output by using a collaborative denoising scheme. Lastly, we introduce a general mean squared error estimator, which can handle the collaborative nature of our filter and its nonlinear weights, to automatically set the bandwidth of our regression kernel.

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Raytracing; I.4.3 [Computer Graphics]: Enhancement—Filtering

1. Introduction
In the last few years, rendering in movie production has experienced a dramatic shift [KFF\textsuperscript{*}15], as evidenced by recent developments in production renderers like Renderman, Arnold, Hyperion, and Manuka. Methods based on rasterization and micropolygon rendering—the dominant rendering techniques for decades—are being displaced by variants of path tracing. The potential benefits of this transition are particularly appealing for production, not only due to the increased flexibility and generality of the (quasi) Monte Carlo (MC) integration, but also due to native support for progressive rendering that allows for quick turnaround for rough previews while easily scaling to final quality with more render time.

Unfortunately, MC rendering comes with the drawback that rendered images are noisy, and, moreover, the level of noise decreases only with the square-root of the number of samples, making synthesis of perceptually noise-free images prohibitively costly for movie production. This has made image-based denoising methods a virtual necessity in production, and images are only rendered up until the remaining noise can be eliminated safely (i.e. without introducing artifacts) using less expensive image-based denoising.

Graphics researchers have developed and adapted many denoising techniques from the signal-processing community; see the comprehensive survey by Zwicker and colleagues [ZJL\textsuperscript{*}15]. While these approaches have gained considerable traction in recent years, based on our analysis, none of them seems capable of outperforming its rivals at all fronts.

An ideal, production-ready denoising algorithm should possess the following essential characteristics:

Effectiveness. Produce significant noise reduction without introducing disturbing artifacts or overblurring.

Predictability. Perform consistently well without requiring per-shot parameter tuning or other manual intervention and degrade gracefully when assumptions are not met.

Temporal stability. Produce flicker-free results across frames rendered with varying random samples, camera positions, object positions, or other scene changes.

Ease of adoption. The filter should be easy to integrate into the production environment: these approaches have gained considerable traction in recent years, based on our analysis, none of them seems capable of outperforming its rivals at all fronts.
Speed. The filter should run significantly faster than the time to render equivalent additional samples. Ideally, filtering would be much faster than a typical render so that it can be performed multiple times during a render for adaptive sampling and for viewing denoised progressive rendering.

Memory. Use significantly less memory than the renderer itself to minimize memory cost of adaptive sampling.

In this paper, we pick the small subset of existing algorithms (Table 1) that represent the state of the art with respect to the aforementioned criteria and compare their performance on a diverse set of test scenes. Based on our findings, we then propose a new practical algorithm that combines individual elements of the algorithms in a novel way, yielding comparable or higher-quality results than attainable before.

Our Nonlinearly weighted First-Order Regression (NFOR) leverages the correlation between auxiliary buffers and ground truth images. We additionally incorporate a patch-based weighting kernel into our regression framework to robustly preserve fine details even when features do not provide enough correlation.

2. Previous Work

In the following, we briefly review image-space denoising algorithms that are relevant to our work; a more thorough theoretical and comparative analysis follows in Sections 3 and 4. We focus on a posteriori techniques that rely on empirical (and often statistical) analysis of sampled data without requiring prior knowledge of how the samples were generated. A thorough review of the alternative—a priori methods—can be found in a recent survey [ZJL+15].

Algorithms for denoising MC renderings often build upon techniques developed for processing natural images, such as the bilateral [TM98, PKTD08], cross bilateral [ED04, PSA+04], and NL-Means [BCM05] filters, or wavelet thresholding [Don95]. Examples include a cross bilateral filter that utilizes a Gaussian blurred image as an edge-stopping function [XP05], non-local means filters which employs a dual-buffer variance estimation [RKZ11], or average them over different colors; the latter ensures that edges are well preserved in the filtered image [XP05], or non-local means filter [BCM05] is a generalization of the bilateral filter where the non-linear (e.g. bilateral or NL-Means) filter is challenging as the weights are both noisy and correlated with the input. Following the work of Van de Ville and Kocher [VdVK09], both Li et al. [LWC12] and Rousselle et al. [RMZ13] estimate the MSE using Stein’s Unbiased Risk Estimate [Ste81] (SURE), a powerful technique that can be used to estimate the MSE of non-linear filters. Kalantari et al. [KBS15] proposed an interesting alternative to conventional bandwidth estimation; they used a neural network trained on a diverse set of scenes to set the parameters of cross bilateral and cross-NL-Means filters.

3. Theoretical Background

Image-space denoising estimates the value of a pixel $p$ as a weighted average of its neighborhood:

\[
\hat{c}_p = \frac{1}{C_p} \sum_{q \in N_p} c_q w(p,q),
\]

where $c_p$ and $\hat{c}_p$ are the (noisy) input and filtered color values of a pixel $p$, $N_p$ is the (typically square) neighborhood centered on $p$, $w(p,q)$ is the weight of the contribution of pixel $q$ to the estimate, and $C_p = \sum_{q \in N_p} w(p,q)$ is a normalization factor. We now describe existing approaches for setting $w(p,q)$ that have been used for denoising MC renderings, in order of increasing complexity.

3.1. Zero-order Models

Bilateral Filtering. The bilateral filter [TM98, Eta02] is a popular edge-preserving smoothing filter that leverages a conventional filtering kernel (typically a Gaussian), and applies it to both the spatial and range (i.e. color, also referred to as radiometric) components of the image:

\[
w_{bf}(p,q) = \exp\left(-\frac{||p-q||^2}{2\sigma_1^2}\right) \exp\left(-\frac{||c_p - c_q||^2}{2\sigma_2^2}\right),
\]

where $\sigma_1^2$ and $\sigma_2^2$ are parameters to adjust distance across image and color space, respectively. For example, the above weight will be small if $p$ and $q$ are far from each other spatially or have different colors; the latter ensures that edges are well preserved in the filtered image. A key issue with the bilateral filter, however, is that the range-distance estimation can be sensitive to noise. To make it more stable, one can either compute the range distances on a low-pass-filtered version of the image [XP05], or average them over small neighborhoods around pixels, as described next.

Non-Local Means. The non-local means (NL-Means) filter [BCM05] is a generalization of the bilateral filter where the range distance is computed using small patches $P(\cdot)$ around the pixels:

\[
w_{nlm}(p,q) = \exp\left(-\frac{||p-q||^2}{2\sigma_1^2}\right) \exp\left(-\frac{||P_P - P_q||^2}{k^22\sigma_2^2}\right).
\]

© 2016 The Author(s)

where $k$ is a user parameter controlling the strength of the filter. The squared patch-based range distance is defined as:

$$
\|P_p - P_q\|^2 = \max \left( 0, \frac{1}{|P|} \sum_{p \in P} d(p + n, q + n) \right),
$$

(4)

where $P_p$ is a square patch of size $|P|$ (usually $7 \times 7$ pixels) centered on $p$, and $P_0$ enumerates the offsets to the pixels within a patch. The range distance $d(p, q)$ between two pixels is estimated using the squared difference of their respective colors, i.e. $d(p, q) = \|c_p - c_q\|^2$. It is worth noting that since the colors are noisy, the estimate tends to overestimate the true squared distance; this is referred to as the noise bias. Buades et al. [BCM05] propose to counter the overestimation by subtracting twice the variance of the input colors $\sigma_w^2$:

$$
d(p, q) = \|c_p - c_q\|^2 - 2\sigma_w^2.
$$

(5)

When the pixel variance changes spatially, which is often the case in MC rendering, the previous distance estimate can be further improved by considering the per-pixel color variance as proposed by Rousselle at al. [RKZ12]:

$$
d(p, q) = \|c_p - c_q\|^2 \left( \frac{\text{Var}_p + \text{Var}_{p,q}}{\text{Var}_p + \text{Var}_q} \right),
$$

(6)

where $\text{Var}_p$ is the variance at $p$, $\text{Var}_{p,q} = \min(\text{Var}_p, \text{Var}_q)$, and $\varepsilon$ is a small offset to prevent dividing by zero.

Joint Filtering. If additional per-pixel information is available, the quality of bilateral or NL-Means filtering can be further improved by factoring the information into the weight—the so-called joint filtering:

$$
w_{\text{bd}}(p, q) = w_{bd}(p, q) w_{\text{aux}}(p, q),
$$

(7)

$$
w_{\text{nlm}}(p, q) = w_{\text{nlm}}(p, q) w_{\text{aux}}(p, q),
$$

(8)

where $w_{\text{aux}}(p, q)$ is a weight derived from the additional information. Given a vector of $k$ auxiliary feature buffers, $f = (f_1, ..., f_k)$, we can compute the weight as:

$$
w_{\text{aux}}(p, q) = \prod_{i=1}^k \exp(-d_{f,i}(p, q)),
$$

(9)

where:

$$
d_{f,i}(p, q) = \left\| f_p - f_q \right\|^2 / 2\sigma^2_{f,i},
$$

(10)

and $\sigma^2_{f,i}$ is the bandwidth of feature $i$. Joint filtering is typical for denoising MC renderings where feature buffers (e.g. normal, albedo, depth) can be obtained inexpensively as a byproduct of rendering the image.

Zero Order Regression. It can be shown that all the aforementioned filters are solutions to a weighted, local, zero-order regression, i.e. they model the pixel neighborhood as a constant function. Consider the following general formulation of zero-order regression:

$$
\hat{c}_p = \arg\min_{\hat{c}_p} \sum_{q \in N_p} (c_q - \hat{c}_p)^2 \hat{w}_x(p, q)
$$

(11)

Table 1: Acronyms for the previous works considered in this analysis, and corresponding regression order.

<table>
<thead>
<tr>
<th>Tag</th>
<th>Algorithm</th>
<th>Order</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBF</td>
<td>Learning-Based filter</td>
<td>0</td>
<td>[KBS15]</td>
</tr>
<tr>
<td>NLM</td>
<td>Non-Local Means</td>
<td>0</td>
<td>[RKZ12]</td>
</tr>
<tr>
<td>RDFC</td>
<td>Robust Denoising – Features &amp; Color</td>
<td>0</td>
<td>[RMZ13]</td>
</tr>
<tr>
<td>RHF</td>
<td>Ray Histogram Fusion</td>
<td>0</td>
<td>[DMB+14]</td>
</tr>
<tr>
<td>WLR</td>
<td>Weighted Linear Regression</td>
<td>1</td>
<td>[MCY14]</td>
</tr>
</tbody>
</table>

where $x = (x_1, ..., x_D)$ is a $D$-dimensional feature vector storing the spatial, radiometric, and auxiliary features; and $\hat{w}_x$ is the weight calculated from $x$. Takeda et al. [TFM07, Eq. 45], show that if the feature space consists of pixel coordinates and pixel colors, i.e. $\hat{w}_x(p, q) = w_{\text{bd}}(p, q)$, then the solution to Equation (11) is the bilateral filter. By extension, the joint bilateral and NL-Means filters are also solutions of weighted, zero-order regressions, but using a higher dimensional feature vector $x$ to set the weights. In the case of the joint NL-Means filter, the feature vector $x$ contains the coordinates of pixel $p$, the colors of the pixels in the patch $P_p$, and the auxiliary feature buffers.

3.2. First-order Models

A natural extension to zero-order models is to consider a first-order regression:

$$
\hat{c}_p, \nabla \hat{c}_p = \arg\min_{\hat{c}_p, \nabla \hat{c}_p} \sum_{q \in N_p} (c_q - \hat{c}_p - \nabla \hat{c}_p \cdot (y_q - y_p))^2 \hat{w}_x(p, q)
$$

(12)

where $y$ is, similarly to $x$, a 1D vector corresponding to a point in the high dimensional feature space. While we may have $y = x$, it is not a requirement, that is, we may choose to use a different set of features to define the regression weights, than to perform the regression itself. We will detail the algorithms which use first-order regression [BEM11, MCY14] in the next section.

4. Comparative Analysis

In the previous section, we showed how various methods for denoising MC renderings can be formulated as either a zero- or first-order regression. Besides the order, the main differentiating characteristic is how these methods compute the regression weights $\hat{w}_x$. There are three elements involved in this computation: the kernel function (uniform, Gaussian, Epanechnikov, etc.), the pixel features considered (color, normal, albedo, etc.), and the kernel bandwidth along each dimension of the feature space.

In this section we cover techniques used in recent works to increase the robustness of the regression weights computation. For the purpose of our comparative analysis, we consider recent methods, listed in Table 1, that encompass the current state of the art for denoising Monte Carlo renderings. We use a test bench composed of 21 scenes covering a wide range of light transport effects, and provide full resolution results and error metrics (MSE, relative MSE, PSNR, and SSIM) for all methods in our supplemental material, along with a web-based interactive viewer that allows for close inspection of the results.

© 2016 The Author(s)
For personal use only. Definitive version at http://diglib.eg.org/ and http://onlinelibrary.wiley.com/
We first present the recently proposed histogram-based distance, then the use of auxiliary buffers in joint filtering schemes, and the approaches used to automatically set kernel bandwidths. We will then consider the specific case of first-order regressions, and finally conclude with some observations that will instruct the design of our proposed filter.

4.1. Mean-vs. Histogram-based Distance

Denoising techniques generally define a pixel feature as the mean of all the corresponding samples. For instance, the pixel color is the sum of all radiance samples, weighted by the pixel filter function. RHF builds on the observation that considering only the sample mean and variance implicitly discards a wealth of information, and instead considers the actual sample distributions. Since storing individual samples is prohibitively expensive, the algorithm bins them in per-pixel histograms, and the distance between two pixels is computed as the chi-square distance of their histograms. This amounts to replacing the 3d feature space of the color mean, with the 60d space of the color histogram (using 20 bins for the histogram), with the hope that it is easier to discriminate neighbors in this high-dimensional space.

Discussion. In our experiments, we have found that the histogram-based distance can indeed allow for discriminating pixels, even when the color-mean difference is within the noise range. Also, an outlier sample can greatly affect the sample mean, but can only offset the sample count of a histogram bin by one, making the histogram-based distance more robust to fireflies. In practice, however, these improvements are not consistent; we illustrate this in Figure 1 by comparing the outputs of NLM, which considers the sample mean and variance, to RHF. The reliability of the histogram-based metric largely depends on having a good binning strategy, which is not known a priori, and the metric degenerates if all samples are concentrated in one or two bins.

4.2. Joint Filtering

With the notable exception of RHF, all recent denoising methods leverage auxiliary buffers to define the regression weights. Auxiliary buffers allow for a more robust discrimination since they are generally significantly less noisy than the color buffer. As illustrated in Figure 2 using a Joint NL-Means filter (JNLM), zero-order methods [SD12, WC12, RMZ13, KBS15] greatly benefit from this improved discrimination.

Discussion. The main drawback of using auxiliary buffers to define the regression weights lies in the ill-defined relation between the feature distance and the filtered output MSE, which makes it difficult to set the feature bandwidths. Consider the case in Figure 2: the tight kernel bandwidth robustly preserves the geometric edges, but leaves distracting banding artifacts on the lamp shade. In the next section, we present common approaches for estimating appropriate bandwidths on a per-pixel basis.

4.3. Bandwidth Estimation

The kernel bandwidth has to balance two conflicting goals: reducing the variance and preventing bias (over-blurring). A larger bandwidth increases the effective size of the neighborhood $N_p$, which reduces the variance, but tends to increase the bias. The kernel bandwidth must therefore strike a good balance between residual variance and bias, which amounts to minimizing the MSE.

As outlined in the introduction, requiring the user to manually tune the bandwidth is not acceptable in production environments. Furthermore, a global bandwidth is likely to be sub-optimal locally; see the lamp in Figure 2 for an illustration of resulting artifacts. Thus, most methods automatically adjust the bandwidth locally, using one of the following two approaches: 1) a selection-based estimate, where the filter is run using different predefined bandwidths, and the one resulting in the lowest estimated MSE is used on a per-pixel basis [RKZ11, RKZ12, RMZ13, BEEM15]; 2) a direct estimate, using either statistical analysis [SD12], or machine learning [KBS15]. Moon et al. [MCY14] combine both approaches: they factor the feature bandwidth as the product of a shared bandwidth, $h$, and a per-feature bandwidth, $h_s$, and use a selection-based estimate for the shared bandwidth, and a direct estimate for the per-feature bandwidth.
order regression often produced better results, both subjectively and 
perform a first-order regression. However, filters performing a zero-
curved surface of a teapot using RDFC (zero-order) and WLR 
neighborhood data. We illustrate this in Figure 4 by filtering the 
since it directly leverages the correlation between the auxiliary 
iary buffers that can be linearly combined, allowing for a far richer 
First-order regression uses a linear approximation instead of a con- 
Features are present.
Discussion. While setting the bandwidth automatically on per-
pixel level is crucial for good performance, as we illustrate in Figure 
video and error-prone, and all existing methods have some limitations. For instance, LBF uses a neural network trained in an offline process to harness the potential of their regression ker-
while the third candidate (blue) better preserves the geometric edges cap-
tured in the auxiliary buffers. The combined reconstruction offers a 
reasonable compromise on a per-pixel basis.
4.4. Zero- vs. First-order Regression
First-order regression uses a linear approximation instead of a con-
stant to represent the pixel neighborhood, solving for both the cen-
ter pixel value and its first partial derivatives. In practice, this addi-
tional degree of freedom is significant, since we have many auxili-
ary buffers that can be linearly combined, allowing for a far richer 
model of the pixel neighborhood than a single constant.
Discussion. For zero-order regressions, the kernel bandwidth 
must sometimes be drastically reduced to enforce the uniformity as-
sumption of the model. A first-order regression is less constrained,
since it directly leverages the correlation between the auxiliary 
buffers and the color buffer, which allows for a better use of the 
neighborhood data. We illustrate this in Figure 4 by filtering the 
curved surface of a teapot using RDFC (zero-order) and WLR 
(first-order). The improved model of first-order regressions can of-
fer a reasonable prediction of the whole pixel neighborhood, a fact 
that was exploited by Moon et al. [MIGYM15] to perform a sparse 
reconstruction in order to achieve interactive filtering performance.
4.5. Observations and Proposed Method
From a theoretical point of view, we would have expected WLR 
to be a clear winner in these tests, since it is the only method to 
perform a first-order regression. However, filters performing a zero-
order regression often produced better results, both subjectively and 
umerically. We believe there are three key reasons for this result. 
First, RDFC and LBF prefilter the auxiliary buffers to ensure they 
are nearly noise-free, whereas WLR relies on a Truncated Singular 
Value Decomposition to handle noisy features. This truncation is 
less robust in practice, since it can only remove noisy features; in 
contrast, prefiltering can extract useful information even when the 
features are noisy.
Second, WLR uses only the auxiliary features to shape the re-
gression kernel and does not utilize the color buffer, and can pro-
vide suboptimal results when the auxiliary features have low cor-
relation with the input. In contrast, the NL-Means kernel used by 
RDFC provides reasonable results even when no useful auxiliary 
features are present.
Third, WLR needs to estimate bandwidths for each feature, and 
does so on a continuous scale, a challenging problem given the di-
imensionality of the reduced feature space. In contrast, RDFC only 
considers three bandwidth configurations known to give generally 
useful results, and solves the simpler problem of picking between 
these three configurations. While this approach has less potential, 
it restricted scope makes it more robust.
Solving the first problem is trivial, since WLR can use prefiltered 
features. We show in Figure 5 that this can greatly improve the filter 
behavior. For the second problem, we propose to use the weights of 
a standard NL-Means filter as the regression kernel. The NL-Means 
kernel provides two advantages: 1) it provides a reasonable base 
when the auxiliary buffers do not provide any useful information; 2) 
it has only one bandwidth parameter to set. While the NLM result 
in Figure 2 suggests that the NL-Means kernel is a poor choice, 
we show in Figure 6 that, when paired with a first-order model, 
it can produce results of very high quality. Using the NL-Means kernel also allows us to use a similar selection-based bandwidth 
estimation as RDFC.
Based on these observations, our proposed method uses:
- feature prefiltering,
- a first order regression,
- NL-Means to compute the regression weights,
- collaborative filtering,
- and a selection-based bandwidth estimation.

© 2016 The Author(s)
Our regression model assumes that the input auxiliary buffers are noise-free, which is generally not the case. In particular, the visibility buffer is often quite noisy; see Figure 7. As proposed by Rousselle et al. [RKZ12], we therefore prefilter all auxiliary buffers using an NL-Means filter. Because we use a first-order model, any residual filtering artifacts in the feature buffers propagate directly to the filtered output, and we have found it beneficial to employ a two-step filtering scheme to minimize residual noise.

Our prefiltering scheme leverages the two statistically independent half buffers in a cross-filtering scheme, which was proposed in an earlier work by Rousselle et al. [RKZ12]. We use the NL-Means weights computed from one half buffer to filter the other and vice-versa, which has the advantage of decorrelating the noise of the weights from the noise of the data. The squared difference of the two filtered half buffers gives an estimate of their residual variance, which we use in a second filtering pass to further remove residual noise. In Figure 7, we compare the outputs of single-pass filtering to our two-step cross-filtering, and show the improvement on the resulting regression output.

### 5.2. Collaborative First-order Regression

Our filter builds upon a first-order model, using the auxiliary buffer data to predict the denoised color output. To this end, we consider the following 10-dimensional feature space: pixel coordinates (2D), normal (3D), albedo (3D), visibility (1D), and depth (1D). We can further enrich this set of features as needed. For instance, LBF uses a secondary albedo buffer that encodes the albedo of the second ray intersection, and when running our filter on the LBF input data, we also include this secondary albedo in our feature set.

#### Feature Cross-Filtering

Our prefiltering of the auxiliary buffers is effective at addressing the mid to high frequencies of the noise, but not its low frequencies. Since the noise in the feature buffers is generally correlated with the noise in the color buffer, the regression will tend to reconstruct the low frequency noise of the color buffer in the filtered output. To address this issue, we again take advantage of the two half buffers. We filter the two half buffers separately, and use the feature vector of the first buffer when fitting the second buffer and vice-versa. This decorrelates the noise in the features from the noise in the color buffer during the regression.
In many cases however, a larger bandwidth is preferable. Consequently, whereas ours targets higher quality output.

Window size. As such, their collaborative scheme targets performance, whereas our filtered MSE and selection map to ground truth in Figure 8. Similarly to Rousselle et al. [RMZ13], we perform a second filtering pass on the color buffer to remove residual noise artifacts. We consider using SURE [Ste81], but this approach requires the filter derivative to be known, which is challenging to do for our collaborative filter. The direct variance and bias estimation used by WLR is also not an option in our case, since its variance estimate assumes the filter weights to be noise-free, which is not the case for our method. Instead we propose a more general scheme leveraging our two half buffers.

Consider the squared difference between the filtered output, $F$, and the noisy input, $C$, minus the variance of the input. Its expected value is

$$E[(C - F)^2 - \text{Var}_C] = \text{Bias}_F^2 + \text{Var}_F - 2 \text{Cov}_{C,F}.$$ 

If the input and the filtered result are uncorrelated, then we can trivially use this result to create an unbiased estimator of the true MSE. This is not the case for our filter, but we can again leverage the two uncorrelated half buffers to sidestep this problem: We compute the squared difference of the filtered first half buffer to the second input half buffer and vice-versa, and subtract the variance of the input. If the two filtered half buffers are uncorrelated, this gives an unbiased estimator of the MSEs of each half buffer with

$$\text{MSE}_{F1} = (F_1 - C_2)^2 - \text{Var}_{C_1},$$
$$\text{MSE}_{F2} = (F_2 - C_1)^2 - \text{Var}_{C_1},$$

where $\text{Var}_{C_1} = \text{Var}_{C_2} = 2 \text{Var}_C$, where $\text{Var}_C$ is the sample mean variance of all color samples. After the MSE estimation, we average the two filtered half buffers to obtain $F = (F_1 + F_2)/2$. The expected bias of $F$ remains unaffected, but its variance is halved; consequently, we estimate the MSE of the averaged half buffers,

$$\text{MSE}_F = \frac{\text{MSE}_{F1} + \text{MSE}_{F2}}{2} - \text{Var}_F.$$ 

This estimator is unbiased under the assumption that our filtered half buffers are statistically independent. Our cross-filtering scheme introduces some correlation, but in practice, this correlation is sufficiently weak to produce a useful MSE estimate. Similar to Rousselle et al. [RMZ13], we then filter the estimated MSE and use it to generate selection maps for each filter bandwidth. We compare our filtered MSE and selection map to ground truth in Figure 8.

Second Regression Pass

Similarly to Rousselle et al. [RMZ13], we perform a second filtering pass on the color buffer to remove residual noise artifacts. We need the variance of the first pass output to compute our regression weights, which we estimate as the variance across our filtered half buffers. We do not perform cross-filtering for the second pass however, but directly refilter the average of the filtered half buffers.
6. Implementation Details
We implemented our algorithm on the CPU using C++11 and the Eigen 3 library for efficient matrix operations.

Feature Normalization. In order to improve numerical conditioning, we locally normalize the auxiliary features to occupy the same range, similarly to Moon et al. [MCY14]. Specifically, we offset and scale each feature within our filtering window such that it occupies the $[-1,1]$ interval.

Space-Time Filtering. Residual low-frequency noise in the filtered output causes disturbing flickering artifacts in animations. Similarly to WLR, we mitigate these artifacts by exploiting temporal coherence. We expand our regression window to the spatio-temporal domain by including noisy pixels and features from the two previous and the two next frames. We include the frame index as an additional auxiliary vector, and compute the regression weights using a spatio-temporal NL-Means kernel [BCM07, RKZ12]. The rest of our algorithm remains the same. We include a video demonstrating the temporal stability of our spatio-temporal filter in our supplementary material.

7. Results
We use a test bench composed of 21 scenes (see Figure 9) to evaluate our method, and compare its output to four state-of-the-art methods: RDFC [RMZ13], RHF [DMB14], WLR [MCY14], and LBF [KBS15]. All results were obtained with a fixed set of parameters for all denoisers. Our set of scenes feature volumes, hair and fur, interiors lit by an environment map, as well as diffuse, glossy, and specular reflections. Full resolution images at 16, 64, 256, and 1024 spp can be found in the supplemental material. Additionally, we provide denoised outputs using the standard NL-Means filter, as well as a multiscale NL-Means variant using the hierarchical scheme used by Delbracio et al. [DMB14] in the RHF algorithm. Lastly, our supplemental material also provides a comparison with the RenderMan denoiser, which is used in production.

Some of these previous methods (RDFC, WLR) are integrated into the PBRT code base, which we did not use for rendering. Consequently, we modified the public implementation of these methods to inject our own data in their pipeline. For RHF, we output the required color histograms from our renderer and use the public implementation to process it. For LBF, we similarly modified the public implementation to inject our own data. We note however that our auxiliary buffers differ from those used by Kalantari et al. [KBS15]: we use depth instead of world positions, and we do not export a secondary albedo, but instead have a buffer that stores the albedo of the first non-specular surface along the path. Since LBF employs a trained neural network to set its filter parameters, it is not clear how it can generalize to our data set. For a fair comparison to LBF, we rendered the SAN MIGUEL scene using the LBF code base, and passed its input data (including the secondary albedo) to the other denoisers. The SAN MIGUEL scene uses low discrepancy sampling, while all the other scenes were rendered using independent samples. For low discrepancy samples, we use the variance across the pairs of half buffers as an estimate of the sample mean variance. To clearly verify denoising performance of all tested methods, we have used the same input color and auxiliary buffers for all denoisers. Since the WLR method does not perform any prefiltering of the auxiliary buffers, we ran it twice, once with the noisy buffers (WLR), and once with the prefiltered buffers produced by our filter (WLR-PF). Results with both filtered and unfiltered inputs are provided in our supplemental material.

We used four different metrics to evaluate the output of the various denoisers: MSE, relative MSE, SSIM and PSNR, and provide the relative MSE and SSIM heatmaps for all results in our supplemental material. In Figure 10, we show the overall statistics of these denoisers on our test scenes. The thin blue lines correspond to the metric on a given scene, while the dots indicate the average metric value over all scenes. The error bars show the standard deviation of each method to the best performing technique for each scene. Therefore, a large error bar suggests that a method output quality varies greatly from scene to scene. In the case of RHF, the large error bars are due partly to the fact that this method relies on hand-tuning of its parameter, whereas we used a constant setting of $\kappa = 0.7$, which we found to be a good compromise over all scenes. Our filter exhibits both the best average and the smallest error bars across all metrics, indicating that it is consistently on par with the best alternative approach, and often substantially better.

In Figure 11, we show results from all denoisers on a selection of our test scenes rendered at 256 samples per pixel (BATHROOM,
Horse Room, Living Room, Red Room) and 64 samples per pixel (San Miguel). On these scenes, RHF yields blurry results (although it performs better on hair data, shown in our supplemental material). LBF tends to be quite aggressive in its filtering, resulting in a loss of local details, but robustly preserves geometric and texture edges. RDFC and WLR give overall good results, but still suffer from noticeable artifacts. In contrast, our filter reliably preserves input detail without overblurring, while still effectively removing noise and introducing few to no artifacts.

8. Discussion

In its current state, our filter offers a good match for the characteristics of the ideal production denoiser outlined in the introduction. In particular, our filter offers a significant and stable improvement across a wide range of scenes, as demonstrated in our supplemental material. Similar to previous works, we leverage common auxiliary buffers that most renderers can readily provide, and operate in a post-process, allowing for a straightforward inclusion in the production pipeline. There are however a number of limitations with our filter, which we will now cover.

Computational overhead. Our current implementation is CPU-based and significantly slower than previous works; see timings in Table 2. While we would expect a GPU implementation to drastically reduce these timings, it remains that our filter is only suitable for offline rendering. We note however that rendering times of multiple core hours per frame are common in production, and that our filter can offer substantial improvements in this context. The main cost of our filter is the regression itself, which we perform three times (twice when filtering the half buffers in the first pass and once in the second pass), and the overall cost of our filter could be greatly reduced by a sparse collaborative reconstruction, instead of the dense one we currently use.

Memory overhead. In order to better parallelize the filtering operations, we compute the full set of regression weights for the entire image buffer before fitting the data using our first-order model. Given that we use a window of 19 × 19 pixels, this represents a significant amount of memory, which could prove problematic for 4k renderings; see Table 2. As with the computational overhead, using a sparse collaborative reconstruction would drastically reduce the memory consumption of our technique.

Adaptive sampling. While our filter does not assume a uniform sampling distribution, we currently do not provide a way of guiding sampling based on the estimated MSE of our filter output.

Residual low-frequency noise. Large-scale temporal flickering due to residual low-frequency noise is a major challenge for image-space denoising techniques. While our spatio-temporal filtering results are on par with the current state of the art, they still leave room for improvement. We do not use motion vector to align frames temporally in our current implementation, which should offer some improvement. We have however experimented with the multiscale reconstruction proposed by Delbracio et al. [DMB+14], but have not been able to get consistently good results with it. In our supplemental material, we provide results using a standard NL-Means filter both without and with a multiscale reconstruction, which show the type of improvement this reconstruction can provide.

9. Conclusion

We have presented a comparative analysis of recent denoising methods, in order to assess their respective strength and weaknesses in the context of high-end production requirements. Based on this analysis, we identified a set of useful techniques that we combined in a novel filter, that consistently produces results on par with the current state of the art, and often provides substantial improvements. In its current implementation, our filter has a significant computational cost that restricts its use to high-quality off-line
Figure 11: We show results from our technique (large image) compared to four baseline methods. For each technique, we show two zoom-in locations and the rMSE (in parenthesis) to the reference; please refer to the supplemental material for full-resolution images and comprehensive comparison across all test scenes.
rendering, and we have performed some promising early experiments using sparse reconstruction, which we intend to pursue further. Lastly, our work has not considered the important question of adaptive sampling, and we expect that a similar study of existing techniques could lead to interesting developments.

Acknowledgements

We thank the following blendswap.com artists: thecali (CAR and SPACESHIP), McChimp2313 (HOUSE), UPSD (LAMP), SlykDrako (BEDROOM), nacinus (BATHROOM), Delarotnic (DRAGON), Wig42 (HORSE ROOM, DINING ROOM, RED ROOM, STAIRCASE), Jay-Artist (LIVING ROOM), and NickWoelk (SHAVING KIT). We also thank Cem Yuksel (CURLY HAIR), Alvaro Luna Bautista and Joel Anderson (MUSEUM), Guillaume M. Leal Laguna (SAN MIGUEL), Chris Harvey (DINOSAUR), Chris Scoville (SHEEP), and Maurizio Nitti (ROBOT). The BUNNY CLOUD comes from the OpenVDB website, and the geometry in the F ROOM scene from Minecraft. We thank Susan Harden for shepherd the internal approval process. This project was supported in part by Innovate UK (project #101858).

References


[SD12] Sen P., Darah S.: On filtering the noise from the random parameters in Monte Carlo rendering. ACM Trans. Graph. 31, 3 (June 2012), 18:1–18:15. 2, 4


© 2016 The Author(s)

For personal use only. Definitive version at http://dl.acm.org/ and http://onlinelibrary.wiley.com/