Denoising vs. Deblurring: HDR Imaging Techniques Using Moving Cameras

Li Zhang   Alok Deshpande   Xin Chen
University of Wisconsin, Madison   ACM
http://www.cs.wisc.edu/~lizhang/projects/motionhdr/

Abstract

New cameras such as the Canon EOS 7D and Pointgrey Grasshopper have 14-bit sensors. We present a theoretical analysis and a practical approach that exploit these new cameras with high-resolution quantization for reliable HDR imaging from a moving camera. Specifically, we propose a unified probabilistic formulation that allows us to analytically compare two HDR imaging alternatives: (1) deblurring a single blurry but clean image and (2) denoising a sequence of sharp but noisy images. By analyzing the uncertainty in the estimation of the HDR image, we conclude that multi-image denoising offers a more reliable solution. Our theoretical analysis assumes translational motion and spatially-invariant blur. For practice, we propose an approach that combines optical flow and image denoising algorithms for HDR imaging, which enables capturing sharp HDR images using handheld cameras for complex scenes with large depth variation. Quantitative evaluation on both synthetic and real images is presented.

1. Introduction

High Dynamic Range (HDR) Imaging has been an active topic in vision and graphics in the last decade. Debevec and Malik [14] developed the widely-used approach that combines multiple photos with different exposure to create an HDR image. This approach is well suited to early digital cameras, which often have 8-bit Analog-to-Digial conversion (ADC). Today, many consumer SLRs or machine vision cameras have higher resolution ADC; for example, Canon EOS 7D and Point Grey Grasshopper have 14-bit ADC, and many others have at least 12-bit ADC. In this paper, we present an effective approach that exploits new cameras with high-resolution ADC to widen the operating range of HDR imaging.

The inconvenient requirement of [14] is that the camera must remain still during the image acquisition and the scene must be static. The requirements of a still camera and scene are due to the need for long-exposure shots to record dark image regions accurately. Any motion of the camera or of the scene will introduce blur in the image. This requirement will not be simply relieved by using a 14-bit sensor, because the lower bits of each pixel only encode the noise accurately.

To capture a good HDR image in a flexible setting, without assuming stationary scenes or cameras, we have to either accumulate more photons using a long exposure and later remove the motion blur, or accumulate less photons using a short exposure and later remove the noise. Since the second approach takes less time, within a fixed time budget, we can take more images for better noise reduction. In this paper, we present a probabilistic formulation that allows us to compare which of denoising and deblurring can produce better HDR images.

Specifically, we compare the following HDR imaging choices:

- Deblurring a single blurry but clean image captured with a long exposure time Δ and a low ISO setting;
- Denoising a series of sharp but noisy images, each captured with a high ISO, together captured within time Δ.

We note that a high-resolution ADC is *essential* for both the procedures to succeed, in particular for denoising, because the noise must be digitized accurately to be averaged out among the multiple frames. Our contributions include:

- We propose a novel probability formulation that unifies both single-image deblurring and multi-image denoising. These two problems are formulated differently in the literature; comparing their solutions analytically is difficult.
- Using variational inference with motion as hidden variables, we derive the approximate uncertainty in the estimation of HDR images analytically for both imaging procedures. Our conclusion is that denoising is a better approach for HDR imaging.
- To put our analytical insight to practical use, we present a novel approach that combines existing optical flow and image denoising techniques for HDR imaging. This approach enables capturing sharp HDR images using handheld cameras for complex scenes with large depth variation. Such scenes cause spatially-varying motion blur for handheld cameras, which cannot be handled by the latest HDR imaging method [22].

Large depth-of-field, high dynamic range, and small motion blur are three of the major goals of computational camera research. Our work shows that, if a camera has high-resolution ADC, high frame rate, and high ISO, it is possible to achieve all the three goals through computation without resorting to specialized optical designs. This feature makes our approach suitable to micro-cameras with simple optics, such as those found in cellphones or used in performing surgeries.

2. Related Work

Our work is related to the recent research combining multiple images of different exposure to produce a sharp and clean image. Yuan *et al*. [27] and Tico and Vehvilainen [24] combined a noisy and blurry image pair, and Agrawal *et al*. [3] combined multiple blurry image with different exposure; all this research is limited to spatially-invariant blur.

One approach to address this limitation is to use video denoising techniques on multiple noisy images. In particular, our work is inspired by Boracchi and Foi [6], who combined a state-of-the-art video denoising method, VBM3D [12], and homography-based alignment for multi-frame denoising. They compared debluring a noisy and blurry image pair and

denoising two noisy images, but found no clear winner: on one hand, denoising produces good results without many artifacts; on the other hand, deblurring better preserves details but may occasionally introduce ringing artifacts. This observation motivates our work, which, for the first time, theoretically predicts when multi-frame denoising performs better than deblurring.

Motion-compensated filtering for video denoising [5, 12, 11] has existed for three decades [18]. Optical flow, however, is difficult to compute for complex motion. Several recent works exploiting temporal information for image enhancement assume simplified transformation between frames, such as translation in [5] and homography in [10, 23]. Handling more complex motion requires user assistance as in [5]. Indeed, state-of-the-art video denoising methods [8, 12] rely on block matching and argue that accurate motion estimation is unnecessary. We show that such an argument may be premature and accurate flow estimation significantly boosts the performance of denoising.

Debevec and Malik [14] introduced the classic method of combining multiple photos to create an HDR image, assuming a fixed camera and a static scene. Subsequent works generalize it to varying viewpoints [21, 25] and dynamic scenes [19]; none of the works considered motion blur during long exposure. Lu *et al.* [22] recently combined deblur and HDR creation, but their method is limited to spatially-invariant kernel. Our approach is the first, to the best of our knowledge, that demonstrates how to automatically create *sharp* HDR images using a handheld camera for scenes with complex geometry, for which the spatially-invariant motion blur assumption is often violated. Our approach is built upon an existing flow algorithm [7] and we describe techniques to deal with flow error.

Bennett and McMillan's work [5] is probably the closest work to ours, because it also seeks to create HDR images from a noisy video. Their method is based upon a denoising approach that is similar to the non-local mean [9] combined with a global translation estimation. We now know that non-local mean does not perform as well as BM3D denoising methods [20], and we demonstrate our approach is more effective than BM3D video denoising [12] for creating HDR images.

Finally, Hasinoff [16] presented the first framework that models the tradeoff between denoising and removing defocus blur for a stationary camera. Our analysis can be viewed as a first step to study the tradeoff between denoising and removing motion blur for a moving camera.

3. Problem Formulation: Deblur or Denoise?

Given a fixed time interval Δ, consider the following two alternative imaging procedures for the same scene: (1) take a single photo B with exposure time Δ, and (2) take a sequence of N photos $\{I_k\}_{k=1}^N$, each with a shorter exposure time $\tau = (1 - \epsilon)\frac{\Delta}{N}$ but a higher ISO, where ϵ is the *camera overhead* for saving images from the sensor to storage. During the imaging process, if the camera may move, the first procedure often produces a blurry image, while the second captures

a sequence of sharper but noisier images.

We next present a common probabilistic model that describes both procedures. The main goal of this model is to provide analytical insight regarding which procedure leads to a more reliable estimation of the underlying *sharp and clean* image J. To simplify this theoretical analysis, we assume that the motion is global translation and the blur is spatially-invariant. In Section 4, we describe an approach that deals with more general spatially-varying image motion in practice.

3.1. Known Image Motion

We start with the simpler case in which camera motion is known. Let $\mathbf{u}_k$ be the translation motion of the kth noisy image I_k with respect to the clean image J; we model I_k as

$$I_k = \delta_{\mathbf{u}_k} J + n_{I_k}, \tag{1}$$

where $\delta_{\mathbf{u}_k}$ represents the linear transformation that shifts an image using the global motion vector $\mathbf{u}_k$. In principle, the noise n_{I_k} has spatially-varying, signal-dependent variance, which is important to model for optimal noise reduction [28]. As a simplification, we assume n_{I_k} is a Gaussian noise whose variance is spatially constant but depends on the mean image intensity of I_k, as in [16].

We model the blurry image B as

$$B = F_{\{\mathbf{u}_k\}} J + n_B \tag{2}$$

where $F_{\{\mathbf{u}_k\}}$ is the linear blur filter induced by the motion trajectory $\{\mathbf{u}_k\}_{k=1}^N$ during exposure and can be modeled as

$$F_{\{\mathbf{u}_k\}} = \frac{1}{N} \sum_{k=1}^N \delta_{\mathbf{u}_k} \tag{3}$$

and n_B is a Gaussian noise whose spatially-constant variance depends on the mean image intensity of B.

We want to decide which of the following two gives a better estimation of J: N number of Eq. (1), each with a large noise; or a single equation of Eq. (2) with a small noise. Since the motion $\{\mathbf{u}_k\}$ is assumed to be known in this subsection, the answer is easy to see. We can compute a sharp and less noisy image $\bar{I}_{\{\mathbf{u}_k\}}$ by averaging all noisy images along motion trajectory $\{\mathbf{u}_k\}$ as

$$\bar{I}_{\{\mathbf{u}_k\}} = \frac{1}{N} \sum_{k=1}^N \delta_{-\mathbf{u}_k} I_k \tag{4}$$

and $\bar{I}_{\{\mathbf{u}_k\}}$ deviates from J as

$$\bar{I}_{\{\mathbf{u}_k\}} = J + n_{\bar{I}} \tag{5}$$

where $n_{\bar{I}} = \frac{1}{N} \sum_{k=1}^N \delta_{-\mathbf{u}_k} n_{I_k}$ is the averaged noise. If the noise n_{I_k} in each noisy image I_k has variance of $\sigma_{\mathtt{n}}^2$, the averaged noise $n_{\bar{I}}$ has a variance of $\frac{1}{N} \sigma_{\mathtt{n}}^2$.

Comparison between $\sigma_{\mathtt{n}}^2$ and $\sigma_{\mathtt{b}}^2$ When capturing noisy images with a short exposure time, we use a high ISO to stretch the intensity range so that the noise can be accurately digitized and later effectively averaged out in Eq. (4). Without loss of generality, we assume *unity gain* ISO: one photoelectron

corresponds to one integer increment of the pixel value—any higher ISO is unnecessary.[1] In this setting,

$$\frac{1}{N}\sigma_{\mathtt{n}}^2 = \frac{1}{N}(J + \rho_0^2 + \rho_1^2) \tag{6}$$

where ρ_0^2 and ρ_1^2 are the read noise variance before and after the gain amplifier, respectively.

When capturing the blurry image, we use a longer exposure time Δ with a lower ISO, which corresponds to a gain factor $g = \frac{\Delta}{\tau}$. In this setting,

$$\sigma_{\mathtt{b}}^2 = \frac{1}{g}J + \frac{1}{g^2}\rho_0^2 + \rho_1^2 \tag{7}$$

where the shot noise is reduced (relative to the same intensity range) due to the increased incoming light, the pre-amplifier read noise is reduced due to the gain factor, and the post-amplifier read noise is unchanged.

In modern cameras, read noise can be made extremely low; for example, Canon 5D has $[\rho_0, \rho_1] \approx [3.4, 1.9]$ and Canon 1D Mark III has $[\rho_0, \rho_1] \approx [3.9, 1.2]$, both normalized for a 12-bit ADC.[2] Therefore, the read noise is about 4 electrons at the unity gain ISO and can be neglected even for an underexposed shot with mean intensity 160 (out of $2^{12} - 1 = 4095$).

Conclusion Building on the analysis above, if we neglect read noise ρ_0^2 and ρ_0^2 and ignore the camera overhead ϵ, then $\sigma_{\mathtt{b}}^2 = \frac{1}{N}\sigma_{\mathtt{n}}^2$ because $g = N$; therefore $n_{\bar{I}}$ and n_B have the same noise variance $\sigma_{\mathtt{b}}^2$. Now, comparing Eq. (2) and Eq. (5), we see that their noise variance is the same, but J can be more reliably estimated from Eq. (5) because it does not involve the additional blurring operation $F_{\{\mathbf{u}_k\}}$ as in Eq. (2).

In practice, to instrument this comparison, we have to assume N is not too large, because we need to ensure (1) the noisy images are reasonably bright compared to the read noise and (2) the pixel has a large enough *Full Well Capacity*[3] to hold at least NJ photons for the blurry image and (3) there is a low ISO with $g = \frac{\Delta}{\tau}$. However, in theory, we can *always* argue that, capturing a sequence of not-too-dark noisy images is better than a blurry image even if the blurry image is captured by a hypothetical ideal camera with infinitely low ISO and infinitely large full well capacity, and therefore is better than a blurry image captured by a real non-ideal camera.

3.2. Unknown Image Motion

Now we consider the case where the image motion is unknown. We estimate the clean image J from the noisy image sequence $\{I_k\}_{k=1}^N$ by maximizing the posterior probability $P(J|\{I_k\}_{k=1}^N, \sigma_{\mathtt{n}}^2)$. Similarly, we estimate the clean image J from the blurry image B by maximizing $P(J|B, \sigma_{\mathtt{b}}^2)$.

To compare which estimation is more reliable, we evaluate the Hessian matrices of $\log P(J|\{I_k\}_{k=1}^N, \sigma_{\mathtt{n}}^2)$ and $\log P(J|B, \sigma_{\mathtt{b}}^2)$ with respect to J. From an optimization point of view, the Hessian matrix with a better condition number

will give rise to a more reliable estimation of J. From a statistical perspective, the Hessian matrix serves as the precision (inverse covariance) matrix of the Laplacian (local Gaussian) approximation of the distribution of J and therefore reveals the uncertainty associated with the estimation of J [4].

3.2.1 Approximate Hessian of $\log P(J|B, \sigma_{\mathtt{b}}^2)$

The blurry image B is related to J through blurring operation $F_{\{\mathbf{u}_k\}}$ induced by motion trajectory $\{\mathbf{u}_k\}$. To evaluate $P(J|B, \sigma_{\mathtt{b}}^2)$, we marginalize over all possible motion trajectories $\{\mathbf{u}_k\}$. Using Bayesian and total probability rules, we have

$$P(J|B, \sigma_{\mathtt{b}}^2) \propto \sum_{\{\mathbf{u}_k\}} P(B|J, \{\mathbf{u}_k\}, \sigma_{\mathtt{b}}^2)P(\{\mathbf{u}_k\})P(J) \tag{8}$$

where the left and right hand sides differ by a constant factor $\frac{1}{P(B|\sigma_{\mathtt{b}}^2)}$. On the right hand side, $P(\{\mathbf{u}_k\})$ is the prior probability for the motion sequence $\{\mathbf{u}_k\}$, and $P(J)$ is the prior probability for the clean image J. Since we assume imaging noise is Gaussian, $P(B|J, \{\mathbf{u}_k\}, \sigma_{\mathtt{b}}^2) = \mathcal{N}(B; F_{\{\mathbf{u}_k\}}J, \sigma_{\mathtt{b}}^2)$.

Since Eq. (8) involves summation of Gaussians, computing its $\log$ analytically is difficult. However, we view it as a mixture of Gaussian distribution, in which each possible motion sequence $\{\mathbf{u}_k\}$ corresponds to one Gaussian whose center is a $|J|$-dimensional vector $F_{\mathbf{u}_k}J$; there are a total of $(2U+1)^{2N}$ number of such Gaussians where $[-U, U] \times [-U, U]$ is the range of the motion. In this view, $\{\mathbf{u}_k\}$ is the hidden variable and the motion prior $P(\{\mathbf{u}_k\})$ is the mixture proportion.

Viewing Eq. (8) as a $|J|$-dimensional mixture of Gaussian allows us to apply the variational inference technique to approximate its $\log$ with an exact lowerbound. Specifically, at a particular J, we compute a q-distribution over all possible motion paths $\{\mathbf{u}_k\}$ as

$$q(\{\mathbf{u}_k\}) \propto P(\{\mathbf{u}_k\}) \exp(-\frac{1}{2\sigma_{\mathtt{b}}^2}\|B - F_{\{\mathbf{u}_k\}}J\|^2) \tag{9}$$

$q(\{\mathbf{u}_k\})$ describes the likelihood of $\{\mathbf{u}_k\}$ at this particular J. With this $q(\{\mathbf{u}_k\})$, we compute the lower bound $L_{\mathtt{b}}(J; q)$ to approximate $\log P(J|B, \sigma_{\mathtt{n}}^2)$

$$L_{\mathtt{b}}(J; q) \triangleq \sum_{\{\mathbf{u}_k\}} -\frac{q(\{\mathbf{u}_k\})}{2\sigma_{\mathtt{b}}^2}\|F_{\{\mathbf{u}_k\}}J - B\|^2 + \log P(J) - \mathtt{C}_1$$

where $L_{\mathtt{b}}(J; q) \leq \log P(J|B, \sigma_{\mathtt{n}}^2)$, and "=" holds when q is computed using Eq. (9).[4]

Note that $L_{\mathtt{b}}(J; q)$ consists of a summation of quadratic terms. Using the definition of $F_{\{\mathbf{u}_k\}}$ in Eq. (3), we simplify $L_{\mathtt{b}}(J; q)$ to compute its Hessian and gradient as

$$\begin{aligned}
L_{\mathtt{b}}(J; q) &= -\frac{1}{2\sigma_{\mathtt{b}}^2}J^{\mathsf{T}}H_{\mathtt{b}}J + \frac{1}{\sigma_{\mathtt{b}}^2}g_{\mathtt{b}}^{\mathsf{T}}J + \log P(J) - \mathtt{C}_2 \\
H_{\mathtt{b}} &= \frac{1}{N}\mathbf{I} + \frac{1}{N^2}\sum_{k \neq l}\sum_{\mathbf{u}_k, \mathbf{u}_l} q(\mathbf{u}_k, \mathbf{u}_l)\delta_{\mathbf{u}_l - \mathbf{u}_k} \\
g_{\mathtt{b}} &= \frac{1}{N}\sum_k \sum_{\mathbf{u}_k} q(\mathbf{u}_k)\delta_{-\mathbf{u}_k}J
\end{aligned} \tag{10}$$

where $q(\mathbf{u}_k)$ and $q(\mathbf{u}_k, \mathbf{u}_l)$ are marginal distribution of $\mathbf{u}_k$ and $(\mathbf{u}_k, \mathbf{u}_l)$, respectively, for the joint distribution $q(\{\mathbf{u}_k\})$.[5]

---

[1] For example, Canon 5D and 1D Mark III have the unity gain at ISO 1600 and 1900, respectively, as per Fig. 6a of "Camera Sensor Performance" in [1].

[2] The values are calculated from the read noise v.s. ISO plots in [2] and the unity gain ISO data in [1].

[3] The amount of photons that an individual pixel can hold before saturating.

[4] $\mathtt{C}_1 = \log P(B|\sigma_{\mathtt{b}}^2) + \frac{|J|}{2}\log(2\pi\sigma_{\mathtt{b}}^2) + \mathtt{KL}(q_{\{\mathbf{u}_k\}} \| P_{\{\mathbf{u}_k\}})$

[5] $\mathtt{C}_2 = \mathtt{C}_1 + \frac{\|B\|^2}{2\sigma_{\mathtt{b}}^2}$

In Eq. (10), we see that the Hessian for the lower bound $L_{\mathtt{b}}(J;q)$ is $-\frac{1}{2\sigma_{\mathtt{b}}^2}H_{\mathtt{b}}$ plus the Hessian of $\log P(J)$. $H_{\mathtt{b}}$ is a convex combination of permutation matrices ($\mathbf{I}$ and all $\delta_{\mathbf{u}_l-\mathbf{u}_k}$); therefore it is a doubly-stochastic matrix, in which each element is within $[0,1]$ and the sum of each row and the sum of each column is 1.

3.2.2 Approximate Hessian of $\log P(J|\{I_k\}_{k=1}^N, \sigma_{\mathtt{n}}^2)$

Our derivation in this case is very similar to Section 3.2.1. We evaluate $P(J|\{I_k\}_{k=1}^N, \sigma_{\mathtt{n}}^2)$ by marginalizing over all the possible motion trajectories. Using Bayesian rule and total probability rule, we have

$$P(J|\{I_k\}, \sigma_{\mathtt{n}}^2) \propto \sum_{\{\mathbf{u}_k\}} P(\{I_k\}|J, \{\mathbf{u}_k\}, \sigma_{\mathtt{n}}^2)P(\{\mathbf{u}_k\})P(J) \quad (11)$$

where the left and right hand sides differ by a constant factor $\frac{1}{P(\{I_k\}|\sigma_{\mathtt{n}}^2)}$. Since we assume imaging noise is Gaussian and independent among different images, $P(\{I_k\}|J, \{\mathbf{u}_k\}, \sigma_{\mathtt{n}}^2) = \prod_k \mathcal{N}(I_k; \delta_{\mathbf{u}_k}J, \sigma_{\mathtt{n}}^2)$.

As in Section 3.2.1, we view Eq. (11) as a mixture of Gaussian distribution, in which each possible motion sequence $\{\mathbf{u}_k\}$ corresponds to one Gaussian whose center is formed by concatenating all $\{\delta_{\mathbf{u}_k}J\}_{k=1}^N$ as a $N|J|$-dimensional vector; there are also a total of $(2U+1)^{2N}$ number of such Gaussians as in Section 3.2.1. In this view, $\{\mathbf{u}_k\}$ is the hidden variable and the motion prior $\{\mathbf{u}_k\}$ serves as the mixture proportion.

With this view, we apply the variational inference technique to approximate its log with an exact lowerbound. Specifically, at a particular J, we compute a q-distribution over all possible motion paths $\{\mathbf{u}_k\}$ as

$$q(\{\mathbf{u}_k\}) \propto P(\{\mathbf{u}_k\})\exp(-\frac{1}{2\sigma_{\mathtt{n}}^2}\sum_k \|I_k - \delta_{\mathbf{u}_k}J\|^2) \quad (12)$$

With this $q(\{\mathbf{u}_k\})$, we compute the variational lower bound $L_{\mathtt{n}}(J;q)$ to approximate $\log P(J|\{I_k\}, \sigma_{\mathtt{n}}^2)$ as

$$L_{\mathtt{n}}(J;q) \triangleq \sum_{\{\mathbf{u}_k\}} -\frac{q(\{\mathbf{u}_k\})}{2\sigma_{\mathtt{n}}^2}\sum_k \|\delta_{\mathbf{u}_k}J - I_k\|^2 + \log P(J) - \mathtt{C}_3$$

where $L_{\mathtt{n}}(J;q) \leq \log P(J|\{I_k\}, \sigma_{\mathtt{n}}^2)$, and the "=" holds when q is computed using Eq. (12).[6] $L_{\mathtt{n}}(J;q)$ can be simplified as

$$L_{\mathtt{n}}(J;q) = -\frac{N}{2\sigma_{\mathtt{n}}^2}\|J - \bar{I}_q\|^2 + \log P(J) - \mathtt{C}_4 \quad (13)$$

where $\bar{I}_q = \sum_{\{\mathbf{u}_k\}} q(\{\mathbf{u}_k\})\bar{I}_{\{\mathbf{u}_k\}}$ is a weighted sum of motion-compensated average images $\bar{I}_{\{\mathbf{u}_k\}}$ defined in Eq. (4).[7]

From Eq. (13), we see that the Hessian for the variational lower bound $L_{\mathtt{n}}(J;q)$ is the negative of a scaled identity matrix $-\frac{N}{2\sigma_{\mathtt{n}}^2}\mathbf{I}$ plus the Hessian of $\log P(J)$.

3.2.3 Comparison of Hessians in Eq. (10) and Eq. (13)

Comparing Eq. (10) and Eq. (13), we make two observations about their Hessian matrices.

---

[6]$\mathtt{C}_3 = \log P(\{I_k\}|\sigma_{\mathtt{n}}^2) + \frac{N|J|}{2}\log(2\pi\sigma_{\mathtt{n}}^2) + \mathtt{KL}(q_{\{\mathbf{u}_k\}} \| P_{\{\mathbf{u}_k\}})$
[7]$\mathtt{C}_4 = \mathtt{C}_3 + \frac{1}{2\sigma_{\mathtt{n}}^2}\sum_k \|I_k\|^2 - \frac{N}{2\sigma_{\mathtt{n}}^2}\|\bar{I}_q\|^2$

First, from an information-theoretical perspective, if we set aside $\log P(J)$, $\frac{N}{2\sigma_{\mathtt{n}}^2}\mathbf{I}$ is the precision matrix for J in Eq. (13). When the camera overhead ϵ is negligible, $\sigma_{\mathtt{b}}^2 = \frac{1}{N}\sigma_{\mathtt{n}}^2$, then the determinant of this precision matrix, $(\frac{1}{2\sigma_{\mathtt{b}}^2/N})^{|J|}$, is an upper bound for the determinant of the matrix $\frac{1}{2\sigma_{\mathtt{b}}^2}H_{\mathtt{b}}$—the precision matrix for J in Eq. (10), because $H_{\mathtt{b}}$ is a doubly-stochastic matrix of which the maximum possible determinant is 1 [26]. Since the determinant of the precision matrices is related to the entropy of the local Gaussian distributions we use to approximate the posteriors of J, without a strong prior $P(J)$, Eq. (13) provides an estimation for J with less uncertainty than Eq. (10). Their difference only becomes less when the prior $\log P(J)$ dominates the estimation.

Second, from a numerical optimization point of view, $\frac{N}{2\sigma_{\mathtt{n}}^2}\mathbf{I}$ in Eq. (13) has the best possible condition number 1. Even if $P(J)$ is non-informative (flat), the Hessian of $L_{\mathtt{n}}(J;q)$ is well-conditioned. $\frac{1}{2\sigma_{\mathtt{b}}^2}H_{\mathtt{b}}$ in Eq. (10) typically has a condition number less than 1; its condition number is 1 only if it is an identity matrix, which requires $q(\mathbf{u}_k, \mathbf{u}_l) = 0$ for all $\mathbf{u}_k \neq \mathbf{u}_l$ (meaning that there is no motion during exposure). Therefore, the Hessian of $\log P(J)$ must be needed as a pre-conditioner to reliably estimate J from Eq. (10).

Conclusion Based on these two observations, we make two conclusions. First, estimating clean image J by denoising multiple images is always more reliable than deblurring a single blurry image. This conclusion is independent of the type of motion $P(\{\mathbf{u}_k\})$ and image prior $P(J)$. Second, conventional HDR approach using shots with varying exposures is less reliable than multi-frame denoising for a moving camera, because shots with long exposure need to be deblurred.

One special case is when there is no motion. In this case, read noise determines which imaging choice is better. From Eq. (6) and Eq. (7), the read noise in denoising and deblurring is $\frac{1}{N}(\rho_0^2 + \rho_1^2)$ and $\frac{1}{N^2}\rho_0^2 + \rho_1^2$, respectively. It is easy to show that the former is less than the latter if $\frac{\rho_1^2}{\rho_0^2} > \frac{1}{N}$. For example, this condition requires $N > 4$ and $N > 11$ for Canon 5D and 1D Mark III, respectively, using the data in Section 3.1.

Our conclusion in this special case seems contradictory to the observation in [16], which suggests too many noisy shots are bad because each will incur read noise. We point out that Hasnoff *et al.* [16] capture and compare multiple shots and a single shot at the *same* ISO setting while we capture noisy images at a higher ISO. As a result, our denoising is an averaging operation and theirs is a summation operation. If the post-amplifier read noise is not trivially small compared to the pre-amplifier one, our averaging operation will result in a cleaner image than the blurry image, when N is large.

Our analytical conclusion is based on two simplifying assumptions: (1) approximating a log posterior by its variational lower bound and (2) neglecting the camera overhead. While the lower bound nicely aggregates the uncertainties in the motion estimation, the exact differences between the Hes-

sian of the original posteriors and their lower bounds are hard to quantify analytically. It is therefore desirable to empirically verify our conclusion by running simulations that estimate clean image J by optimizing $\log P(J|\{I_k\}_{k=1}^N, \sigma_\mathtt{n}^2)$ and $\log P(J|B, \sigma_\mathtt{b}^2)$, respectively, and comparing the result. Furthermore, through the simulations, we can also examine the effect of camera overhead ϵ on the estimation of J.

3.3. Simulation Algorithms

In this subsection, we present algorithms that estimate J from the two posteriors in Eq. (8) and Eq. (11). In principle, alternating between evaluating Eq. (9) and maximizing Eq. (10) is a deblurring algorithm; similarly, alternating Eq. (12) and Eq. (13) is a denoising algorithm. However, both algorithms are impractical, as their $q(\{\mathbf{u}_k\})$ are defined over an exponential number of states, $(2U + 1)^{2N}$. We next make approximations to make the estimation algorithms efficient.

3.3.1 Estimating J by Denoising

We assume that the motion prior $P(\{\mathbf{u}_k\})$ is independent among each $\mathbf{u}_k$; that is $P(\{\mathbf{u}_k\}) = \prod_k P(\mathbf{u}_k)$. Under this assumption, $q(\{\mathbf{u}_k\})$ in Eq. (12) must have a factorized form as $q(\{\mathbf{u}_k\}) = \prod_k q_k(\mathbf{u}_k)$ and

$$q_k(\mathbf{u}_k) \propto P(\mathbf{u}_k) \exp(-\frac{1}{2\sigma_\mathtt{n}^2}\|I_k - \delta_{\mathbf{u}_k}J\|^2) \qquad (14)$$

With this factorized form, the computation of $\bar{I}_q$ in Eq. (13) is simplified as

$$\bar{I}_q = \frac{1}{N}\sum_k \sum_{\mathbf{u}_k} q_k(\mathbf{u}_k)\delta_{-\mathbf{u}_k}I_k \qquad (15)$$

Iterating between Eq. (14) and Eq. (13) is an exact EM algorithm for estimating J from the noisy sequence $\{I_k\}$.

3.3.2 Estimating J by Deblurring

Similarly, we assume independence among $\{\mathbf{u}_k\}$ for $P(\{\mathbf{u}_k\})$ and restrict $q(\{\mathbf{u}_k\})$ to have a factorized form $q(\{\mathbf{u}_k\}) = \prod_k q_k(\mathbf{u}_k)$.[8] Under these two assumptions,

$$q_k(\mathbf{u}_k) \propto P(\mathbf{u}_k) \exp(\frac{1}{N\sigma_\mathtt{b}^2}(\delta_{\mathbf{u}_k}J)^{\mathsf{T}}(B - \mathbf{Q}_kJ)) \qquad (16)$$

where $\mathbf{Q}_k = \frac{1}{N}\sum_{l \neq k}\sum_{\mathbf{u}_l} q_l(\mathbf{u}_l)\delta_{\mathbf{u}_l}$. In practice, for each k, the probability mass of q_k is often concentrated on a particular $\mathbf{u}_k$. In this case, the variational lower bound $L_\mathtt{b}(J; q)$ in Eq. (10) can be approximated as

$$L_\mathtt{b}(J; q) \approx -\frac{1}{2\sigma_\mathtt{b}^2}\|B - \mathbf{Q}J\|^2 + \log P(J) - \mathtt{C}_1 \qquad (17)$$

where $\mathbf{Q} = \frac{1}{N}\sum_k\sum_{\mathbf{u}_k} q_k(\mathbf{u}_k)\delta_{\mathbf{u}_k}$

Iterating between evaluating Eq. (16) and maximizing Eq. (17) is a variational EM algorithm for estimating J from the blurry image B.

---
[8]In the case of deblurring, the factorized form of q does not simply follow the independence assumption of $P(\{\mathbf{u}_k\})$. We need to explicitly make an assumption that restricts the form of q.
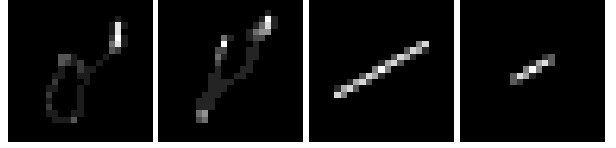

Figure 2. Four motion trajectories used in our synthetic experiments.

3.4. Simulation Results

We have compared the estimation of J using the denoising algorithm and the deblurring algorithm on synthetic data. Since both are EM algorithms, which are sensitive to initial values, we start with a ground truth image and motion, to factor out the local minimum issue when evaluating the quality of the results. We used four ground truth images; the last column of Figure 1 shows a patch from two of the four. We also generated four motion trajectories, each has a duration of $N = 100$ frames. The blur kernels corresponding to these trajectories are shown in Figure 2. **Please refer to our supplemental material for the complete set of experiment results.**

Deblurring vs Denoising For each ground truth image, for each motion trajectory, we generate $N = 100$ images moving along the trajectory; each image is corrupted by Gaussian noise with a standard deviation $\sigma_\mathtt{n} = 20$. We also use the the corresponding motion kernel to generate a single blurry image, corrupted by Gaussian noise with $\sigma_\mathtt{b} = \frac{\sigma_\mathtt{n}}{\sqrt{N}} = 2$.

When applying the deblurring algorithm, we need a prior model for $\log P(J)$. Since our goal is to compare which of the two types of image inputs gives better estimation for J, any image prior can be used, so long as we use the same prior for both denoising and deblurring. To simplify the maximization step, we assume a *weak* quadratic prior as follows

$$\log P(J) = -\frac{1}{2\sigma_0^2}\|J - J_0\|^2 \qquad (18)$$

where J_0 is the ground truth corrupted by a large Gaussian noise with $\sigma_0 = 40$. Strictly speaking, Eq. (18) is equivalent to having an additional noisy observation rather than being a prior; but we can treat it as a prior to regularize deconvolution.

Figure 1 shows the estimation results. The first column shows the blurry image. The second column shows one of the $N = 100$ noisy images. The third column shows the deblurring result, which contains certain amount of noisy. This phenomena is because $\mathbf{Q}$ in Eq. (17) is a low-pass filter and therefore a noisier J decreases its distance to J_0, increases $\log P(J)$, but does not change the value of the product $\mathbf{Q}J$ much; In the end, a noisier image is more preferred than the ground truth image. This phenomena verifies that the information in a single blurry image is low and the estimation is heavily influenced by the prior model. For a better prior, we replace J_0 with its denoised version $\tilde{J}_0$ using a state-of-the-art single-image denoising algorithm [13], and define $\log P(J)$ as

$$\log P(J) = -\frac{1}{2\tilde{\sigma}_0^2}\|J - \tilde{J}_0\|^2 \qquad (19)$$

where $\tilde{\sigma}_0^2$ is set to be smaller than σ_0^2 as $\tilde{J}_0$ is closer to the ground truth than J_0 is. The fourth column of Figure 1 shows the improved deblurring results using this prior.

| A Blurry Image | A Noisy image | 24.56 dB | 30.62 dB | 41.78 dB | Ground Truth |

| A Blurry Image | A Noisy Image | 26.26 dB | 31.69 dB | 41.79 dB | Ground Truth |

Figure 1. A comparison between single-image deblurring and multi-image denoising. From left to right: A blurry image B, $\sigma_{\mathrm{b}} = 2$; One of the 100 noisy images, $\sigma_{\mathrm{n}} = 20$; Deblurring using Eq. (18) as a weak prior; Deblurring using Eq. (19) as a stronger prior; Denoising using Eq. (18) as a weak prior. Even with a weak prior, multi-image denoising performs much better than single-image deblurring with a stronger prior. Please refer to our supplemental material for full resolution results. **Best viewed electronically.**

We can employ a more sophisticated prior to further enhance the deblurring results. Excellent results are shown in [27] using a pair of noisy and blurry images, where the prior is difficult to be described by a single analytical expression, rather implemented as a sequence of heuristic steps. This again supports our claim that a single blurry image contains very limited information: the quality of the result is heavily influenced by the image prior. As the the prior becomes increasingly complicated, generalizing it to handle spatially-varing blur is difficult.

On the other hard, as the fifth column shows, denoising with multiple noisy images produces very good results. The method is not sensitive to initial value. Using the noisy image J_0 as initial value also works well. Furthermore, it is easy to extend to spatially-varying motion, as Section 5 shows.

Effect of Camera Overhead We now evaluate the effect of camera overhead on denoising performance. The camera overhead ϵ reduces the exposure time for each noisy image by a factor of $1 - \epsilon$ and increases its noise standard deviation by a factor of $1/\sqrt{1 - \epsilon}$. We tested multi-image denoising with increased noise for a series of $\epsilon = 0, 0.1, 0.2, \cdots, 0.9$ and plot the PSNR of the results in Figure 3. As expected, the performance degrades as the camera overhead increases. However, even at $\epsilon = 0.9$, multi-image denoising still has noticeably better PSNR than single image deblurring using the same weak prior, and on par with it if a stronger prior is used.

4. Handling Spatially Varying Motion

It is straightforward to extend the multi-image denoising algorithm in Section 3.3.1 to handle spatially varying motion by computing optical flow. In doing so, we approximate the distribution of optical flow by its most likely sample. The idea of filtering noisy video along flow has been known for three decades [18]. Issues with this idea include occlusion and flow accumulation error. For the purpose of estimating flow induced by hand motion, which is typically not too violent, we
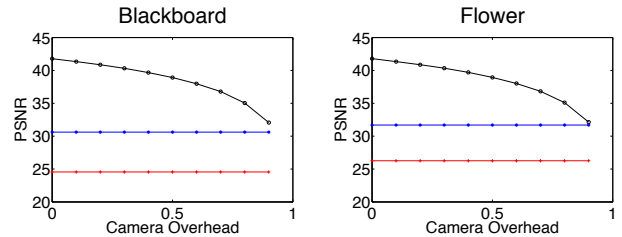

Figure 3. The performance of multi-image denoising as the camera overhead varies (top curve). The horizontal lines at the bottom and in the middle indicate the PSNR of single image deblurring using a weak prior, Eq. (18), and a strong prior, Eq. (19), respectively. Multi-image denoising performs better than single-image deblurring, even when the camera overhead is 90%; Beyond 90%, the performance drops dramatically.

found that state-of-the-art flow algorithms, *e.g.* [7], often work quite well, even if the input sequence is noisy. Specifically, we use [7] to compute flow between neighboring frames and then use the result as an initial solution to solve for flow from reference frame to the rest of the sequence. We combine three known techniques to handle the gross or sub-pixel flow errors.

Robust Temporal Averaging After registering all the frames to the reference frame, we only average pixels that are within $\pm 3\sigma$ of the reference pixel I. We pre-calibrate the sensor noise using an affine model $\sigma^2 = \tau^2 + \kappa J$ [17], where J is the ground truth intensity and we use its noisy observation to approximate it. We found this technique is effective to handle mis-registration due to occlusions or gross flow errors.

Temporal Denoising using PCA Optical flow may drift over a large number of frames. Such a drift blurs subtle image details in averaging. We notice that a pair of slightly mis-registered image patches I and J can be modeled as $I(x, y) \approx J(x, y) + J_x \Delta x + J_y \Delta y$, where $[J_x, J_y]$ is image gradient and $[\Delta x, \Delta y]$ is the drift. Therefore, a collection of slightly mis-registered patches approximately stay in a subspace spanned by J, J_x, J_y and PCA can be used to re-

move the noise in the patch collection [28]. Specifically, using the robust averaging result as the reference image, for every 4 pixel, we define a patch (8x8) centered around the pixel and collect patches along optical flow that are similar to the reference patch. We apply PCA to denoise this collection, and combine denoised patches for all reference patches as in [28].

Spatial Denoising using BM3D   After temporal denoising, the resulting image may still be a little grainy, because there may not be enough pixels or patches available for denoising due to outlier rejection. Such graininess is more visible in uniform regions but less so in textured areas. We use a state-of-the-art single-image denoising method [13] to remove the graininess while keeping the sharp details; this works well because the graininess is much smaller than the original noise.

5. HDR results

We have experimented the method in Section 4 on several scenes. We use Point Grey Grasshopper 14S3C color video camera (1384 × 1032@21FPS, 14-bit) in our experiments. **Please refer to our supplemental material for the complete set of experiments.**

For all experiments, we use small aperture (F8), short exposure time (0.56 millisecond) and the highest gain setting to acquire 100 noisy images with minimal defocus and motion blur. For static scenes, we first put the camera on a tripod and capture 1000 images, from which we compute the ground truth by averaging. After that, we release the camera from the tripod and hand shake it around the viewpoint from which we take the ground truth. Our testing images include one image from the ground truth sequence as the reference image and all the 99 shaky images taken afterwards. Doing so allows us to compute the PSNR of our results for quantitative evaluation.

HDR Imaging by Denoising   Our first scene consists of a set of books, ranging from 1 meter to 2 meters from the camera, with a few surrounding objects in a dark room (Figure 4, the first row). Such a scene introduces spatially-varying motion in the image plane, which can not be handled by the latest HDR imaging method [22]. The input images are sharp but noisy (first column); The noise is especially high in dark regions as shown in the tone mapped image (second column). The third column is our result, in which object details in dark regions are revealed. We use the `tonemap` function in Matlab with default parameters to compute the tonemapped images.

The second scene has many objects occluding each other with very detailed objects like hair (Figure 4, the second row). Our approach produces a HDR image that well preserves these details. The third scene (the third row) simulates a birthday party environment. Because of the fluttering flame on the cake, we did not record the ground truth. In our result, we see clearly the texture on the table surface and the birthday card.

Comparison with State-of-the-Art Video Denoising   We have also compared our approach with the state-of-the-art video denoising [12], VBM3D, for HDR estimation. Since VBM3D only applies to grayscale images; we convert our

Raw Noisy Images

Tone-Mapped Noisy Images

Our Results

Tone-Mapped Ground Truth

Figure 4. Computing HDR images from a sequence of 100 noisy images captured by a 14-bit handheld moving camera for three different scenes. The noise in the input images is higher in dark regions, as shown in the tone-mapped images. Our approach produces sharp and clean HDR images, and works for complex scenes with large depth variation. The cake scene has dynamic flames, and therefore does not have a ground truth. **Best viewed electronically.**

input images to gray scale and apply both our method and VBM3D on them. Figure 5 shows the comparison. Our approach works noticeably better because we use a global flow algorithm which registers images more accurately than the block matching technique used by VBM3D. Without accurate matching, temporal data may not be exploited as effectively as possible—the same observation was also made in [28].

Other Comparisons   We have compared denoising results using 8-bit vs. 14-bit quantization for HDR imaging [15], as well as temporal denoising using robust averaging vs. PCA. Please visit our project website for the results due to the lack of space.

| VBM3D (43.23 dB) | Our Result (45.01 dB) | Ground Truth |

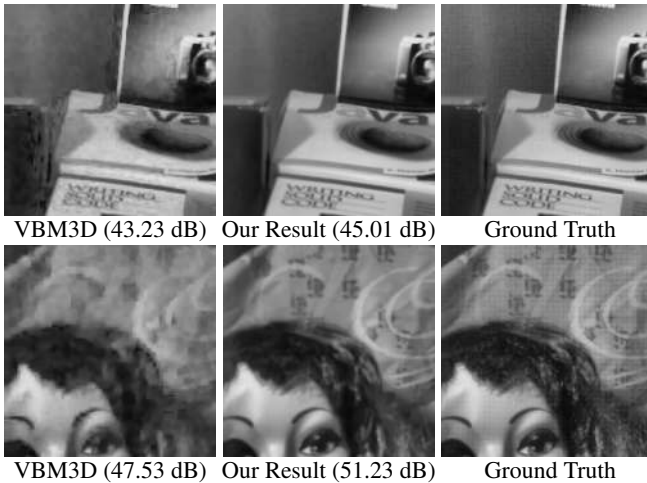| VBM3D (47.53 dB) | Our Result (51.23 dB) | Ground Truth |

Figure 5. A comparison between our approach and using VBM3D [12] for HDR imaging. Our approach more effectively removes noise in uniform regions (top row) while preserving details (bottom row), such as hair and backdrop texture. **Best viewed electronically.**

6. Discussion

In this paper, we argue that denoising is a more reliable way than deblurring to exploit new cameras with high resolution ADC for flexible HDR photography from a moving camera. Our approach enables capturing sharp HDR images for complex scenes of large depth variation using a handheld camera. There are several interesting future research directions.

The Optimal Number Needed We used 100 images in all of our experiments because many SLR cameras today can take dozens of images in burst mode. It is desirable to more carefully model the performance curve of HDR imaging and derive the optimal N as in [16].

High Speed Cameras for Consumer Photography Image resolution has increased tremendously for consumer cameras; however, the frame rate has not been changed as much. This paper demonstrates that high frame rate benefits flexible HDR capture. We are interested in exploring other aspects of photography that can benefit from fast cameras.

Acknowledgement

References

[1] *Clark Vision*. http://www.clarkvision.com/articles. 3

[2] *Noise, Dynamic Range and Bit Depth in Digital SLRs*. http://theory.uchicago.edu/ ejm/pix/20d/tests/noise/. 3

[3] A. Agrawal, Y. Xu, and R. Raskar. Invertible motion blur in video. *ACM Trans. Graph.*, 28(3), 2009. 1

[4] F. D. Anat Levin, William Freeman. Understanding camera trade-offs through a bayesian analysis of light field projections. In *ECCV*, 2008. 3

[5] E. P. Bennett and L. McMillan. Video enhancement using per-pixel virtual exposures. In *SIGGRAPH*, 2005. 2

[6] G. Boracchi and A. Foi. Multiframe raw-data denoising based on block-matching and 3-d filtering for low-light imaging and stabilization. In *Proc. Int. Workshop on Local and Non-Local Approx. in Image Processing*, 2008. 1

[7] A. Bruhn, J. Weickert, and C. Schnorr. Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *IJCV*, 61(3):211231. 2, 6, 8

[8] A. Buades, B. Coll, and J. Morel. Denoising image sequences does not require motion estimation. In *IEEE Conf. on Advanced Video and Signal Based Surveillance*, pages 70–74, 2005. 2

[9] A. Buades, B. Coll, and J. M. Morel. A review of image denoising algorithms, with a new one. *Simulation*, 4, 2005. 2

[10] T. Buades, Y. Lou, J. Morel, and Z. Tang. A note on multi-image denoising. In *Proc. Int. Workshop on Local and Non-Local Approx. in Image Processing*, 2009. 2

[11] J. Chen and C.-K. Tang. Spatio-temporal markov random field for video denoising. In *CVPR*, 2007. 2

[12] K. Dabov, A. Foi, and K. Egiazarian. Video denoising by sparse 3d transform-domain collaborative filtering. In *Proc. 15th European Signal Processing Conference*, 2007. 1, 2, 7, 8

[13] K. Dabov, R. Foi, V. Katkovnik, K. Egiazarian, and S. Member. Image denoising by sparse 3d transform-domain collaborative filtering. *TIP*, 16:2007, 2007. 5, 7

[14] P. E. Debevec and J. Malik. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH*, 1997. 1, 2

[15] M. Grossberg and S. Nayar. High Dynamic Range from Multiple Images: Which Exposures to Combine? In *ICCV Workshop on Color and Photometric Methods in Computer Vision*. 7

[16] S. W. Hasinoff, K. N. Kutulakos, F. Durand, and W. T. Freeman. Time-constrained photography. In *ICCV*, 2009. 2, 4, 8

[17] G. Healey and R. Kondepudy. Radiometric ccd camera calibration and noise estimation. *TPAMI*, 16(3), 1994. 6

[18] T. Huang. *Image Sequence Analysis*. 2, 6

[19] S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High dynamic range video. *ACM Trans. Graph.*, 22(3), 2003. 2

[20] V. Katkovnik, K. E. A. Foi, and J. Astola. From local kernel to nonlocal multiple-model image denoising. *IJCV*, 2009. 2

[21] S. Kim and M. Pollefeys. Radiometric self-alignment of image sequences. In *CVPR*, 2004. 2

[22] P.-Y. Lu, T.-H. Huang, M.-S. Wu, Y.-T. Cheng, and Y.-Y. Chuang. High dynamic range image reconstruction from handheld cameras. In *CVPR*, 2009. 1, 2, 7

[23] J. Telleen, A. Sullivan, J. Yee, P. Gunawardane, O. Wang, I. Collins, and J. Davis. Synthetic shutter speed imaging. In *EUROGRAPHICS*, 2007. 2

[24] M. Tico and M. Vehvilainen. Motion deblurring based on fusing differently exposed images. In *Proceedings of the SPIE on Digital Photography III*, volume 6502, 2007),. 1

[25] A. Troccoli, S. B. Kang, and S. Seitz. Multi-view multi-exposure stereo. In *3DPVT*, 2006. 2

[26] J. von Below and S. Renier. Even and odd diagonals in doubly stochastic matrices. *Discrete Mathematics*, 308:3917, 2008. 4

[27] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum. Image deblurring with blurred/noisy image pairs. *SIGGRAPH*, 26(3), 2007. 1, 6

[28] L. Zhang, S. Vaddadi, H. Jin, and S. Nayar. Multiple view image denoising. In *CVPR*, 2009. 2, 7