Active Short-Long Exposure Deblurring

Dan Yang, Samu Koskinen, and Joni-Kristian Kämäräinen
Huawei Finland & Tampere University
Email: https://research.tuni.fi/vision

***Abstract*—Mobile phones can capture image bursts to produce high quality still photographs. The simplest form of a burst is two frame short-long (S-L) exposure. S-L exposure is particularly suitable in low light conditions where short exposure frames are sharp but noisy and dark, and long exposure frames are affected by motion blur but have better scene chromaticity and luminance. In this work, we take a step further and define *active short-long exposure deblurring* where the viewfinder frames before the burst are used to optimize the S-L exposure parameters. We introduce deep architectures and data generation for active S-L exposure deblurring. The approach is experimentally validated with realistic data and it shows clear improvements. For the most difficult scenes (worst 5%) the PSNR is improved by +1.39dB.**

I. Introduction

An alternative approach to the conventional single frame photography is multi-frame (burst) photography where multiple frames are captured and fused to form a single high quality image. Burst photography has been used in a number of imaging problems such as denoising [1], [2], [3], high dynamic range imaging [4], [5], [6], [7], and deblurring [8], [9], [10]. In this work, we focus on burst deblurring in low light.

The simplest form of a burst is short-long exposure. In low light the short exposure frame is dim and contains more noise, but is less affected by motion blur than the long exposure frame, and therefore the two carry complementary information. A number of short-long exposure deblurring methods have been proposed [11], [12], [10]. These methods are *passive* in the sense that the burst capture parameters are fixed or unknown. In this work, we investigate *active* setting where optimal burst parameters are actively set before capture.

In this work, we introduce *active short-long exposure deblurring*. Active model estimates optimal exposure parameters before capturing the S and L frames. Estimation is based on the viewfinder frames just before the S-L burst. The novel contributions are **1)** a novel active approach for short-long exposure deblurring; **2)** practical formulation of the S-L exposure parametrization for mobile phone cameras; **3)** a deep architecture that contains separate modules for S-L exposure optimization f_p and deblurring f_f; **4)** generation of realistic viewfinder and S-L frames from high frame rate videos.

II. Background and related work

Single image deblurring is the problem of recovering the original sharp image x from the blurry observation y which has been distorted by convolution, $y = k * x + n$, with the blur kernel k and noise n. The problem is ill-posed since there are infinite combinations of x and k that generate y. The conventional deconvolution approaches utilize natural image priors to recover the sharp image [13], [14], [15]. The more recent methods use deep architectures since it is easy to generate training data for the problem [16], [17], [18], [19], [20], [21]. For example, Nah et al. [17] propose a multi-scale and Tao et al. [18] a recurrent architecture that progressively reconstruct the sharp image. Rotation and object motion generate spatially varying blur, that is handled in [16] by first estimating an optical flow map and then by recovering a sharp image that is consistent with the flow map. Jin et al. [22] reconstruct video frames that together form the observed blurry image.

Burst photography. Multiple frames, a burst, are used for various imaging tasks. For example, [23] introduces a framework to analyze an optimal time-slice strategy to capture multiple photos at different focus settings to reduce optical blur. Another suitable task is high dynamic range (HDR) photography. A number of works propose optimal selection of exposure parameters to produce HDR output [5], [4]. [24] merges under-exposed photos to reconstruct a HDR photo in low light. Liba et al. [25] propose a method to estimate the suitable exposure time and gain from gyroscope based flow and stability measurements. Their image processing pipeline is adopted from [24]. A recent approach is to turn the camera to an autonomous agent that learns to operate through Reinforcement Learning [7]. The problem is divided to two parts, exposure parameter selection (bracketing) and HDR fusion. RL is used to search for a policy that selects exposure parameters for HDR.

Burst deblurring. Similar to HDR, burst deblurring uses multiple frames to obtain better quality than single capture. [26] use a short-long pair to estimate the blur kernel and then the residual images are iteratively minimized. Delbracio and Sapiro propose a method using the weighted average of burst images in the Fourier domain [27]. The weights are computed from the Fourier spectrum magnitudes. The network in [9] takes a burst of images which are processed by a copy of the same reconstruction network with shared weights and maximum pooling among the copies produce the final output. The LSD^2 network in [10] achieves SotA results by jointly denoising and deblurring the short-long exposure inputs.

III. Short-Long Exposure Photography

Photography of moving objects or with moving camera in dim light is balancing between motion blur and sensor noise. The balance is set by the exposure parameters that are discussed next.

A. Camera exposure

Camera exposure is defined by [28]:

$$\frac{{N_f}^2}{t_e} = \frac{L_L S_I}{K} \ , \tag{1}$$

where N_f is the f-number, t_e is the exposure time (s), S_I is the ISO speed, K is the light meter calibration constant and L_L is the average scene luminance (cd/m^2). The calibration constant K is a factory setting and thus there are only four free parameters in (1). Moreover, mobile phone camera photography is even more limited than DSLR photography due to practical limitations of its size. Mobile camera sensors have a fixed physical aperture with a typical value $N_f = 1.8$. Hence, the remaining three parameters in (1) are *luminance* L_L, *exposure time* t_e and *ISO speed* S_I.

Practical exposure model. For practical photography it is more convenient to use *lux* E instead of luminance L_L since lux express the amount of light falling onto a surface. The change of unit affects to the light meter calibration constant K that needs to be calibrated. We calibrated a Huawei P20 rear camera using a fixed gray target and a fixed illuminant. For our device, we selected a 18% flat grey target and a standard D65 laboratory illuminant. We iteratively measured the lux level for each ISO speed and increased the exposure time until the brightest color channel (green) reached the same 18% exposure as the test chart reflectance. Finally, the practical exposure time derived from (1) is

$$t_e = 1300/E/S_I \cdot 1000 \ , \tag{2}$$

where E is the illuminance in lux and t_e is the exposure time in milliseconds. Notably, in (2) the only free parameter is the ISO speed S_I for a given lux level E.

B. Camera Noise

The factors that set the practical limits to Eq. 2 are *sensor noise*, *clipping* and *quantization*. For our experiments we produce realistic sensor noise, clipping and quantization as part of the noise generation model.

Raw-data sensor model. The selected sensor model suits for raw RGB readings of camera sensors. The final sRGB image is processed by a number of ISP algorithms such as image denoising, demosaicing and color transform, but their effects are camera and scene specific and thus difficult to model. We study deblurring as a low level raw RGB problem which is a valid choice as multi-frame algorithms such as the short-long exposure deblurring should be done early in the camera ISPs.

The raw-data sensor model is [29]:

$$z(x) = y(x) + \sigma(y(x))\xi(x) \tag{3}$$

where x is the pixel location, z is the observed signal, and y is the true signal. ξ is 0-mean and 1-standard deviation random noise and σ is a function of y that gives the total measurement noise. This formulation is convenient as the expectation $E\{z(x)\} = y(x)$ and $std\{z(x)\} = \sigma\left(E\{z(x)\}\right) = \sigma(y(x))$. For the raw-data imaging sensor model the noise term is composed of two parts

$$\sigma(y(x))\xi(x) = \eta_p(y(x)) + \eta_g(x) \tag{4}$$

where η_p is a Poissonian signal-dependent noise component $\mathcal{P}(\frac{1}{a}y(x))$ and η_g is a Gaussian signal-independent component $\mathcal{N}(0, b)$. The terms a and b were chosen as the overall noise variance in (3) has the affine form

$$\sigma^2(y(x)) = ay(x) + b \Rightarrow \sigma(y(x)) = \sqrt{ay(x) + b} \ . \tag{5}$$

Noise calibration. The actual sensor reading in (3) and its variance in (5) depend on multiple aspects of the sensor hardware. The elementary aspects are [29] i) *quantum efficiency* ($\chi = a^{-1}$), ii) *pedestal parameter* (p_0 that constitutes an offset-from-zero of the output data), and iii) *analog gain*. Quantum efficiency and pedestal parameter have less affect, but analog gain plays a predominant role as large gains of short exposures lead to worse signal-to-noise (SNR) ratio. Therefore small gains are preferred. Lower gain values may need longer exposure that may produce motion blur which is at the other end of the problem.

In digital cameras, the analog gain (θ) is usually controlled by the ISO sensitivity setting; ≥ 800 being large (less motion blur) and ≤ 200 small values (less noise). The relation between the model parameters a and b and the gain θ is

$$a = \chi^{-1}\theta, \ \ b = \theta^2 var\left\{\hat{\eta}_g(x)\right\} + var\left\{\hat{\hat{\eta}}_g(x)\right\} \tag{6}$$

where $\hat{\eta}_g$ is the signal dependent part of the Gaussian noise before the gain and $\hat{\hat{\eta}}_g$ is the signal independent part of the Gaussian noise before the gain and assuming that the pedestal shift is zero. (6) shows that signal dependent component in (5) has linear relationship and signal independent quadratic relationship to the analog gain (ISO value).

We resolved the noise model parameters a and b for P20 in our laboratory by using a calibration pattern containing solid gray regions from black to white. The pattern was captured with the standard ISOs 50, 100, 200, 400, 800, 1600 and 3200.

Practical noise model. By exploiting the normal approximation of the Poisson distribution,

$$\mathcal{P}(\lambda) \approx \mathcal{N}(\lambda, \lambda), \tag{7}$$

we obtain the following normal approximation of the errors

$$\sigma(y(x))\xi(x) = \sqrt{ay(x) + b}\xi(x) \approx \mathcal{N}(0, ay(x) + b) \ . \tag{8}$$

The result in (8) can be used to generate realistic noise of any calibrated device with known a and b and by drawing random numbers from $\mathcal{N}(0, 1)$. In order to constraint the values to $[0, 1]$ the noisy images were quantized to 10-bit and clipped after adding the noise.

IV. Active Deblurring Architecture

The main modules of the proposed active short-long exposure deblurring are: 1) *S-L exposure parameter optimization* f_p and 2) *S-L fusion* f_f. The modules are implemented as deep neural networks whose parameters θ_p and θ_f are optimized using generated data (Sec. V).

A. S-L exposure optimization

The optimal parameter estimation network f_p takes a sequence of view finder frames $I_v^{(t)}$, $t = -1, \ldots, -N$, and outputs the short and long exposure parameters, $\boldsymbol{p} = \langle \boldsymbol{p}_S, \boldsymbol{p}_L \rangle$. The network can be defined as

$$\boldsymbol{p} = f_p \left(\left\{ I_v^{(t)} \right\}; \theta_p \right) \tag{9}$$

where θ_p are the network weights. In our work the time stamp $t = 0$ defines the shot moment (ground truth) when the photographer presses the shutter.

In practice, exposure is defined by (2) in Sec. III-A where the only unknowns are the ISO speeds I_I^S and I_I^L. The speeds must be selected from the standard settings $S_I^S, S_I^L \in \{50, 100, 200, 400, 800, 1600, 3200\}$ that results to 28 valid S-L combinations. In our experiments we use only the last two view finder frames, $I_v^{(-1)}$ and $I_v^{(-2)}$, which leads to the following definition of S-L exposure parameter estimation:

$$\langle S_I^S, S_I^L \rangle = f_p \left(\langle I_v^{(-1)}, I_v^{(-2)} \rangle; \theta_p \right) \ . \tag{10}$$

The viewfinder frames can be captured arbitrarily, but the fixed ISO speed of 800 was used in the experiments.

Network architecture. For f_p we adopt a variant of the AlexNet architecture [30]. Inputs are two 270×480 viewfinder frames. The network consists of five convolutional layers: (1) 11x11 Conv-Relu-Pool (96 outputs), (2) 5x5 Conv-Relu-Pool (256 outputs), (3) 3x3 Conv-Relu (384 outputs), (4) 3x3 Conv-Relu (384 outputs), and (5) 3x3 Conv-Relu (256 output). The convolutional layers are followed by 3 fully connected layers with dropouts. The output layer uses softmax and outputs 28 probability values for each valid S-L ISO pair. The loss is

$$\mathcal{L}_p = -\sum_{i=1}^{49} y_i \times log \left(f_p \left(\langle I_v^{(-1)}, I_v^{(-2)} \rangle_i \right) \right) \tag{11}$$

where y is a 28×1 one-hot encoded ground truth vector.

B. S-L fusion

f_f should fuse information from the short and long exposure frames to produce a high quality noise and blur free image I. A suitable architecture is the recent LSD^2 [10] that was particularly designed for short-long deblurring. The original network was trained with blurry images generated from a Flickr image dataset [31]. Their scenes were static, but we train the network with images generated from the Sony Slow Motion Video dataset [32] that are more realistic (Sec. V).

Network architecture. The backbone of LSD^2 is U-Net [33] that implements image-to-image transfer for deblurring. The original network was optimized for PSNR that does not always match with human perception and thus we wanted to experiment with loss functions that better reflect Human Visual System (HVS) preferences. In addition to the MSE loss of the original LSD^2,

$$\mathcal{L}_{\text{MSE}} = ||\text{U-Net}\left(\langle I^S, I^L \rangle\right) - I^{gt}||_2^2 \ , \tag{12}$$

we add a number of HVS-inspired loss terms

$$\mathcal{L} = \mathcal{L}_{\text{MSE}} + \alpha_1 \mathcal{L}_{\text{DSSIM}} + \alpha_2 \mathcal{L}_{\text{edge}} + \alpha_3 \mathcal{L}_{\text{GAN}} \tag{13}$$

with the adjustable weights α_i.

$\mathcal{L}_{\text{DSSIM}}$ is a differentiable version of the well-known full-reference image quality metric, Structural Similarity Index Measure (SSIM) [34]. Inspired by [35] we use the following form of the SSIM which produces 1.0 as the best score:

$$\mathcal{L}_{\text{DSSIM}} = \frac{1 - \text{SSIM}(\text{U-Net}\left(\langle I^S, I^L \rangle\right), I^{gt})}{2} \ . \tag{14}$$

$\mathcal{L}_{\text{edge}}$ loss term is added to improve recovery of high frequency details (edges) in images. The original LSD^2 recovers poorly image edges that are important for human quality experience. In order to provide better restoration we introduce a loss that measures recovery of edges detected by the Canny edge detector [36]:

$$\mathcal{L}_{\text{edge}} = ||\text{Canny}\left(\text{U-Net}\left(\langle I^S, I^L \rangle\right)\right) - \text{Canny}\left(I^{gt}\right)||_2^2 \ . \tag{15}$$

$\mathcal{L}_{\text{GAN}}$ is an adversarial loss term that learns to detect whether the U-Net produces images that are natural or not. This loss term was motivated by the finding that neural networks produce artifacts that are easily spotted by human observers [37]. We adopt their network that uses 8 convolutional layers and 2 fully-connected layers. The final activation is a logistic function that encodes real vs. fake classification. The GAN loss terms follows the definitions in [38], [37]:

$$\mathcal{L}_{\text{GAN}} = -\log \text{Discriminator}\left(\text{U-Net}\left(I^S, I^L\right)\right) \ . \tag{16}$$

In the first experiments we provide ablation study on the loss weights α_1, α_2 and α_3.

V. Short-Long Exposure Data

Suitable training data is needed to opimize the network weights θ_p and θ_s of the optimal parameter estimation network f_p (Sec. IV-A) and the image fusion network f_s (Sec. IV-B). The previous works use blur kernels and generate data from still photographs [39], [9], [16], [10], but such data is not realistic as it lacks moving objects and realistic camera shake. Therefore, we generate more realistic data from real captured fast frame rate videos.

Dataset. Realistic motion blur can be generated by averaging high frame rate videos that contain fast moving objects and camera shake. For that purpose, the recent Sony Slow-Motion Video dataset (SONY) [32] was selected. SONY contains high quality videos captured at 250 fps. The dataset contains 63 video clips that were randomly split to 53 training and 10 test clips. The original clips were further divided to smaller clips. The total number of unique training instances is 1,184 and test instances 238. The fusion network (Sec. IV-B) was trained using the generated short-long exposure frames of all 28 possible ISO value pairs meaning the total of 28 × 1,184 = 33,152 training samples.

Data generation. For each training and test sample the first frame of the corresponding original high frame rate video was

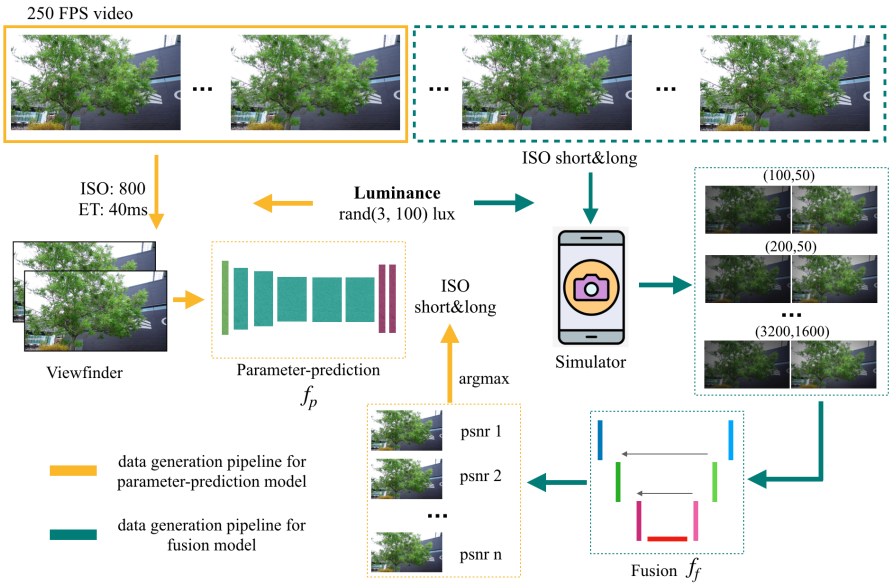

Fig. 1. Active short-long exposure data generation from the Sony Slow-Motion Video dataset (Section V)

used as the ground truth sharp image. This means that the shot moment $t = 0$ is when the burst capture starts. This setting is more challenging than using the middle frame and is more suitable for practical photography.

For each training and test clip a random scene illuminance value was picked from $[3, 100]$ lx where 3 lx corresponds to the full moon and 100 lx to a dark overcast day. Bright daylight is approximately 10,000 lx. The viewfinder frames I_v^{-1} and I_v^{-2} were generated by averaging the number of frames that corresponds to the fixed viewfinder exposure time 40 ms. Frame luminance was adjusted according to Huawei P20 exposure in (2) and noise was added using the ISO setting and illuminance value in (8) where a and b were fitted for the same mobile phone model. The viewfinder ISO was fixed to 800. The data generation pipeline is depicted in Figure 1 and a number of generated images in Figure 2.

VI. Experiments

A. Settings

Experiments were conducted with the Sony Slow Motion Video dataset (SONY) from which viewfinder and S-L bursts were generated as described in Section V. The f_p and f_s networks were trained until the training loss did not improve which occured at 40 epochs.

As the performance metrics the following standard image quality metrics were used: *Peak Signal-to-Noise Ratio* (PSNR) and *Structural Similarity Index Measure* (SSIM) [34]. The PSNR and SSIM metrics correlate on many types of distortions, such Gaussian blur and additive Gaussian noise [40], but SSIM matches better with the image quality as perceived by human observers. During the experiments we found that PSNR and SSIM sometimes conflict with human observations, but they provide indicative numbers to compare different methods.

B. Weighted HVS loss for S-L fusion

In the first experiment the proposed LSD^2-HVS with the human visual system inspired loss terms (Section IV-B) was compared to the original LSD^2 that uses only the standard MSE loss [10]. The different variants of LSD^2-HVS were trained and tested with the generated SONY training and test sets and the results reflect overall performance over all illuminance and ISO values.

The results in Table I show that the proposed loss terms improve the LSD^2 performance. The both performance metrics, perceptual SSIM and PSNR, improve using the HVS loss terms. PSNR is improved by 0.48dB. Note that the generative loss weight was fixed to $\alpha_3 = 10^{-3}$ following [37]. Examples are shown in Figure 3 where the original LSD^2 misses details (see the closeups) and distorts the output (for example the printed word "DAY").

TABLE I
Results for the variants of the proposed LSD^2-HVS using different weights for the loss terms (Section IV-B). All networks use the same training and test data (SONY)

Method					*Performance* SONY	
	MSE	α_1	α_2	α_3	SSIM	PSNR
LSD^2 [10]	1.0	-	-	-	0.924	32.94
LSD^2-HVS	1.0	0.0	0.0	10^{-3}	0.924	**33.42**
	1.0	0.5	0.0	10^{-3}	**0.925**	33.01
	1.0	1.0	0.0	10^{-3}	**0.925**	33.03
	1.0	0.5	0.5	10^{-3}	**0.925**	32.47
	1.0	0.5	1.0	10^{-3}	0.923	32.38

C. Fusion with fixed S-L ISO values

In this experiment the SONY test images were fused to deblurred output using fixed S-L ISO values to study the

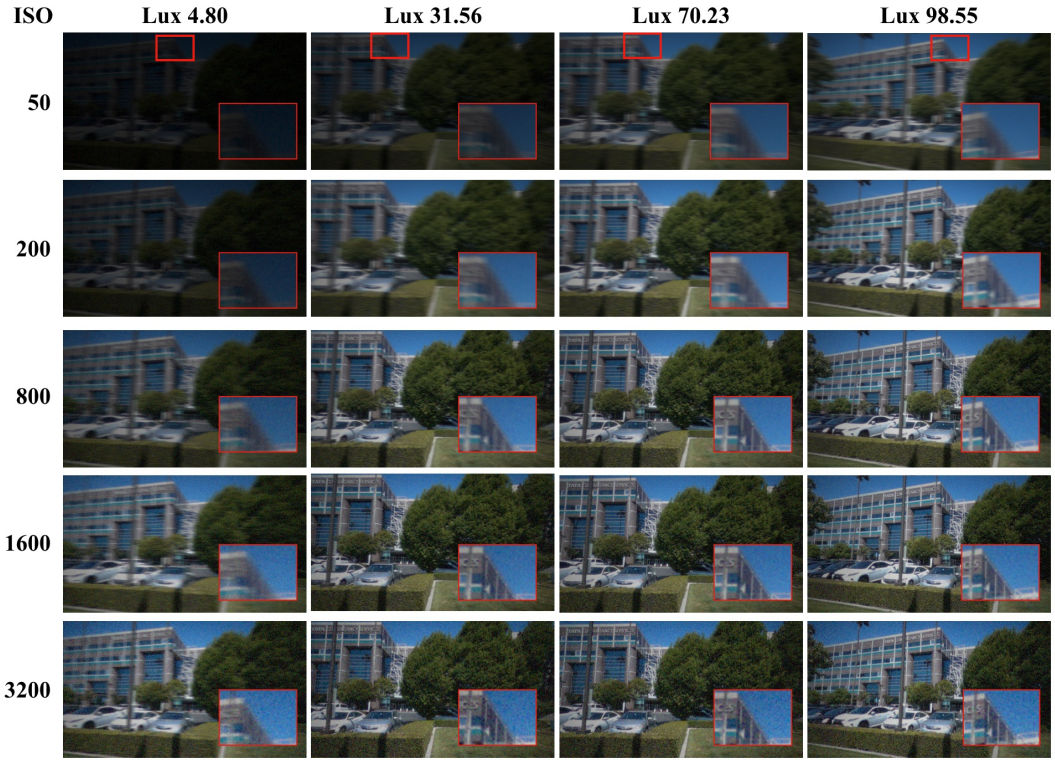

Fig. 2. Generated examples using different lux levels and ISO speeds

effect of exposure parameters. All 28 valid combinations were tested. The results are shown in Table II separately for different lux levels from the dimmest (0-30 lx) to the brightest (60-100 lx), and average over all lux values. The results provide the following interesting findings: **1)** the ISO setting clearly affects to the deblurring performance; **2)** for brighter scenes the tendency is as expected toward smaller ISO values that produce less sensor noise and allow faster shutter speed; **3)** overall the best ISO pairs are large numbers indicating that fast shutter speeds are preferred that. The last finding indicates that image denoising is an easier problem than motion deblurring.

D. Active short-long exposure deblurring

In the last experiment, we benchmarked the proposed active short-long exposure deblurring against various baselines and the SotA method in [10]. "Do nothing" means that the short or long exposure image is compared to the sharp ground truth. The other tested settings were i) average over all ISO pairs, ii) using the best fixed ISO pair, and iii) active deblurring that uses the ISO estimation network f_p. "Best ISO" is the ideal case where the ISO setting producing the best PSNR was used for each test images and thus represents the best achievable numbers. "Best fixed ISO" on the other hand represents the best passive deblurring method using the same fusion network.

The results are summarized in Table III and provide interesting findings: **1)** the original LSD^2 short-long exposure deblurring is clearly better than doing nothing and thus verifies good performance of the fusion network; **2)** the proposed LSD^2-HVS is clearly better than the prior art not using the proposed HVS loss terms; **3)** ISO optimization provides the best results - in particular, the worst-5% case performance is

TABLE II
LSD^2-HVS RESULTS FOR THE SONY TEST SET AND USING FIXED ISO SETTINGS. THE RESULTS ARE FIRST SHOWN SEPARATELY FOR DIFFERENT LUX LEVELS AND THEN FOR ALL. THE BEST NUMBERS ARE **bolded**.

ISO		0-30 lx		30-60 lx		60-100 lx		All	
L	S	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
50	50	0.801	26.09	0.856	28.03	0.851	28.20	0.838	27.56
50	100	0.825	26.75	0.865	28.53	0.861	29.24	0.852	28.34
50	200	0.837	27.33	0.880	29.93	0.907	31.70	0.880	29.97
50	400	0.850	28.25	0.917	32.00	0.942	33.63	0.909	31.66
50	800	0.870	29.36	0.942	33.67	0.957	34.91	0.929	32.99
50	1600	0.887	30.03	0.950	34.54	0.961	35.71	0.937	33.78
50	3200	0.894	30.24	0.942	34.41	0.951	35.17	0.932	33.57
100	100	0.830	27.19	0.874	29.56	0.869	29.84	0.860	29.02
100	200	0.844	27.54	0.888	30.97	0.915	32.23	0.887	30.56
100	400	0.858	28.68	0.926	32.98	0.948	34.35	0.916	32.37
100	800	0.882	30.26	0.950	34.70	0.963	35.90	0.936	33.97
100	1600	0.903	31.40	0.957	35.56	0.964	36.33	0.945	34.72
100	3200	0.912	31.99	0.949	35.40	0.955	35.59	0.941	34.53
200	200	0.852	28.16	0.893	31.21	0.918	32.72	0.893	31.01
200	400	0.865	29.22	0.931	33.55	0.953	35.05	0.922	32.99
200	800	0.887	30.84	0.953	35.33	0.966	36.60	0.940	34.62
200	1600	0.910	32.40	0.960	36.11	0.967	36.96	0.949	35.44
200	3200	0.921	32.81	0.953	35.97	0.958	36.21	0.946	35.18
400	400	0.870	29.78	0.935	33.83	0.955	35.37	0.925	33.36
400	800	0.892	31.27	0.956	35.65	0.967	36.86	0.943	34.94
400	1600	0.916	33.01	0.962	36.33	**0.968**	37.29	0.952	35.81
400	3200	0.928	33.31	0.955	36.20	0.960	36.38	0.950	35.47
800	800	0.896	31.68	0.957	35.72	**0.968**	37.11	0.945	35.19
800	1600	0.921	33.07	**0.962**	**36.76**	**0.968**	**37.49**	**0.952**	**35.94**
800	3200	0.932	33.77	0.956	36.19	0.961	36.68	0.952	35.72
1600	1600	0.923	33.23	0.960	36.55	0.965	37.17	0.952	35.89
1600	3200	**0.933**	33.82	0.953	36.10	0.959	36.65	0.950	35.70
3200	3200	0.930	**33.87**	0.945	35.67	0.951	35.82	0.943	35.23

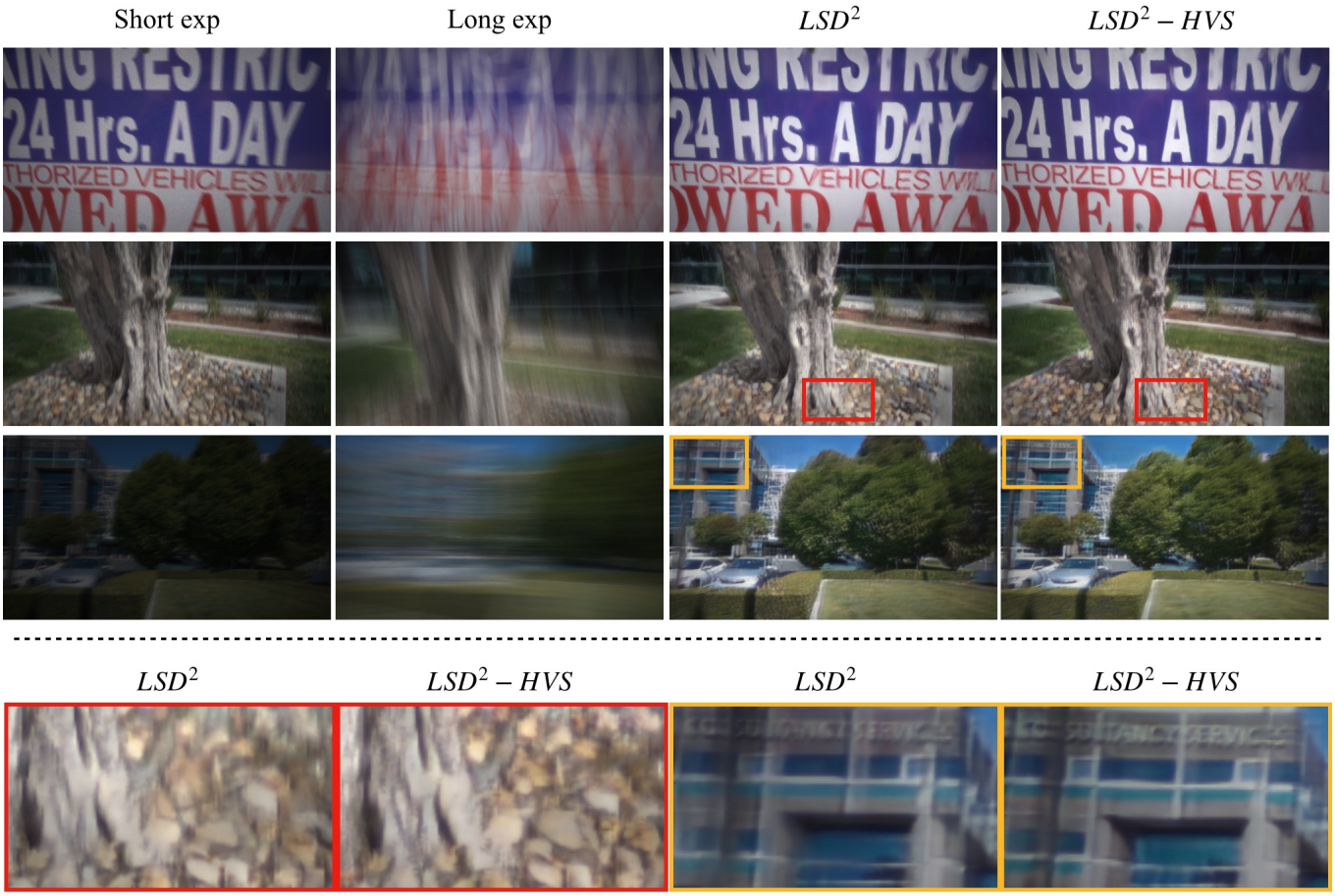

Fig. 3. High frequency details are better recovered by the proposed LSD^2-HVS than the original LSD^2 [10]. LSD^2-HVS uses human visual system inspired loss terms are used (here α_1=0.5, α_2=1.0 and $\alpha_3 = 0.001$).

TABLE III
ACTIVE VS. PASSIVE S-L DEBLURRING. "AVG ISO" IS AVERAGE OVER ALL ISO PAIRS; "BEST FIXED ISO" USES THE BEST FIXED ISO (1600-800); "W/ f_p" IS THE ACTIVE DEBLURRING THAT SELECTS ISO VALUES AUTOMATICALLY. "WORST-10%" AND "WORST-5%" ARE THE ERRORS AT 90% AND 95% QUANTILE, RESPECTIVELY.

Method	Avg.		Worst-10%		Worst-5%	
	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
Do nothing (short)	0.735	26.49	0.567	21.00	0.530	18.99
Do nothing (long)	0.683	24.51	0.370	16.98	0.297	14.81
LSD^2 [10] avg ISO	0.924	32.94	0.821	25.72	0.738	23.52
LSD^2-HVS avg ISO	0.924	33.42	0.820	25.98	0.735	23.67
LSD^2-HVS best fixed ISO	0.952	35.94	**0.929**	31.40	0.873	28.78
LSD^2-HVS w/ f_p	**0.955**	**36.00**	0.927	**32.00**	**0.901**	**30.17**
LSD^2-HVS best ISO†	0.958	37.12	0.929	33.43	0.912	31.46

† Uses oracle to select the best ISO for each test image (ideal perf.)

improved by +1.39dB (PSNR) as compared to the best fixed ISO (4.5% better SSIM).

VII. CONCLUSIONS

This work introduces a novel approach to burst imaging based deblurring: active short-long exposure deblurring. The method differs from the prior art in the sense that it estimates the best capture parameters before a short-long burst is captured. Exposure parameter estimation is based on viewfinder frames before the burst shot. These viewfinder frames observe scene illumination and motion and thus help to find suitable exposure parameters. We propose deep architectures for exposure parameter optimization and short-long fusion and data generation for them. With realistic data from the SONY dataset the proposed active short-long exposure deblurring achieved substantial improvements as compared to passive S-L exposure deblurring that uses fixed exposure settings. The improvements were particularly clear for the most difficult test scenes (worst 5%) for which the improvement was $\geq$1.39dB (PSNR).

REFERENCES

[1] G. Boracchi and A. Foi, "Multiframe raw-data denoising based on block-matching and 3-d filtering for low-light imaging and stabilization," in *Int. Workshop on Local and Non-Local Approximation in Image Processing*, 2009.

[2] B. Mildenhall, J. Barron, J. Chen, D. Sharlet, R. Ng, and R. Carroll, "Burst denoising with kernel prediction networks," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[3] C. Godard, K. Matzen, and M. Uyttendaele, "Deep burst denoising," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 538–554.

[4] O. Gallo, M. Tico, R. Manduchi, N. Gelfand, and K. Pulli, "Metering for exposure stacks," in *Eurographics*, 2012.

[5] S. W. Hasinoff, F. Durand, and W. T. Freeman, "Noise-optimal capture for high dynamic range photography," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2010.

[6] R. Yu, W. Liu, Y. Zhang, Z. Qu, D. Zhao, and B. Zhang, "Deepexposure: Learning to expose photos with asynchronously reinforced adversarial learning," in *NeurIPS*, 2018.

[7] Z. Wang, J. Zhang, M. Lin, J. Wang, P. Luo, and J. Ren, "Learning a reinforced agent for flexible exposure bracketing selection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 1820–1828.

[8] P. Wieschollek, M. Hirsch, B. Scholkopf, and H. Lensch, "Learning blind motion deblurring," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 231–240.

[9] M. Aittala and F. Durand, "Burst image deblurring using permutation invariant convolutional neural networks," in *European Conference on Computer Vision (ECCV)*, 2018, pp. 731–747.

[10] J. Mustaniemi, J. Kannala, J. Matas, S. Sarkka, and J. Heikkila, "LSD_2 - joint denoising and deblurring of short and long exposure images with convolutional neural networks," in *BMVC*, 2020.

[11] M. Tico, M. Trimeche, and K. Pulli, "Motion blur identification based on differently exposed images," in *ICIP*, 2006.

[12] M. Tico and K. Pulli, "Image enhancement method via blur and noisy image fusion," in *ICIP*, 2009.

[13] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image restoration by sparse 3d transform-domain collaborative filtering," in *Image Processing: Algorithms and Systems VI*, vol. 6812. International Society for Optics and Photonics, 2008, p. 681207.

[14] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2009, pp. 1964–1971.

[15] Y. Liu, W. Dong, D. Gong, L. Zhang, and Q. Shi, "Deblurring natural image using super-gaussian fields," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 452–468.

[16] D. Gong, J. Yang, L. Liu, Y. Zhang, I. Reid, C. Shen, A. van den Hengel, and Q. Shi, "From motion blur to motion flow: A deep learning solution for removing heterogeneous motion blur," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[17] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[18] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, "Scale-recurrent network for deep image deblurring," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8174–8182.

[19] H. Chen, J. Gu, O. Gallo, M.-Y. Liu, A. Veeraraghavan, and J. Kautz, "Reblur2deblur: Deblurring videos via self-supervised learning," in *International Conference on Computational Photography (ICCP)*. IEEE, 2018, pp. 1–9.

[20] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "Deblurgan: Blind motion deblurring using conditional adversarial networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 8183–8192.

[21] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 8878–8887.

[22] M. Jin, G. Meishvili, and P. Favaro, "Learning to extract a video sequence from a single motion-blurred image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 6334–6342.

[23] S. W. Hasinoff, K. N. Kutulakos, F. Durand, and W. T. Freeman, "Time-constrained photography," in *International Conference on Computer Vision (ICCV)*. IEEE, 2009, pp. 333–340.

[24] S. W. Hasinoff, D. Sharlet, R. Geiss, A. Adams, J. T. Barron, F. Kainz, J. Chen, and M. Levoy, "Burst photography for high dynamic range and low-light imaging on mobile cameras," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 6, p. 192, 2016.

[25] O. Liba, K. Murthy, Y.-T. Tsai, T. Brooks, T. Xue, N. Karnad, Q. He, J. T. Barron, D. Sharlet, R. Geiss *et al.*, "Handheld mobile photography in very low light," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 6, pp. 1–16, 2019.

[26] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum, "Image deblurring with blurred/noisy image pairs," *ACM Transactions On Graphics (TOG)*, vol. 26, no. 3, p. 1, 2007.

[27] M. Delbracio and G. Sapiro, "Removing camera shake via weighted fourier burst accumulation," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3293–3307, 2015.

[28] International Organization for Standardization, *ISO 2720:1974 Photography — General purpose photographic exposure meters*, 1974.

[29] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical poissonian-gaussian noise modeling and fitting for single-image raw-data," *IEEE Trans. on Image Processing*, 2008.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[31] M. Huiskes, B. Thomee, and M. Lew, "New trends and ideas in visual concept detection: the MIR Flickr retrieval evaluation initiative," in *ACM Int. Conf. on Multimedia Information Retrieval*, 2010.

[32] M. Jin, Z. Hu, and P. Favaro, "Learning to extract flawless slow motion from blurry videos," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8112–8121.

[33] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[34] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[35] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on computational imaging*, vol. 3, no. 1, pp. 47–57, 2016.

[36] J. Canny, "A computational approach to edge detection," *IEEE Transactions on pattern analysis and machine intelligence*, no. 6, pp. 679–698, 1986.

[37] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[38] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *NeurIPS*, 2014.

[39] G. Boracchi and A. Foi, "Uniform motion blur in poissonian noise: Blur/noise tradeoff," *IEEE TIP*, vol. 20, no. 2, 2011.

[40] A. Hore and D. Ziou, "Image quality metrics: Psnr vs. ssim," in *2010 20th international conference on pattern recognition*. IEEE, 2010, pp. 2366–2369.