Burst Imaging for Light-Constrained Structure-From-Motion

Ahalya Ravendran, Mitch Bryson, Donald G. Dansereau

Abstract—Images captured under extremely low light conditions are noise-limited, which can cause existing robotic vision algorithms to fail. In this paper we develop an image processing technique for aiding 3D reconstruction from images acquired in low light conditions. Our technique, based on burst photography, uses direct methods for image registration within bursts of short exposure time images to improve the robustness and accuracy of feature-based structure-from-motion (SfM). We demonstrate improved SfM performance in challenging light-constrained scenes, including quantitative evaluations that show improved feature performance and camera pose estimates. Additionally, we show that our method converges more frequently to correct reconstructions than the state-of-the-art. Our method is a significant step towards allowing robots to operate in low light conditions, with potential applications to robots operating in environments such as underground mines and night time operation.

I. INTRODUCTION

Current and emerging robots use vision sensors for a broad range of tasks including simultaneous localisation and mapping (SLAM), navigation, pose estimation, depth estimation and 3D reconstruction. State-of-the-art methods for structure-from-motion (SfM) and vision-based reconstruction (e.g. [1], [2]) perform well under good lighting conditions but fail to reconstruct in low light for tasks such as autonomous driving, drone surveillance and underground mining. When using light-constrained images, state-of-the-art methods [1] yield incorrect 3D shape estimation, inaccurate camera trajectories and in some scenes even fail to converge for light-constrained SfM. This results from low signal-to-noise ratio (SNR) in images, where true features are not detected and/or spurious features are detected and matched with features in other images for reconstruction.

Burst photography is an established mobile photography technique which uses a series of consecutive frames with small exposure time to produce a single image with an improved SNR upon merging [3]–[6]. Burst imaging has been shown to improve the SNR in images without the need for additional ambient light sources [4]. However, prior works and on-going developments on burst imaging are steered towards mobile photography, where the primary objective is to produce convincing content for human visual perception while accounting for camera motion due to handshake and scene motion within a single burst.

We propose to reconstruct light-constrained scenes by adapting burst imaging for reconstruction. We describe an imaging pipeline that captures multiple bursts of images and accounts for motion variation within each burst using hierarchical tile-based alignment. We temporally merge each aligned burst to a single image using a voting scheme and then use these merged images in feature-based SfM to reconstruct scenes (see Fig. 1). We also consider adding spatial filtering separately on temporally denoised images to evaluate hybrid denoising.

Our key contributions are:

• We establish the viability of using burst imaging to improve robotic vision in low light, and provide a set of recommendations for adopting this approach in reconstruction tasks such as SfM;
• We evaluate different approaches to burst imaging in robotics applications and show that burst capture with merge offers significant advantages in both computational requirements and performance; we offer a theo-
retical explanation and experimental evidence showing why this is so; and

- We demonstrate the proposed method improving low-light SfM by yielding more precise 3D points, more true features, fewer spurious features, more precise camera trajectories and an ability to operate in lower light than was previously possible.

Our approach assumes that the level of noise present in images is not so great as to make information unrecoverable (e.g. the underlying signal being suppressed by quantization noise). Additionally, we assume that there exists sufficient overlap between subsequent sets of images bursts for registration, which is commonly available for mobile platforms moving at moderate speeds relative to their environment.

To validate our method, we mounted a monocular machine vision camera on a UR5e robotic arm, and captured burst imagery of real scenes in an illumination-controlled environment over different exposure times. The images are captured over 20 scenes with objects of various size, shape and colour as 22 bursts for each scene, yielding 6160 images in 880 bursts. We are releasing a dataset containing raw Bayer images, captured over the exposure time of 1ms and 0.1ms for a repeated trajectory. Code and dataset are available at https://roboticimaging.org/Projects/BurstSfM/.

To evaluate, we compare our results against existing alternative approaches depicted in Fig. 2: burst without merge which includes every image in each burst, and conventional single-image capture. We demonstrate our proposed method outperforms the alternative approaches by reconstructing more true features with high feature localization accuracy and less spurious features as shown in Fig. 1. We also show our method yields accurate estimation of camera pose trajectory while minimising the failure cases of convergence compared to the existing alternative approaches.

This work opens the way for a broad range of applications in which low light commonly complicates vision such as autonomous driving and delivery drones.

II. RELATED WORK

When optimising camera exposure settings for low SNR images, there exist many trade-offs in the final image quality. Conventional cameras can increase SNR by either widening the aperture which reduces depth of field or increasing the exposure time, which for dynamic scenes, increases motion blur.

Considering alternative exposure schemes, depicted in Fig. 2, capturing images as a sequence over a long period of time yields a video. However, it is computationally intensive to process images at full video rates and in low light, each of these frames have low SNR.

A number of existing computational imaging approaches successfully break discussed imaging trade-offs and take advantages of the alternative exposure schemes as demonstrated in coded aperture [7], flexible depth of field photography [8], flutter shutter [9] and motion-invariant photography [10]. Unfortunately they require hardware modification to imaging sensors/cameras whereas our focus is on using existing monocular cameras to get quality images in low light for robotic vision.

Learning-based techniques show promising results for denoising [11]–[13], deblurring [14]–[16], dehazing [17]–[19] and image enhancement [20]–[22]. However, they are not shown to generalize outside the training domain [23] and need large datasets and long training times. There is a factor of uncertainty [24], [25] in terms of interpretability and transparency, which imposes additional constraints on the use of deep learning approaches in robotics applications. Many of these techniques are also trained on simulated images rather than captured images and thus, often yield unreliable results for previously unseen terrain and environmental conditions [26], [27].

Single-image based approaches are fundamentally limited by the amount of information in a conventional single image. They also have a tendency to produce visually pleasing results and not necessarily accurate information through learned image priors [28]. This affects their reliability for applications such as reconstruction and tracking.

Hasinoff et al. [29] exploit the time slice advantage and use burst imaging as a solution for night photography by capturing multiple underexposed images and merging them to get a single higher SNR image. This is implemented as night sight in Google Pixel mobile phones [4], astrophotography mode [5], multi-frame super resolution [6] and using learning approaches [30]. It has also been adapted as quanta burst photography [31] for single-photon sensors to produce
high quality images. This and other previous work in mobile photography has focused mainly on producing visually pleasing still images from bursts with little camera motion. In this work we adapt burst imaging for 3D reconstruction with extended camera motion in low-light scenes.

Our work also considers considerably more challenging imagery than what is typically employed in existing burst photography work. Noisy images that are captured as a burst in mobile photography applications still have acceptable SNR regimes for human visual systems and thus, the robots have the ability to perform reconstruction using existing state-of-the-art methods [1]. However, robots need the ability to visualize beyond this SNR regime in operating environments such as mining quarry. As reconstruction in robotics has a wide coverage with large motions within a single trajectory unlike prior work which focuses on a burst, an efficient way of using multiple bursts for 3D image-based reconstruction applications is also an open question.

In this work, we adapt burst imaging for 3D reconstruction in low light. We do this by combining feature-based SfM and burst photography, exploiting the advantages of each. By capturing rapid successions of frames, we enable the use of direct methods for image registration [32], exploiting the small camera motion between frames to yield a strong SNR advantage with moderate computational expense. After extracting features from the merged burst images, we apply feature-based SfM [1] to handle large camera motions between bursts. Feature-based methods handle large camera motions [33] while benefiting from the improved feature quality associated with the burst-merged images.

Because our method compresses each burst into a single denoised image, total computation is lower during reconstruction compared with an approach employing all measured frames. Furthermore, the higher SNR in our imagery can reduce the numbers of spurious features, again lowering computational requirements during reconstruction as there are fewer outliers to detect and reject.

III. BURST-BASED SFM

We capture multiple images as a burst and perform hierarchical tile-based alignment to compute the dense correspondence between every image in a single burst stack to a common image. We temporally merge the aligned stack with a voting scheme [4] to get one denoised image per burst, and we evaluate different means of obtaining it: temporal, temporal with Wiener filtering [4] and temporal with bilateral filtering. In the following sub-sections we outline how we capture bursts of images, align and merge each to a common image and use reconstruction pipeline.

A. Image Acquisition for Reconstruction

We capture $N$ frames over time $T_b$, before ceasing image acquisition for time $T_a$, and repeating this process, where $t$ is the time taken for capturing single image as in Fig. 2. We consider $N$ frames captured over time $T_a$ as a single burst and capture multiple bursts for a single trajectory.

Capturing a greater number of images per burst (i.e. larger $T_a$) increases the resulting SNR in merged images, provided there is a sufficient overlap. However, more images are also more computationally expensive for capturing, buffering and processing, so there is a trade-off in selecting the number of images in a burst between quality and computation. This is application-dependent: in our experiments, we use bursts of $N=7$ frames based on empirical evaluation.

B. Hierarchical Tile Alignment

To address non-uniform apparent motion between frames in a burst, we perform coarse-to-fine alignment on multi level Gaussian pyramids of single channel images [4]. Using the initial estimates from the coarser scale, we compute pairwise-tile-based alignment at each pyramid level by minimizing the distance measure between the common tile and the corresponding candidate tile of each alternate frame within a burst, similar to the approach in [4].

Directly minimizing an image-based misregistration measure based on pixel intensity errors is ideal here, because we capture consecutive images with a constant exposure and illumination, which is unlikely to change quickly over the duration of a single burst. To address illumination changes between bursts in a single trajectory, we use feature-based reconstruction.

We balance computational effort and the quality of alignment by fine tuning our alignment for different requirements that are common in reconstruction such as the number of channels in an input image, number of levels in an image pyramid and size of the input images in each pyramid level. Having high values for the aforementioned factors increases the computation at alignment. The number of levels in an image pyramid is also limited by the size of an input image.

There is a trade-off between the number of pyramid levels and the motion between the frames. If there is a larger motion between frames, we need larger search distances which require more pyramid levels or larger down sampling factors of input images on pyramid levels. This will reduce the likelihood of the search getting trapped in a local minimum and allow large motions between images. We empirically select the aforementioned factors to balance misalignment and motion between the frames.

C. Robust Temporal Merge

We merge all the images in the burst together, following the pairwise method developed in [4]. For a static scene, capturing and merging $N$ images ideally reduces random noise by $\sqrt{N}$. Merging method must also be robust to alignment failures that may occur due to changes in lighting, unaccounted scene motion at motion boundaries and occlusions.

To increase robustness to motion, we take the approach proposed in [4] in which temporal filtering incorporates a controllable degree of contribution from each image:

$$T_k(\omega) = \frac{1}{N} \sum_{z=1}^{N} T_z(\omega) + A_z(\omega)[T_k(\omega) - T_z(\omega)].$$ (1)
We improve the noise in the captured images by aligning every image in a burst with the chosen reference image of the burst. We merge the aligned images in temporal direction with a voting scheme to avoid misalignment. We use Wiener and bilateral filtering on temporally merged images. We give the outputs of this pipeline to COLMAP as three different inputs: burst with merge, burst with merge and Wiener filtering and burst with merge and bilateral filtering for reconstruction.

$R_\omega$ is the temporally filtered tile in frequency domain, $T_z(\omega)$ is the computed 2D discrete Fourier transform (DFT) of the corresponding alternative tile and $T_z(\omega)$ is the computed 2D DFT of the corresponding common tile. $z$ is the frame index and $N$ is the number of frames within a burst.

\[
A_z(\omega) = \frac{|D_z(\omega)|^2}{[D_z(\omega)]^2 + \sigma^2^2},
\]

where $D_z(\omega) = R_\omega(\omega) - T_z(\omega)$. $\sigma^2$ is the noise variance and $c$ is the degree of contribution that increases noise reduction at the expense of robustness.

We have control over the degree of contribution $c$ and the tile size used in filtering for fine tuning merged images. Ringing artifacts that are commonly associated with frequency domain filters are avoided with a windowing approach while other distortions that are not caused by the noise in the images are treated as misalignment in the frequency domain by the robust filtering method.

We also perform hybrid denoising by adding spatial filters separately on temporally merged images as depicted in Fig. 3. We filter high spatial frequency content aggressively to maximize subject quality with a noise shaping Wiener filter in 2D DFT domain similar to [4]. Separately, we apply bilateral filter as an edge-preserving filter on the temporally merged images in spatial domain.

Fig. 4 illustrates the performance of the align-and-merge approach, comparing with “conventional” noisy image, which corresponds to a single raw frame taken from a burst. Naïve averaging the images in a burst to a common image without alignment fails due to motion within the burst. This also demonstrates similarity towards capturing an image with longer exposure. Burst with merge demonstrates denoised image with higher SNR after successful tile-based hierarchical alignment and robust temporal merge.

D. Reconstruction Pipeline

We use COLMAP [1], an end-to-end feature-based state-of-the-art reconstruction pipeline to extract, match and geometrically verify features between sparse camera poses. We demonstrate sparse reconstruction by triangulating scene points and refining via bundle adjustment. We evaluate features following the procedures used in [34], [35].

IV. RESULTS

In the following we first evaluate improvements in feature detection and accuracy afforded by our method in noisy imagery using synthetic image sequences. Then we quantitatively evaluate our method in an SfM pipeline, comparing against conventional image capture and direct use of all images in the burst. We consider both 3D reconstruction performance and camera trajectory accuracy.

A. Feature Performance in Noise

We use a synthetic set of scenes with known feature locations to demonstrate feature performance in different noise levels. The synthetic images, shown in Fig. 5, have 25 disks at varying scales. Every image has an apparent motion of 12 or less pixels between each other. The contrast between the disks and background is 0.1. We introduce moderate noise with variance 0.03 (top) and strong noise with variance 0.1 (bottom).

We extract scale invariant feature transform (SIFT) features at a peak threshold of 0.015. It is evident from Fig. 5 that burst with merge performs better than burst without merge and conventional noisy with no spurious features for the moderate noise case ($\sigma = 0.03$). While our method extracts all true features in strong noise, the conventional method fails to extract all true features and burst without merge is overwhelmed by spurious features.

We demonstrate the measurement of true positive (TP) rate and false positive (FP) count quantitatively on these synthetic images with varying noise levels and peak detection thresholds as shown in Fig. 6 for both the proposed and conventional methods. The top row shows the TP rate and FP count at peak detection thresholds of 0.006 and 0.01 for a range of noise levels.

An appropriate peak threshold gives negligible FP count and higher TP rate at lower noise level. As the amount of noise increases, our method outperforms the conventional method yielding more true positive features and less false positive count.
We employ COLMAP with default settings, meaning a constant feature peak threshold of 0.0066 is employed for all images. To more fairly compare methods, we also repeat the experiment with peak thresholds adjusted to suit the different noise levels yielded by each method. Because unmerged images have more noise, they call for a more selective peak threshold. We empirically select peak thresholds of 0.001 and 0.005 for the proposed and conventional methods, respectively, as these yield similar levels of spurious feature detection.

Finally, we repeated the same experiments with COLMAP configured to be more permissive of images with few inlier feature matches, allowing as few as 15. This is useful when dealing with challenging scenes and allowed more images to be successfully incorporated into the SfM solution. Following the feature comparison approach in [34], we evaluate reconstruction performance as shown in Tab. I in terms of numbers of keypoints per image, putative feature matches per image, number of putative matches classified as inliers, match ratio: the proportion of detected features yielding putative matches, precision: the proportion of putative matches yielding inlier matches, matching score: the proportion of detected features yielding inlier matches, the...
mean number of 3D points in the reconstructed models and the mean number of 3D points per image.

At a moderate noise level, i.e., for images captured over 1 ms, our method outperforms alternative approaches across all metrics, by reconstructing all images passed for all scenes at default settings.

At a higher noise level, i.e., for images captured over 0.1 ms, not all images are reconstructed with default settings, and thus, we evaluate by tuning peak threshold values. The proposed method successfully reconstructed all the scenes with the strongest putative matches per image, inferior matches per image, match ratio, match score and 3D points per image. Our method outperforms the conventional approach or shows competitive results across all metrics.

While burst without merge uses 154 images and takes 49.58 minutes on average for reconstruction of a scene, our method uses only 22 images and takes 3.25 minutes on average for reconstruction on an Intel i7-9700 at 4.70 GHz. We extract features using NVIDIA GTX 770 for reconstruction. Our method also reduces the failure rate of reconstruction due to non-convergence and thus reconstructs more scenes than alternative approaches in strong noise.

C. Camera Trajectory Accuracy

We evaluate the accuracy of the camera trajectory estimate by using a robotic arm to collect ground truth poses. We aligned the camera poses to the ground truth poses as there is an arbitrary scale factor involved in monocular SfM. The arbitrary scale we use when reporting results is determined by the distance between the first pair of registered images. Fig. 8 shows how our method reconstruct more accurate camera poses than the competing method. We compute absolute instantaneous error and relative pose error between reconstructed camera poses and ground truth poses for translation and rotation as shown in Tab. II. Our method is competitive with moderate noise and outperforms alternative approaches with strong noise.

V. DISCUSSION

Our key finding is that using sparse higher SNR merged images is better than using lower SNR unmerged images for light-constrained SfM. We recommend using burst with merge for light-constrained SfM to increase feature performance, camera trajectory accuracy, and rate of success in incorporating challenging low-light images into the SfM solution. Although our approach involves increased computation in merging images within a burst, this is offset by the decreased rate of false positive feature detection, reducing the burden of extraneous feature matching and rejection.

In adapting burst photography for reconstruction in low light, there exists many trade-offs between the number of images within a single burst, gain, exposure time, motion blur, speed of the robot, image pyramid designs for motion alignment and noise in the image. In the following, we provide a set of recommendations on how to use and tune multiple bursts with merging in reconstruction tasks for common robotics applications:

- Increase exposure time as high as possible without getting deleterious motion blur. This yields higher-SNR input images and ultimately better reconstruction performance.
- Fixed-pattern noise can easily become the dominant source of noise in capturing low light images. This can be addressed by subtracting an average of multiple dark frames taken with the same gain, sensor temperature, and exposure time as the intended images within a burst.
- Increase gain as high as possible without yielding excessive saturation. This amplifies both signal and noise, but is important in overcoming the quantisation limit of the camera.
- Maximize the number of images in each burst to maximise SNR in the merged images. The image count is ultimately limited by availability of compute, and the rate of apparent motion of the scene. Motion relative to the reference frame should not exceed 1/2 of the total frame. For this reason we also recommend minimizing the maximum apparent motion relative to the reference frame by employing the centre most pose as the reference.
- Maximize the number of pyramid level and minimize the search window size to avoid search getting trapped in local minimum for images with strong noise. We recommend having the minimum pyramid level higher than 3 and image size higher than 1/16 of the original image at the coarse level of the pyramid for robust alignment.
- For images with motion relative to the reference frame larger than 1/4 of the total frame, minimize the image size at each pyramid level not lower than 1/16 of the original image and maximize search extend to locate corresponding pixel coordinates of aligned tiles not higher than 32 pixels for robust alignment.

VI. CONCLUSIONS

We adapted burst photography, commonly used in mobile photography for reconstruction in low light. We enabled
the use of direct methods for image registration within the burst and used feature-based SFM to handle the sparsity between bursts for reconstruction. We demonstrated successful reconstruction with decreased failure cases due to non-convergence compared to conventional methods. We also demonstrated improved performance relative to conventional imaging in true features, spurious features, putative matches, inlier matches, 3D points per images and accurate camera pose estimation.

Our method showed faster reconstruction with lower overall computational requirements compared to conventional methods. We expect that in more challenging low light conditions our method can improve the performance of 3D reconstruction and expand the range in which feature-based reconstruction can be applied.

This work is a first step toward solving the problem of 3D reconstruction in low light. For future work we anticipate employing adaptive sampling schemes, in which the parameters of burst capture and processing are dynamically tuned employing adaptive sampling schemes, in which the parameters are adjusted based on the illumination conditions of the scene to maximize reconstruction performance.

**TABLE I**

| Reconstruction Modes | Method | No. of images | No. of images pass | % images pass | Key Points/ Image | Putative Matches/ Image | Inlier Matches/ Image | Match Ratio | Precision | Match Score | 3D Points/ Image |
|----------------------|--------|---------------|-------------------|--------------|------------------|------------------------|-----------------------|-------------|-----------|-------------|----------------|
| Burst without merge  | 1ms    | 154           | 152               | 98.70        | 9110            | 681                    | 629                   | 7.50E-02    | 0.92      | 6.90E-02    | 39790          |
| Conventional noisy  | 1ms    | 22            | 21.6              | 98.13        | 9100            | 656                    | 603                   | 7.20E-02    | 0.92      | 6.40E-02    | 8270           |
| Burst with merge + Wiener | 1ms | 22            | 22               | 100          | 1052           | 866                    | 801                   | 8.20E-02    | 0.93      | 7.60E-02    | 11130          |
| Burst with merge + bilateral | 1ms | 22            | 22               | 100          | 10300          | 907                    | 842                   | 8.80E-02    | 0.93      | 8.20E-02    | 11387          |

**TABLE II**

| Trajectory Evaluation | Reconstruction Modes | Burst without Merge | Burst with Merge | Burst with Merge + Wiener | Burst with Merge + Bilateral |
|-----------------------|----------------------|--------------------|-------------------|---------------------------|----------------------------|
| Absolute              | Default (1ms)        | 0.73               | 0.395             | 0.82                      | 0.505                     |
|                       | Default (0.1ms)      | 2.83               | 0.715             | 4.51                      | 0.775                     |
| Instantaneous         | Permissive Mapping   | 2.77               | 0.499             | 5.24                      | 1.040                     |
|                       | Peak Threshold       | 2.46               | 0.771             | 5.43                      | 1.155                     |
|                       | Peak Threshold + Permissive Mapping | 2.75 | 0.487             | 2.92                      | 0.615                     |

| Pose                  | Default (1ms)        | 2.70               | 0.004             | 2.25                      | 0.005                     |
|                       | Default (0.1ms)      | 4.28               | 0.052             | 3.9                       | 0.032                     |
| Error                 | Permissive Mapping   | 4.99               | 0.035             | 7.15                      | 0.126                     |
|                       | Peak Threshold       | 3.36               | 0.021             | 5.66                      | 0.044                     |
|                       | Peak Threshold + Permissive Mapping | 4.66 | 0.039             | 4.71                      | 0.091                     |
chosen to suit the situation. We also expect the fusion of complementary sensors to yield interesting results.

REFERENCES

[1] J. L. Schonberger and J.-M. Frahm, “Structure-from-Motion Revisited,” in Computer Vision and Pattern Recognition (CVPR), 2016, pp. 4104–4113.

[2] X. Han, H. Laga, and M. Bennamoun, “Image-based 3D Object Reconstruction: State-of-the-Art and Trends in the Deep Learning Era,” IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), vol. 43, no. 5, pp. 1578–1604, 2021.

[3] Z. Liu, L. Yuan, X. Tang, M. Uyttendaele, and J. Sun, “Fast Burst Images Denoising,” ACM Transactions on Graphics (TOG), vol. 33, no. 6, pp. 1–9, 2014.

[4] S. W. Hasinoff, D. Sharlet, R. Geiss, A. Adams, J. T. Barron, F. Kaiz, 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, pp. 1–12, 2016.

[5] 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.

[6] B. Wronski, I. Garcia-Dorado, M. Ernst, D. Kelly, M. Krainin, C.-K. Liang, M. Levoy, and P. Milanfar, “Handheld Multi-Frame Super-Resolution,” ACM Transactions on Graphics (TOG), vol. 38, no. 4, pp. 1–18, 2019.

[7] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, “Image and Depth from a Conventional Camera with a Coded Aperture.” ACM Transactions on Graphics (TOG), vol. 26, no. 3, pp. 70–ex, 2007.

[8] H. Nagahara, S. Kuthirummal, C. Zhou, and S. K. Nayar, “Flexible Depth of Field Photography,” in European Conference on Computer Vision (ECCV), 2008, pp. 60–73.

[9] R. Raskar, A. Agrawal, and J. Tumblin, “Coded Exposure Photography: Motion Deblurring using Fluttered Shutter.” ACM Transactions on Graphics (TOG), vol. 25, no. 3, pp. 795–804, 2006.

[10] A. Levin, P. Sand, T. S. Cho, F. Durand, and W. T. Freeman, “Motion-Invariant Photography.” ACM Transactions on Graphics (TOG), vol. 27, no. 3, pp. 1–9, 2008.

[11] Y. Quan, M. Chen, T. Pang, and H. Ji, “Self2Self With Dropout: Learning Self-Supervised Denoising From Single Image.” in Computer Vision and Pattern Recognition (CVPR), 2020, pp. 1890–1899.

[12] M. Tassano, J. Delon, and T. Veit, “FastDVDeNet: Towards Real-Time Deep Video Denoising Without Flow Estimation,” in Computer Vision and Pattern Recognition (CVPR), 2020, pp. 1534–1563.

[13] H. Yue, C. Cao, L. Liao, R. Chu, and J. Yang, “Supervised Raw Video Denoising With a Benchmark Dataset on Dynamic Scenes,” in Computer Vision and Pattern Recognition (CVPR), 2020, pp. 2301–2310.

[14] Y. Nan, Y. Quan, and H. Ji, “Variational-EM-Based Deep Learning for Noise-Blind Image Deblurring,” in Computer Vision and Pattern Recognition (CVPR), 2020, pp. 3623–3632.

[15] J. Pan, H. Bai, and J. Tang, “Cascade Deep Video Deblurring Using Temporal Sharpness Prior,” in Computer Vision and Pattern Recognition (CVPR), 2020, pp. 3043–3051.

[16] Y. Yuan, W. Su, and D. Ma, “Efficient Dynamic Scene Deblurring Using Spatially Variant Deconvolution Network With Optical Flow Guided Training,” in Computer Vision and Pattern Recognition (CVPR), 2020, pp. 3555–3564.

[17] H. Dong, J. Pan, L. Xiang, Z. Hu, X. Zhang, F. Wang, and M.-H. Yang, “Multi-Scale Boosted Dehazing Network with Dense Feature Fusion,” in International Conference on Computer Vision (ICCV), 2020, pp. 2157–2167.

[18] Y. Li, Q. Miao, W. Ouyang, Z. Ma, H. Fang, C. Dong, and Y. Quan, “LAP-Net: Level-Aware Progressive Network for Image Dehazing,” in International Conference on Computer Vision (ICCV), 2019, pp. 3276–3285.

[19] X. Liu, Y. Ma, Z. Shi, and J. Chen, “GridDehazeNet: Attention-Based Multi-Scale Network for Image Dehazing,” in International Conference on Computer Vision (ICCV), 2019, pp. 7314–7323.

[20] Y. S. Chen, C. Y. Wang, M. H. Kao, and Y. Y. Chuang, “Deep Photo Enhancer: Unpaired Learning for Image Enhancement from Photographs with GANs,” in Computer Vision and Pattern Recognition (CVPR), 2018, pp. 6306–6314.