



DeblurExpandNet: image motion deblurring network aided by inertial sensor data

Shuang Zhang¹ · Ada Zhen¹ · Robert L. Stevenson¹

Received: 24 May 2021 / Revised: 19 October 2021 / Accepted: 23 October 2021 / Published online: 10 January 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2021

Abstract

In this paper, we propose a convolutional neural network for the removal of spatially varying motion blur from captured images with the assistance of inertial sensor data. In the proposed system, both the image and motion data are captured simultaneously and passed to a network for processing. The proposed network adopts three parallel nets to extract image features and a per-pixel concatenation to tightly integrate motion homographies estimated from inertial sensor data with the input degraded image. This unique network design facilitates the use of homographies which describes the motion blur kernel more accurately. Compared to the recently proposed image deblurring networks, the proposed network is found to produce restored images that have fewer artifacts and provide quantifiable and subjective improvement.

Keywords Image deblurring · Computer vision · Deep learning · Inertial sensor

1 Introduction

Motion blur is a common artifact in captured images and it occurs when there is relative motion between a camera and the scene being captured during the exposure time. The captured blurry image is typically modeled as a latent sharp image convolved with an unknown blur kernel plus noise. The goal of image deblurring is to estimate the sharp image from the blurry observation [3,10,19]. This is an ill-posed inverse problem which requires additional information or assumptions for the development of a stable and reasonable solution [3]. This paper addresses the situation where blur mainly comes from camera shake and auxiliary motion data from inertial sensors is available. This situation is found in most modern cellular smartphones which contain both image sensors and inertial sensors and are commonly used to capture images when held in the hand. Smartphone Application Programming Interfaces (APIs) are also available for developers to control image capture configurations and inertial data collection.

With significant progress in convolution neural networks, end-to-end image deblurring networks have been developed

to directly infer the latent sharp image without explicitly estimating the blur kernel [15,20,23,25]. Though these image deblurring networks achieve much shorter processing time and better generalization to most types of blur kernels, poor deblurring performance still occurs in challenging situations. This is because these networks only use a blurry image as input. More additional information or assumptions are required to better address the challenging situations.

Thus, researchers have explored the possibility of employing auxiliary information. Built-in inertial sensors can be utilized to track camera motion during the exposure time [5,17,21,29]. In the deep learning field, the inertial sensor data-aided deblurring scheme has achieved less success. The major challenge of this scheme is how to fuse the 2D image data with the 1D inertial sensor data using the network structure. To our best knowledge, the work of Mustaniemi et al. [14] is the only deep learning network that utilizes gyroscope data for single image deblurring. In the work of [14], only the gyroscope data is considered and the blur kernel shape is simplified to a straight line. Therefore, the estimated blur field is far from accurate. As a result, the potential of inertial sensor data is not fully realized. How to fuse different types of data is still an open problem.

In this paper, we propose an image deblurring network aided by inertial sensor data to address spatially varying motion blur. The proposed network is a two-input-single-output convolutional neural network. Its two inputs consist

✉ Shuang Zhang
szhang15@nd.edu

¹ Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA

of a blurry image and a stack of homographies computed from inertial sensor data. And its output is a restored sharp image. In the proposed network, an advanced deblurring Expand-Net [13] is tailored to merge information from the inertial sensor data with image data. Additional parallel multi-scale structures, residual blocks, and a global skip connection are also incorporated to improve overall image deblurring performance.

To train and test the proposed network, a synthetic dataset is developed in our previous conference paper [28]. The dataset contains gyroscope and accelerometer data as well as sharp/blurry image pairs. The dataset also considers the realistic error effects, including noise, rotation center shift and the rolling shutter effect. To evaluate the deblurring performance of the proposed approach, comparisons with several recent image deblurring networks are conducted. Quantitative, as well as visual results, are provided for both synthetic data and real data collected by a smartphone application.

The main contribution of this paper is as follows:

- A novel image deblurring network is proposed to address spatially varying motion blur by integrating the homographies generated from inertial sensor data with a simultaneously captured input blurry image,
- An ablation study is conducted to discuss usages of each component of the proposed network, and
- A comprehensive comparison with state-of-the-art image deblurring networks is provided to demonstrate the advantages of the proposed network.

The rest of the manuscript is organized as follows: Sect. 2 introduces the geometric blur model and the proposed dataset. Detailed network architecture and its training loss are described in Sect. 3. Section 4 investigates the usage of each component in the network in Sect. 4.2 Ablation Study and conducts comparisons with state-of-the-art image deblurring networks on the proposed synthetic dataset in Sect. 4.3 and the captured real data in Sect. 4.4. Finally, a conclusion is drawn in Sect. 5.

2 Blur model and DeblurIMU dataset

This section introduces the geometric blur model and its usage for generating the proposed synthetic dataset that contains sharp/blurry image pairs and simultaneously recorded gyroscope data and accelerometer data.

2.1 Geometric blur model

The geometric blur model has been widely used for inertial sensor-aided image deblurring algorithms [5,29] as it builds the connection between the motion blur and the inertial sen-

sor data that records the camera motion when the shutter is open. This blur model formulates the blurry image as an integration of the sharp images undergoing a sequence of projective motions during the exposure time [22]:

$$\mathbf{I}_B = \sum_{i=1}^{N_p} w_i \mathbf{I}_S(\mathbf{H}_i \mathbf{x}) + N, \quad (1)$$

where \mathbf{I}_B and \mathbf{I}_S refer to the blurry image and its latent sharp image, respectively. Noise N is often formulated as additional white Gaussian noise. The 3×1 vector \mathbf{x} denotes the homogeneous pixel coordinate. $\mathbf{I}_S(\mathbf{H}_i \mathbf{x})$ can be treated as an intermediate transformed frame captured by the camera at one pose i which is characterized by the 3×3 homography matrix \mathbf{H}_i and the corresponding weight w_i . w_i is proportional to exposure time at that pose. N_p refers to the total number of camera poses during the camera exposure time. The homography matrix \mathbf{H}_i can be decomposed into a rotation matrix \mathbf{R}_i , a translation vector \mathbf{t}_i , a normal vector \mathbf{n}_i , scene depth d and an intrinsic matrix $\mathbf{\Pi}$ [2]:

$$\mathbf{H}_i = \mathbf{\Pi} \left(\mathbf{R}_i + \frac{\mathbf{t}_i \mathbf{n}_i^T}{d} \right) \mathbf{\Pi}^{-1}, \quad (2)$$

To simplify the model, the depth is set to 1 and the normal vector assumes to be vertical to the image plane. And the intrinsic matrix $\mathbf{\Pi}$ is represented by focal length f and camera optical center (o_x, o_y) as the following:

$$\mathbf{\Pi} = \begin{bmatrix} f & 0 & o_x \\ 0 & f & o_y \\ 0 & 0 & 1 \end{bmatrix}. \quad (3)$$

where the focal length is set to 50 mm and the pixel size is $2.44\text{e-}6$ m/pixel in data synthesis. Therefore, $f = 50\text{e-}3/2.44\text{e-}6 = 20492$ pixels.

Camera rotations \mathbf{R}_i and translations \mathbf{t}_i can be inferred from the measurement of gyroscopes and accelerometers, respectively. The gyroscope measures rotation rates around the xyz- axis. In the proposed work, Android camera API 2 is adopted to directly produce linear acceleration that eliminates the gravity using other sensor data. In the following, linear acceleration instead of the original acceleration is assumed. The measured 3-axis angular velocity and the linear acceleration at the pose i are denoted as $\boldsymbol{\omega}_i = [\omega_{ix}, \omega_{iy}, \omega_{iz}]$ and $\mathbf{a}_i = [a_{ix}, a_{iy}, a_{iz}]$.

Given the sampling interval Δt and the rotation matrix \mathbf{R}_i at pose i , the rotation matrix at next pose $i + 1$ can be approximated as [21]:

$$\mathbf{R}_{i+1} = \mathbf{R}_i + \Delta t * \frac{d\mathbf{R}_t}{dt} = \begin{bmatrix} 1 & -d\phi_{iz} & d\phi_{iy} \\ d\phi_{iz} & 1 & -d\phi_{ix} \\ -d\phi_{iy} & d\phi_{ix} & 1 \end{bmatrix} \mathbf{R}_i, \quad (4)$$

where $d\phi_i = [d\phi_{ix}, d\phi_{iy}, d\phi_{iz}] = \boldsymbol{\omega}_i * \Delta t$. Then, the camera translation \mathbf{t}_i can be derived from accelerations \mathbf{a}_i using twice integration [29]:

$$\begin{aligned} \mathbf{v}_i &= \mathbf{v}_{i-1} + (\mathbf{R}_{i-1}^{-1} \mathbf{a}_{i-1} + \mathbf{R}_i^{-1} \mathbf{a}_i) * \Delta t / 2 \\ \mathbf{t}_i &= \mathbf{t}_{i-1} + (\mathbf{v}_{i-1} + \mathbf{v}_i) * \Delta t / 2 \end{aligned} \quad (5)$$

where the initial velocity \mathbf{v}_0 and initial translation \mathbf{t}_0 is assumed to be $\mathbf{0}$.

2.2 DeblurIMU dataset

To train and test the proposed image deblurring network aided by inertial sensor data, a novel dataset is developed based on the geometric blur model and GOPRO Large all dataset in our previous conference paper [28]. The GOPRO Large all dataset consists of successive sharp frames taken by a GOPRO4 Hero Black camera in more than 30 different scenes [15].

As suggested in our previous paper [28], the first step of generating a synthetic instance is to simulate data points collected from gyroscopes and accelerometers. Gyroscopes and accelerometers record angular velocity and acceleration of smartphone motion along three-axis, respectively. The angular velocity and the acceleration of each axis are modeled as a Gaussian distribution with zero means. After obtaining the measurement samples of gyroscope and accelerometer, Eqs. (2) to (5) are applied to generate a sequence of homographies.

To generate a blurry image, a sharp image is first picked from the GOPRO Large all dataset [15]. Then, the sequence of homographies generated from the previous step is applied to the sharp image using Eq. (1). White Gaussian noise with the randomly sampled standard deviation σ_r is also added to the blurry image. And the sharp frame is treated as the ground-truth image.

Though image noise is considered in Eq. (1), this geometric blur model is a relatively ideal blurry image model. It implicitly assumes that the smartphone rotation center locates at the optical center of the camera and that the image sensor employs a global shutter to capture the entire frame all at once. To simulate more realistic situations, the practical error effects, including rotation center shift and rolling shutter effect, are considered.

The overall data generation steps for the DeblurIMU dataset are summarized in Algorithm 1 in the **Supporting Information**. The value of parameters can be found in [28].

The proposed dataset consists of 2264 sets for training and 1221 sets for testing. Each set contains a ground-truth

sharp frame, an intermediate blurry image without errors, a blurry image with error effects, and inertial sensor data. The resolution of the images is 720×1280 .

Figure 1 illustrates a blurry example taken from the proposed DeblurIMU dataset and its point spread function (PSF). In this example, the motion blur is caused by rotation and its rotation center locates in the upper left area of the image. The resulting blur is spatially varying.

In the proposed network, a blurry image with error effects, its corresponding sharp image, and inertial sensor data are taken from the DeblurIMU dataset to train and test the network. The camera rotation center is also assumed to be known or already estimated from pre-processing steps described in Park et al. [17] or Hu et al. [5].

3 Proposed network

The proposed image deblurring network, DeblurExpandNet, aims to restore a sharp clear image from a blurry noisy image and the corresponding camera built-in inertial sensor data. Its overall architecture is presented in Fig. 2. Its input consists of an image corrupted by motion blur and noise and a stack of homographies calculated from inertial sensor data. Before being fed into the network, the homographies are zero-padded from size 3×3 to 16×16 and are stacked together in time order.

3.1 Network architecture

As illustrated in Fig. 2, the network architecture is composed of four branches: local, medium, dilation, and global. Their output feature maps are concatenated together and then passed into the following fusion, residual, and end blocks. Different branches are marked as different colors. All the convolution blocks have 64 channels. These blocks (except for the end block) represent image feature maps generated after a convolution layer and an activation layer, Scaled Exponential Linear Unit (SELU) [9]. SELU is a good replacement of the batch normalization [6] that automatically converges the input feature maps towards zero mean and unit variance.

The blurry image \mathbf{I}_B is accepted by local, medium, and dilation branches separately. These branches are designed to obtain image features at different scales. A global skip connection from input to output is also employed. This helps the network focus on training the residual and therefore suppresses the ringing artifacts.

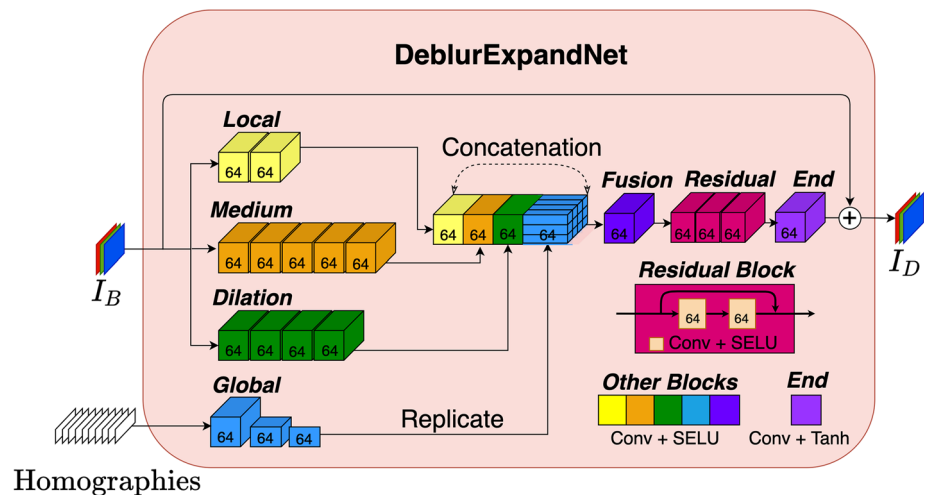
3.1.1 Local branch

The local branch adopts two convolution blocks with kernel size 3×3 , stride size 1, padding size 1, and dilation size 1.



Fig. 1 Spatially varying motion blur. **a** Image degraded by non-uniform motion blur taken from the DeblurIMU dataset [28]. **b** Point spread function (PSF) of (a)

Fig. 2 Network architecture inside DeblurExpandNet. The DeblurExpandNet accepts a blurry image, I_B , and a stack of homographies calculated from inertial sensor data to restore a sharp clear image I_D . The network architecture is composed of four branches: local, medium, dilation, and global. Their output feature maps are concatenated together and then passed into the following fusion, residual, and end blocks. Different branches are marked as different colors. All the convolution blocks have 64 channels



The receptive field of the local branch is 5×5 pixels and thus this branch aims to extract high-frequency image features.

3.1.2 Medium branch

The medium branch is designed to fill the gap between the local and dilation branch to generate medium level features with a receptive field of 11×11 pixels. It consists of five consecutive convolution blocks with kernel size 3×3 , stride size 1, padding size 1, and dilation size 1.

3.1.3 Dilation branch

The dilation branch achieves a larger receptive field of 17×17 pixels through four convolution blocks with kernel size 3×3 , stride size 1, padding size 2, and dilation size 2. Though with one convolution block fewer than the medium branch, the dilation branch has a wider receptive field. It is because the adopted dilated convolution layers can provide exponential expansion of the receptive field without loss of resolution or coverage [24].

3.1.4 Global branch

In parallel with the three image branches mentioned above, the global branch deals with inertial sensor data. Unlike the original version in ExpandNet [13], the global branch in the proposed DeblurExpandNet accepts a sequence of padded homographies rather than images as its input. It has three convolution blocks with kernel size 3×3 , stride size 2, padding size 1, and dilation size 1. The global branch downsamples input padded homographies with size 16×16 to a 1×1 feature vector with 64 channels. In projective geometry, a homography is an invertible structure-preserving mapping of two projective spaces [1]. And in the proposed network, the homography indicates how each pixel is moving during capture time to yield the motion blur since it is generated from inertia sensor data that records the camera movement. The generated feature vector of the global branch is replicated to match the height and width of the input image so that each pixel of the feature maps from the image branches is appended with a homography feature vector. By applying the feature vector of the homography to each pixel of high-level feature maps of the image branches, the proposed network

mimics the image warping process using homography. And the global branch plays an important role in inverse the homographies and extract useful projective mapping information.

3.1.5 Fusion block

The fusion block contains one convolution block with kernel size 3×3 , stride size 1, padding size 0, and dilation size 1. The size 1 stride ensures that the output feature maps have the same height and width as the input image I_B .

3.1.6 Residual block

Three residual blocks are appended to the fusion block to improve the sharpness of the deblurred image (See Ablation Study). The residual block is proposed to increase network depth and improve overall performance [4]. It consists of two convolution layers and a skip connection that adds the input to the output. Image deblurring networks like [15] and [23] add a batch normalization layer [6] after each convolution layer inside the residual block. In each residual block of the proposed network, the batch normalization layer is removed and the activation function Rectified Linear Units (RELU) [16] is replaced by SELU.

3.1.7 End block

End block is a convolution layer with kernel size 1×1 , stride size 1, padding size 0. An activation layer Tanh is added to make sure the output pixel value is within the range of $[-1, 1]$.

3.2 Loss functions

The loss function for training the proposed network is defined as:

$$\mathcal{L} = \lambda_2 \mathcal{L}_2 + \lambda_{grad} \mathcal{L}_{grad}, \quad (6)$$

where \mathcal{L}_2 and \mathcal{L}_{grad} refer to L_2 loss and gradient loss, respectively. And λ_2 and λ_{grad} denote the weights assigned for each loss term, where $\lambda_2 = 100$ and $\lambda_{grad} = 20$. These two coefficients, λ_2 and λ_{grad} , are picked to make sure the value of the two losses is on a similar scale (see Fig. S2 in the Supporting Information). In this way, the two losses make a similar contribution in the training process.

The loss \mathcal{L}_2 in Eq. (6) is the distance between the network output I_D and the ground-truth sharp image I_S in L_2 norm:

$$\mathcal{L}_2 = \mathbb{E}[\|I_S - I_D\|_2]. \quad (7)$$

The gradient loss proposed by the work of [7] has been proven to effectively suppress the pattern artifacts in

deblurred images [27]. The gradient loss \mathcal{L}_{grad} in Equation is defined as the distance between gradient maps of I_D and I_S in L_1 norm:

$$\mathcal{L}_{grad} = \mathbb{E}[\|\nabla_h(I_S) - \nabla_h(I_D)\|_1 + \|\nabla_v(I_S) - \nabla_v(I_D)\|_1], \quad (8)$$

where ∇_h and ∇_v denote the horizontal and vertical gradient operators which are approximated by applying the Sobel filter.

4 Experimental results

4.1 Training details

All the training and testing of the proposed approach are conducted on a NVIDIA GeForce GTX 1080 Ti GPU using Pytorch framework [18]. Adam optimizer [8] with initial learning rate $lr = 0.0002$, exponential decay rate $\beta_1 = 0.5$, $\beta_2 = 0.999$ and $\epsilon = 1e - 08$ is adopted to train the proposed network. All trainable weights are initialized using a Gaussian distribution with zero mean and standard deviation 0.02 and bias are initialized as 0.0. The network is trained for 125 epochs with batch size 2.

In the proposed network, the input blurry/ground-truth training pair is randomly cropped into size 720×720 in the original scale. In this way, the input image and the homographies are matched with each other. In total 10 homographies are computed using Eq. (2) to (5) and each homography is zero-padded to size 16×16 before fed into the network.

4.2 Ablation study

The original ExpandNet was proposed for the image super-resolution problem. It consists of three parallel branches: local, dilation, and global, and these three branches are followed by a fusion net to merge the high-level feature maps into a single image [13]. To deal with spatially varying motion blur, in this paper, unique structures are designed and implemented, including a new global branch with homographies as input, an extra image branch (medium branch), residual blocks, and global skip connection.

As described in Table 1, four baseline models are designed to evaluate these network structures in the proposed Deblur-ExpandNet, where the baseline model Model0 refers to the original ExpandNet [13] where only the blurry image is considered and the input for the global branch is also the blurry image that is resized into 256×256 . All the five models are trained on the train dataset and tested on the test dataset of the DeblurIMU dataset [28].

The results (shown in Table 1 and Fig. 3) demonstrate that the original ExpandNet (Model0) presents the lowest quan-

Table 1 Ablation study on DeblurIMU dataset

	Description	Homography	Medium branch	Global skip connection	Residual block	PSNR	SSIM
Model0	Original ExpandNet					23.63	0.826
Model1	Model0 + Homography	✓				26.28	0.885
Model2	Model1 + Medium branch	✓	✓			26.93	0.896
Model3	Model2 + Global skip	✓	✓	✓		27.29	0.898
Model4	Proposed	✓	✓	✓	✓	27.19	0.894

**Fig. 3** Ablation study. **a** Input Blurry. Deblurred results of **(a)** processed by: **b** Model0. **c** Model1. **d** Model2. **e** Model3. **f** Model4

titative metrics and the most blurry visual result, because ExpandNet is designed for the image super-resolution problem, not for image deblurring. By incorporating the information from homography, Model1 achieves a significant bump on the PSNR and SSIM, and its visual result is also sharper than Model0. Integrated with the medium branch, Model2 attains a much sharper result compared to Model1. It's because the medium branch fills the gap of the receptive field between the local and dilation branch. However, ringing artifacts occur in the results of Model2. The input-to-out design in Model3 suppresses the ringing artifacts by letting the network focus on the residual between the input and the out image and achieves the highest PSNR and SSIM among all models. However, its visual result appears less sharp. Eventually, three residual blocks are inserted between the fusion block and the end block in the proposed network (Model4) and the sharpest visual result is therefore obtained.

4.3 Comparisons on synthetic dataset

The proposed DeblurExpandNet is compared with previous state-of-the-art image deblurring networks on the dataset DeblurIMU, including the work of Sim et al. [20], Tao et al. [23], Zamir et al. [25] and Mustaniemi et al. [14].

Sim et al. [20] trained a per-pixel blur kernel map and applied it to a residual image to restore a latent sharp frame from a blurry one. Tao et al. [23] proposed an image deblurring network that achieves very sharp visual results and comparatively high Peak-to-Noise Ratio (PSNR) and Structure Similarity (SSIM). Zamir et al. [25] developed a multi-stage architecture that progressively learns restoration

Table 2 Average PSNR, SSIM and processing time tested on DeblurIMU dataset

Method	PSNR	SSIM	Processing time
Zhang [26]	18.37	0.707	1.2 s
DeblurGAN [11]	20.84	0.756	1.8 s
Zamir [25]	21.63	0.762	2.5 s
Tao [23]	21.80	0.761	1.6 s
Sim [20]	21.88	0.767	0.8 s
DeblurGAN_v2 [12]	22.45	0.836	3.1 s
Mustaniemi [14]	22.72	0.812	1.4 s
DeblurExpandNet	27.19	0.894	0.6 s

functions for the blurry input. Similar to the DeblurExpandNet, the work of Mustaniemi *et al.* [14] adopts gyroscope data as auxiliary information.

Table 2 presents average PSNR and SSIM on the DeblurIMU dataset, respectively. In this comparison, the official implementation and pretrained model released by authors are adopted. The results (shown in Table 2) demonstrate that the proposed DeblurExpandNet achieves the highest average PSNR & SSIM, and the shortest processing time among all image deblurring networks.

Figure 4 illustrates the visual results of these networks. The deblurred results of Sim [20], Tao [23] and Zamir [25] suffer from artifacts, for example, the local images of the wheel and shadows in Fig. 4b–d. These areas either appear to have undesired texture or are degraded by spatially varying blur. Without the information provided by the inertial sensor data, it is hard for these networks to tell the difference

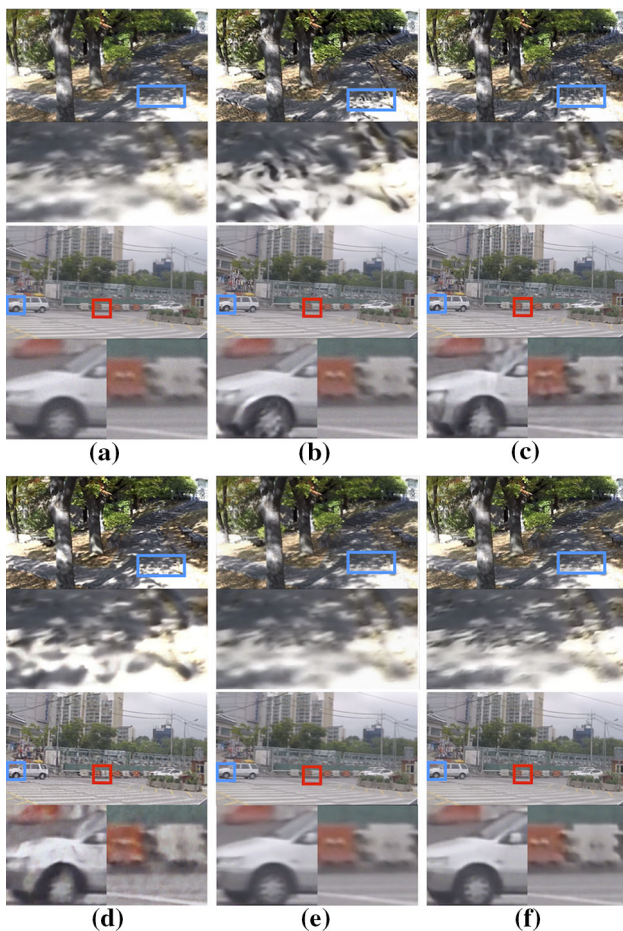


Fig. 4 Comparison tested on the proposed DeblurIMU dataset. **a** Input blurry. Deblurred results of **a** processed by: **b** Sim [20]. **c** Tao [23]. **d** Zamir [25]. **e** Mustaniemi [14]. **f** Proposed DeblurExpandNet

between motion blur and image texture or deal with spatially varying blur. This results in the visible artifacts that are seen. DeepGyro proposed by Mustaniemi et al. [14] has no problem handling the texture and non-uniform blur, because it accepts a blur field generated from gyroscope data. However, its image deblurring performance is limited and the zoomed-in patches remain blurry and less sharp compared to the proposed DeblurExpandNet.

4.4 Comparisons on real images and inertial sensor data

Comparisons are also conducted on real noisy blurry images and correspondent inertial sensor data that are captured and collected by a hand-held smartphone LG Nexus 5. An android application [30] is developed and adopted to acquire a blurry image, gyroscope and accelerometer data during the exposure time.

The deblurred results on the real data are shown in Fig. 5. The inertia sensor data is assumed to be accurate and no

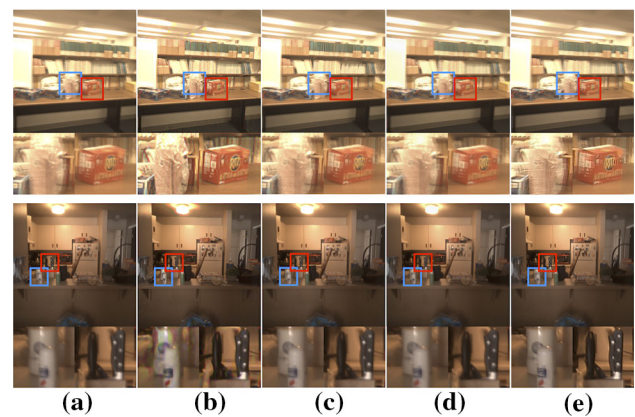


Fig. 5 Comparison on real data. **a** Input Blurry. Deblurred results of **a** processed by: **b** Tao [23]. **c** Zamir [25]. **d** Mustaniemi [14]. **e** Proposed.

pre-processing steps are added. Similar to the observation on the synthetic data, the deblurred images of Tao [23] present some artifacts, such as on the plastic wrap and the candle in Fig. 5b. And the results of Mustaniemi [14] and Zamir [25] in Fig. 5c, d remain blurry. The proposed DeblurExpandNet has the sharpest output with the least artifacts among all the compared networks.

5 Conclusion

In this paper, an advanced ExpandNet is proposed to address the problem of spatial varying blur through the capture of one blurry image and inertial sensor data. The spatial varying blur across the blurry image is learned from a sequence of homographies that are calculated from the inertial sensor data. To integrate the homographies with the input blurry image, the high-level feature vector of the homographies is concatenated with each pixel in multi-level feature maps of the input. Additional structures, such as medium branch and residual blocks, are also considered to improve sharpness and suppress artifacts. Compared to the state-of-the-art deblurring networks, the proposed DeblurExpandNet achieved the highest PSNR and SSIM as well as the shortest processing time. And the visual results of the proposed method presented the least artifacts and the sharpest performance.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11760-021-02067-1>.

References

- Berger, M., Cole, M., Cole, M., Levy, S.: Geometry. No. v. 2 in Geometry I [-II]. Springer (1987). <https://books.google.com/books?id=MoYZnQAACAAJ>
- Faugeras, O.D., Lustman, F.: Motion and structure from motion in a piecewise planar environment. *IJPRAI* **2**, 485–508 (1988)

3. Fergus, R., Singh, B., Hertzmann, A., Roweis, S.T., Freeman, W.T.: Removing camera shake from a single photograph. *ACM Trans. Graph.* **25**(3), 787–794 (2006)
4. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp. 770–778 (2016). <https://doi.org/10.1109/CVPR.2016.90>
5. Hu, Z., Yuan, L., Lin, S., Yang, M.: Image deblurring using smartphone inertial sensors. In: Proceedings of the 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp. 1855–1864 (2016). <https://doi.org/10.1109/CVPR.2016.205>
6. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd international conference on international conference on machine learning, Vol 37, ICML'15, pp. 448–456. JMLR.org (2015)
7. Jiao, J., Tu, W.C., He, S.W.H., Lau, R.: Formresnet: formatted residual learning for image restoration. In: Proceedings of the 2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW), pp. 1034–1042 (2017). <https://doi.org/10.1109/CVPRW.2017.140>
8. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. *CoRR* **abs/1412.6980** (2015)
9. Klambauer, G., Unterthiner, T., Mayr, A., Hochreiter, S.: Self-normalizing neural networks. *CoRR* **abs/1706.02515** (2017)
10. Krishnan, D., Tay, T., Fergus, R.: Blind deconvolution using a normalized sparsity measure. In: CVPR 2011, pp. 233–240 (2011). <https://doi.org/10.1109/CVPR.2011.5995521>
11. Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., Matas, J.: DeblurGAN: blind motion deblurring using conditional adversarial networks. In: Proceedings of the 2018 IEEE/CVF conference on computer vision and pattern recognition, pp. 8183–8192 (2018)
12. Kupyn, O., Martyniuk, T., Wu, J., Wang, Z.: Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In: The IEEE International Conference on Computer Vision (ICCV) (2019)
13. Marnerides, D., Bashford-Rogers, T., Hatchett, J., Debattista, K.: Expandnet: a deep convolutional neural network for high dynamic range expansion from low dynamic range content. *Comput. Graph. Forum* **37**(2), 37–49 (2018). <https://doi.org/10.1111/cgf.13340>
14. Mustaniemi, J., Kannala, J., Särkkä, S., Matas, J., Heikkilä, J.: Gyroscope-aided motion deblurring with deep networks. In: Proceedings of the 2019 IEEE winter conference on applications of computer vision (WACV) pp. 1914–1922 (2018)
15. Nah, S., Kim, T.H., Lee, K.M.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proceedings of the 2017 IEEE conference on computer vision and pattern recognition (CVPR) pp. 257–265 (2017)
16. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In: J. Fürnkranz, T. Joachims (eds.) ICML, pp. 807–814. Omnipress (2010)
17. Park, S.H., Levoy, M.: Gyro-based multi-image deconvolution for removing handshake blur. In: Proceedings of the 2014 IEEE conference on computer vision and pattern recognition, pp. 3366–3373 (2014). <https://doi.org/10.1109/CVPR.2014.430>
18. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch (2017)
19. Shan, Q., Jia, J., Agarwala, A.: High-quality motion deblurring from a single image. In: ACM SIGGRAPH 2008 Papers, SIGGRAPH '08, pp. 73:1–73:10. ACM, New York, NY, USA (2008). <http://doi.acm.org/10.1145/1399504.1360672>
20. Sim, H., Kim, M.: A deep motion deblurring network based on per-pixel adaptive kernels with residual down-up and up-down modules. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 2140–2149 (2019). <https://doi.org/10.1109/CVPRW.2019.00267>
21. Sindelar, O., Sroubek, F.: Image deblurring in smartphone devices using built-in inertial measurement sensors. *J. Elect. Imag.* **22**, 011003 (2013)
22. Tai, Y., Tan, P., Brown, M.S.: Richardson-lucy deblurring for scenes under a projective motion path. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(8), 1603–1618 (2011). <https://doi.org/10.1109/TPAMI.2010.222>
23. Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.: Scale-recurrent network for deep image deblurring. In: Proceedings of the 2018 IEEE conference on computer vision and pattern recognition, CVPR 2018, Salt Lake City, UT, USA, pp. 8174–8182 (2018). <https://doi.org/10.1109/CVPR.2018.00853>
24. Yu, F., Koltun, V.: Multi-scale context aggregation by dilated convolutions. In: International conference on learning representations (ICLR) (2016)
25. Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.H., Shao, L.: Multi-stage progressive image restoration. In: CVPR (2021)
26. Zhang, H., Dai, Y., Li, H., Koniusz, P.: Deep stacked hierarchical multi-patch network for image deblurring. In: The IEEE conference on computer vision and pattern recognition (CVPR) (2019)
27. Zhang, S., Zhen, A., Stevenson, R.L.: Deep motion blur removal using noisy/blurred image pairs (2019)
28. Zhang, S., Zhen, A., Stevenson, R.L.: A dataset for deep image deblurring aided by inertial sensor data. In: Fast track article for IS&T International Symposium on Electronic Imaging 2020: Computational Imaging XVIII proceedings. pp. 379–1–379–6(6) (2020). <https://doi.org/10.2352/ISSN.2470-1173.2019.13.COIMG-136>
29. Zhen, R., Stevenson, R.: Inertial sensor aided multi-image nonuniform motion blur removal based on motion decomposition. *J. Elect. Imag.* **27**(5), 053026 (2018)
30. Zhen, R., Stevenson, R.L.: Multi-image motion deblurring aided by inertial sensors. *J. Elect. Imag.* **25**, 013027 (2016). <https://doi.org/10.1117/1.JEI.25.1.013027>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.