Single Image Non-uniform Blur Kernel Estimation via Adaptive Basis Decomposition

Guillermo Carbajal
Universidad de la República
Uruguay

Patricia Vitoria
Universitat Pompeu Fabra
Spain

Mauricio Delbracio
Universidad de la República
Uruguay

Pablo Musé
Universidad de la República
Uruguay

José Lezama
Universidad de la República
Uruguay

Abstract

Characterizing and removing motion blur caused by camera shake or object motion remains an important task for image restoration. In recent years, removal of motion blur in photographs has seen impressive progress in the hands of deep learning-based methods, trained to map directly from blurry to sharp images. Characterization of motion blur, on the other hand, has received less attention and progress in model-based methods for restoration lags behind that of data-driven end-to-end approaches. In this paper, we propose a general, non-parametric model for dense non-uniform motion blur estimation. Given a blurry image, we estimate a set of adaptive basis kernels as well as the mixing coefficients at pixel level, producing a per-pixel map of motion blur. This rich but efficient forward model of the degradation process allows the utilization of existing tools for solving inverse problems. We show that our method overcomes the limitations of existing non-uniform motion blur estimation and that it contributes to bridging the gap between model-based and data-driven approaches for deblurring real photographs.

1. Introduction

Motion blur results from the relative motion between the camera and the scene, which is determined by the interaction of three elements: the motion of the camera or ego-motion, the three-dimensional geometry of the scene, and the motion of objects in the scene. When the exposure time is large compared to the relative motion, the camera sensor at each point receives and accumulates light coming from different sources, producing different amounts of blur.

Psychophysical and neurological evidence show that motion blur provides important cues for visual perception, scene understanding and locomotion [4, 17, 40]. Besides deblurring, motion blur estimation has been successfully applied to different tasks such as scene interpretation, structure from motion, image segmentation, and uncertainty characterization of the observation [11, 19, 28].

Most non-uniform motion blur estimation methods assume a parametric model of the motion field, either by considering a global parametric form induced by camera motion [16, 18, 41, 47], or by locally modeling the motion field with linear kernels, parameterized by the length of the kernel support and its orientation [15, 23, 41, 44]. In most situations, for instance under camera shake from hand tremor, those models are not adapted to real case scenarios [13].

To overcome these limitations, we propose a novel approach for non-parametric, dense, spatially-varying motion blur estimation based on an efficient low-rank representation of the pixel-wise motion blur kernels. More precisely, for each blurred image, a neural network estimates an image-specific set of kernel basis functions, as well as a set of pixel-wise mixing coefficients, cf. Figure 1. In this way, for each pixel a unique motion blur kernel is assigned, given by the corresponding linear combination of the image-specific kernel basis functions. We show that this procedure allows to generate a wide range of complex motion blur kernels that are well adapted to real acquisition scenarios. To the best of our knowledge, the proposed approach is the first dense non-parametric non-uniform motion blur estimation method.

To further validate our method, we apply our estimated motion blur fields to two tasks: model-based image deblurring [6, 25, 26, 49, 47] and blur detection [14, 29, 42]. We show that in both cases we achieve results that are comparable to those of state-of-the-art end-to-end deep learning methods in standard benchmarks of real blurred images, therefore contributing to bridge the gap between model-based and data-driven approaches.

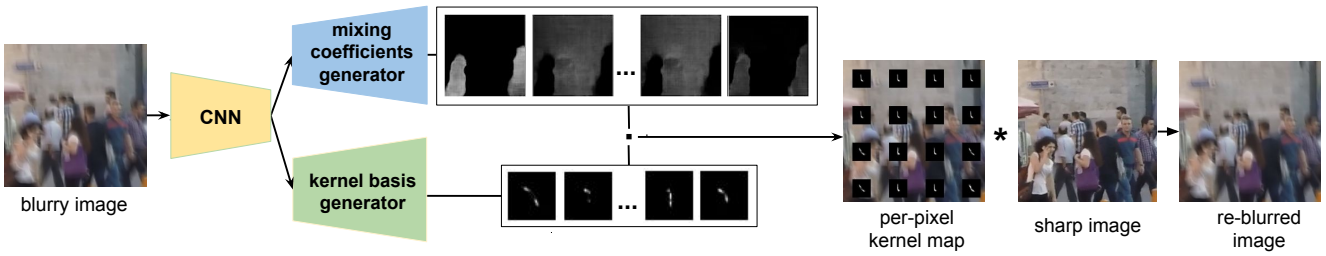

Figure 1: **Overview of the proposed method.** Given a blurry image, a two-headed neural network predicts the kernel basis and the pixel-wise mixing coefficients. This efficient representation allows re-blurring the corresponding sharp image by first convolving it with each basis kernel and performing a weighted sum of the results using the mixing coefficients. The CNN and generators are trained by comparing the re-blurred image to the original blurry image.

Code and pre-trained model weights will be available upon acceptance.

2. Related Work

Single Image Non-Uniform Motion Blur Estimation Early methods attempting to estimate spatially-varying motion blur kernels consider that such non-uniformity is mainly caused by 3D camera tilts or rotations [16, 45, 47]. In this setting, the blurred image results from the integration of the intermediate images that are varying perspective projections of the scene. By assuming that the focal length is sufficiently long or the scene is far enough to be considered as planar, the transformations are reduced to homographies. This leads to the so-called Projective Motion Blur Model (PMBM) [45]. These methods achieve impressive results under these conditions, but fail when the scene is close to the camera [16, 47]. Moreover, these methods suffer from high computational cost, since the optimization involves a large number of homographies that must be computed for the intermediate estimated images.

In order to reduce the computational cost of PMBP approaches, Hirsch *et al*. [18] propose a position dependent combination of a set of localized blur kernels. Thus, they are able to express smoothly varying blur using the structural constraints of the PMBM model while still being linear in its parameters. To that end, the local blur kernels at the patch-level are defined as linear combinations of a set of pre-computed homographies. However, the weights of these linear combinations are optimized for the full image and therefore are global and not pixel-specific.

To deal with spatially-varying blur due to depth and moving objects, methods like [22, 36] propose to segment the image in a reduced number of layers according to a metric representing the amount of blur. These methods are sensitive to the segmentation of the blurred image, and while they are well adapted to scenes with moving objects, they do not behave well when the camera is close to a scene presenting complex 3D structure.

A different approach that is well adapted to both scene depth variations and moving objects consists in predicting motion blur locally, at the patch level. Sun *et al*. [44] propose a deep learning approach to predict the probabilistic distribution of motion blur at the patch level using a Convolutional Neural Network (CNN). To this end, they consider a set of pre-defined basis motion kernels. These kernels are linear, and are parameterized by their lengths and orientations. The network estimates the probability of each kernel for a given patch. This leads to a non-dense motion blur kernels field, which is later made dense using a Markov random field model enforcing motion smoothness. Gong *et al*. [15] directly estimate a dense motion flow from the blurred image using a fully-convolutional deep neural network (FCN). To train the FCN, they simulate motion flows to generate synthetic blurred image / motion flow pairs. As in [44], the predicted motion blur kernels are linear and parameterized by their horizontal and vertical components, each of them defined on a discrete set. Both methods propose to apply their local motion blur kernel estimates to image deblurring, using an L^2 data fitting term and combined with EPLL image prior [55].

Another interesting parameterization in the context of video appears in [5, 20] where per-pixel motion blur kernels are modeled as the combination of two line segments, obtained from the next- and previous-frame optical flow estimations. In [5], kernels are inferred by a neural network that predicts indexes in a pre-computed look-up table.

Kernel Prediction Networks Recently, Kernel Prediction Networks (KPN) have been proposed for low-level vision tasks such burst denoising [31, 48], optical flow estimation, frame interpolation [34, 35], stereo and video prediction [21], among others. Several works have used KPNs in the context of burst denoising. Mildenhall *et al*. [31] produce denoised estimates at each pixel as a weighted average of observed noisy intensities in a window around that

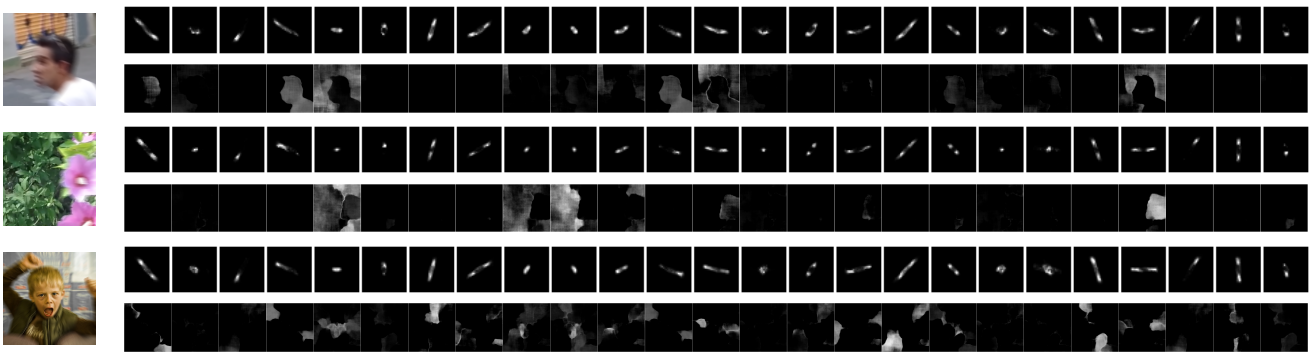

Figure 2: **Examples of generated kernel basis $\{\mathbf{k}^b\}$ and corresponding mixing coefficients $\{\mathbf{m}^b\}$** predicted from the blurry images shown on the left. The adaptation to the input is more notorious for the elements that have significant weights.

pixel location in all frames. These averaging weights, or kernels, are allowed to vary from pixel to pixel to implicitly account for motion and image discontinuities, and are predicted from the noisy input burst using a KPN. To improve the computational efficiency of this procedure, Xia *et al*. [48] propose a basis prediction network that, given an input burst, predicts a set of global basis kernels — shared within the image — and the corresponding mixing coefficients, which are specific to individual pixels.

Relation with the Proposed Approach In this work we propose dense, spatially-varying motion blur estimation based on an efficient low-rank representation of the pixel-wise motion blur kernels. As in [5, 15, 20, 44] the per-pixel kernels are modeled as linear combinations of base elements. However, contrary to their approaches, we use a KPN to infer kernel base elements that are specific to each input image, and that have no fixed parametric form. Instead of learning the kernels to solve directly the inverse problem, like in [48], we learn them to fit the forward model.

3. Method for Non-Uniform Blur Estimation

3.1. Image Degradation Model

We model non-uniform motion blur as the convolution[1] of a sharp image with a spatially varying filter, the *motion blur kernel*. This simple model represents the integration, at each pixel, of photons arriving from different sources due to relative motion between the camera and the scene. Formally, given a sharp image $\mathbf{u}$ of height H and width W, and a set of blur kernels $\mathbf{k}_{i,j} \in [0,1]^{K\times K}$, for $i = 1 \dots H, j = 1 \dots W$, the blurry image $\mathbf{v}$ is the result of applying the per-pixel operation:

$$v_{i,j} = \langle \bar{\mathbf{u}}_{i,j}, \mathbf{k}_{i,j} \rangle + n_{i,j}, \tag{1}$$

[1] In general, we refer to convolution in the deep learning sense, except when it is clear from the context.

where $\bar{\mathbf{u}}_{i,j}$ is a window of size $K \times K$ around pixel (i, j) in image $\mathbf{u}$ and $n_{i,j}$ is independent zero-mean white Gaussian noise. In our model, we assume conservation of energy by imposing $||\mathbf{k}_{i,j}||_1 = 1$.

Estimating a unique blur kernel per pixel $\mathbf{k}_{i,j}$ becomes computationally impractical for large images and large kernel sizes. To mitigate this problem, we propose an efficient representation of the per-pixel kernels $\mathbf{k}_{i,j}$ using an adaptive basis decomposition. A set of B image-dependent basis motion kernels $\{\mathbf{k}^b\}_{b=1,\dots,B}$ is computed, together with the corresponding pixel-wise mixing coefficients $\{\mathbf{m}^b\}$. The mixing coefficients are normalized so that they sum to one at each pixel location. Thus, the per-pixel kernels $\mathbf{k}_{i,j}$ result from the convex combination of the basis kernel, conservation of energy is guaranteed, and the degradation model becomes:

$$v_{i,j} = \langle \bar{\mathbf{u}}_{i,j}, \sum_{b=1}^{B} \mathbf{k}^b m^b_{i,j} \rangle + n_{i,j}, \tag{2}$$

We use a deep neural network to estimate, from a given input blurry image, both the dictionary of B basis motion kernels $\{\mathbf{k}^b\}_{b=1,\dots,B}$ and the mixing coefficients $\{\mathbf{m}^b\}_{b=1,\dots,B}$. Building upon recent work in kernel prediction networks [48], the network is composed by a shared backbone CNN and two generator heads. We refer to Figure 1 for an overview of the proposed method. The first generator outputs a global kernel basis of size $K \times K \times B$ (i.e., B basis elements of size $K \times K$). The second generator outputs B maps of mixing coefficients of the same spatial size as the input image, thus, the resulting size is equal to $H \times W \times B$. In our experiments we used $K = 33$, and the number of basis kernels $B = 25$ was set by analyzing the reconstruction cost of the low-rank decomposition for typical rotation, zoom and object motion blur fields.

Figure 2 shows several examples of the set of kernel basis and corresponding mixing coefficients predicted for different images. Note how the basis is image-dependent, and

Figure 3: **Visual comparison of non-uniform motion blur kernel estimation.** From top to bottom: Gong *et al*. [15], Sun *et al*. [44] and our proposed approach. From left to right: two examples from CUHK blur detection dataset [42], two from GoPro [33] and one from REDs [32]. Best viewed in electronic format.

this adaptation is more notorious for the kernel basis that are active (i.e. the corresponding mixing coefficients have high values throughout the scene). Normalization of the blur kernels and the mixing coefficients is achieved by using Softmax layers. Architectural details of the network are presented in the Appendix.

Limitations of the model The main limitations of our image degradation model is that the motion fields that can be captured are limited by the size of the kernel support K. This size is limited for computational reasons, and because a larger kernel dimension would require a larger number of base elements to capture the complexity of the motion flow, i.e. for the low-rank approximation to be accurate. One idea to overcome this limitation is to extend the proposed approach to a multi-scale setting.

Also, our model cannot cope with with saturated pixels, since in these pixels the energy conservation is not satisfied. Alternatives to deal with this limitation are proposed in [47, 7, 38]. Having the ability to deal with saturated pixels may improve kernel estimation, since usually motion fields observed at point light sources are very well defined.

3.2. Objective Function

We propose two reconstruction losses to train the generators of basis and mixing coefficients.

Reblur Loss Given corresponding blurry and sharp images, we first apply the predicted motion blur field to the sharp image. The re-blurring of the sharp image can be done efficiently by first convolving it with each of the kernels in the base, and then doing an element-wise product of the B resulting images with the corresponding mixing coefficients, and then adding the results. More precisely, given a blurry image $\mathbf{v}^{GT}$, we aim to find the global *kernel basis* $\{\mathbf{k}^b\}$ and mixing coefficients $\{\mathbf{m}^b\}$ that minimize

$$\mathcal{L}_{reblur} = \sum_i \sum_j w_{i,j}(v_{i,j} - v_{i,j}^{GT})^2, \tag{3}$$

where the $v_{i,j}$ are computed using (2), and $w_{i,j}$ is a scalar used to weigh different regions in the image. The effect of these weights will become more clear in Section 3.3, when training on synthetic data.

Kernel Loss When available, ground truth pixel-wise motion blur kernels are compared to the predicted per-pixel kernels. This is the case when using synthetic blurry images, as described in Section 3.3. Given a ground truth per-pixel blur kernel $\{\mathbf{k}_{i,j}^{GT}\}$, the computed kernel basis $\{\mathbf{k}^b\}$ and mixing coefficients $\{m_{i,j}^b\}$, the *kernel loss* is defined as:

$$\mathcal{L}_{kernel} = \sum_i \sum_j w_{i,j} \left\| \sum_{b=1}^{B} m_{i,j}^b \mathbf{k}^b - \mathbf{k}_{i,j}^{GT} \right\|_p, \tag{4}$$

where the weights $w_{i,j}$ are defined as in the *reblur loss*.

3.3. Synthetic Dataset Generation

Our synthetically blurred dataset consists of 5,888 images from the ADE20K semantic segmentation dataset [54]. To generate random motion kernels, we use a camera-shake kernel generator [13, 9] based on physiological hand tremor data and pre-compute 100,000 kernels with an exposure

time of 1s. In our experiments, we observed that training on this synthetic data generalizes remarkably well to real photographs with different types of scenes and motion.

More specifically, for a random sharp image $\mathbf{u}$, we performed a convolution of the image by a random kernel $\mathbf{k}$. Additionally, each segmented part in the image annotation (if any), is independently convolved with a different random kernel. Finally, for each image we obtain a tuple $\left(\mathbf{u}^{GT}, \mathbf{v}^{GT}, \{\mathbf{k}\}^{GT}, \{\mathbf{m}\}^{GT}\right)$ containing the sharp image, blurry image and the pairs of ground truth kernels and masks applied to generate the blurry image.

In order to have a soft transition between different blurry regions, each segmentation mask was convolved with its corresponding kernel. To simplify, a maximum of three segmentation masks with a minimum size of 400 pixels are considered for each image. Also, to prevent a single kernel from dominating the losses (3) and (4), weights $w_{i,j}$ are computed as the inverse of the number of pixels which belong to the same segment. We refer to the Appendix for more details and examples from the dataset.

3.4. Model Training

We train the network using the sum of the *Reblur loss* (3) and *Kernel loss* (4) with equal weights. Training only with the *Reblur* term would make the problem more challenging. Adding the *Kernel loss* improves the convergence.

Training a model to predict a per-pixel kernel estimation is a difficult task. Moreover, in our case the model needs to figure out an image-specific low-rank decomposition in order to approximate all the kernels present in the image. This difficulty was noticeable in our experiments, where we observed very slow convergence and only started to see well-shaped kernels after around 200 epochs. In our experiments we observed that an L^2-norm on the kernels (4) was adequate to find a first approximation of the model. After 350 epochs, we switched to the more robust L^1-norm, which is harder to optimize but allows to recover sharper kernels. In total we trained our model for 900 epochs using image patches of 256×256 pixels. Additional details of the training procedure and hyper-parameters can be found in the Appendix.

3.5. Qualitative Results

Figure 3 shows some examples of non-uniform blur kernel estimation obtained by our method. We visually compare them with the results of two other existing deep learning-based non-uniform motion blur estimation methods, Gong *et al*. [15] and Sun *et al*. [44]. Despite being trained on synthetically blurred images, the method generalizes remarkably well to real blurry images (first two columns) as well as blurry images synthesized from video sequences as in GoPro [33] and REDs [32] datasets.

Our model is able to characterize different types of camera and objects motion, including rotations and zoom-in, and shows some degree of global reasoning of the scene when estimating motion in texture-less regions. Note also that the motion blur kernels estimated by the compared methods tend to correlate with the image structure instead of the underlying motion. Moreover, our model predicts continuous free-form motion kernels, whereas [44] and [15] are restricted to linear ones. Further qualitative estimation results are shown in Figure 6.

4. Applications

In this section we validate our estimated non-uniform motion blur kernels with two applications: non-blind image deblurring and blur detection.

4.1. Image Deblurring

Image deblurring methods can be broadly classified in two types: classical variational methods and deep learning methods. The former solve the deblurring inverse problem at the same time as estimating the motion blur kernel or forward model [12, 37, 6], and typically excel when the motion blur is uniform across the scene. Recently, deep learning based approaches have outperformed classical methods by directly estimating the transformation that maps blurry images to sharp images. One of the reasons for their success is the ability of neural networks to learn to both resolve the deconvolution as well as remove any artifacts of that process.

Early deep learning methods sought to simply minimize the L^2-norm between the sharp image and the output of the model. This might introduce a blurring effect due to the regression-to-the-mean problem [46, 50], motivating the use of Generative Adversarial Networks (GANs) to obtain more realistically looking restorations [25, 26]. However, GAN-based approaches introduce the potential pitfall of hallucinating image content [3]. To this purpose, another advantage of knowing the forward model is that it can be used to impose consistency between the restored and input blurry images [5].

Here we aim at leveraging the advantages of both data-driven and model-based approaches. Once the spatially-varying dense motion kernels have been estimated by our deep model, we can obtain a precise formulation of the inverse problem. This allows both a useful analysis of the scene, and a more controlled solution of the inverse problem, using variational methods for maximum a-posteriori (MAP) estimation, compared to the black-box one-to-one mapping of purely data-driven approaches.

4.1.1 Formulation

Resorting to classical non-blind deconvolution *maximum-a-posteriori* (MAP) estimation, we proceed as follows. Given

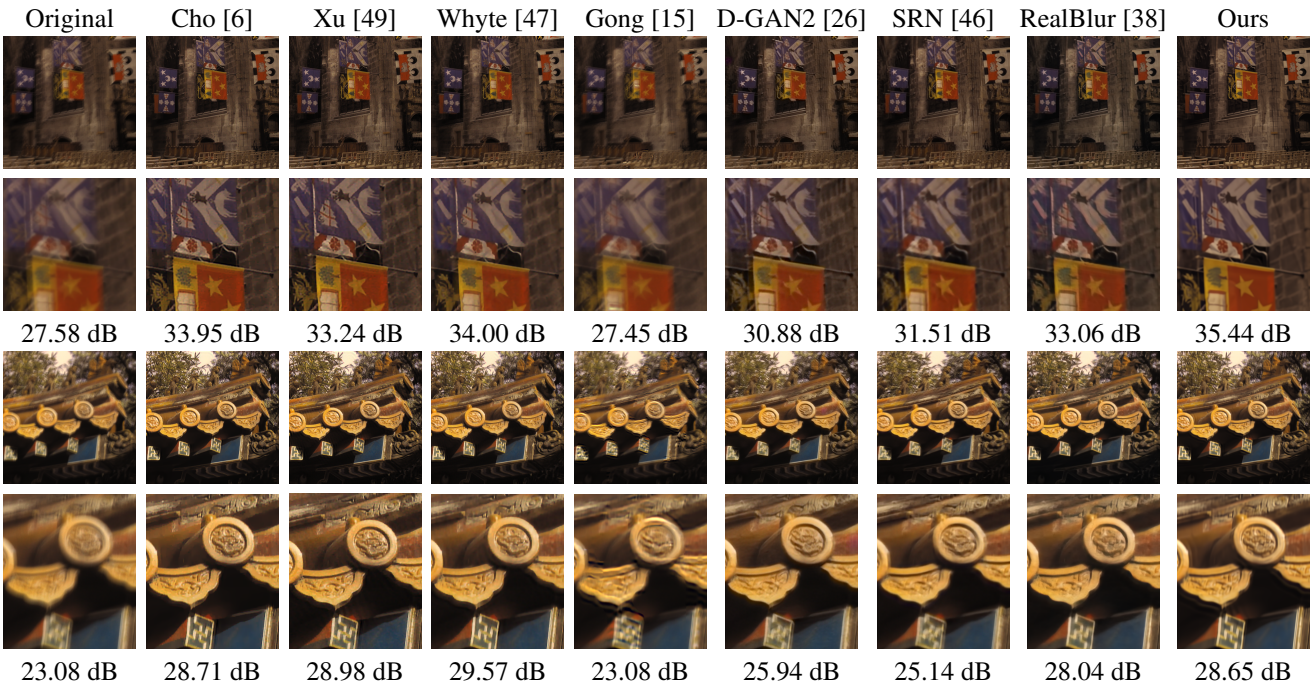

Figure 4: **Qualitative comparison of different deblurring methods in Köhler Dataset [24].** The complete, full resolution images are available in the Appendix.

an input blurry image $\mathbf{v}$, and the estimated kernel basis $\{\mathbf{k}^b\}$ and mixing coefficients $\{\mathbf{m}^b\}$, we search for the corresponding sharp image $\hat{\mathbf{u}}$ that minimizes the reblur loss $\mathcal{L}_{reblur}$ (3) and is not far away from the manifold of natural images. Note that $\mathcal{L}_{reblur}$ is a function of the image $\hat{\mathbf{u}}$, for fixed $\{\mathbf{k}^b\}$, $\{\mathbf{m}^b\}$ and $\mathbf{v}$, per Equation 2. As explained in Section 3, the blurring of the sharp image with the non-uniform blur field can be computed efficiently with B convolutions and one mixing operation.

We represent the manifold of natural images by means of a Gaussian denoising prior, as proposed in methods such as PnP [53, 52] and RED [39], and we use the denoiser proposed by [52]. More specifically, we look for a restored image $\hat{\mathbf{u}}$ which minimizes $\mathcal{L}_{reblur}$ and is a fixed point of the Gaussian denoiser with noise level σ^2, H_σ, i.e.

$$\hat{\mathbf{u}} = \underset{\mathbf{u}=H_\sigma(\mathbf{u})}{\arg\min}\ \mathcal{L}_{reblur} \tag{5}$$

To solve this problem we perform 30 iterations of a hybrid steepest descent method (HSD) [2, 8]. Following Zhang *et al*. [53, 52], we anneal the noise level σ^2 of the denoiser with an exponential decay rate from 49 to 7.65.

4.1.2 Comparison on Real Blurry Images

In this section we evaluate the ability of our deblurring procedure to generalize to real photographs containing real motion blur. To compare on this scenario, two standard datasets in the literature are used: Khöler [24] and Lai [27].

Quantitative results for Khöler dataset [24] are presented in Table 1 and qualitative results in Figure 4. Our method compares favorably to state-of-the-art end-to-end deep learning methods, that fail to generalize from the synthetic dataset they were trained on. Our method also outperforms the non-uniform motion blur estimation proposed by Sun *et al*. [44] and Gong *et al*. [15]. When compared to classic model-based methods (Table 2), our approach performs on par, although it suffers from the limitation in estimated kernel size, especially for blur kernels #8, #9, #10 and #11, which are bigger than our maximum support of 33×33. One solution to overcome this issue is to process the images at half resolution, however the size of the kernels for those images still falls outside of our hypothesis. Note also that the Köhler dataset consists of planar scenes since these are pictures of photographs; methods such as [47] are specifically designed to these conditions.

Lai [27] dataset is another standard benchmark that contains real blurry images with very non-uniform motion blur. The dataset has no corresponding ground truth, so it only allows for visual comparison, which we show in Figure 5 and in the Appendix. Note that our model-based method is competitive with state-of-the-art end-to-end approaches, outperforming most of them except for the very recent [38], which was specifically trained to restore saturated images.

Method	Dataset	Khöler	GoPro	DVD	RealBlurJ
DeblurGAN [25]	GoPro	26.05/0.75	27.92/0.84	28.27/0.84	(27.97/0.83)
GoPro K=2 [33]	GoPro	(26.02/0.81)	(29.23/0.92)	-	27.87/0.83
SRN [46]	GoPro	27.18/0.79	**30.72/ 0.91**	29.80/0.88	28.56/0.87
DMPHN 1-2-4 [50]	GoPro	25.69/0.75	29.98/0.90	28.28/0.84	27.80/0.85
DeblurGANv2 Inc. [26]	GoPro	27.25/0.79	29.49/0.88	29.55/0.93	28.69/0.87
DeblurGANv2 Mob. [26]	GoPro	25.88/0.74	27.40/0.83	28.70/0.85	28.09/0.84
RealBlurJ (SRN) [38]	RealBlurJ	(26.57/0.80)	(26.68/0.84)	-	(31.02/0.90)
RealBlurJ (SRN) [38]	GoPro, BSD, RealBlurJ	27.85/**0.81**	30.30/0.90	29.98/0.89	**31.38/0.91**
Sun *et. al* [44]	VOC2010	(25.22/0.77)[1]	(24.64/0.84)[2]	-	-
Gong [15]	BSD500	25.23/0.74	(26.06/0.86)[3]	-	-
Ours	ADE20K	**28.23/0.81**	27.51/0.84	28.23/0.85	27.86/0.82

Table 1: **Quantitative comparison for image deblurring (PSNR/SSIM).** When possible, we reproduced the results using their available code, otherwise parenthesis are used. [1] values extracted from [33]. [2] values extracted from [50]. [3] values extracted from [51]

Method / Dataset	Köhler	Köhler (except #8, #9, #10, #11)
Cho *et al*. [6]	28.98	31.09
Whyte *et al*. [47]	28.07	**31.71**
Xu *et al*. [49]	**29.53**	31.69
Ours	28.02	31.23

Table 2: **Quantitative comparison with classic methods over the Köhler dataset [24].** Unlike Table 1, no homography was used while computing the PSNR to compare to the originally reported values.

4.1.3 Comparison on Synthetic Blurry Images

State-of-the-art deblurring networks are typically trained with datasets that synthesize motion blur by averaging several short exposure frames. The GoPro dataset [33] is widely used both for training and benchmarking. The DVD dataset [43] reduces *ghosting* artifacts thanks to a proper alignment of frames and the generation of new intermediate frames before averaging. Recently, the carefully designed RealBlur dataset [38] was presented, containing low-light static scenes with lower illumination and more saturated regions. In [38], authors also proposed to train on synthetic images generated from the BSD dataset [30].

Following [38], we perform a quantitative evaluation of our method using the PSNR and SSIM metrics. To do so, the deblurred and sharp ground truth image are aligned using an homography estimated by the enhanced correlation coefficients method [10]. method [10]. The comparison is shown in Table 1. Our method falls behind when the comparison is done on the same dataset used for training, but performs comparably for the more challenging cross-dataset scenario. Note that end-to-end deep learning methods learn to both solve the deconvolution and remove remaining artifacts. Qualitative comparisons on representative examples of these datasets can be found in the Appendix.

4.1.4 Blurring to Deblur

Chen *et al*. [5] proposed to impose cycle-consistency to deblurring models using a forward model, learned from consecutive frames of a video. The motivation was to prevent a deblurring conditional GAN [25] from hallucinating image content. In the same spirit, and to validate our estimated kernels, we fine-tuned a pre-trained DeblurGAN [25] network, using our estimated kernels for imposing the forward-model consistency. Results shown in Table 3 prove that our fine-tuning is useful to improve a DeblurGAN model, slightly outperforming [5]. Also note that different to [5], our method works with single images instead of videos.

Network / Test Dataset	Gopro	
	PSNR	SSIM
DeblurGAN	27.25	0.81
DeblurGAN (resume training)	27.57	0.83
DeblurGAN + [5]	28.03[1]	0.90[1]
DeblurGAN + ours	28.06	0.85

Table 3: **Blurring to deblur.** Comparison between Reblur2Deblur [5] and the proposed method. The incorporation of a reblur loss in training produces better results than just resuming training. [1] No code available.

4.2. Blur Detection

Blur region detection aims at segmenting the blurred areas of a given image [1, 42, 14]. Recently, the use of synthetic datasets [1] for training blur detection networks has allowed deep learning methods surpass the performance of

Figure 5: **Deblurring examples on real blurry images from Lai dataset [27].** Our model-based approach is competitive with state-of-the-art data-driven methods. Best appreciated in electronic format.

CUHK [42]	HiFST [14]	Ma *et al*. [29]	Self-sup. [1]	Ours
0.6944	0.7484	(0.784)	(0.838)	0.8199

Table 4: **Comparison on the CUHK blur detection dataset [42], motion blur category.** Following [1], we report mean average precision across images of the evaluation split. Values in parenthesis are taken from [1].

methods based on local features [42, 14]. This is a straightforward application of our dense kernels estimation method. To build a segmentation mask, we group all the mixing coefficient images corresponding to kernels with an L^2-norm lower than a threshold (0.25 in this paper). Low L^2-norm indicates a more spread kernel. We evaluate on the standard CUHK blur detection dataset [42], under the motion blur category. Figure 6 shows that our approach can effectively segment regions of the image with motion blur. Quantitatively, following [1] we measure the mean average precision across the evaluation dataset, shown in Table 4. Despite not being trained for this task, our method is competitive with existing methods.

5. Conclusions

We revisited the problem of non-uniform kernel estimation and proposed a method to predict a dense map of kernels by decomposing it into a basis of image-specific kernel elements and corresponding per-pixel mixing coefficients.

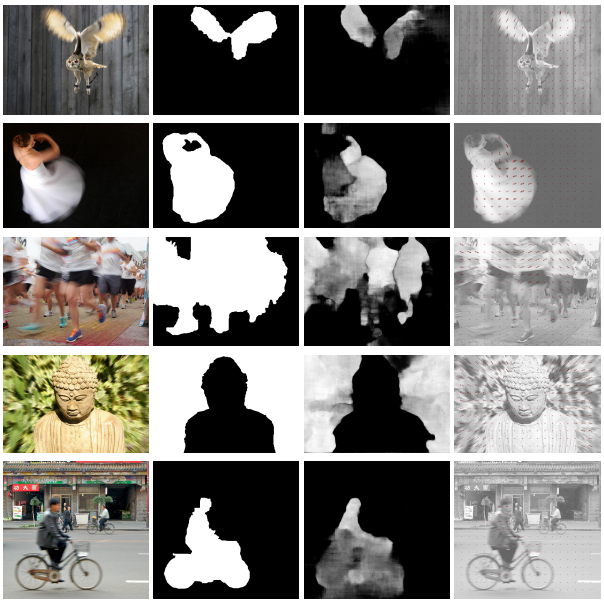

Figure 6: **Blur segmentation.** We use the norm of the predicted non-uniform motion blur kernels to detect regions with motion blur. Images from the CUHK blur detection dataset [42].

This results in a compact but non-parametric definition of the non-uniform motion field. Qualitative results validate

the estimated kernels and show that the model generalizes well to real blurry images with different types of relative camera motions, outperforming existing methods for non-uniform blur estimation. Additionally, we validated our model estimations in two applications: image deblurring and blur detection, and achieved results that are competitive with both deep learning-based methods and classical variational methods.

Acknowledgements

GC was supported partially by Agencia Nacional de Investigación e Innovación (ANII, Uruguay) grant POS_FCE_2018_1_1007783 and PV by the MICINN/FEDER UE project under Grant PGC2018-098625-B-I0; H2020-MSCA-RISE-2017 under Grant 777826 NoMADS and Spanish Ministry of Economy and Competitiveness under the Maria de Maeztu Units of Excellence Programme (MDM-2015-0502). The experiments presented in this paper were carried out using ClusterUY (site: https://cluster.uy) and GPUs donated by NVIDIA Corporation. We also thanks Juan F. Montesinos for his help during the experimental phase.

References

[1] Aitor Alvarez-Gila, Adrian Galdran, Estibaliz Garrote, and Joost Van de Weijer. Self-supervised blur detection from synthetically blurred scenes. *Image and Vision Computing*, 92:103804, 2019.

[2] Heinz H Bauschke, Regina S Burachik, Patrick L Combettes, Veit Elser, D Russell Luke, and Henry Wolkowicz. *Fixed-point algorithms for inverse problems in science and engineering*, volume 49. Springer Science & Business Media, 2011.

[3] Y. Blau and T. Michaeli. The perception-distortion tradeoff. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6228–6237, 2018.

[4] E.A. Burton, J. Wattam-Bell, G.S. Rubin, J. Atkinson, O. Braddick, and M. Nardini. The effect of blur on cortical responses to global form and motion. *Journal of vision.*, 15(15):1–14, November 2015.

[5] H. Chen, J. Gu, O. Gallo, M. Liu, A. Veeraraghavan, and J. Kautz. Reblur2deblur: Deblurring videos via self-supervised learning. In *2018 IEEE International Conference on Computational Photography (ICCP)*, pages 1–9, Los Alamitos, CA, USA, may 2018. IEEE Computer Society.

[6] Sunghyun Cho and Seungyong Lee. Fast motion deblurring. In *ACM SIGGRAPH Asia 2009 papers*, pages 1–8. 2009.

[7] Sunghyun Cho, Jue Wang, and Seungyong Lee. Handling outliers in non-blind image deconvolution. In *2011 International Conference on Computer Vision*, pages 495–502. IEEE, 2011.

[8] Regev Cohen, Michael Elad, and Peyman Milanfar. Regularization by denoising via fixed-point projection (red-pro). *arXiv preprint arXiv:2008.00226*, 2020.

[9] Mauricio Delbracio and Guillermo Sapiro. Removing camera shake via weighted fourier burst accumulation. *IEEE Transactions on Image Processing*, 24(11):3293–3307, 2015.

[10] Georgios D Evangelidis and Emmanouil Z Psarakis. Parametric image alignment using enhanced correlation coefficient maximization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1858–1865, 2008.

[11] Paolo Favaro and Stefano Soatto. A variational approach to scene reconstruction and image segmentation from motion-blur cues. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, pages I–I. IEEE, 2004.

[12] Rob Fergus, Barun Singh, Aaron Hertzmann, Sam T Roweis, and William T Freeman. Removing camera shake from a single photograph. In *ACM SIGGRAPH 2006 Papers*, pages 787–794. 2006.

[13] Fabien Gavant, Laurent Alacoque, Antoine Dupret, and Dominique David. A physiological camera shake model for image stabilization systems. In *SENSORS, 2011 IEEE*, pages 1461–1464, 2014.

[14] S Alireza Golestaneh and Lina J Karam. Spatially-varying blur detection based on multiscale fused and sorted transform coefficients of gradient magnitudes. In *CVPR*, pages 596–605, 2017.

[15] Dong Gong, Jie Yang, Lingqiao Liu, Yanning Zhang, Ian Reid, Chunhua Shen, Anton Van Den Hengel, and Qinfeng Shi. From motion blur to motion flow: a deep learning solution for removing heterogeneous motion blur. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2319–2328, 2017.

[16] Ankit Gupta, Neel Joshi, C. Lawrence Zitnick, Michael Cohen, and Brian Curless. Single image deblurring using motion density functions. In *Proceedings of the 11th European Conference on Computer Vision: Part I*, ECCV'10, page 171–184, Berlin, Heidelberg, 2010. Springer-Verlag.

[17] Thomas L. Harrington and Marcia K. Harrington. Perception of motion using blur pattern information in the moderate and high-velocity domains of vision. *Acta Psychologica*, 48(1):227 – 237, 1981.

[18] M. Hirsch, C. J. Schuler, S. Harmeling, and B. Schölkopf. Fast removal of non-uniform camera shake. In *2011 International Conference on Computer Vision*, pages 463–470, 2011.

[19] Armin Hornung, Maren Bennewitz, and Hauke Strasdat. Efficient vision-based navigation: Learning about the influence of motion blur. *Autonomous Robots*, 2010.

[20] Tae Hyun Kim and Kyoung Mu Lee. Generalized video deblurring for dynamic scenes. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5426–5434, 2015.

[21] Xu Jia, Bert De Brabandere, Tinne Tuytelaars, and Luc V Gool. Dynamic filter networks. In *Advances in neural information processing systems*, pages 667–675, 2016.

[22] T. H. Kim, B. Ahn, and K. M. Lee. Dynamic scene deblurring. In *2013 IEEE International Conference on Computer Vision*, pages 3160–3167, 2013.

[23] T. H. Kim and K. M. Lee. Segmentation-free dynamic scene deblurring. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2766–2773, 2014.

[24] Rolf Köhler, Michael Hirsch, Betty Mohler, Bernhard Schölkopf, and Stefan Harmeling. Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In *European conference on computer vision*, pages 27–40. Springer, 2012.

[25] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8183–8192, 2018.

[26] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 8878–8887, 2019.

[27] Wei-Sheng Lai, Jia-Bin Huang, Zhe Hu, Narendra Ahuja, and Ming-Hsuan Yang. A comparative study for single image blind deblurring. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recog.*, pages 1701–1709, 2016.

[28] Jun Ma and S. I. Olsen. Depth from zooming. *J. Opt. Soc. Am. A*, 7(10):1883–1890, Oct 1990.

[29] Kede Ma, Huan Fu, Tongliang Liu, Zhou Wang, and Dacheng Tao. Deep blur mapping: Exploiting high-level semantics by deep neural networks. *IEEE Transactions on Image Processing*, 27(10):5155–5166, 2018.

[30] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423. IEEE, 2001.

[31] Ben Mildenhall, Jonathan T Barron, Jiawen Chen, Dillon Sharlet, Ren Ng, and Robert Carroll. Burst denoising with kernel prediction networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2502–2510, 2018.

[32] Seungjun Nah, Sungyong Baik, Seokil Hong, Gyeongsik Moon, Sanghyun Son, Radu Timofte, and Kyoung Mu Lee. Ntire 2019 challenge on video deblurring and super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.

[33] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene deblurring. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[34] Simon Niklaus, Long Mai, and Feng Liu. Video frame interpolation via adaptive convolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 670–679, 2017.

[35] Simon Niklaus, Long Mai, and Feng Liu. Video frame interpolation via adaptive separable convolution. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 261–270, 2017.

[36] Jin-shan Pan, Zhe Hu, Zhixun Su, Hsin-Ying Lee, and Ming-Hsuan Yang. Soft-segmentation guided object motion deblurring. In *2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016*, pages 459–468. IEEE Computer Society, 2016.

[37] Jinshan Pan, Zhe Hu, Zhixun Su, and Ming-Hsuan Yang. Deblurring text images via l0-regularized intensity and gradient prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2901–2908, 2014.

[38] Jaesung Rim, Haeyun Lee, Jucheol Won, and Sunghyun Cho. Real-world blur dataset for learning and benchmarking deblurring algorithms. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020.

[39] Yaniv Romano, Michael Elad, and Peyman Milanfar. The little engine that could: Regularization by denoising (red). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844, 2017.

[40] Constance S. Royden and Kathleen D. Moore. Use of speed cues in the detection of moving objects by moving observers. *Vision Research*, 59:17 – 24, 2012.

[41] Shengyang Dai and Ying Wu. Motion from blur. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[42] Jianping Shi, Li Xu, and Jiaya Jia. Discriminative blur detection features. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2965–2972, 2014.

[43] Shuochen Su, Mauricio Delbracio, Jue Wang, Guillermo Sapiro, Wolfgang Heidrich, and Oliver Wang. Deep video deblurring for hand-held cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1279–1288, 2017.

[44] Jian Sun, Wenfei Cao, Zongben Xu, and Jean Ponce. Learning a convolutional neural network for non-uniform motion blur removal. In *Proceedings of the IEEE Conference on Comput.er Vis.*, 2015.

[45] Y. Tai, P. Tan, and M. S. Brown. Richardson-lucy deblurring for scenes under a projective motion path. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(8):1603–1618, 2011.

[46] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiaya Jia. Scale-recurrent network for deep image deblurring. In *Proceedings of the IEEE Conference on Comput.er Vis.ion and Pattern Recog. Worksh.*, 2018.

[47] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce. Non-uniform deblurring for shaken images. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 491–498, 2010.

[48] Zhihao Xia, Federico Perazzi, Michaël Gharbi, Kalyan Sunkavalli, and Ayan Chakrabarti. Basis prediction networks for effective burst denoising with large kernels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11844–11853, 2020.

[49] Li Xu and Jiaya Jia. Two-phase kernel estimation for robust motion deblurring. In *European conference on computer vision*, pages 157–170. Springer, 2010.

[50] Hongguang Zhang, Yuchao Dai, Hongdong Li, and Piotr Koniusz. Deep stacked hierarchical multi-patch network for im-

age deblurring. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 200319.

[51] Jiawei Zhang, Jinshan Pan, Jimmy Ren, Yibing Song, Linchao Bao, Rynson WH Lau, and Ming-Hsuan Yang. Dynamic scene deblurring using spatially variant recurrent neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2521–2529, 2018.

[52] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *arXiv preprint arXiv:2008.13751*, 2020.

[53] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3929–3938, 2017.

[54] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[55] Daniel Zoran and Yair Weiss. From learning models of natural image patches to whole image restoration. In *2011 International Conference on Computer Vision*, pages 479–486. IEEE, 2011.