

Restoration of Fast Moving Objects

Jan Kotera¹, Member, IEEE, Jiří Matas², and Filip Šroubek¹, Member, IEEE

Abstract—If an object is photographed at motion in front of a static background, the object will be blurred while the background sharp and partially occluded by the object. The goal is to recover the object appearance from such blurred image. We adopt the image formation model for fast moving objects and consider objects undergoing 2D translation and rotation. For this scenario we formulate the estimation of the object shape, appearance, and motion from a single image and known background as a constrained optimization problem with appropriate regularization terms. Both similarities and differences with blind deconvolution are discussed with the latter caused mainly by the coupling of the object appearance and shape in the acquisition model. Necessary conditions for solution uniqueness are derived and a numerical solution based on the alternating direction method of multipliers is presented. The proposed method is evaluated on a new dataset.

Index Terms—Blind deconvolution, motion deblurring, object deblurring, motion estimation, image matting, shape estimation, alternating direction method of multipliers.

I. INTRODUCTION

MANY applications produce images in which moving objects appear blurred and partially blended with the background. Sports videos provide diverse examples with various ball-like objects flying at high speed; surveillance cameras record e.g. fast-driving cars; for other examples, see Fig. 1. To invert the image formation process and recover the sharp object appearance, we must consider the role of the background, object trajectory, appearance, and shape. One particular property of the problem is that the area where foreground and background are both visible and blended together is proportional to the ratio of trajectory length and object size, for large blurs or small objects the image then cannot be hard-segmented as foreground/background and treated separately.

The notion of the *fast moving object* (FMO) was introduced in [1] in the context of object tracking. FMO was defined as an object that moves over a distance exceeding its size during the camera exposure time. The authors of [1] define the image



Fig. 1. Examples of objects that appear blurred due to their intra-frame motion: (left-to-right, top-to-bottom) car, motorcycle, model airplane, skier, volleyball, ping-pong ball.

formation model and present an algorithm for the inverse problem of estimating the object appearance from a single image if the background image, and object shape and trajectory are known. The inverse problem was later generalized by another tracking algorithm [2] that simultaneously estimates the object trajectory, shape, and appearance. The estimation task is called *deblatting* to emphasize that two inverse problems are being solved: motion *deblurring* and image *matting*. FMO tracking methods are based on the novel idea that blur is a cue for motion, and they focus on estimating the blur kernel, which defines the trajectory, and not the object appearance, as the tracked object is assumed to be known in standard tracking methodologies, e.g. it is specified by the user in one video frame as the region of interest.

In this work, we study the deblatting inverse problem from the perspective of robust estimation of the latent object shape (a binary mask) and appearance (a color image) without any knowledge of the object blur. Similarities to blind deconvolution (BD) are evident, yet the task differs in several important ways. The target scenario of most BD methods is a static scene blurred by camera motion or incorrect focus. In the deblatting problem, the object image and background are blended together as the background is partially occluded by the moving object. More importantly, the unknown object is not represented by a standard rectangular image, as it is in BD, but can have arbitrary shape which must be estimated from the blurred input. In its full generality, the deblatting problem deals with 3D object motion (including rotation), which cannot be modeled by convolution. All these differences make the deblatting problem a challenging endeavor as well as a separate research topic rather than an application of BD.

The 3D case requires full 3D modeling of the object and the camera projection onto the image plane, which is beyond the scope of this work. Instead, we consider 2D deblatting and extend our preliminary results in [3] by fully blind shape estimation. We assume 2D objects undergoing arbitrary 2D motion and rotation. The 2D case is general enough to model some interesting 3D cases as well; in particular, 3D objects

Manuscript received July 19, 2019; revised May 18, 2020 and July 10, 2020; accepted August 4, 2020. Date of publication August 19, 2020; date of current version August 31, 2020. The work of Jan Kotera and Jiří Matas was supported by the Czech Science Foundation under Grant GA18-05360S. The work of Filip Šroubek was supported in part by the Czech Science Foundation under Grant GA18-05360S and in part by the Praemium Academiae. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Zhen He. (*Corresponding author: Jan Kotera.*)

Jan Kotera and Filip Šroubek are with the Institute of Information Theory and Automation, Czech Academy of Sciences, 18208 Praha, Czech Republic (e-mail: honzakotera@gmail.com).

Jiří Matas is with the Department of Cybernetics, Czech Technical University, 121 35 Prague, Czech Republic.

This article has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the authors.

Digital Object Identifier 10.1109/TIP.2020.3016490

moving in a plane perpendicular to the camera optical axis with imperceptible rotation or with in-plane rotation. We also make an assumption that the background is known, which is not restrictive if we use video sequences. The FMO tracking methods [1], [2] show that the 2D approximation of many real scenarios is locally valid and that the background image can be inferred from a video sequence with sufficient accuracy.

The contributions of this work are:

- 1) We adopt the image formation model for moving objects and propose a solution of the corresponding blind inverse problem of deblatting restricted to 2D motion. From a single image and known background we estimate the object shape and appearance as well as its motion blur including rotation. The proposed method is quantitatively evaluated on a new dataset of real images acquired by a highspeed camera.
- 2) We identify necessary conditions for solution uniqueness, propose novel regularization terms, and discuss similarities and differences with BD in terms of problem solvability.

Three observations follow from the necessary conditions for solution uniqueness that we demonstrate on examples: (i) deblatting for color images is similar to multiframe BD and is thus better posed than the same problem for grayscale images, (ii) small blur can be detrimental to successful restoration, which is in a stark contrast to BD, and (iii) uniform background causes solution ambiguity.

In the rest of the paper, we discuss related work in detail, formulate the problem and propose a solution, analyze and discuss necessary conditions for uniqueness, and finally evaluate and compare the proposed solution.

II. RELATED WORK

The problem of BD has received considerable attention by the image processing and computer vision community; see surveys in [4], [5]. Most methods consider the space-invariant case where one unknown blur degrades the whole image. The key idea of robust algorithms is to tackle the ill-posedness of BD by Variational Bayesian inference [6], [7] and marginalization [8]. Coordinate descent methods searching for the classical maximum a posteriori were repeatedly proposed in parallel [9]–[12] and if augmented by various edge-sharpening steps they tend to converge to the correct solution. Later, unnaturally sparse priors were applied to images and robustness of BD algorithms was improved even more [13]–[17]. BD methods were extended in [18]–[20] to cases where the convolution model is partially violated, e.g. in saturated regions. Learning-based methods [21] indicate that convolutional neural networks (CNN) can be trained to remove unknown blurs directly from the blurred image. However, sufficiently general learning-based methods for arbitrary blurs and images are still missing.

More complex problems arise in the case of space-variant BD, when the convolution kernel is a function of position. Conventional approaches are based on the assumption that space-variant blur can be efficiently parameterized. For example, blur caused by camera motion is limited to six degrees of freedom of rigid body rotation and translation [17], [22], and

the blurring operator is expressed as a linear combination of a few basis vectors [23]. The space-variant case is currently dominated by CNN approaches, e.g. [24]–[27]. Even more general scenario is the so-called dynamic-scene deblurring, where the degradation is caused by camera motion as well as motion of the individual objects in the photographed scene. This problem was initially addressed by an extension of classical deblurring methods [28]–[30] but more recently using end-to-end CNNs prevails [26], [31], [32].

Another studied scenario is the multiframe (sometimes called multichannel) BD [33]–[35], which assumes multiple images of the same scene each degraded by an unknown yet different convolution kernel. Unlike in the single-frame BD case, multiple observations provide the necessary constraint to make the image restoration task well posed as was proved independently in [36] and [37]. The main drawback is that in many practical scenarios multiple observations of the same scene are not available and even if they are, the images must be registered first, which is difficult to achieve especially for blurred data. The problem of misregistration was partially alleviated in [38] by introducing robustness to translation.

Methods specifically designed for motion deblurring exist yet they generally assume that the degradation is modeled locally by a linear motion (straight line). One category of such methods exploit the foreground-background blending and work with the object blurred transparency map (alpha matte) [39]. Blind deconvolution of the transparency map is better posed, since the latent sharp map is a binary image. The same idea applied to deblurring of moving objects was proposed in [40], but the method is limited to pure rotation and requires considerable user interaction. Another variation was proposed in [41], where linear motion blur is estimated locally using a relation similar to optical flow. The main drawback of these methods is that an accurate estimation of the transparency map using alpha matting algorithms such as [42] is necessary and this is difficult to achieve, especially without user input. A similar background-blending model but applied to images with out-of-focus foreground was recently proposed in [43]. Several methods specifically targeting restoration of motion-blurred objects were proposed in [28]–[30], but these methods do not fully model the foreground-background blending and assume a binary transition between the (blurred) object and the background, they are therefore applicable only for small blurs. Some methods for motion deblurring exploit the fact that autocorrelation increases in the direction of blur. This was for example applied to deblurring objects moving over a static background in [29] or for detecting camera motion from the blurred image in [44]. Methods based on autocorrelation require a relatively large neighborhood to estimate blur parameters, which means that they are not suitable for small moving objects or if the background-blending is significant due to large object motion.

III. PROBLEM FORMULATION

The image formation model for 2D deblatting assumes an object moving in a plane perpendicular to the camera optical axis and over a static background b . The object in

the formation model is defined by its projection onto the 2D image plane with a pair of variables (f, m) , where f is the object 2D *appearance* and m is the object opacity channel – real-valued *mask* $m \in (0, 1)$. The appearance f is the image of the object placed over a totally black (zero) background and it is assumed to be fixed in the given image or video frame. The opacity m allows us to correctly model pixels on object boundaries (so called mixed pixels), semi-transparent objects and even object shadow if it is modeled as a semi-transparent shape cast by the object itself. For example, if the object is an opaque sphere then f is a 2D disk and m has ones inside, zeros outside and intermediate values on the disk boundary. If the object is semi-transparent then m has values less than one inside. For the purpose of initial formulation we will consider grayscale images and translational object motion. Color images and rotation will be discussed later. The acquired image or video frame g is then expressed using convolution as

$$g = (1 - h * m)b + h * f, \quad (1)$$

where h is the object motion trace (projection of the object trajectory on the image plane). In this rotation-free case, the trace is equivalent to the convolution kernel of the object blur. We refer to the inverse problem of estimating h , f , and m knowing only g and b as the *deblatting problem*. A visual interpretation of the formation model is given in Fig. 2. Notice that there is no restriction on the shape of h and therefore the model allows complex trajectories, such as parabolic curves due to gravity and non-smooth curves due to bounces.

Let us rewrite the above equation in a more general vector-matrix notation. The convolution $h * f$ is replaced by a multiplication of f by a general blur operator (matrix) \mathcal{H} determined by h , which performs the considered blur – in our case caused by translational motion and 2D rotation. We will occasionally use the notation $h * f = Hf$ to denote pure motion blur without rotation, in this case H denotes convolution matrix. The roles of f and h are interchangeable and the blur operation can be equivalently written as $\mathcal{H}f \equiv \mathcal{F}h$ where the operator \mathcal{F} is determined by f and performs blurring of (a fixed) f according to the motion parameters h . The parametrization of rotation blur h and its implementation is described in the next subsection. For brevity, we also introduce a background-subtraction image $d \equiv g - b$ and $\bar{f} \equiv [f^T, m^T]^T$, which is the object appearance and opacity combined into one column vector. The model (1) then rewrites as

$$d = (\mathcal{F} - B\mathcal{M})h = \mathcal{H}f - B\mathcal{H}m = A\bar{f}, \quad (2)$$

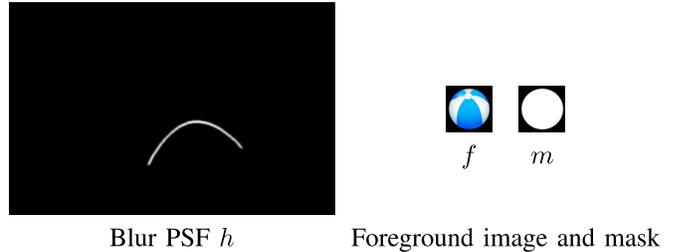
where \mathcal{M} is a blur operator analogous to \mathcal{F} but corresponding to m , B is a diagonal matrix with b on the main diagonal (pixel-wise multiplication with b), and $A \equiv [\mathcal{H}, -B\mathcal{H}]$.

Deblatting is an ill-posed problem requiring regularization and feasibility constraints. We use total variation for the object appearance, which enforces sparsity on image gradients $\nabla f = [\nabla_x^T, \nabla_y^T]^T f$. The blur h is of the same size as the input g and is nonzero only along the object motion trace, which also suggests sparsity promoting regularization. We assume that g is not saturated and clipped. The variables h and f and m are non-negative and f is bounded from above by the maximum

Input to deblatting:

Input image g Background image b

Output of deblatting:

Blur PSF h

Foreground image and mask

Image formation model:

The term $h * f$ in eq. (1)The term $(1 - h * m)b$ in (1)

Fig. 2. Illustration of the acquisition model of the deblatting problem. **Top:** Input to deblatting – image with blurred foreground object and separate background image. **Middle:** Output of deblatting – motion blur (displayed as simple motion PSF without rotation for simplicity), sharp foreground object, and shape mask. **Bottom:** Components of the image formation model (1) – blurred foreground and partially occluded background. All images are in proper relative scale. Notice that the blur has the same size as the input image and effectively encodes position of the object in the input. The foreground object and shape mask are smaller than the input. Source images [45].

image intensity which we assume to be 1. In addition f and m are coupled, since m is the image opacity and f is the image intensity of the object if perceived over a zero background. Above constraints are combined in our proposed set of admissible values for f and m as

$$S = \{(f, m) : m \in (0, 1), 0 \leq f \leq m\} \quad (3)$$

and illustrated in Fig. 3a. Note that the set is convex, which is important. The deblatting problem formulated as a constrained minimization then becomes

$$\begin{aligned} \min_{h, f, m} & \frac{1}{2} \|(\mathcal{F} - B\mathcal{M})h - d\|_2^2 + \alpha_h \|h\|_1 + \alpha_f \|\nabla f\|_1 \\ \text{s.t.} & h \geq 0, (f, m) \in S. \end{aligned} \quad (4)$$

This optimization problem is efficiently solved using the ADMM algorithm [46] as explained in the next section.

The variables can represent both grayscale and color images. In the case of color images, f consists of three concatenated color channels $[f_1^T, f_2^T, f_3^T]^T$ and so does b and g , while m and h are single-channel even in the color case.

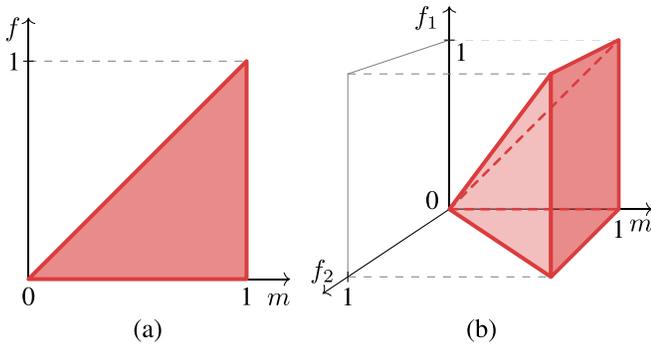


Fig. 3. The set S from eq. (3) of admissible values of pixel intensity f and opacity m . (a) Single-channel grayscale case and (b) two-channel color case. In the three-channel (RGB) color case, the set is an oblique polyhedral pyramid in \mathbb{R}^4 .

The combined variable \bar{f} will have the form $[f_1^T, f_2^T, f_3^T, m^T]^T$. All operators in (4) must be modified accordingly to operate on color images. The admissible set S for the color case is an oblique polyhedral pyramid in \mathbb{R}^4 , which can not be visualized, yet Fig. 3b shows the reduced two-channel case.

A. Extension to 2D Rotation

The acquisition model is extendable to 2D in-plane rotation without introducing substantial conceptual changes. In this section we describe how the motion and rotation blur is parametrized and modeled so that the corresponding blur operator remains linear.

Pure motion blur is traditionally modeled using convolution, which is essentially a weighted superposition of finite number of possible displacements of the sharp image f . In other words, motion blur can be written as $h * f = \sum_i h_i T^i f$ where T^i is a (fixed) translation operator that shifts f to position with offset i . The weights h_i correspond to the intensities of the convolution kernel (PSF) h and are estimated in the blur estimation phase, whereas the family of admissible translations $\{T^i\}$ is fixed and, in the traditional convolutional representation, determined by the dimensions of the (PSF) including its natural discretization – the admissible displacements of the object are constrained to the pixel grid of the image. The model is therefore linear in h_i .

We adopt the same strategy for representing motion-and-rotation blur. The space of admissible rotations is discretized into finite number of rotation angles, indexed by j , and we construct a (fixed) family of corresponding rotation operators $\{R^j\}$. The general blur operator is then decomposed as a weighted combination of rotated and shifted copies of the sharp image f as

$$\mathcal{H}f = \sum_{ij} h_{ij} T^i R^j f. \quad (5)$$

Motion and rotation blur of f is thus equivalent to pure motion blur of pre-rotated f 's. The weights h_{ij} correspond to intensities of the generalized blur PSF h and are subject to estimation whereas the set of rotations angles is fixed and determined beforehand – in this case the discretization of the admissible rotations is less natural and must be determined ad-hoc. The extended model therefore remains linear in h_{ij}

and from mathematical perspective is equivalent, the only difference are its numerical properties, increased dimensionality, and in the implementation, because \mathcal{H} is no longer a block circulant matrix and cannot be evaluated as efficiently.

In contrast, extension to 3D rotation is a significant conceptual change that would require modeling of f and m in 3D and introducing their projection onto the image plane, which is beyond the scope of this article.

IV. PROBLEM SOLUTION

Minimization of (4) is done in a coordinate-descent manner by alternating between two minimization steps: h -step and fm -step (f and m are estimated simultaneously as the combined variable \bar{f}). ADMM allows splitting each of these optimization problems into several easier-to-solve subproblems by introducing substitutions by auxiliary variables and Lagrange multipliers for the corresponding constraints. In the case of minimization w.r.t. h (h -step), there is an auxiliary variable $v = h$ and the corresponding Lagrange multiplier a . For the minimization w.r.t. \bar{f} (fm -step), we have two auxiliary variables $w = \nabla f$ and $z = \bar{f}$ with Lagrange multipliers c and e , respectively. In both steps we then minimize an augmented cost function:

1) h -step:

$$\min_{h,v} \frac{1}{2} \|(\mathcal{F} - B\mathcal{M})h - d\|_2^2 + \alpha_h \|v\|_1 + \frac{\beta_h}{2} \|h - v - a\|_2^2 \quad (6)$$

s.t. $h \leq 0$.

2) fm -step:

$$\min_{\bar{f}, w, z} \frac{1}{2} \|A\bar{f} - d\|_2^2 + \alpha_f \|w\|_1 + \frac{\beta_f}{2} \|\nabla f - w - c\|_2^2 + \phi(z) + \frac{\beta_{\bar{f}}}{2} \|\bar{f} - z - e\|_2^2. \quad (7)$$

Minimization in the h -step consists of three update equations:

$$h \leftarrow [(\mathcal{F} - B\mathcal{M})^T (\mathcal{F} - B\mathcal{M}) + \beta_h I]^{-1} [(\mathcal{F} - B\mathcal{M})^T d + \beta_h (v + a)] \quad (8)$$

$$v \leftarrow \max([h - a] - \frac{\alpha_h}{\beta_h}, 0) \quad (9)$$

$$a \leftarrow a - h + v \quad (10)$$

The first update equation (8) is a system of linear equations that we solve with the conjugate gradient (CG) method. The remaining two update equations are solved element-wise, where (9) is the proximal mapping of ℓ_1 norm and (10) updates the Lagrange multipliers.

Minimization in the fm -step introduces $\phi(\cdot)$, which is a function enforcing the constraint $(f, m) \in S$ where S is the set of admissible values of \bar{f} in (3). It is defined as

$$\phi(z) = \begin{cases} 0 & z \in S, \\ \infty & \text{otherwise.} \end{cases} \quad (11)$$

Since the constraint is defined pixel-wise, minimization w.r.t. z amounts to projection of a point onto the set S . We use a fast version of Dykstra's algorithm [47] to implement

the projection, as S is an intersection of simple convex sets (half-spaces).

Minimization in the f -step then consists of five update equations:

$$\bar{f} \leftarrow \left(A^T A + \begin{bmatrix} \beta_{\bar{f}} I + \beta_f \nabla^T \nabla & 0 \\ 0 & \beta_{\bar{f}} I \end{bmatrix} \right)^{-1} \times \left(A^T d + \beta_f \begin{bmatrix} \nabla^T \\ 0 \end{bmatrix} (w + c) + \beta_{\bar{f}} (z + e) \right) \quad (12)$$

$$w \leftarrow \frac{\nabla f - c}{|\nabla f - c|} \max(|\nabla f - c| - \frac{\alpha_f}{\beta_f}, 0) \quad (13)$$

$$z \leftarrow \text{proj}_S(\bar{f} - e) \quad (14)$$

$$c \leftarrow c - \nabla f + w \quad (15)$$

$$e \leftarrow e - \bar{f} + z \quad (16)$$

The first update equation (12) is a system of linear equation that we again solve with the CG method. The second and third update equations are for auxiliary variables and they are solved element-wise; (13) is the proximal mapping of ℓ_1 norm and (14) is the proximal mapping of $\phi(\cdot)$, which is the projection onto S . The last two equations update the corresponding Lagrange multipliers.

The algorithm contains several parameters related to regularization and feasibility constraints. We chose their values by a search using a small set of test images and for all experiments we use the following values: $\alpha_f = 10^{-3}$, $\alpha_h = 1$, $\beta_f = 10^{-2}$, $\beta_h = 10^3$, $\beta_{\bar{f}} = 10^{-3}$.

V. UNIQUENESS OF SOLUTION

Blind image deblurring (BD) is a long-studied problem in image processing. Due to the image-blur coupling in the image formation model, it is non-convex with a generally non-unique solution. Deblatting is naturally similar to BD, though with some interesting differences. In this section we discuss and state some theoretical properties of the deblatting problem related to solution uniqueness and we frequently compare deblatting to the BD problem to highlight both similarities and differences.

In our analysis we focus on the data-fidelity term of (4) alone, as it corresponds to the physical properties of the problem while the regularizers and constraints are added mainly to improve its numerical properties. We consider the formation model (1) in the discrete space and every variable has a bounded domain denoted by dom and support denoted by supp , e.g. $\text{supp } f \subseteq \text{dom } f$. We say that an image f is *compactly supported* within its domain if increasing its domain results in simple zero padding (i.e. the support can be overestimated). We say that a blur h is *irreducible* if factorization is possible only with a delta function δ , i.e. $h = s * h'$ only for $s \propto \delta$.

Note that the domain of variables is either determined by the size of input data, e.g. in the case of the acquired image g , or set by a user, e.g. in the case of the object appearance f as it is not known in advance. For simplicity, we index pixels in the image domain by a single index, e.g. the j -th pixel of f is $f(j)$, and $|\text{dom } f|$ denotes the number of pixels in the domain.

Further, we make the following assumptions.

- (a1) g , b , and h have the same bounded rectangular domain $\text{dom } g$, f and h are compactly supported with $|\text{dom } f| \ll |\text{dom } h|$ and $\text{supp } (h * f) \subset \text{dom } g$, and all variables satisfy $0 \leq g, b, h, f, m \leq 1$, $f \leq m$,
- (a2) h preserves energy, i.e. $\sum_{j=1}^{|\text{dom } h|} h(j) = 1$,
- (a3) h , f , and m are irreducible.

A tuple $(\hat{f}, \hat{m}, \hat{h})$, also shortly (\hat{f}, \hat{h}) , is called a solution for the given b and g if it satisfies (1). The first assumption (a1) sets realistic domains for all variables. The second assumption (a2) eliminates trivial equivalent solutions of the form $(\alpha \hat{f}, (1/\alpha) \hat{h})$. The third assumption (a3) eliminates inherent ambiguity in the solution as it would be impossible to determine whether the factor s belongs to h or \bar{f} .

A. No-Blur Solutions

We start the discussion assuming only grayscale images and then conclude with color images. First notice that if the background is zero (black), deblatting reduces to BD

$$g = h * f \quad (17)$$

as follows from (1) after substituting $b = 0$. In the standard BD problem, f is the unknown sharp image with the domain similar to the input g and its support is not compact, since g is typically a cropped observation of some larger scene. The blur h corresponding to, e.g. camera motion or out-of-focus blur, is assumed to have compact support of size smaller than $\text{dom } f$. In the deblatting problem the roles of f and h are somewhat swapped. Image f is a moving object, which is compactly supported (a1) and covers only part of the input $\text{supp } f \subset \text{dom } g$. The same holds for m . The blur h encodes the motion as well as position of the object within the input frame, its domain is of the same size as the input $\text{dom } h = \text{dom } g$, yet also compactly supported according to (a1), i.e. the object trajectory is not cropped.

The standard BD is prone to spurious solutions, of which the most notable is the so-called *no-blur* solution [15], where the estimated sharp image is the input blurred image and the estimated blur is a delta function. Note that the no-blur solution satisfies (17) exactly. In deblatting the situation is similar, yet due to the compact support of both f and h , we have two no-blur solutions summarized in the following proposition.

Proposition 1: The deblatting formation model (1) has two unique no-blur solutions: $(\hat{f}, \hat{m}, \hat{h}) =$

$$1^{\text{st}} \text{ kind: } (f * h, m * h, \delta), \quad (18)$$

$$2^{\text{nd}} \text{ kind: } (\delta, \delta, \frac{(f * h - b(m * h))}{(1 - b)}). \quad (19)$$

See Appendix for the proof. Note that these solutions satisfy the positivity constraints.

Fig. 5 middle rows illustrate both no-blur solutions for a grayscale input image in Fig. 4. The no-blur solution of the first kind (18) contains the blurred appearance $f * h$ in the estimated \hat{f} . This only occurs for vastly overestimated $\text{dom } \hat{f}$ that is large enough to accommodate $\text{supp } (f * h)$.



Fig. 4. Examples of an input image in color (left) and grayscale (right); best viewed in color.

Setting $\text{dom } \hat{f}$ by the user close above the true $\text{supp } f$ prevents this no-blur solution. However, if $\text{dom } \hat{f}$ is overestimated even slightly, other spurious solutions exist that are not exact yet they might be potentially dangerous. They are of the form $(\hat{f}, \hat{m}, \hat{h}) = (s * f, s * m, s^\dagger * h)$ for some s of sufficiently small support such that $\text{supp } (s * f) \subset \text{dom } \hat{f}$ and s^\dagger being an approximate inverse of s such that $\text{supp } (s^\dagger * h) \subset \text{dom } \hat{h}$. If the compactly supported approximation s^\dagger is replaced with the exact inverse s^{-1} , i.e. $s^{-1} * s = \delta$, then the solution $(s * f, s * m, s^{-1} * h)$ will be exact. However, compactly supported s might not have s^{-1} and even if it does then it has infinite support, which will not fit into our bounded rectangular domain. With the compactly supported approximation s^\dagger , $(s * f, s * m, s^\dagger * h)$ is no longer an exact solution to (1) yet it could still present a local minimum of (4).

The no-blur solution of the second kind (19) is another potential problem as it cannot be eliminated even by properly setting $\text{dom } \hat{f}$. The object appearance \hat{f} is a single pixel, surely fitting inside $\text{dom } \hat{f}$, and the estimated blur \hat{h} contains the whole blurred input which also fits in $\text{dom } \hat{h}$ as $\text{dom } \hat{h} = \text{dom } g$.

The regularizers and constraints in (4) help to avoid the no-blur solution of the second kind (19) and the inexact solutions of the form $(s * f, s * m, s^\dagger * h)$. Sparsity-promoting regularization $\|h\|_1$ penalizes the less sparse solution \hat{h} while at the same time the bound constraints $f, m \leq 1$ prevent the tendency to decrease intensities in \hat{h} and compensate with increased intensities in \hat{f} . If $s \geq 0$ then s^\dagger is bound to have negative values or vice versa, which implies that either \hat{f} or \hat{h} will contain negative values and this is prohibited by the bound constraints $f, h \geq 0$.

To illustrate the phenomenon of multiple solutions, we conducted an experiment similar to [12] but modified for deblatting. We consider a simplified fm -step (7), in which we assume that m is known and fixed, and define a cost function of h as a marginal minimization of the data term in (4)

$$E(h) = \min_f \|h * f - b(h * m) - d\|_2^2 \quad \text{s.t. (a1) and (a2),} \quad (20)$$

where the constrained minimization w.r.t. f is carried out by the ADMM algorithm similarly to the original fm -step. To be able to visualize this cost function, we assume a 1D toy case with h of length four, $h = [h(1), h(2), h(3), h(4)]$ and f of length two. Assumption (a2) reduces dimensionality of the problem by one since $h(4) = 1 - \sum_{k=1}^3 h(k)$ and so $E(h)$ is really a function of three variables. Final reduction to a 2D function is done by plotting $\min_{h(3)} E(h)$.

Fig. 6a shows an example of $E(h)$ plotted for deblatting with $h = [0.7, 0.2, 0.1, 0]$, $f = [0.5, 0.5]$ and zero background $b = [0, 0, 0, 0]$. The correct solution is marked by circle at location $[h(1), h(2)] = [0.7, 0.2]$. Another exact solution is at $[\hat{h}(1), \hat{h}(2)] = [0.35, 0.45]$, which corresponds to the no-blur solution of the second kind $\hat{h} = f * h = [0.35, 0.45, 0.15, 0.05]$. The no-blur solution of the first kind is not present as $|\text{dom } \hat{f}| = 2$, which is the correct length of f . The inexact solutions, if present, would cluster along the valley connecting the correct solution and the no-blur solution of the second kind.

When the acquisition is in color, we claim that the deblatting problem is better posed than the classical BD. As in the grayscale case, first notice that for a zero background the acquisition model (17) reduces to

$$g_i = f_i * h, \quad (21)$$

where g_i 's denote different color channels. This is a multiframe blind deconvolution (MBD) problem, in which we observe repeatedly a single h each time blurred by a different kernel f_i . In our case, the observed image is the object trajectory and the blur kernels correspond to color channels of the object appearance. In this aspect, deblatting fundamentally differs from the classical BD. Each channel g_i in BD brings the same number of unknowns as equations since $|\text{dom } f_i| = |\text{dom } g_i|$. Whereas in deblatting, the number of new unknowns for every new color channel is less since f_i is compactly supported with $|\text{dom } f_i| < |\text{supp } (f_i * h)|$ (a1). The advantage of MBD is that it is better posed [36] and the no-blur solution of the second kind is not present.

Proposition 2: The deblatting formation model (1) with $K > 1$ color channels and object appearance $[f_1, \dots, f_K]$ with $f_i \neq f_j$ for every $i \neq j$ has only a no-blur solution of the first kind $([\hat{f}_1, \dots, \hat{f}_K], \hat{m}, \hat{h}) = ([f_1 * h, \dots, f_K * h], m * h, \delta)$. See Appendix for the proof.

The no-blur solution for the color case is illustrated in Fig. 5 bottom row. We repeated the same 1D toy experiment in color and calculated $E(h)$ for f having two color channels, $f_1 = [0.5, 0.5]$ and $f_2 = [0.8, 0.2]$; see Fig. 6b. The plot shows a unique minimum at the correct location without the spurious no-blur solution of the second kind.

B. Image-Mask Ambiguity

Deblatting differs from classical BD by the presence of the object shape mask m which is estimated together with the object appearance f and causes ambiguity in the fm -step.

Recall that the acquisition model (2) of pure motion blur (without rotation) using the combined variable $\bar{f} = [f^T, m^T]^T$ is $d = A\bar{f} = [H, -BH]\bar{f}$, where H is a convolution matrix corresponding to h (a rotation-free special case of \mathcal{H}). Without regularization, the fm -step (12) has a unique solution if the matrix A is full column rank since then $A^T A$ is regular. Let us consider only square image domains to simplify the notation and denote the number of pixels in h (as well as b, g) and f (as well as m) as $L^2 = |\text{dom } h|$ and $P^2 = |\text{dom } f|$, respectively. The following propositions define two notable situations when A is column-rank deficient.

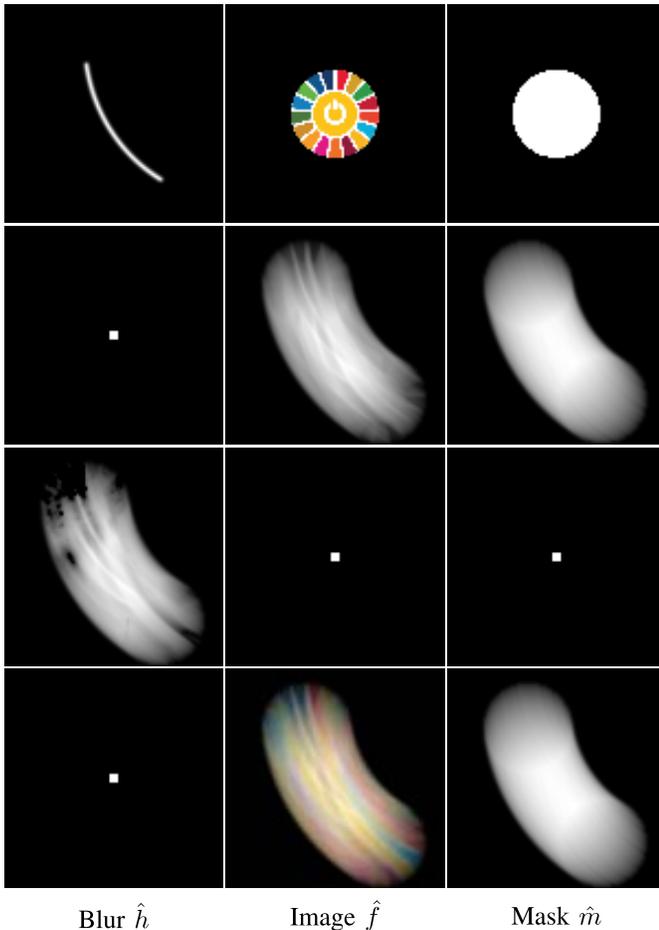


Fig. 5. No-blur solutions for a input image in Fig. 4: (1. row) correct solution, (2. row) grayscale no-blur solution of the first kind, (3. row) grayscale no-blur solution of the second kind, (4. row) color no-blur solution of the first kind.

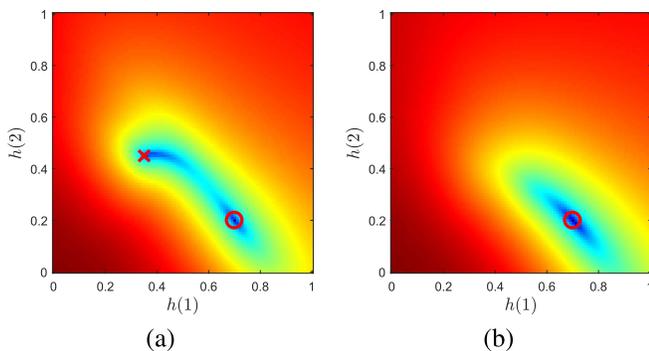


Fig. 6. Toy experiment with two-pixel image and four-pixel blur, plot of the cost function $E(h)$ marginalized over $h(3)$, $h(4)$, and image f : (a) grayscale image f , (b) color (two-channel) image f . Red circle indicates the correct solution, red cross indicates no-blur solution of the second kind present only in the grayscale case.

Proposition 3: Assuming a object with a rectangular appearance f of size $P \times P$ moving along a continuous linear trajectory h of length L , the matrix $A = [H, -BH]$ is column-rank deficient if $L \leq P$ in the grayscale case and $L \leq P/3$ in the three-channel color case.

Proposition 4: The matrix $A = [H, -BH]$ is column-rank deficient if the background is uniform. See Appendix for the proofs.

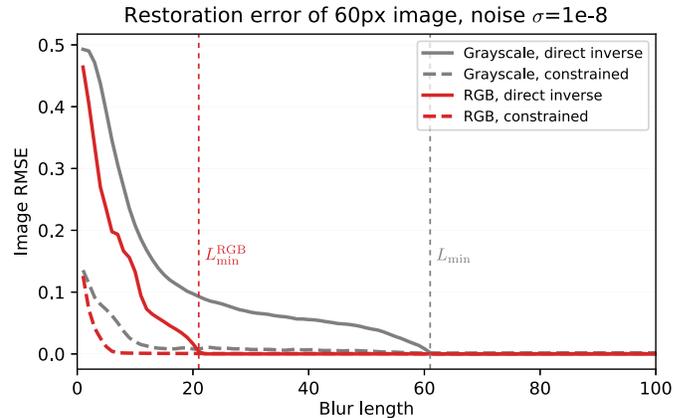


Fig. 7. Image reconstruction error of direct inversion of fm -step without regularization, averaged over 100 different images and backgrounds synthetically blurred by linear motion blur; image range is $[0, 1]$. In grayscale deblatting (solid gray) the blur length must exceed the object size (60px) for both the appearance f and shape mask m to be unambiguously inferable. In three-channel color case (solid red) one-third of the object size suffices. When regularization (3) is used, the restoration is successful for much shorter blurs (corresponding dashed lines).

The first proposition states that the necessary condition for the fm -step to be well conditioned is to have linear blur extending over the entire object size. In the color case, the condition is weaker with the minimum blur length being over one third of the object size. For general motion and/or rotation blur the condition cannot be written in such simple form, because the matrix rank depends on the motion structure as well as on its length, e.g. discontinuous “stroboscopic” blur has better conditionality than continuous motion blur of the same length. The blur must be sufficiently long or complex and thus provide enough snapshots of the object against different parts of the background to have unambiguous estimation of f and m . The problem is partially alleviated if the bound constraints (3) are considered in the fm -step.

We conducted a synthetic experiment to verify these conclusions. We blurred an image by a motion blur of increasing length and then performed estimation solely based on the data-term of (4) by choosing the minimum-norm solution (i.e. using pseudoinverse in the resulting linear system). The results of image reconstruction error (averaged over 100 different images and backgrounds) for varying blur length are in Fig. 7. The size of f is 60×60 px ($P = 60$). In the grayscale case (solid gray) the error decreases slowly as the blur length increases but does not drop to zero until the the minimum blur length condition is met and the system matrix becomes full-rank. In the RGB case (solid red), the error vanishes at roughly $60/3+1 = 21$ px as stated in Proposition 3. We repeated the same experiment under the bound constraints defined in (3), i.e. we solved $A^T A \bar{f} = A^T d$, s.t. $\bar{f} \in S$; the results correspond to the dashed lines in Fig. 7. Especially in the RGB case (red dashed line) we see that the estimation is successful for much shorter blurs (~ 5 px), the ambiguity has been reduced by the bound constraints. The ambiguity in deblatting for small blurs is caused by the f - m coupling, a phenomenon that does not exist in standard BD.

The estimation of f and m is also ambiguous when the condition of Proposition 4 holds, i.e. when the background is

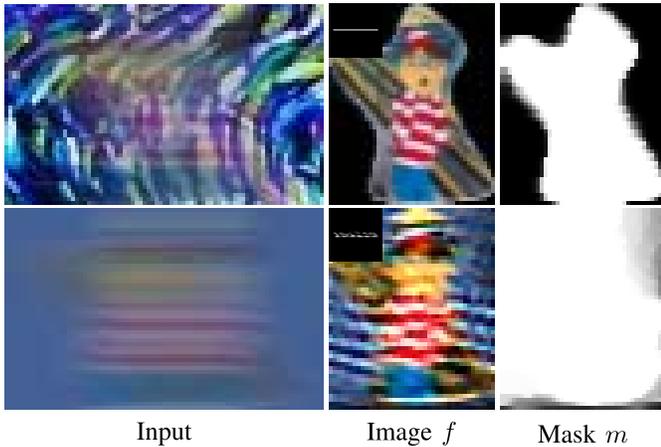


Fig. 8. Effect of background content in blind deblurring. Top row: natural image as background and corresponding estimation of motion-blurred image and mask. Bottom row: the same image and blur on uniform background (background from the top row averaged) – the results is significantly worse due to ambiguity between f and m .

uniform. Although this analysis addresses the fm -step alone, the necessary conditions translate to the fully blind estimation. Fig. 8 shows a synthetic example of how uniform background negatively influences the blind case.

We discussed two canonical cases of the system matrix rank deficiency – due to H and B . Since the two are aggregated in A , any combination of the two cases can happen in practice, for example when only part of the the object motion is over nearly uniform background or when the background contains repetitive structures in the direction of the motion. Naturally, if the system matrix is rank-deficient, there is infinite number of possible solutions and the result will be determined by the regularizers and initialization. Our intention was to provide a deeper insight into the less intuitive property of deblurring that a seemingly easier problem – small blur over simple background – is in fact ill-posed when the shape of the object is unknown. The proposed regularization partially alleviates the problem in practice.

VI. EXPERIMENTS

A. Quantitative Experiments on the Test Dataset

To evaluate the proposed method, we created a dataset containing images obtained by sliding a picture on a table while taking a photograph from above. For the acquisition we used a high-speed video camera with short exposure and zero gap between consecutive frames (“360° shutter”). The obtained video frames were then summed, without any other processing, to simulate taking a single photograph of the object motion with long exposure. This procedure is physically equivalent to taking a long-exposure photograph with the advantage that we also get the ground-truth unblurred appearance of the object at all orientations, which we can use in the evaluation of the deblurring results. Five physically different pictures (see the second left column in Fig. 10) ranging in size from $150 \times 100\text{px}$ to $240 \times 170\text{px}$ were each photographed 15 times while undergoing different motion ranging from about 10px to 130px in length; the dataset thus consists of 75 input images

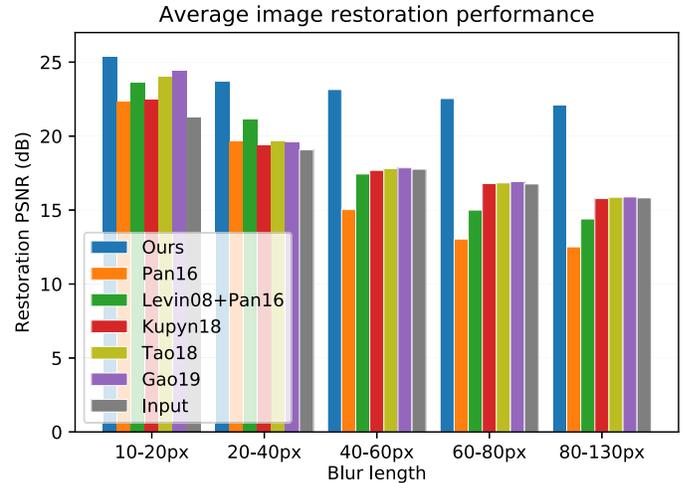


Fig. 9. Results of blind object restoration on the dataset of real images plotted as average PSNR over 5 images and 3 blur realizations.

in total. The error between the reconstructed picture and the ground truth is measured in terms of PSNR and is evaluated invariant to linear intensity transform, shift, and rotation (all rotations present in the high-speed sequence were considered in the PSNR calculation).

We quantitatively compared our method with several recent state-of-the-art methods for blind image deblurring: the optimization-based method Pan16 [19], and the deep learning methods Kupyn18 [24], Tao18 [26], and Gao19 [32]. Pan16 is a robust method for uniform motion deblurring with extra handling of outliers. Kupyn18 (“DeblurGAN”) is for general blind deblurring and uses the GAN framework to produce high-quality natural-looking images. Tao18 (“SRN-DeblurNet”) and Gao19 are based on multiscale processing of the input and both address the problem of dynamic scene deblurring, i.e. blur due to complex unrestricted motion of objects as well as camera, which is very similar to the scenario that we target with the proposed method, although we consider much faster object motion.

To achieve maximum performance we adjusted the input images for each method to suit its particular needs. Pan16 assumes that the input is uniformly blurred in the whole area, this assumption is violated by the surrounding and partially blended background. We therefore included this method in two different variants. In the first case we used this method only for blur estimation and cropped the input to the central part containing mostly only the blurred foreground (while still providing sufficiently large image for the given blur length). To estimate the full-size foreground image we then used the non-blind deblurring part of the method [3] with the blur and shape mask fixed; this result is listed as Pan16. In the second case we estimated the alpha matte of the foreground object using modified method [42] (with maximally informative input scribbles provided with the help of the high-speed sequence) and then extracted only the blurred foreground from the composite input, which results in image similar to Fig. 2 bottom-left. We then used Pan16 to deblur this image; this result is listed as Levin08+Pan16. Due to

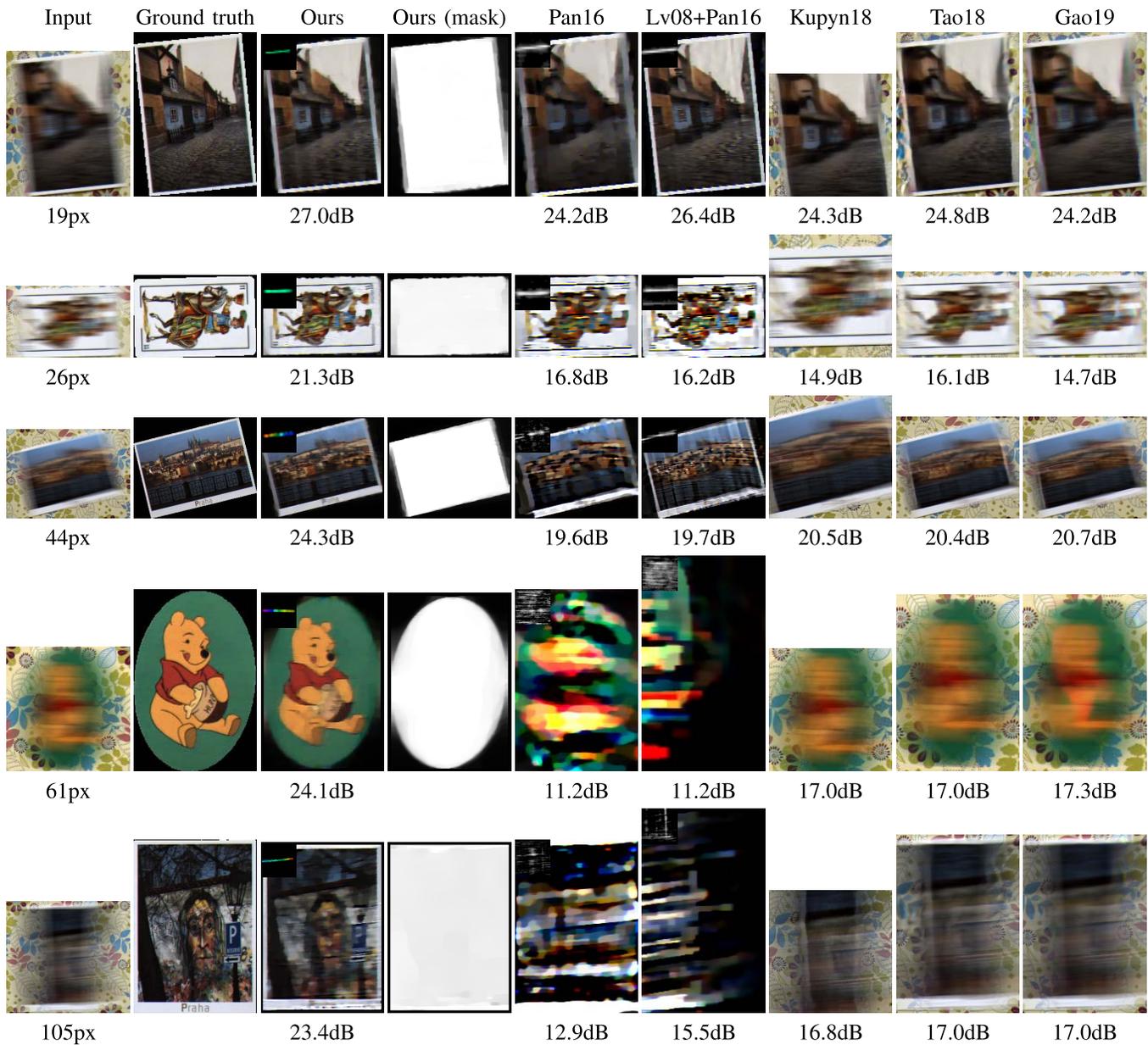


Fig. 10. Examples of images from the dataset and results of the evaluated methods with corresponding restoration PSNR below. The leftmost column shows the input image and the motion blur length, which increases from top to bottom. Results of Pan *et al.* and the proposed method include estimated blur in top left. In the case of the proposed method, the color in blur indicates estimated rotation.

extra outlier handling the method has a good chance of being successful even if the preprocessing of the input is not perfect.

The deep learning methods skip explicit blur estimation and output the sharp image directly. In this case we also cropped the input to limit the amount of background while adhering to the size requirements of each method. The image restoration error was measured only in the central part of the foreground, so that the results are not affected by any boundary effects.

The results in terms of restoration PSNR averaged over all 5 different picture cards are in Fig. 9, corresponding results per each foreground image are in the supplementary. The results are divided into 5 bins based on the blur length. For easier comparison the chart also contains PSNR measured directly on the blurred input (in gray). We can see that for small

blurs, all methods manage to improve the input to a certain degree, the proposed method being superior. As the blur size increases, the performance of Pan16 drops rapidly – it fails to estimate the blur PSF correctly and the restored image then contains increasing amount of deblurring artifacts, rendering it worse than the blurred input itself. The combination of matting and deblurring separately (Levin08+Pan16) performs better for short and medium-length blurs, but the result strongly depends on accuracy of the matting. To test the limits of this approach we used the high-speed sequence as an oracle to provide the best possible input and regularization for the matting method. Regardless of the oracle, for longer blurs with more prominent foreground-background blending the matting fails a consequently so does the deblurring.

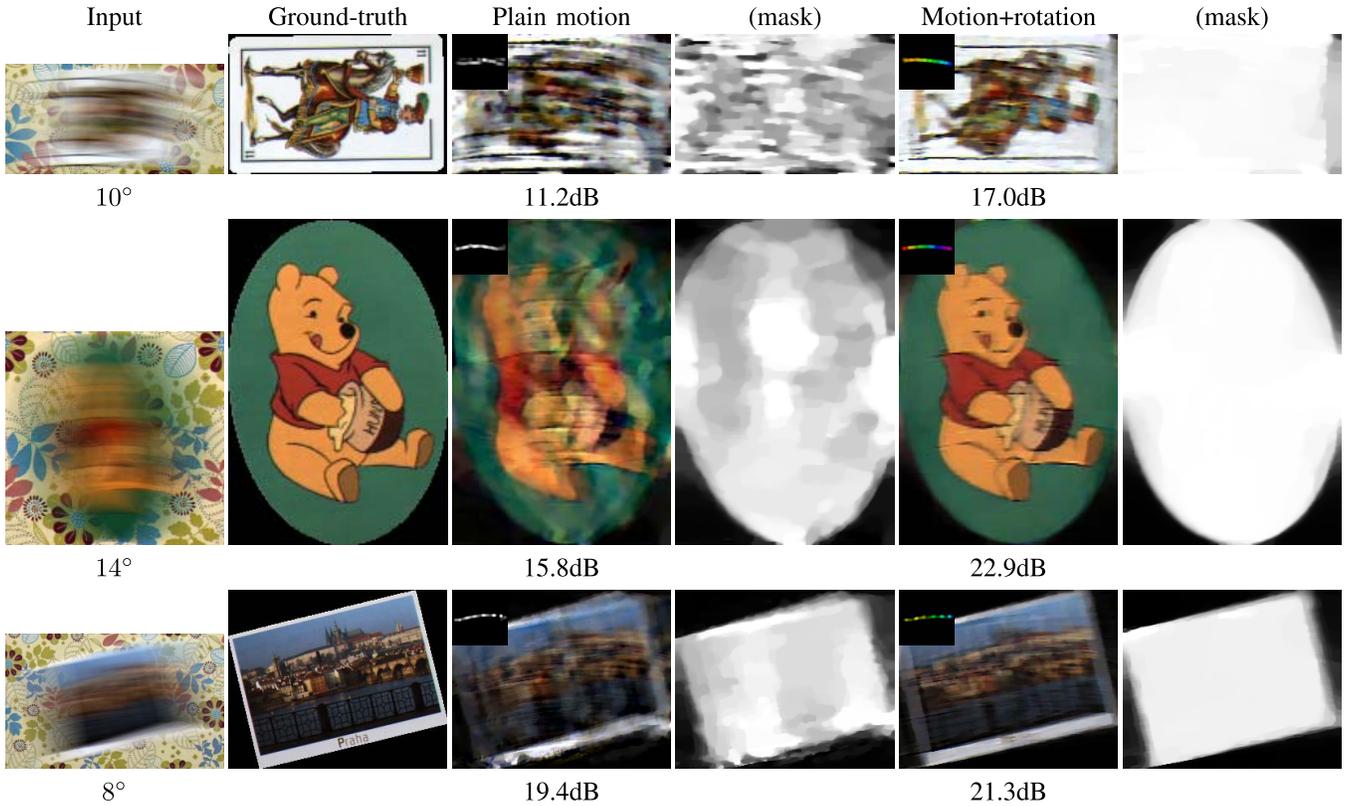


Fig. 11. Comparison of the model with translational motion only and the model with both translation and rotation. From left to right: input with the amount of rotation present, ground-truth image, result of the pure-translation model with PSNR and corresponding mask, result of the full model and corresponding mask. The blur caused by rotation of the object is small compared to the translation blur, yet the performance is significantly weaker when rotation is neglected.

The performance of the deep learning methods deteriorates less because they do not perform explicit algebraic inversion and therefore escape the typical pitfall of classical methods, but they are not trained to perform on such severe blurs and as the blur length increases they start to return the blurred input as output. The proposed method is visibly superior and although its performance also gets progressively worse with blur length, the result is still much improved over the input. Fig. 10 contains several examples of results for different input images from the dataset with varying blur lengths. It best illustrates the behavior of each method and demonstrates the effectiveness of the proposed method compared to the other approaches. In conclusion, the evaluation shows that the task of fast moving object deblurring is not correctly solved by existing deblurring approaches and a specialized deblurring method is necessary.

Our method estimates not only the object translational motion but also its rotation. Fig. 11 demonstrates that in practice this is a necessity. Results where rotation is included in the model (two rightmost columns) are significantly better than if the motion is approximated by pure translation even if the total image rotation in the input is relatively small.

B. Qualitative Real-World Examples

We also tested the proposed method in a fully real-world setting without known ground truth. As inputs we used either still images from a DSLR camera or frames of 25fps or 30fps

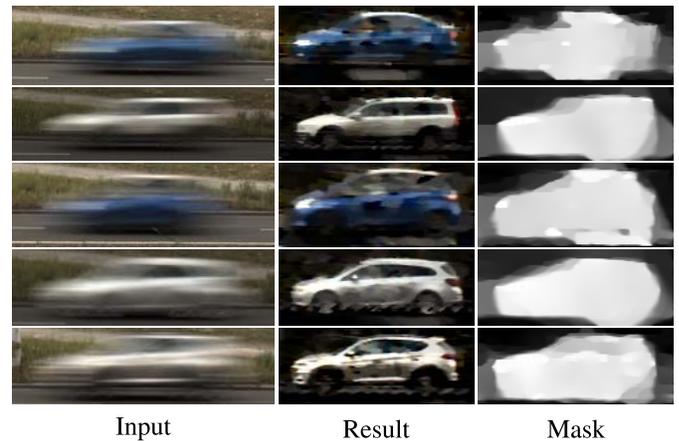


Fig. 12. Motion deblurring results in real-world setting. From left to right: input image, estimated appearance f and shape mask m .

sequences of a video camera (DSLR or GoPro) without any preprocessing. The results of the first such experiment can be seen in Fig. 12. The left column contains long exposure pictures of cars driving by and the left two columns contain restored images of the cars and corresponding masks, respectively. The images suffer from deblurring artifacts but nevertheless bring undeniable value over the inputs, for example when attempting to recognize the make and model of the cars. Similar examples are in Fig. 13 containing images

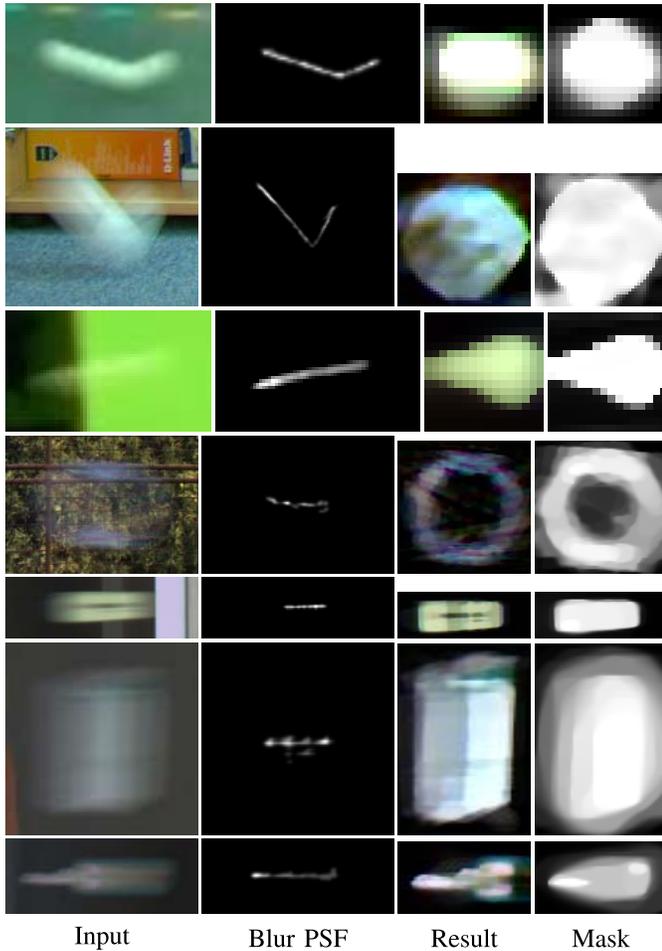


Fig. 13. Other real-world deblurring examples – balls with nonlinear trajectories (pingpong, floorball) or noncircular shapes (badminton, aerobicie) and falling objects (highlighter, box, key; rotated for visualization purposes).

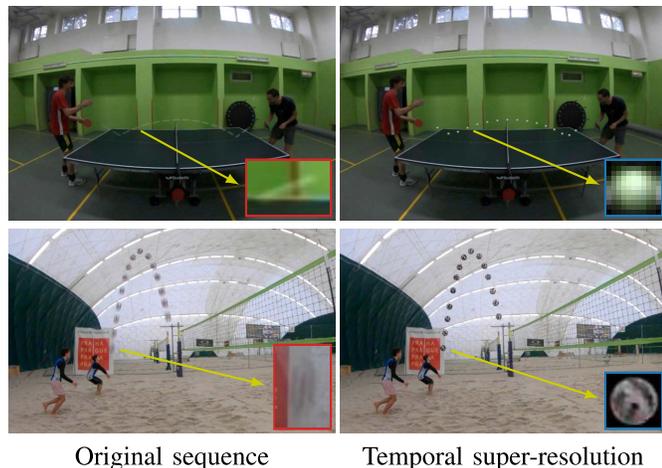


Fig. 14. Temporal super-resolution of two video sequences. Left: Original sequence (only selected frames displayed). Right: Temporal super-resolution (selected frames). In the bottom-right corners of the pictures is a closeup on one of the input frames (red outline) and results of deblurring (blue).

of various objects in flight. Notice the estimated trajectories or shapes of the objects (in case of badminton shuttlecock or aerobicie). The objects undergo 3D rotation, which is

not included in our model, yet the reconstruction and motion estimation are relatively successful in these cases, because angular velocity is small compared to the camera shutter speed.

Final deblurring results are provided in Fig. 14 in the form of a temporal super-resolution of a video sequence. The images on the left are combined selected frames from a video with a fast moving object (ball). Each frame was processed by the proposed deblurring method and using the estimated trajectory and appearance of the ball, the sequence was resynthesized to simulate acquisition at higher framerate (with shorter exposure). Closeups on the bottom right show one of the input blurred frames (red outline) and result of deblurring (blue outline).

VII. CONCLUSION

The problem of deblurring (deblurring and matting) is an ill-posed problem that consists of estimating object shape, appearance, and motion from an image in which the object is blurred and blended with the background due to its motion. We showed that the deblurring problem differs from the standard blind deconvolution in several important aspects and consequently cannot be solved by existing blind deconvolution methods. The principal difference from blind deconvolution lies in the shape estimation and the shape-appearance coupling in the acquisition model. We derived necessary conditions for solution uniqueness and defined the theoretical limitations of deblurring that were demonstrated on synthetic experiments.

We proposed an effective optimization method for solving the deblurring problem with arbitrary 2D translation and rotation. Performance of the presented method was experimentally validated on a new dataset, both quantitatively in comparison with state-of-the-art BD methods and qualitatively on several scenarios of real-world 3D objects with a limitation that the motion is in a 2D plane. Extending the proposed method to arbitrary 3D motion increases its practical usefulness and is therefore of primary interest in our future work.

APPENDIX

Proofs of propositions.

Proposition 1: No-blur of the first kind – Direct substitution of $(\hat{f}, \hat{m}, \hat{h}) = (f * h, m * h, \delta)$ into the image formation model (1) results in $(1 - \hat{h} * \hat{m})b + \hat{h} * \hat{f} = (1 - h * m)b + h * f = g$, which concludes that the solution is exact.

No-blur of the second kind – Substituting $\hat{f} = \delta$ and $\hat{m} = \delta$ into the image formation model (1) results in $(1 - \hat{h} * \hat{m})b + \hat{h} * \hat{f} = (1 - \hat{h})b + \hat{h}$. Being an exact solution $(1 - \hat{h})b + \hat{h} = (1 - h * m)b + h * f$ from which follows that $\hat{h} = (h * f - (h * m)b) / (1 - b)$.

Proposition 2: No-blur of the first kind is proved similarly as in Proposition 1 by direct substitution of $([\hat{f}_1, \dots, \hat{f}_K], \hat{m}, \hat{h}) = ([f_1 * h, \dots, f_K * h], m * h, \delta)$ into (1).

The variable h is a single-channel image and f is a K -channel image. The no-blur solution of the second kind is not present if $f_i \neq f_j$ since $(h * f_i - (h * m)b_i) \neq (h * f_j - (h * m)b_j)$ except a pathological case when $(h * f_i - (h * m)b_i) = 0 \forall i$, which corresponds to an object of the same color as the background.

Proposition 3: Grayscale case – Matrix $A = [H, -BH]$, where H is the convolution matrix of size $L^2 \times P^2$ and $L \gg P$ according to (as1). A is of size $L^2 \times 2P^2$ and has more rows than columns. If h is a linear motion blur of length L then the number of nonzero rows in H is $(L - 1)P + P^2$. Matrix A is column-rank deficient if the number of nonzero rows is less than the number of columns, i.e. $(L - 1)P + P^2 < 2P^2$ from which follows $L < P + 1$.

Color case ($K = 3$) –

$$A = \begin{bmatrix} H & 0 & 0 & -B_1 & H \\ 0 & H & 0 & -B_2 & H \\ 0 & 0 & H & -B_3 & H \end{bmatrix}$$

is of size $3L^2 \times 4P^2$ and is column-rank deficient if $3((L - 1)P + P^2) < 4P^2$ and thus $L < P/3 + 1$.

Proposition 4: If the background is uniform with color ζ then A of size $L^2 \times 2P^2$ becomes $[H, -\zeta H]$ and has rank at most P^2 . Since $L > P$, A is column-rank deficient.

REFERENCES

- [1] D. Rozumnyi, J. Kotera, F. Sroubek, L. Novotny, and J. Matas, “The world of fast moving objects,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4838–4846.
- [2] J. Kotera, D. Rozumnyi, F. Sroubek, and J. Matas, “Intra-frame object tracking by deblatting,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 2300–2309.
- [3] J. Kotera and F. Sroubek, “Motion estimation and deblurring of fast moving objects,” in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 2860–2864.
- [4] P. Campisi and K. Egiazarian, *Blind image Deconvolution: Theory and Applications*. Boca Raton, FL, USA: CRC Press, 2016.
- [5] P. Ruiz, X. Zhou, J. Mateos, R. Molina, and A. K. Katsaggelos, “Variational Bayesian blind image deconvolution: A review,” *Digit. Signal Process.*, vol. 47, pp. 116–127, Dec. 2015.
- [6] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman, “Removing camera shake from a single photograph,” in *Proc. ACM SIGGRAPH Papers SIGGRAPH*, 2006, pp. 787–794.
- [7] R. Molina, J. Mateos, and A. K. Katsaggelos, “Blind deconvolution using a variational approach to parameter, image, and blur estimation,” *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3715–3727, Dec. 2006.
- [8] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, “Understanding blind deconvolution algorithms,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2354–2367, Dec. 2011.
- [9] N. Joshi, R. Szeliski, and D. J. Kriegman, “PSF estimation using sharp edge prediction,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
- [10] Q. Shan, J. Jia, and A. Agarwala, “High-quality motion deblurring from a single image,” in *Proc. ACM SIGGRAPH Papers SIGGRAPH*, 2008, pp. 1–10.
- [11] L. Xu and J. Jia, “Two-phase kernel estimation for robust motion deblurring,” in *Proc. 11th Eur. Conf. Comput. Vis., I ECCV*. Berlin, Germany: Springer, 2010, pp. 157–170.
- [12] D. Perrone and P. Favaro, “A clearer picture of total variation blind deconvolution,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 6, pp. 1041–1055, Jun. 2016.
- [13] S. D. Babacan, R. Molina, M. N. Do, and A. K. Katsaggelos, “Bayesian blind deconvolution with general sparse image priors,” in *Computer Vision—ECCV*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds. Berlin, Germany: Springer, 2012, pp. 341–355.
- [14] L. Xu, S. Zheng, and J. Jia, “Unnatural l0 sparse representation for natural image deblurring,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 1107–1114.
- [15] D. Wipf and H. Zhang, “Revisiting Bayesian blind deconvolution,” *J. Mach. Learn. Res.*, vol. 15, no. 111, pp. 3775–3814, 2014.
- [16] D. Perrone and P. Favaro, “A logarithmic image prior for blind deconvolution,” *Int. J. Comput. Vis.*, vol. 117, no. 2, p. 159, Apr. 2016.
- [17] J. Pan, D. Sun, H. Pfister, and M.-H. Yang, “Deblurring images via dark channel prior,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2315–2328, Oct. 2018.
- [18] S. Cho, J. Wang, and S. Lee, “Handling outliers in non-blind image deconvolution,” in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 495–502.
- [19] J. Pan, Z. Lin, Z. Su, and M.-H. Yang, “Robust kernel estimation with outliers handling for image deblurring,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 2800–2808.
- [20] J. Kotera, V. Smidl, and F. Sroubek, “Blind deconvolution with model discrepancies,” *IEEE Trans. Image Process.*, vol. 26, no. 5, pp. 2533–2544, May 2017.
- [21] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Scholkopf, “Learning to deblur,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 7, pp. 1439–1451, Jul. 2016.
- [22] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce, “Non-uniform deblurring for shaken images,” *Int. J. Comput. Vis.*, vol. 98, no. 2, pp. 168–186, Jun. 2012.
- [23] M. Hirsch, C. J. Schuler, S. Harmeling, and B. Scholkopf, “Fast removal of non-uniform camera shake,” in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 463–470.
- [24] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, “DeblurGAN: Blind motion deblurring using conditional adversarial networks,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8183–8192.
- [25] M. Jin, G. Meishvili, and P. Favaro, “Learning to extract a video sequence from a single motion-blurred image,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6334–6342.
- [26] X. Tao, H. Gao, X. Shen, J. Wang, and J. Jia, “Scale-recurrent network for deep image deblurring,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8174–8182.
- [27] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, “DeblurGAN-v2: Deblurring (orders-of-magnitude) faster and better,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 8877–8886.
- [28] T. H. Kim, B. Ahn, and K. M. Lee, “Dynamic scene deblurring,” in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 3160–3167.
- [29] T. H. Kim and K. M. Lee, “Segmentation-free dynamic scene deblurring,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2766–2773.
- [30] J. Pan, Z. Hu, Z. Su, H.-Y. Lee, and M.-H. Yang, “Soft-segmentation guided object motion deblurring,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 459–468.
- [31] S. Nah, T. H. Kim, and K. M. Lee, “Deep multi-scale convolutional neural network for dynamic scene deblurring,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 257–265.
- [32] H. Gao, X. Tao, X. Shen, and J. Jia, “Dynamic scene deblurring with parameter selective sharing and nested skip connections,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3843–3851.
- [33] F. Šroubek and J. Flusser, “Multichannel blind iterative image restoration,” *IEEE Trans. Image Process.*, vol. 12, no. 9, pp. 1094–1106, Sep. 2003.
- [34] A. Rav-Acha and S. Peleg, “Two motion-blurred images are better than one,” *Pattern Recognit. Lett.*, vol. 26, no. 3, pp. 311–317, Feb. 2005.
- [35] H. Zhang, D. Wipf, and Y. Zhang, “Multi-observation blind deconvolution with an adaptive sparse prior,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1628–1643, Aug. 2014.
- [36] G. Hari Kumar and Y. Bresler, “Perfect blind restoration of images blurred by multiple filters: Theory and efficient algorithms,” *IEEE Trans. Image Process.*, vol. 8, no. 2, pp. 202–219, Feb. 1999.
- [37] G. B. Giannakis and R. W. Heath, “Blind identification of multichannel FIR blurs and perfect image restoration,” *IEEE Trans. Image Process.*, vol. 9, no. 11, pp. 1877–1896, Nov. 2000.
- [38] F. Sroubek and J. Flusser, “Multichannel blind deconvolution of spatially misaligned images,” *IEEE Trans. Image Process.*, vol. 14, no. 7, pp. 874–883, Jul. 2005.
- [39] J. Jia, “Single image motion deblurring using transparency,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2007, pp. 1–8.
- [40] Q. Shan, W. Xiong, and J. Jia, “Rotational motion deblurring of a rigid object from a single image,” in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.
- [41] S. Dai and Y. Wu, “Motion from blur,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.
- [42] A. Levin, D. Lischinski, and Y. Weiss, “A closed-form solution to natural image matting,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 228–242, Feb. 2008.
- [43] C. Dong, S. Barbero, and J. Portilla, “Nonconvex Bayesian restoration of blurred foreground images,” in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2019, pp. 754–758.

- [44] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a convolutional neural network for non-uniform motion blur removal," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 769–777.
- [45] Freepik.com. (2020). *Image Resources From Freepik.Com*. [Online]. Available: <https://www.freepik.com/>
- [46] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.
- [47] R. Tibshirani, "Dykstra's algorithm, ADMM, and coordinate descent: Connections, insights, and extensions," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 517–528.



and computer vision, particularly blind deblurring and image restoration.

Jan Kotera (Member, IEEE) received the master's and Ph.D. degrees in mathematical modeling and computer science from Charles University, Prague, Czech Republic, in 2011 and 2014, respectively, where he is currently pursuing the Ph.D. degree in computer science. Since 2011, he has been with the Institute of Information Theory and Automation, Czech Academy of Sciences. He is the author of several journal and conference papers on blind image deblurring and related topics. His research interests include various aspects of digital image processing



pattern recognition, and RANSAC-type optimization methods. He is on the editorial board of the IJCV and was an Associate Editor-in-Chief of the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI).

Jiří Matas received the Ph.D. degree from the University of Surrey, Guildford, U.K., in 1995. He is a Professor with Center for Machine Perception, Czech Technical University, Prague, Czech Republic. He has published more than 200 articles in refereed journals and conferences. His publications have approximately 34,000 citations in Google Scholar and 13,000 in the Web of Science. His H-index is 65 (GS) and 43 (WoS), respectively. His research interests include visual tracking, object recognition, image matching and retrieval, sequential



and teaches at Charles University. He is the author of eight book chapters and over 90 journal and conference papers on image fusion, blind deconvolution, super-resolution, and related topics.

Filip Šroubek (Member, IEEE) received the M.S. degree in computer science from Czech Technical University, Prague, Czech Republic, in 1998, and the Ph.D. degree in computer science from Charles University, Prague, in 2003. From 2004 to 2006, he was on a postdoctoral position with the Instituto de Optica, CSIC, Madrid, Spain. From 2010 to 2011, he was the Fulbright Visiting Scholar with the University of California at Santa Cruz, Santa Cruz. He is currently with the Institute of Information Theory and Automation, Czech Academy of Sciences, and