Video Deblurring via Semantic Segmentation and Pixel-Wise Non-Linear Kernel

Wenqi Ren[1,2], Jinshan Pan[3], Xiaochun Cao[1,4]*, and Ming-Hsuan Yang[5]

[1]State Key Laboratory of Information Security (SKLOIS), IIE, CAS

[2]Tencent AI Lab

[3]School of Computer Science and Engineering, Nanjing University of Science and Technology

[4]School of Cyber Security, University of Chinese Academy of Sciences

[5]Electrical Engineering and Computer Science, University of California, Merced

Abstract

Video deblurring is a challenging problem as the blur is complex and usually caused by the combination of camera shakes, object motions, and depth variations. Optical flow can be used for kernel estimation since it predicts motion trajectories. However, the estimates are often inaccurate in complex scenes at object boundaries, which are crucial in kernel estimation. In this paper, we exploit semantic segmentation in each blurry frame to understand the scene contents and use different motion models for image regions to guide optical flow estimation. While existing pixel-wise blur models assume that the blur kernel is the same as optical flow during the exposure time, this assumption does not hold when the motion blur trajectory at a pixel is different from the estimated linear optical flow. We analyze the relationship between motion blur trajectory and optical flow, and present a novel pixel-wise non-linear kernel model to account for motion blur. The proposed blur model is based on the non-linear optical flow, which describes complex motion blur more effectively. Extensive experiments on challenging blurry videos demonstrate the proposed algorithm performs favorably against the state-of-the-art methods.

1. Introduction

The recent years have witnessed significant advances in image deblurring with numerous applications [26, 43]. However, most deblurring methods are developed for single images [3, 23, 41] and considerably less attention has been paid to videos [13, 33, 39], where the blur is caused by camera shakes, object motions, and depth variations, as illustrated by an example in Figure 1. Due to interacting and complex motions, video deblurring cannot be modeled well by conventional uniform [8] or non-uniform blur [38] models. On the other hand, as most existing methods for

(a) Frame t (b) Frame $t+1$

(c) Initial segmentation [9]   (d) Our segmentation

(e) Optical flow [13]   (f) Our optical flow

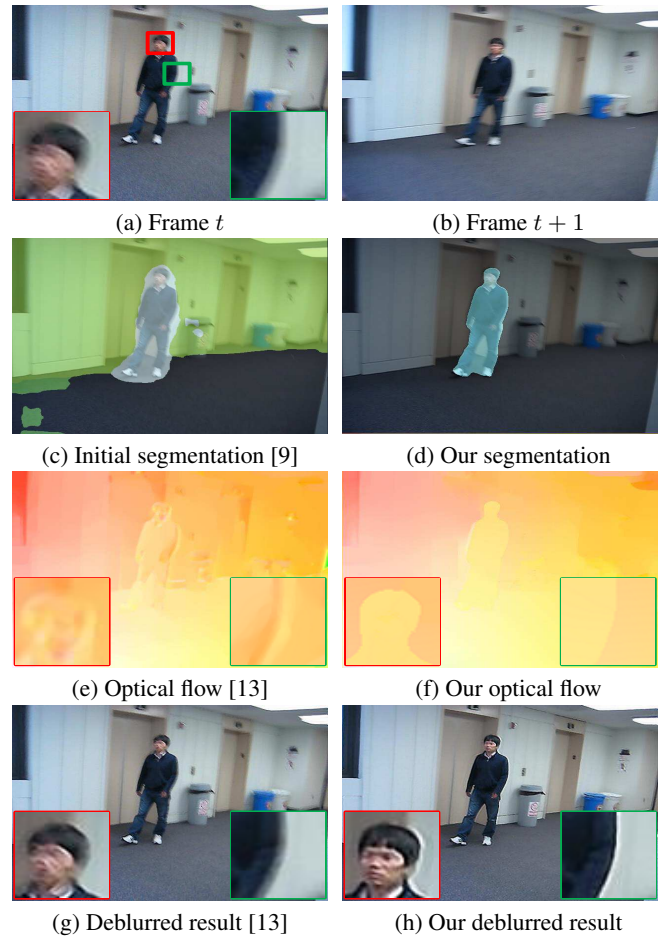(g) Deblurred result [13]   (h) Our deblurred result

Figure 1. (a)-(b) Consecutive frames. (c) Semantic segmentation by [9]. (d) Segmentation of by the proposed algorithm, which is more accurate at the object boundary. (e) Optical flow by [13] from frame t to $t+1$. (f) Optical flow by the proposed algorithm, which is more accurate around the object and background. (g) Deblurred result by [13]. (h) Deblurred image by the proposed algorithm.

---

*Corresponding author.

video deblurring assume that the captured scenes are static [16, 24], these approaches do not handle blurs caused by abrupt motions and usually generate deblurred results with significant artifacts.

To address these issues, deblurring algorithms based on segmentation [1, 5] and motion transformation [22, 6] have been proposed. However, segmentation based algorithms [1, 5] require accurate object segments for kernel estimation. In addition, transformation based methods [22, 6] depend heavily on whether sharp image patches can be extracted across frames for restoration. Recently, Kim and Lee [13] use the bidirectional optical flow to estimate pixelwise blur kernels, which is able to handle generic blur in videos. However, the deblurred results still contain some artifacts which can be attributed to two reasons. First, the estimated optical flow may contain significant errors, particularly due to large displacements or blurred edges [25, 36]. Second, the pixel-wise linear blur kernel is assumed to be the same as the bidirectional optical flow. This assumption does not usually hold for real images as illustrated in Figure 2.

In this work, we propose an efficient algorithm to estimate optical flow and semantic segmentation for video deblurring. If the semantic segmentation of the scene is known, optical flow within the same object should be smooth but flow across the boundary needs not be smooth, and such constraints facilitate accurate blur kernel estimation. On the other hand, accurate optical flow and segmentations are crucial to restore sharp frames. Hence, accurate semantic segmentations and optical flow facilitate to recover accurate sharp frames and vice versa. In addition, as blur kernel is caused by a complicated combination of camera shakes and objects motions, it is different from the estimated linear optical flow as shown in Figure 2. Although some non-linear optical flow methods [42] have been developed, these approaches focus on restoring complex flow structure, *e.g.*, vortex and vanishing divergence, and the estimated optical flow is still a straight line for each pixel. To deal with various blurs in real scenes, we propose a motion blur model using a quadratic function to model optical flow and approximate the pixel-wise blur kernel based on the non-linearity assumption. Extensive experiments on challenging blurry videos demonstrate the proposed algorithm performs favorably against the state-of-the-art methods.

The contributions of this work are summarized as follows. First, we propose a novel algorithm to solve semantic segmentation, optical flow estimation, and video deblurring simultaneously in a unified framework. Second, we exploit semantic segmentation to account for occlusions and blurry edges for accurate optical flow estimation. Third, we propose a pixel-wise non-linear kernel (PWNLK) model to approximate motion trajectories in videos, where the blur kernel is estimated from optical flow under the non-linearity
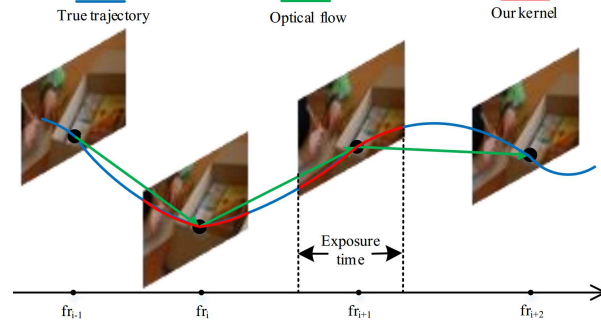

Figure 2. Video motion blur. The green line represents the true motion blur trajectory of the highlighted pixel. The blue line denotes the estimated optical flow. The ground truth motion blur trajectory is smooth and different from optical flow. Based on this observation, we approximate the true motion blur trajectory using the PWNLK model (red line) obtained from a quadratic function of optical flow.

assumption. We show that motion blur cannot be simply modeled by optical flow, and the non-linearity assumption of optical flow is important for video deblurring.

2. Related Work

Deblurring based on motion transformation. Video deblurring based on motion transformation detects sharp images or patches by computing the absolute displacements of pixels between adjacent frames, from which the clear contents are restored [15]. Matsushita *et al.* [22] transfer and interpolate sharper image pixels of neighboring frames for deblurring. Clear regions in a blurry video are detected to restore blurry regions of the same content in nearby frames [6]. A multi-image enhancement method based on a unified Bayesian framework is proposed by Sunkavalli *et al.* [32] to establish correspondence among neighboring frames. However, these transformation based methods do not involve deconvolution and rely on sharp patches from nearby frames which may not exist.

Deblurring based on deconvolution. Deconvolution based methods [7] can be categorized into three approaches based on uniform kernel, layered blur model, and pixel-wise kernel. Uniform kernel based methods [2, 29] assume that the blur in each frame is spatial invariant. These methods are less effective for complex scenes with spatially variant blurs.

To deal with complex motion blurs, layered blur model is developed in the deblurring problem to handle locally varying blurs [5, 39]. Cho *et al.* [5] simultaneously estimate multiple object motions, blur kernels, and the associated image segmentations to solve video deblurring problem. Kim *et al.* [11] adopt a nonlocal regularization on the estimated residual and blurred image to handle object segmentation for dynamic scene deblurring. A layered motion

model is proposed by Bar *et al.* [1] to segment images into foreground as well as background layers, and estimate a linear blur kernel for the foreground layer. Wulff and Black [39] extend this layered model to segment images into foreground and background regions from where the global motion blur kernels are estimated based on affine motion. However, these methods depend heavily on whether accurate segments can be obtained or not since each region is deblurred based on the segmentation.

To address this issue, Li *et al.* [17] parameterize the observed frames in a blurry video by homography and recover sharp contents by jointly estimating blur kernels, camera duty cycles, and latent images. In [44], a projective motion path model [34] is used to estimate blur kernels by exploiting inter-frame misalignments between frames. However, blur models based on homography and projection are designed to account for global camera motions, which cannot model complex object motion and depth variations. To solve this problem, Kim and Lee [12] propose a segmentation-free algorithm by using bidirectional optical flow to model motion blurs for dynamic scene deblurring. This method is extended to generalized video deblurring in [13] by alternatively estimating optical flow and latent frames. Although promising results have been obtained, the assumption that motion blur is same as optical flow does not hold in complex scenes as illustrated in Figure 2 especially when the camera duty cycle is large.

Different from these methods, we take scene semantics and objects into account and use the segmentation to improve optical flow estimation rather than direct deblurring. We then use the estimated optical flow to compute pixel-wise kernel based on non-linear assumption.

Deblurring based on deep learning. Recently, image or video restoration algorithms that aim to recover the underlying sharp contents based on convolutional neural networks, have emerged. In [27], deep neural networks are used for single image deblurring using synthetic training data. Su *et al.* [30] propose a deep encoder-decoder network to address real world video deblurring problems. Nevertheless, when images are heavily blurred, this method may introduce temporal artifacts that become more visible after stabilization.

Semantic segmentation. Semantic segmentation [18, 19, 20] aims to cluster image pixels of the same object class with assigned labels. Numerous recent methods use semantic segmentation to resolve ambiguities in road signs detection [21], 3D reconstruction [10], and optical flow estimation by using different motion models at different object regions [28].

3. Proposed Algorithm

The use of semantic information facilitates modeling optical flow for each region and results in better estimates of

pixel movements, especially at motion boundaries. In addition, the proposed PWNLK model is designed to estimate blur kernels more accurately. In this section, we analyze the relationship between optical flow and motion blur trajectory, and present a video deblurring algorithm based on semantic segmentation and non-linear kernels.

3.1. Motion Blur Model from Optical Flow

The main challenge of video deblurring is how to estimate pixel-wise blur kernels from images. As shown in Figure 2, optical flow (green line) reflects the moving linear direction of a pixel between adjacent frames which may be different from the motion trajectory (blue line). Thus, it is less accurate to model motion blur using optical flow based on linear assumption. A motion blur trajectory is usually smooth and its shape can be approximated by a quadratic function. To model motion blur trajectories t, we use the following parametric PWNLK model:

$$t(f) = af^2 + bf + c, \tag{1}$$

where $f = (u, v)$ is the estimated optical flow of adjacent frames, and a, b, as well as c are parameters to be determined. We find that the motion blur trajectory can be approximated well with this model as shown in Figure 2. We parameterize each kernel $k_i(x)$ at pixel x of frame i as a quadratic function of bidirectional optical flow [12, 13],

$$k_i(x) =$$
$$\begin{cases} \dfrac{\delta(uv_{i,i+1} - vu_{i,i+1})}{2\tau_i(a_{i,i+1}\|f_{i,i+1}\|^2 + b_{i,i+1}\|f_{i,i+1}\| + c_{i,i+1})} & \text{if } f \in [0, \tau_i f_{i,i+1}], \\ \dfrac{\delta(uv_{i,i-1} - vu_{i,i-1})}{2\tau_i(a_{i,i-1}\|f_{i,i-1}\|^2 + b_{i,i-1}\|f_{i,i-1}\| + c_{i,i-1})} & \text{if } f \in (0, \tau_i f_{i,i-1}], \\ 0, & \text{otherwise.} \end{cases}$$
$$\tag{2}$$

With the blur kernel k_i, the blurry frame y_i can be formulated as

$$y_i = k_i l_i + \varepsilon, \tag{3}$$

where l_i denotes the i-th latent frame, and ε denotes noise. Based on the blur model (3), we present an effective video deblurring method and present detailed analysis of the algorithm in the following sections.

3.2. Proposed Video Deblurring Model

Based on the PWNLK model (1), blur formulation (3) and the standard maximum a posterior framework [14], our video deblurring model is defined as

$$E(l, k, f, s) = \sum_i \{E_d(l_i, k_i, y_i) + E_m(f_{ik}, s_{ik}) \\ + E_t(l_i, f_i, s_i) + E_s(l_i, f_i, s_i)\}, \tag{4}$$

where $f_{ik} = (u_{ik}, v_{ik})$ and s_{ik} denote optical flow and segmentation in the k-th layer of i-th frame, respectively. The

first term E_d in (4) is the data fidelity term, *i.e.*, the deblurred frame l_i should be consistent with the observation y_i. The second term E_m denotes a motion term which encodes two assumptions. First, neighboring pixels should have similar motion if they belong to the same semantic segmentation layer. Second, pixels from each layer k should share a global motion model $f(\theta_{ik})$, where θ_{ik} is parameter that changes over time and depends on the object class k. The third term E_t is the temporal regularization term, which is used to ensure the brightness constancy between adjacent frames. The last term E_s denotes the spatial regularization term of latent images and optical flow. The details of each term in (4) are described below.

Data term based on the PWNLK model. It has been shown that using gradients of latent and blurry images in the data term can reduce ringing artifacts [12, 13]. Thus, our data fidelity term is defined as

$$E_d(l_i, k_i, y_i) = \sum_i \lambda \|\nabla(k_i l_i) - \nabla y_i\|_2^2, \quad (5)$$

As blur kernel k_i is computed according to the motion blur trajectory in (1), the data fidelity term (5) involves parameters a, b, and c. To obtain a stable solution, we need to regularize these motion blur parameters [35]. The Tikhonov regularization has been extensively used in the literature of image deblurring. However, we note that motion blur has similar properties to the optical flow in most examples. For example, the estimated motion blur would have the same property if the estimated optical flow has piece-wise property. That is, if $\nabla f_i = 0$ at some regions, we would have $\nabla(a_i f_i^2 + b_i f_i + c_i) = 0$. Based on this assumption, we have $b_i = -2 a_i f_i$. As $\nabla f_i = 0$, f_i should be a constant C. This property motivates us to use the following regularization on parameters a and b,

$$\sum_i \{\beta \|a_i\|_2^2 + \gamma \|b_i - C\|_2^2\}, \quad (6)$$

where β and γ denote the weights of each term in the regularization terms.

Motion term. The motion term should satisfy: 1) pixels in the same segmentation layer s_{ik} should share a global motion model $f(\theta_{ik})$, 2) neighboring pixels in the same segmentation layer s_{ik} should have similar optical flow. Thus, our motion term is defined as

$$
\begin{aligned}
E_m(f_{ik}, s_{ik}) = \sum_i \{ \sum_x \rho_{\text{aff}}(f_{ik}(x) - f(\theta_{ik})) \\
+ \sum_x \sum_{r \in \mathcal{N}_x} \|f_i(x) - f_i(r)\|_2^2 \delta(s_{ik}(x) = s_{ik}(r))\},
\end{aligned}
\quad (7)
$$

where $\mathcal{N}_x$ denotes the four nearest neighbors of the pixel x, and ρ_{aff} is a robust penalty function which enforces that the

pixels in the same segmentation have the same affine motion model [31]. In addition, $\delta(\cdot)$ denotes the indicator function that is equal to 1 if its expression is true, and 0 otherwise.

Spatial term. The spatial regularization term aims to alleviate the ill-posed inverse problem. We assume that the spaial term should 1) constrain the pixels with similar colors to lie within the same segmentation layer s_{ik}, and 2) enforce spatial coherence in both latent frames and optical flow. With these assumptions, the spatial term is defined by

$$
\begin{aligned}
E_s(l_i, f_i, s_i) = \sum_i \{ |\nabla l_i| + \sum_{n=-N}^{N} g_i(x) |\nabla f_{i,i+n}| \\
+ \sum_x \sum_{r \neq x} \omega_{x,r} \delta(s_{ik}(x) \neq s_{ik}(r))\},
\end{aligned}
\quad (8)
$$

where the weight $g_i(x)$ denotes edge-map [13] to preserve discontinuities in the optical flow at edges. In addition, $\omega_{x,r}$ is a weight which measures the similarity between x and r. Similar to the optical flow estimation method [31], we define it as

$$\omega_{x,r} = \exp\{-\frac{\|x - r\|^2 + \|l_i(x) - l_i(r)\|^2}{\sigma^2}\}, \quad (9)$$

where σ is a constant. For a given pixel x, if we know other neighboring pixels r have similar color as x, we set them with the same segment. The effectiveness of the regularization term is demonstrated in Section 4.1.

Temporal term. Human vision system is sensitive to temporal inconsistencies presented in videos. To improve temporal coherence, we first utilize the optical flow to find the corresponding pixels between neighboring frames in a local temporal window $[i - N, i + N]$ and ensure that the corresponding pixels vary smoothly. We then enforce that corresponding pixels between neighboring frames should belong to the same segment. Thus, the temporal coherence is defined by

$$
\begin{aligned}
E_t(l_i, f_i, s_i) = \sum_i \{ \sum_{n=-N}^{N} \mu_n |l_i(x) - l_{i+n}(x')| \\
+ \sum_{n=-N}^{N} \mu_n |s_i(x) - s_{i+n}(x')|\},
\end{aligned}
\quad (10)
$$

where n denotes the index of neighboring images at frame i and μ_n is a weight for the regularization term. In addition, $x' = x + f_{i,i+n}$ is the corresponding pixel at the next n-th frame for x according to the motion $f_{i,i+n}$. We use the L_1-norm regularization in (10) for robust estimates against outliers and occlusions [13].

3.3. Inference

Based on the above analysis, we obtain the proposed video deblurring model. Although the objective function

is non-convex with multiple variables, we can use an alternating minimization method [13] to solve it.

Latent frames estimation. With the optical flow f, segmentation s, and the parameters a, b and c, the optimization problem with respect to l_i is

$$
\begin{aligned}
\min_{l_i} \lambda \sum_i \{ \| \nabla(k_i l_i) - \nabla y_i \|_2^2 + |\nabla l_i| \\
+ \sum_{n=-N}^{N} \mu_n |l_i(x) - l_{i+n}(x')| \}.
\end{aligned}
\tag{11}
$$

Similar to [13], we optimize the latent frames subproblem (11) using the primal-dual update method [4].

Semantic segmentation. The semantic segmentation estimation can be achieved by solving

$$
\begin{aligned}
\min_{s_i} \sum_i \{ \sum_x \sum_{r \neq x} \omega_{x,r} \delta(s_{ik}(x) \neq s_{ik}(r)) \\
+ \sum_x \sum_{r \in \mathcal{N}_x} \| f_i(x) - f_i(r) \|_2^2 \delta(s_{ik}(x) = s_{ik}(r)) \\
+ \sum_{n=-N}^{N} \mu_n |s_i(x) - s_{i+n}(x')| \}.
\end{aligned}
\tag{12}
$$

We optimize this subproblem (12) using the method in [28]. The semantically segmented regions provide information on a potential optical flow for the motion blurred object, which is used to guide optical flow estimation instead of directly deblurring on each segment [1, 39].

Note that we only refine the segmentation results s_{ik} according to possible moving objects including person, rider, car, etc, as like in Figure 1(d). For other background objects (e.g., road, sky, wall), we do not refine their segmentation since these objects are always smooth and their segmentation results cannot affect our deblurring results.

Optical flow estimation. After obtaining l and s, the optimization problem with respect to f becomes

$$
\begin{aligned}
\min_{f_i} \lambda \sum_i \{ \| \nabla(k_i l_i) - \nabla y_i \|_2^2 + \sum_{n=-N}^{N} g_i(x) |\nabla f_{i,i+n}| \\
+ \sum_x \sum_{r \in \mathcal{N}_x} \| f_i(x) - f_i(r) \|_2^2 \delta(s_{ik}(x) = s_{ik}(r)) \\
+ \sum_x \rho_{\text{aff}}(f_{ik}(x) - f(\theta_{ik})) + \sum_{n=-N}^{N} \mu_n |l_i(x) - l_{i+n}(x')| \}.
\end{aligned}
\tag{13}
$$

We solve (13) using the method in [13] and [31]. After obtaining f_i, we utilize it to estimate the blur kernel based on the non-linearity assumption, instead of directly using the bidirectional optical flow as blur kernel.

Algorithm 1 Proposed video deblurring algorithm

Input: Blurry frames y, duty cycle τ, initialized optical flow f by [37] and semantic segmentation s by [9].
Repeat the following steps from coarse to fine image pyramid level:
1. Solve for parameters a, b and c by minimizing (14).
2. Solve for optical flow f by minimizing (13).
3. Estimate blur kernel based on PWNLK model (2).
4. Solve for latent image l by minimizing (11).
5. Solve for segmentation s by minimizing (12).
Output: latent frames l, blur kernels k, optical flow f and segmentation s.

(a) Input   (b) Flow   (c) 16.05dB   (d) 17.98dB   (e) 18.13dB   (f) Truth

Figure 3. The limitation of linear assumption in [13]. (a) Blurred input. (b) Ground truth optical flow. (c) Deblurred result by segmentation based method [39]. (d) and (e) are the deblurred results using flow in (b) based on linear assumption [15] and our nonlinear model, respectively. (d) Ground truth image.

Motion blur trajectory parameters estimation. For each blurry frame y_i, we obtain its corresponding sharp reference l_i and its bidirectional optical flow f_i. With each image pair and the corresponding optical flow, the parameters of the motion blur kernel a_i, b_i and c_i are solved by

$$
\min_{a,b} \lambda \sum_i \{ \| \nabla(k_i l_i) - \nabla y_i \|_2^2 + \beta \|a_i\|_2^2 + \gamma \|b_i - C\|_2^2 \}.
\tag{14}
$$

This is a least squares minimization problem and we have the closed-form solutions for the parameters a, b and c, respectively.

Similar to the existing methods, we use the coarse-to-fine method with an image pyramid [13] to achieve better performance. Algorithm 1 summarizes the main steps of the proposed video deblurring on one image pyramid level.

4. Experimental Results

In this section, we first analyze and show the effects of the semantic segmentation and PWNLK model. We then evaluate the proposed method on both synthetic and real-world blurry videos. We compare the proposed algorithm with the state-of-the-art methods, based on motion transformation [6], uniform kernel [29], piece-wise kernel [39], and pixel-wise linear kernel by Kim and Lee [13].

Parameter settings. In all experiments, we set the parameters $\lambda = \mu_n = 250$, $\beta = \gamma = 0.5\lambda$, $\sigma = 7$, and $N = 2$. We initialize the parameters of the quadratic bidirectional optical flow as $a = c = 0$ and $b = 1$. For fair comparisons, we
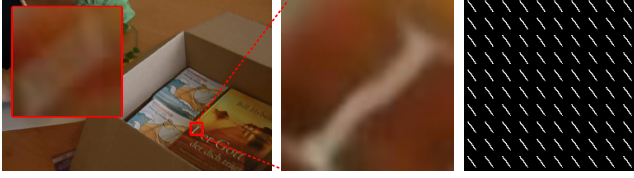
(a) Deblurred result by the linear approximation method [13]
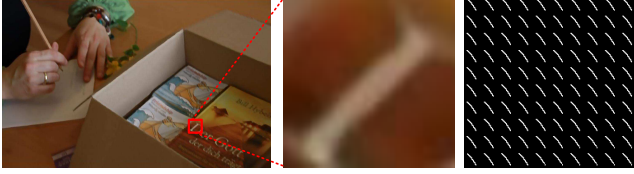
(b) Deblurred result by our non-linear approximation method.

Figure 4. Effects of PWNLK. (a) Deblurred results and estimated kernel by linear approximation method [13]. (b) Deblurred results and estimated kernel by the proposed non-linear approximation approach (2). The highlighted area in the red rectangle is the corresponding blurry input. The recovered kernel in (a) is almost straight, which results in the deblurred result has some distortion artifacts. In contrast, the estimated kernel by the proposed PWNLK model is more close to real situation, and results in the recovered image is visually more pleasing.

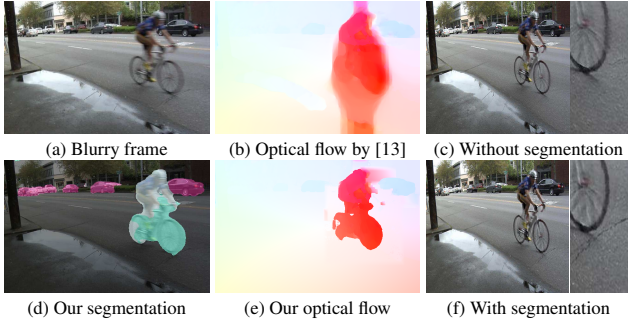| (a) Blurry frame | (b) Optical flow by [13] | (c) Without segmentation |
| (d) Our segmentation | (e) Our optical flow | (f) With segmentation |

Figure 5. Effects of semantic segmentation on deblurring. (a) Blurry input. (b) and (c) are estimated optical flow and deblurred result by [13]. (d) Our segmentation results (semantic color coded using [28]). (e) and (f) are estimated optical flow and deblurred result with the proposed semantic segmentation. The background and road regions in (c) are over-smoothed due to the inaccurate estimated optical flow in (b).

use the TV-ℓ^1 based method [37] to initialize optical flow as like in [13]. We also use the state-of-the-art semantic segmentation method [9] to segment images first, and refine the results based on the proposed algorithm. In addition, we use the method in [13] to estimate the camera duty cycle τ.

4.1. Analysis of Proposed Method

Effects of PWNLK model. We note that [13] directly uses the linear bidirectional optical flow to restore the clear images. As mentioned in Figure 2, this method is less effective since motion trajectories in videos are different from optical flow. Figure 3(a) shows an example where the blurred
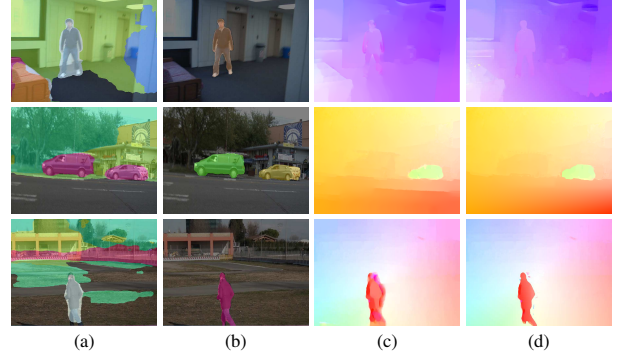

| (a) | (b) | (c) | (d) |

Figure 6. Qualitative analysis of semantic segmentation. (a) Blurry input and initialized segmentation results [9]. (b) Our refined segmentation. (c) Optical flow by [13]. (d) Our optical flow.

image is generated by affine transformation [39]. We first show the deblurred result by the layer based method [39] in Figure 3(c). Note that there are significant artifacts around the *elephant* boundary since the inaccurate segmentation. As shown in Figure 3(d), the restored image generated by the ground truth optical flow (Figure 3(b)) using the pixel-wise linear kernel method [13] contains significant ring artifacts, which demonstrates that the linear bidirectional optical flow cannot model motion blur well.

Figure 4 shows an example which is able to demonstrate the effectiveness of the PWNLK model. We use the same optical flow to estimate the pixel-wise linear and non-linear kernel. We note that the linear assumption of motion blur for each pixel does not hold in real images as shown in Figure 4(a). The estimated motion blur kernel using linear approximation for the zoomed-in region is almost straight and the corresponding deblurred results contain distortion artifacts on the line of letter *D*. The trajectories of the estimated motion kernel by the proposed non-linear approximation method coincide well with the real motion blur trajectories and the corresponding deblurred image is much clearer and contains fewer artifacts as shown in Figure 4(b), which indicate that the proposed blur model (1) can better approximate motion trajectories in real scenes.

Effects of semantic segmentation. Semantic segmentation improves video deblurring in multiple ways as it is used to help estimate optical flow from which the blur kernel is estimated. First, it provides region information about object boundaries. Second, as different objects (layers) move differently, semantic segments are used to constrain optical flow estimation of each region. As shown in Figure 5(b), the estimated optical flow is over-smoothed around the bicycle when semantic segmentation is not used. Consequently, the deblurred results for the background and road regions are over-smoothed. In contrast, the semantic segmentation results by the proposed algorithm describe boundaries well and help generate accurate optical flow. As shown in Fig-
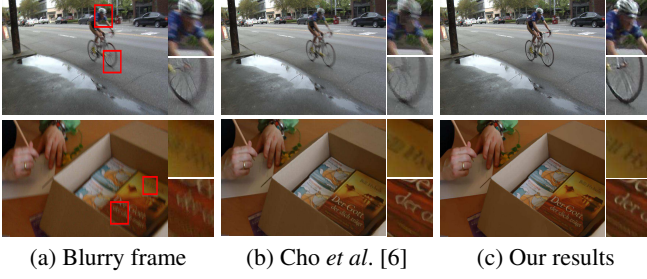
(a) Blurry frame   (b) Cho *et al*. [6]   (c) Our results

Figure 7. Comparisons with transformation based method [6].

(a) Blurry frames   (b) Šroubek and Milanfar [29]   (c) Our results

Figure 8. Comparisons with uniform kernel based method [29].

(a) Blurry frames   (b) Wulff and Black [39]   (c) Our results

Figure 9. Comparisons with piece-wise kernel (segmentation) based video deblurring method [39].

(a) Blurry frames   (b) Kim and Lee [13]   (c) Our results

Figure 10. Comparisons with pixel-wise linear kernel based video deblurring method [13].

ure 5(f), the deblurred images by the proposed algorithm are clear with fine details.

In addition, we carry out more experiments to examine the effects of semantic segmentation for optical flow estimation. Although the initialized segmentations are inaccurate as shown in Figure 6(a), the proposed algorithm can precisely segment the moving objects (Figure 6(b)) and provide more accurate motion boundaries information for optical flow estimation, and thereby facilitates video deblurring.

4.2. Real Datasets

We evaluate the proposed algorithm against the state-of-the-art video deblurring methods [6, 29, 39, 13] on real sequences from [6, 39]. We first compare our algorithm with the transformation based method by Cho *et al*. [6]. As shown in the first row of Figure 7(b), the method [6] does not recover the moving bicycle because the object motion is large and there are no sharp images in the nearby frames. In contrast, the proposed algorithm is able to deal with the blur caused by the moving objects and generates a clear image as shown in the first row of Figure 7(c). The transformation based approach [6] does not handle large camera motion blur as shown in the second row of Figure 7(b). The recovered texts for the *Books* sequence contain significant distortion artifacts since this transformation based method [6] introduces incorrect patch matches if the clear images or sharp patches are not available. In contrast, the proposed method based on the estimated optical flow does not require clear images or patches. The deblurred result is visually more pleasing especially for the texts.
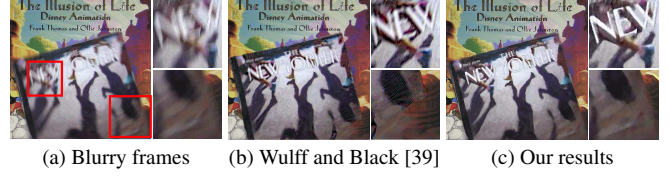
We compare the proposed algorithm with the uniform kernel based multi-image deblurring method [29]. On the *Street* sequence, the sign PAY HERE and the structure of the windows can be clearly recognized from the deblurred image by the proposed algorithm, while the one by the multi-image based method does not recover such details. Furthermore, our method recovers clear edges and details in the *Kid* sequence. However, the multi-image based deblurring method does not generate clear images. The main reason is that the uniform kernels estimated by the multi-image based method do not account for complex scenes with non-uniform blur. In addition, the deblurred results of this multi-image deblurring method depend on whether the alignments of adjacent frames are accurate or not.

We show the deblurred results by the proposed method and segmentation based video deblurring approach [39] in Figure 9. Although the deblurred image by [39] is sharp, it contains some distortion artifacts around the image boundaries due to the inaccurate segmentations (*e.g.*, the boundary of the *Magazine* on the right-bottom corner in Figure 9(b)). In contrast, the deblurred image in Figure 9(c) shows that proposed method is able to recover the clear edge of the *Magazine*. In addition, the recovered text *NEW* in the foreground layer by Wulff and Black [39] is blurry compared to the result generated by the proposed algorithm.

We compare the proposed algorithm with the state-of-the-art video deblurring method based on pixel-wise linear kernel by Kim and Lee [13]. The deblurred results by [13] contain blurry edges and distortion artifacts as shown in Figure 10(b). For example, due to the inaccurate kernel estimation, the deblurred result by [13] has distortion artifacts around the left-bottom corner of the *Sign* in the second row

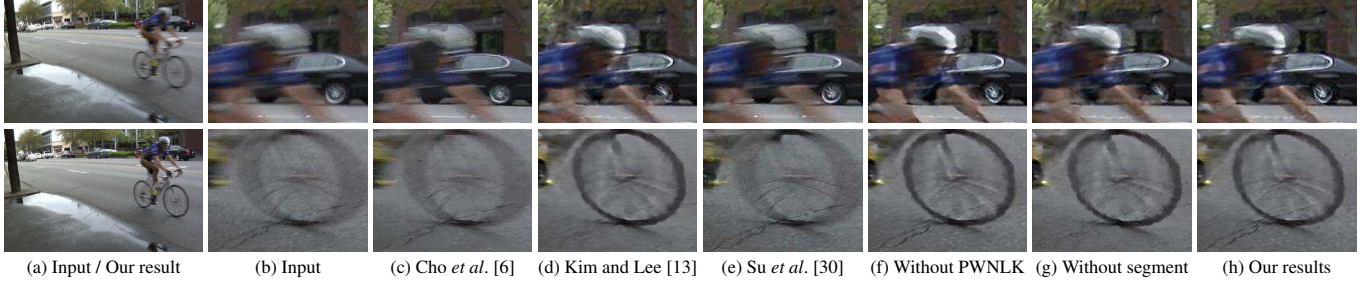| (a) Input / Our result | (b) Input | (c) Cho *et al*. [6] | (d) Kim and Lee [13] | (e) Su *et al*. [30] | (f) Without PWNLK | (g) Without segment | (h) Our results |

Figure 11. Deblurred results with and without the PWNLK model and semantic segmentation.

of Figure 10(b). In contrast, as the proposed motion blur model is able to approximate the true motion blur trajectories, the recovered images contain fine details. Note that in Figure 10(c), the deblurred texts in both first and second rows by the proposed algorithm are clearer and sharper.

Finally, we show the deblurred results with and without the PWNLK model and semantic segmentation, and compare with the state-of-the-art transformation based [6], deconvolution based [13] and deep learning based [30] video deblurring methods in Figure 11. The state-of-the-art video deblurring methods [6, 30] do not generate clear images as shown in Figure 11(c) and (e). Pixel-wise linear kernel based method [13] can generate sharp image, but the road region is over-smoothed as show in the bottom line in Figure 11(d). In Figure 11(f), the road region is successfully recovered, but there are some visual artifacts around the tire due to imperfect kernel estimation. Figure 11(g) shows the deblurred result without performing semantic segmentation. Although the tire is deblurred well, the road region is over-smoothed. Compared to the image shown in (h), the visual quality of (f) and (g) is lower, which indicates the importance of the proposed PWNLK model (1) and semantic segmentation regularization.

4.3. Limitations

Our algorithm does not performs well when the input video contains significant blur along with bad initial segmentations. Figure 12(c) and (d) are the initial segmentation results for the consecutive blurry frame Figure 12(a) and (b), respectively. Since the assumed spatial and temporal constraints in (8) and (10) do not hold in the segmented image, the final segmentation result in Figure 12(e) does not have any semantic information. Thus, our method degenerates to traditional optical flow estimation in [13] and generate similar deblurred results as shown in Figure 12(g) and (h).

5. Conclusions

In this paper, we propose an effective video deblurring algorithm by exploiting semantic segmentation and PWNLK model. The proposed segmentation applies differ-
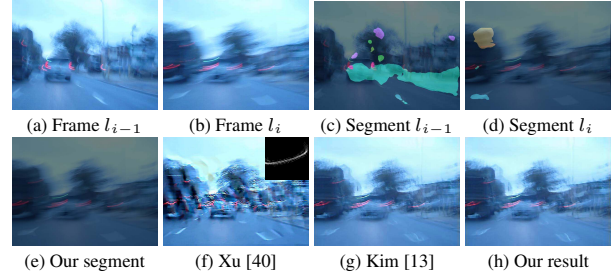

| (a) Frame l_{i-1} | (b) Frame l_i | (c) Segment l_{i-1} | (d) Segment l_i |
| (e) Our segment | (f) Xu [40] | (g) Kim [13] | (h) Our result |

Figure 12. Failure cases. (a) and (b) are blurred inputs l_{i-1} and l_i. (c) and (d) are initialized segmentation results on frames l_{i-1} and l_i. (e) Our final segmentation on frame l_i. (e)-(g) are deblurred results by [40], [13] and our method on frame l_i.

ent motion model to different object layers, which can significantly improve the optical flow estimation, especially at object boundaries. The PWNLK model is based on the non-linear assumption and is able to model the relationship between motion blur and optical flow. In addition, we analyze that conventional uniform, homography, piece-wise, pixel-wise linear based blur kernels cannot model the complex spatially variant blur caused by the combination of camera shakes, objects motions and depth variations. Extensive experimental results on synthetic and real videos show that the proposed algorithm performs favorably in video deblurring against the state-of-the-art methods.

References

[1] L. Bar, B. Berkels, M. Rumpf, and G. Sapiro. A variational framework for simultaneous motion estimation and restora-

tion of motion-blurred video. In *ICCV*, 2007.

[2] J.-F. Cai, H. Ji, C. Liu, and Z. Shen. Blind motion deblurring using multiple images. *Journal of computational physics*, 228(14):5057–5071, 2009.

[3] X. Cao, W. Ren, W. Zuo, X. Guo, and H. Foroosh. Scene text deblurring using text-specific multiscale dictionaries. *TIP*, 24(4):1302–1314, 2015.

[4] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[5] S. Cho, Y. Matsushita, and S. Lee. Removing non-uniform motion blur from images. In *ICCV*, 2007.

[6] S. Cho, J. Wang, and S. Lee. Video deblurring for hand-held cameras using patch-based synthesis. *TOG*, 31(4):64, 2012.

[7] M. Delbracio and G. Sapiro. Hand-held video deblurring via efficient fourier aggregation. *TCI*, 1(4):270–283, 2015.

[8] R. Fergus, B. Singh, A. Hertzmann, S. T. Roweis, and W. T. Freeman. Removing camera shake from a single photograph. *TOG*, 25(3):787–794, 2006.

[9] G. Ghiasi and C. C. Fowlkes. Laplacian pyramid reconstruction and refinement for semantic segmentation. In *ECCV*, 2016.

[10] C. Hane, C. Zach, A. Cohen, R. Angst, and M. Pollefeys. Joint 3d scene reconstruction and class segmentation. In *CVPR*, 2013.

[11] T. H. Kim, B. Ahn, and K. M. Lee. Dynamic scene deblurring. In *ICCV*, 2013.

[12] T. H. Kim and K. M. Lee. Segmentation-free dynamic scene deblurring. In *CVPR*, 2014.

[13] T. H. Kim and K. M. Lee. Generalized video deblurring for dynamic scenes. In *CVPR*, 2015.

[14] D. Krishnan, T. Tay, and R. Fergus. Blind deconvolution using a normalized sparsity measure. In *CVPR*, 2011.

[15] D.-B. Lee, S.-C. Jeong, Y.-G. Lee, and B. C. Song. Video deblurring algorithm using accurate blur kernel estimation and residual deconvolution based on a blurred-unblurred frame pair. *TIP*, 22(3):926–940, 2013.

[16] H. Lee and K. Lee. Dense 3d reconstruction from severely blurred images using a single moving camera. In *CVPR*, 2013.

[17] Y. Li, S. B. Kang, N. Joshi, S. M. Seitz, and D. P. Huttenlocher. Generating sharp panoramas from motion-blurred videos. In *CVPR*, 2010.

[18] X. Liang, S. Liu, X. Shen, J. Yang, L. Liu, J. Dong, L. Lin, and S. Yan. Deep human parsing with active template regression. *TPAMI*, 37(12):2402–2414, 2015.

[19] S. Liu, X. Liang, L. Liu, X. Shen, J. Yang, C. Xu, L. Lin, X. Cao, and S. Yan. Matching-cnn meets knn: Quasi-parametric human parsing. In *CVPR*, 2015.

[20] S. Liu, C. Wang, R. Qian, H. Yu, and R. Bao. Surveillance video parsing with single frame supervision. *arXiv preprint arXiv:1611.09587*, 2016.

[21] S. Maldonado-Bascon, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno, and F. López-Ferreras. Road-sign detection and recognition based on support vector machines. *T-ITS*, 8(2):264–278, 2007.

[22] Y. Matsushita, E. Ofek, X. Tang, and H.-Y. Shum. Full-frame video stabilization. In *CVPR*, 2005.

[23] T. Michaeli and M. Irani. Blind deblurring using internal patch recurrence. In *ECCV*, 2014.

[24] C. Paramanand and A. Rajagopalan. Non-uniform motion deblurring for bilayer scenes. In *CVPR*, 2013.

[25] T. Portz, L. Zhang, and H. Jiang. Optical flow in the presence of spatially-varying motion blur. In *CVPR*, 2012.

[26] W. Ren, X. Cao, J. Pan, X. Guo, W. Zuo, and M.-H. Yang. Image deblurring via enhanced low-rank prior. *TIP*, 25(7):3426–3437, 2016.

[27] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf. Learning to deblur. *TPAMI*, 38(7):1439–1451, 2016.

[28] L. Sevilla-Lara, D. Sun, V. Jampani, and M. J. Black. Optical flow with semantic segmentation and localized layers. In *CVPR*, 2016.

[29] F. Šroubek and P. Milanfar. Robust multichannel blind deconvolution via fast alternating minimization. *TIP*, 21(4):1687–1700, 2012.

[30] S. Su, M. Delbracio, J. Wang, G. Sapiro, W. Heidrich, and O. Wang. Deep video deblurring. *arXiv preprint arXiv:1611.08387*, 2016.

[31] D. Sun, J. Wulff, E. B. Sudderth, H. Pfister, and M. J. Black. A fully-connected layered model of foreground and background flow. In *CVPR*, 2013.

[32] K. Sunkavalli, N. Joshi, S. B. Kang, M. F. Cohen, and H. Pfister. Video snapshots: Creating high-quality images from video clips. *TVCG*, 18(11):1868–1879, 2012.

[33] Y.-W. Tai, H. Du, M. S. Brown, and S. Lin. Image/video deblurring using a hybrid camera. In *CVPR*, 2008.

[34] Y.-W. Tai, P. Tan, and M. S. Brown. Richardson-lucy deblurring for scenes under a projective motion path. *TPAMI*, 33(8):1603–1618, 2011.

[35] J. Tang, X. Shu, G.-J. Qi, Z. Li, M. Wang, S. Yan, and R. Jain. Tri-clustered tensor completion for social-aware image tag refinement. *TPAMI*, 39(8):1662–1674, 2017.

[36] Y.-H. Tsai, M.-H. Yang, and M. J. Black. Video segmentation via object flow. In *CVPR*, 2016.

[37] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers. An improved algorithm for TV-L1 optical flow. In *Statistical and Geometrical Approaches to Visual Motion Analysis*, pages 23–45, 2009.

[38] O. Whyte, J. Sivic, A. Zisserman, and J. Ponce. Non-uniform deblurring for shaken images. *IJCV*, 98(2):168–186, 2012.

[39] J. Wulff and M. J. Black. Modeling blurred video with layers. In *ECCV*, 2014.

[40] L. Xu, S. Zheng, and J. Jia. Unnatural L0 sparse representation for natural image deblurring. In *CVPR*, 2013.

[41] Y. Yan, W. Ren, Y. Guo, R. Wang, and X. Cao. Image deblurring via extreme channels prior. In *CVPR*, 2017.

[42] J. Yuan, C. Schörr, and G. Steidl. Simultaneous higher-order optical flow estimation and decomposition. *SIAM Journal on Scientific Computing*, 29(6):2283–2304, 2007.

[43] T. Yue, S. Cho, J. Wang, and Q. Dai. Hybrid image deblurring by fusing edge and power spectrum information. In *ECCV*, 2014.

[44] H. Zhang and J. Yang. Intra-frame deblurring by leveraging inter-frame camera motion. In *CVPR*, 2015.