Blind Deconvolution
From Model-Based to Deep Learning

Paolo Favaro
Computer Vision Group — University of Bern

Meiguang Jin  Givi Meishvili  Stefan Roth  Zhe Hu  Daniele Perrone

NTIRE 2019 — Long Beach, CA
Motion blur is caused by object and/or camera motion during the exposure interval.
Motion Blur

Motion blur is caused by object and/or camera motion during the exposure interval.
Short Exposure Images

Synthetic Long Exposure Images
Deep learning approach

- Need to collect ground truth data: (blur image, sharp video sequence)
- Use high frame rate cameras, average frames to simulate blurry image, use the average as input and the sharp frames as output
- Need to address temporal ambiguities (e.g., forward or backward ordering yields the same blurry image), otherwise learning cannot succeed
- Use a sequence order-invariant loss function
Jin, Meishvili, Favaro Learning to Extract a Video Sequence from a Single Motion-Blurred Image CVPR 2018
Jin, Meishvili, Favaro Learning to Extract a Video Sequence from a Single Motion-Blurred Image CVPR 2018
Jin, Meishvili, Favaro Learning to Extract a Video Sequence from a Single Motion-Blurred Image CVPR 2018
Jin, Meishvili, Favaro Learning to Extract a Video Sequence from a Single Motion-Blurred Image CVPR 2018
Slow motion & deblurring from a blurry video

input (30 FPS)
Slow motion & deblurring from a blurry video

output (300 FPS)
Slow motion & deblurring from a blurry video

input (30 FPS)
Slow motion & deblurring from a blurry video output (300 FPS)
Deep learning approaches

**pros**

- Can handle scenes of high complexity
- No need to manually design models/priors
- No need to design custom optimization procedures
- Extremely fast execution

**cons**

- Not state of the art in existing datasets (Nah et al @ -2dB PSNR from best model-based)
- No direct control/guarantees on the artifacts
Model-based approaches

- If the camera translates along the X-Y axes and the scene is a fronto-parallel plane (or at infinity) a simple blur model is

\[ f = k * u + n \]
Model-based approaches

- If the camera translates along the X-Y axes and the scene is a fronto-parallel plane (or at infinity), a simple blur model is

\[ f = k \ast u + n \]

- blurry image
- kernel
- sharp image
- noise
Model-based approaches

- If the camera translates along the X-Y axes and the scene is a fronto-parallel plane (or at infinity), a simple blur model is

\[ f = k \ast u + n \]
Model-based approaches

- If the camera translates along the X-Y axes and the scene is a fronto-parallel plane (or at infinity) a simple blur model is

\[ f = k \ast u + n \]
Model-based approaches

- If the camera translates along the X-Y axes and the scene is a fronto-parallel plane (or at infinity), a simple blur model is

\[ f = k \ast u + n \]
Blind deconvolution

• Recover **both** the blur kernel and the sharp image given the blurry image

\[ f = k \ast u + n \]

• By using Maximum a Posteriori it can be posed as an optimization problem with some image prior (eg Total Variation*)

\[ \min_{u,k} \lambda |\nabla u|_{2,1} + \frac{1}{2} |f - k \ast u|^2 \]

*Chan and Wong *Total Variation Blind Deconvolution* TIP 1998 (also You and Kaveh 1996)
A little problem

- The TV prior has a little flaw

\[
\min_{u,k} \lambda |\nabla u|_{2,1} + \frac{1}{2} |f - k \ast u|_2^2
\]

*Levin et al Understanding and evaluating blind deconvolution algorithms CVPR 2009*
A little problem

• The TV prior has a little flaw

\[
\min_{u,k} \lambda |\nabla u|_{2,1} + \frac{1}{2} |f - k \ast u|_2^2
\]

• Compare the true solution \((u, k)\) with the no-blur solution \((f, \delta)\)

\[f \equiv \delta \ast f \equiv k \ast u\]

*Levin et al Understanding and evaluating blind deconvolution algorithms CVPR 2009*
A little problem

• The TV prior has a little flaw

\[
\min_{u,k} \lambda|\nabla u|_{2,1} + \frac{1}{2}|f \ast k \ast u|_2^2
\]

• Compare the true solution \((u, k)\) with the no-blur solution \((f, \delta)\)

\[
f \equiv \delta \ast f \equiv k \ast u
\]

*Levin et al Understanding and evaluating blind deconvolution algorithms CVPR 2009*
A little problem

• The TV prior has a little flaw

\[
\min_{u,k} \lambda |\nabla u|_{2,1} + \frac{1}{2} |f \ast \delta * u|_2^2
\]

• Compare the true solution \((u, k)\) with the no-blur solution \((f, \delta)\)

\[
f \equiv \delta * f \equiv k * u
\]

• Only the image prior is left in the cost, but the prior favors the no-blur solution!

\[
|\nabla f|_{2,1} \leq |\nabla u|_{2,1}
\]

*Levin et al Understanding and evaluating blind deconvolution algorithms CVPR 2009*
Revisiting total variation BD

• The complete problem statement is

$$\min_{u,k} \lambda \left| \nabla u \right|_{2,1} + \frac{1}{2} \left| f - k^* u \right|_2^2$$

s.t. $k \succeq 0, \quad |k|_1 = 1$

where the constraints on the blur kernel ensure that the blur is non negative and adds up to 1 (or, equivalently, its $L_1$ norm is 1)

• The $L_1$ norm constraint fixes the scale ambiguity between $u$ and $k$; without it the minimization would make the scale of $u$ tend to 0 and make the image prior irrelevant
Fixing the scale ambiguity

• The complete problem statement is

\[
\min_{u, k} \lambda |\nabla u|_{2,1} + \frac{1}{2} |f - k^* u|^2_2 \\
\text{s.t. } k \geq 0, \quad |k|_1 = 1
\]

• If all we need is to fix the scale of k, then $L_p$ norms could be used too

• Would $p \neq 1$ make a difference?
L_p normalization

• The new problem statement is

\[
\min_{z,w} \lambda \| \nabla z \|_{2,1} + \frac{1}{2} \| f - w^* z \|_2^2 \\
\text{s.t. } w \geq 0, \quad \| w \|_p = 1
\]
L_p normalization

• The new problem statement is

\[
\min_{z, w} \lambda \| \nabla z \|_{2,1} + \frac{1}{2} \| f - w^* z \|_2^2
\]

s.t. \( w \geq 0, \quad |w|_p = 1 \)

• Now substitute \( k = w / |w|_1 \) and \( u = |w|_1 z \)
L_p normalization

• The new problem statement is

\[
\min_{z,w} \lambda \| \nabla z \|_{2,1} + \frac{1}{2} \| f - w^* z \|_2^2
\]

s.t. \( w \geq 0, \quad |w|_p = 1 \)

• Now substitute \( k = w / |w|_1 \) and \( u = |w|_1 z \)

• Obtain the equivalent formulation

\[
\min_{u,k} \lambda |k|_p \| \nabla u \|_{2,1} + \frac{1}{2} |f - k^* u|_2^2
\]

s.t. \( k \geq 0, \quad |k|_1 = 1 \)
L_p normalization

- The new problem statement is

\[
\begin{align*}
\min_{z,w} \lambda \| \nabla z \|_{2,1} + & \frac{1}{2} \| f - w^* z \|_2^2 \\
\text{s.t. } w \geq 0, & \ |w|_p = 1
\end{align*}
\]

- Now substitute \( k = w/|w|_1 \) and \( u = |w|_1 z \)

- Obtain the equivalent formulation

\[
\begin{align*}
\min_{u,k} \lambda |k|_p \| \nabla u \|_{2,1} + & \frac{1}{2} \| f - k^* u \|_2^2 \\
\text{s.t. } k \geq 0, & \ |k|_1 = 1
\end{align*}
\]

which has a regularization parameter that depends on the blur L_p norm.
L_p normalization

- The equivalent formulation is almost like the previous total variation form

\[
\min_{u, k} \lambda |k|_p |\nabla u|_{2,1} + \frac{1}{2} \left| f - k * u \right|_2^2 \\
\text{s.t. } k \geq 0, \quad |k|_1 = 1
\]
L_p normalization

- The equivalent formulation is almost like the previous total variation form
  \[
  \min_{u,k} \lambda \|k\|_p \|\nabla u\|_{2,1} + \frac{1}{2} \left\| f - k^* u \right\|_2^2
  \]
  s.t. \( k \geq 0, \quad \|k\|_1 = 1 \)

- Let us compare now the true solution \((u,k)\) with the no-blur solution \((f,\delta)\)
  \[\|k\|_p \|\nabla u\|_{2,1} \quad \text{vs} \quad \|\nabla f\|_{2,1}\]
L_p normalization

• The equivalent formulation is almost like the previous total variation form

\[
\min_{u,k} \lambda |k|_p |\nabla u|_{2,1} + \frac{1}{2} \left| f - k^* u \right|_2^2 \\
\text{s.t. } k \geq 0, \quad |k|_1 = 1
\]

• Let us compare now the true solution \((u, k)\) with the no-blur solution \((f, \delta)\)

\[|k|_p |\nabla u|_{2,1} \quad \text{vs} \quad |\nabla f|_{2,1}\]

• When \( p = 2 \) the term \(|k|_p < 1 \) if \( k \neq \delta \) and this makes the LHS term small
Rescuing the TV prior

- **Theorem** Assume the gradients of the true sharp image $u$ to be i.i.d. zero-mean Gaussian and the true blur kernel $k$ to have finite support. Given a blurry image $f = k * u$, the new formulation favors with high probability the true blur/image pair $(u, k)$ over the trivial no-blur pair $(f, \delta)$ for $p \geq 2$. 

Jin, Roth, Favaro Normalized blind deconvolution ECCV 2018
Optimization

- We use the Frank-Wolfe algorithm and alternate between blur and image

- Advantages
  1. For the first time it is possible to optimize the cost function exactly
  2. Coarse to fine scheme is not needed
  3. Careful initialization is not necessary (can start with $k = \delta$)
  4. Regularization parameter is not changed during the iteration time
  5. The formulation is convex separately in each variable
Quantitative evaluation

Table 1: Quantitative comparison on the entire SUN dataset [39] (640 blurry images).

<table>
<thead>
<tr>
<th>Method</th>
<th>mean error ratio</th>
<th>maximum error ratio</th>
<th>failure cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Krishnan et al. [20]</td>
<td>12.015</td>
<td>142.668</td>
<td>475</td>
</tr>
<tr>
<td>Levin et al. [23]</td>
<td>6.695</td>
<td>44.171</td>
<td>357</td>
</tr>
<tr>
<td>Sun et al. [39]</td>
<td>2.581</td>
<td>35.765</td>
<td>44</td>
</tr>
<tr>
<td>Xu &amp; Jia [44]</td>
<td>3.817</td>
<td>75.036</td>
<td>98</td>
</tr>
<tr>
<td>Perrone &amp; Favaro [31]</td>
<td>2.114</td>
<td>8.517</td>
<td>7</td>
</tr>
<tr>
<td>Chakrabarti [4]</td>
<td>3.062</td>
<td>11.576</td>
<td>64</td>
</tr>
<tr>
<td>Michaeli &amp; Irani [24]</td>
<td>2.617</td>
<td>9.185</td>
<td>30</td>
</tr>
<tr>
<td>Pan et al. [29]</td>
<td>1.914</td>
<td>23.279</td>
<td>11</td>
</tr>
<tr>
<td>PN</td>
<td>2.299</td>
<td>6.764</td>
<td>8</td>
</tr>
<tr>
<td>FW</td>
<td>2.195</td>
<td>6.213</td>
<td>8</td>
</tr>
</tbody>
</table>

Sum of squared differences ratio \(\frac{|u - \hat{u}|^2}{|u - \hat{u}_*|^2}\)

\(\hat{u}_*\) estimated with GT kernel
\(\hat{u}\) estimated with estimated kernel
# Quantitative evaluation

## Table 2: Quantitative comparison on the small BSDS dataset [1] (72 blurry images).  

<table>
<thead>
<tr>
<th>Method</th>
<th>mean error ratio</th>
<th>maximum error ratio</th>
<th>failure cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sun et al. [39]</td>
<td>2.648</td>
<td>15.152</td>
<td>12</td>
</tr>
<tr>
<td>Xu &amp; Jia [44]</td>
<td>3.645</td>
<td>22.272</td>
<td>13</td>
</tr>
<tr>
<td>Perrone &amp; Favaro [31]</td>
<td>2.093</td>
<td>7.493</td>
<td>4</td>
</tr>
<tr>
<td>Michaeli &amp; Irani [24]</td>
<td>3.458</td>
<td>23.001</td>
<td>14</td>
</tr>
<tr>
<td>Pan et al. [29]</td>
<td>2.058</td>
<td>13.516</td>
<td>3</td>
</tr>
<tr>
<td>Yan et al. [46]</td>
<td>2.022</td>
<td>12.237</td>
<td>3</td>
</tr>
<tr>
<td>(L^1) normalization</td>
<td>2.211</td>
<td>7.821</td>
<td>3</td>
</tr>
<tr>
<td>weight decay (heuristic)</td>
<td>2.591</td>
<td>8.762</td>
<td>2</td>
</tr>
<tr>
<td>(L^2) blur prior (classic)</td>
<td>2.487</td>
<td>7.953</td>
<td>4</td>
</tr>
<tr>
<td>PN</td>
<td>2.011</td>
<td>4.676</td>
<td>0</td>
</tr>
<tr>
<td>FW</td>
<td>1.983</td>
<td>4.387</td>
<td>0</td>
</tr>
</tbody>
</table>

Sum of squared differences ratio \[ \frac{|u - \hat{u}|^2}{|u - \hat{u}_*|^2} \]  
\(\hat{u}_*\) estimated with GT kernel  
\(\hat{u}\) estimated with estimated kernel

Qualitative comparisons

- **input**
- Xu and Jia 2010
- Chakrabarti 2016
- Sun et al 2013
- Michaeli and Irani 2014
- Pan et al 2016
- Perrone and Favaro 2016
- FW
Worst cases in real images

input  Pan et al 2016  PN
Conclusions

• Deep learning methods will probably prevail in the end

• There are some limitations that might take time to address

• Can we trust that the reconstruction is not a hallucination of the data?
Conclusions

• In contrast model-based methods are easily interpretable

• There is still quite a bit to do even with simple formulations

• It pays to pay attention to the details