Video Reconstruction by Spatio-temporal Fusion of Blurred-Coded Image Pair

Anupama S¹, Prasan Shedligeri², Abhishek Pal³, Kaushik Mitra²

¹Qualcomm India, Bangalore, India
²Department of EE, IIT Madras, Chennai, India
³Mad Street Den, Chennai, India
Video extraction from a single blurred image

Motion ambiguity in the reconstructed video

Video from blurred image

[Jin et al., Purohit et al.]
Video extraction from a coded exposure image

Light inefficient due to exposure sequence

Video from coded exposure
[Raskar et al., Reddy et al., Holloway et al., Liu et al., Yoshida et al., Gupta et al.]
Complementary information from blurred-coded pair

<table>
<thead>
<tr>
<th></th>
<th>Motion information</th>
<th>Light efficiency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blurred Image</td>
<td>Ambiguous</td>
<td>100% light captured</td>
</tr>
<tr>
<td>Coded Image</td>
<td>Unambiguous</td>
<td>~50% light captured</td>
</tr>
</tbody>
</table>

**Objective:** Extract motion information from coded image and use the light efficiency of the blurred image
Coded-blurred image acquisition

Code, C

Coded image + Blurred Image

Complementary Code, 1-C

Coded-2-Bucket sensor

[Sarhangnejad et al. ‘19]
Coded-blurred video reconstruction

Blurred image

Coded image

Our Reconstruction algorithm

Reconstructed video
Video Reconstruction by Spatio-temporal Fusion of Blurred-Coded Image Pair
Video Reconstruction from Coded-Blurred Image Pair

- Coded image
- Inverse using code
- Low-res video
- Shared Encoder
- Extracted feature map
- Attention Block
- UNet
- Predicted high-res video

- Fully-exposed image
- Pixel shuffle
- Shuffled video
- Shared Encoder
- Extracted feature map
Extracting low-resolution videos from coded image

Full-resolution video from single coded image is ill-posed.

So, we make a local homogenous intensity assumption on the predicted video to solve for only low-resolution video sequence
Extracting low-resolution video from blurred image

Rearranging the pixels of an image into a video
Attention Block: Spatio-temporal fusion of blurred-coded pair

\[ \Phi_c \]

- Extracted feature map from coded images

\[ \Phi_f \]

- Extracted feature map from fully-exposed image

Attention Module

Multiply (1 - A)

Attention map (A)

Multiply (A)

\[ \Phi \]
Comparison of blurred vs. coded vs. coded-blurred video reconstruction
Exposure code used to obtain the coded images

We use a sequential impulse code of size 3x3x9 to generate our coded exposure images. The 3x3 code depicted here is repeated to cover the full size of each sharp sub-frame. 1 represents exposed pixel and 0 represents unexposed pixel.
Input Images

Ground truth video

Blurred image as input

Jin et al.
PSNR: 26.23 dB
SSIM: 0.939

Purohit et al.
PSNR: 30.51 dB
SSIM: 0.967

Coded image as input

GMM [Yang et al.]
PSNR: 33.20 dB
SSIM: 0.973

Ours
PSNR: 34.25 dB
SSIM: 0.980

Coded-blurred pair as input

GMM [Yang et al.]
PSNR: 35.22 dB
SSIM: 0.981

Ours
PSNR: 36.16 dB
SSIM: 0.986
<table>
<thead>
<tr>
<th>Input</th>
<th>Blurred Image</th>
<th>Coded Image</th>
<th>Coded + blurred</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algorithm</td>
<td>Jin et al.</td>
<td>Purohit et al.</td>
<td>Ours</td>
</tr>
<tr>
<td>PSNR</td>
<td>22.89</td>
<td>23.48</td>
<td>23.86</td>
</tr>
<tr>
<td>SSIM</td>
<td>0.865</td>
<td>0.879</td>
<td>0.861</td>
</tr>
</tbody>
</table>
Attention maps learned for coded-blurred fusion
Visualizing Learned Attention Maps

Blurred Image  Predicted Video  Predicted Attention Map
Visualizing Learned Attention Maps

- Blurred Image
- Predicted Video
- Predicted Attention Map
Summary

A framework for video recovery from coded-blurred image pairs

Exploiting complementary information from coded-blurred pairs for better video recovery

Attention map module for attending to the complementary information.

Better reconstruction performance than either coded image or blurred image alone