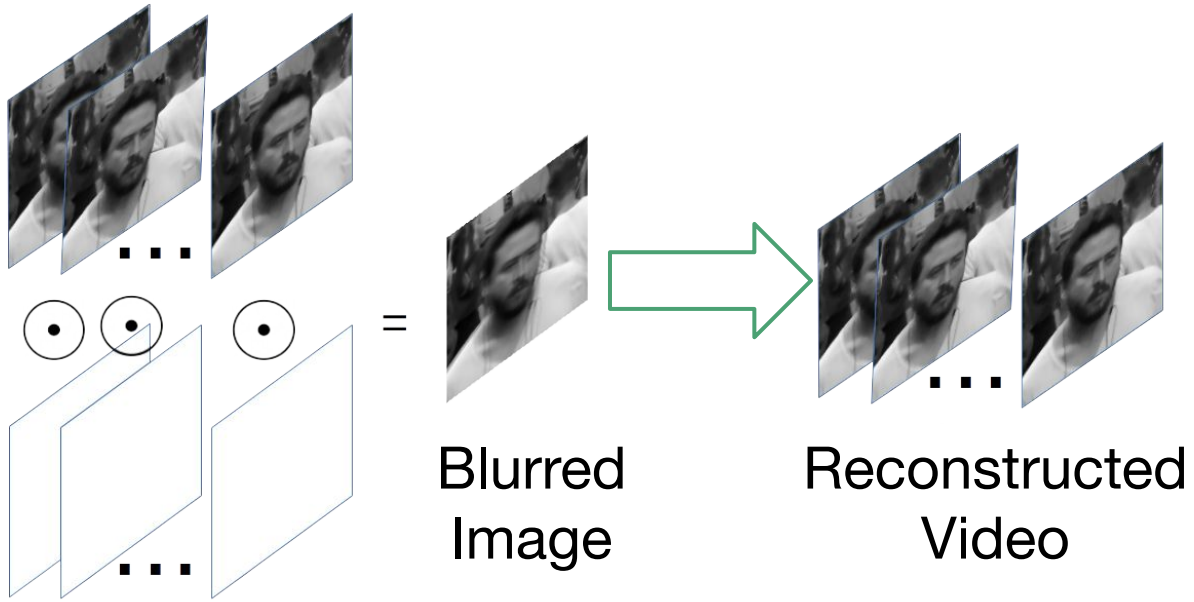
Video Reconstruction by Spatio-temporal Fusion of Blurred-Coded Image Pair

Anupama S[1], Prasan Shedligeri[2], Abhishek Pal[3], Kaushik Mitra[2]

[1]Qualcomm India, Bangalore, India
[2]Department of EE, IIT Madras, Chennai, India
[3]Mad Street Den, Chennai, India

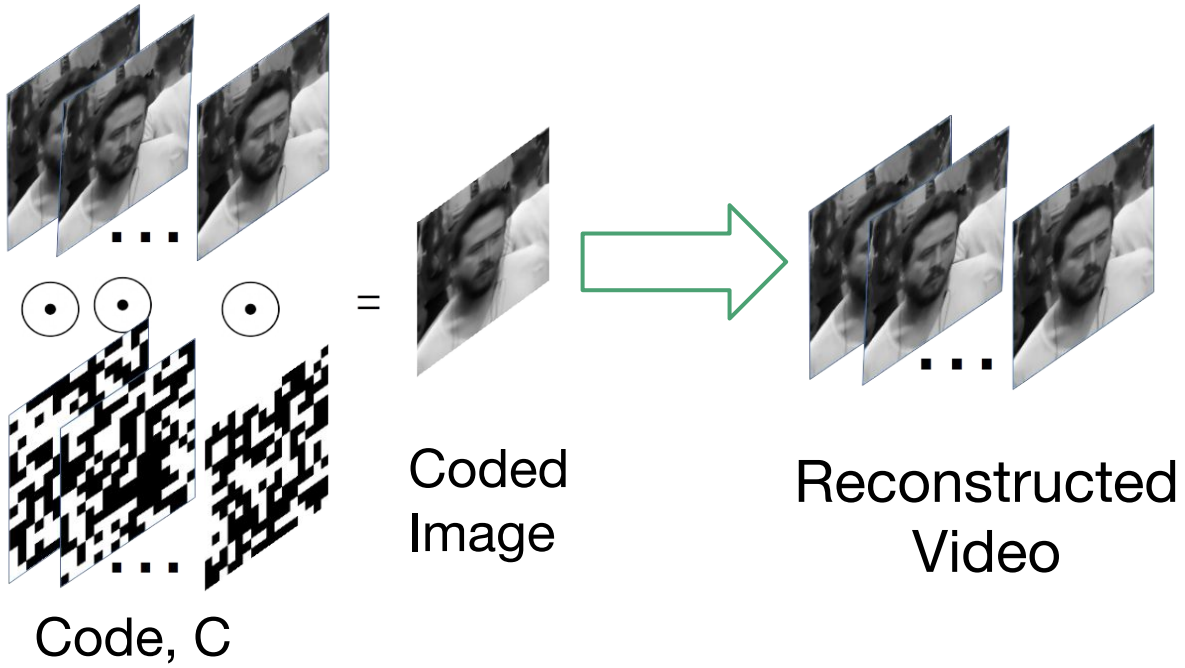
Video extraction from a single blurred image



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

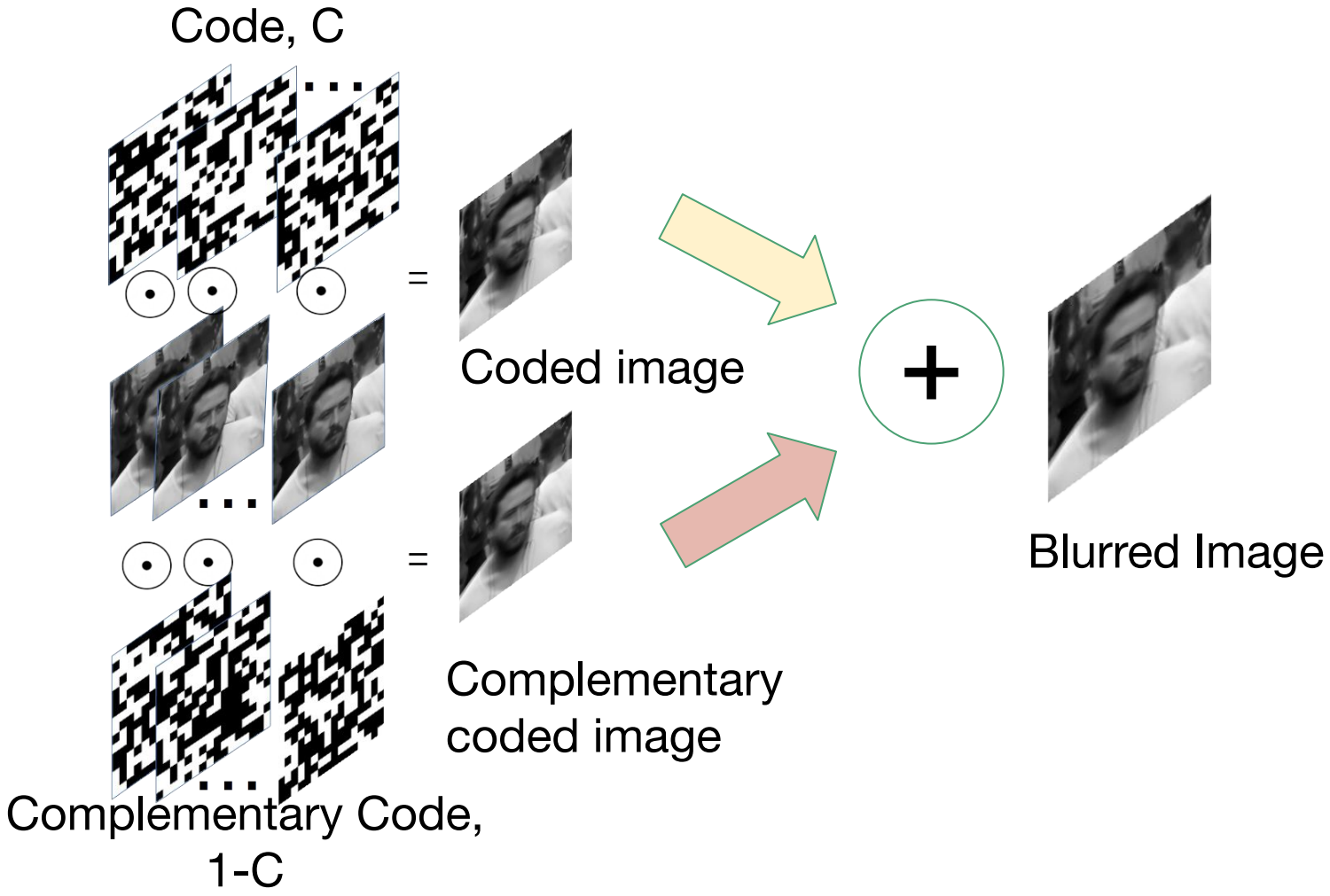
[Raskar et al., Reddy et al., Holloway et al., Liu et al., Yoshida et al., Gupta et al.]

Complementary information from blurred-coded pair

	Motion information	Light efficiency
Blurred Image	Ambiguous	100% light captured
Coded Image	Unambiguous	~50% light captured

Objective: Extract motion information from coded image and use the light efficiency of the blurred image

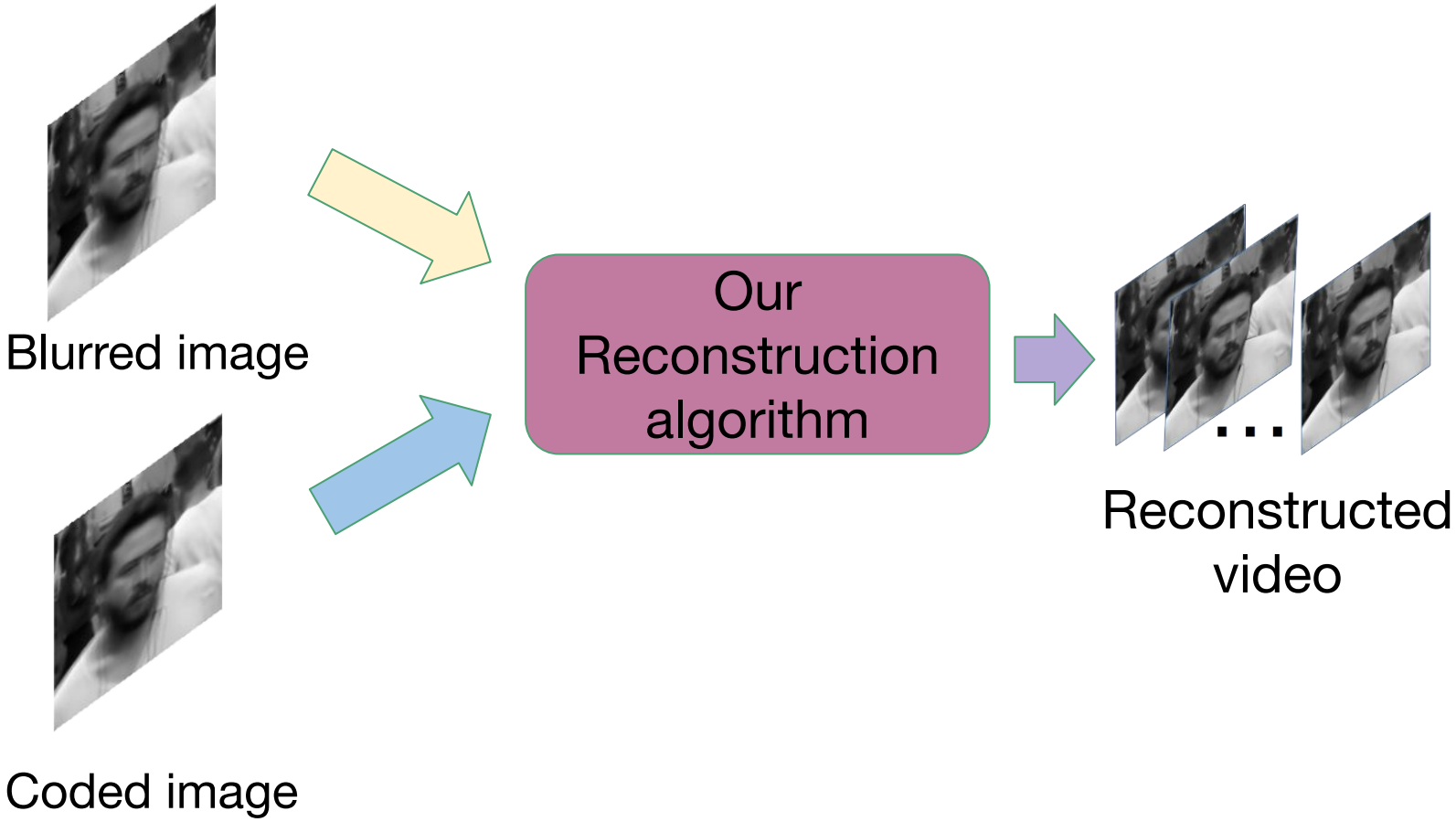
Coded-blurred image acquistion



Coded-2-Bucket sensor

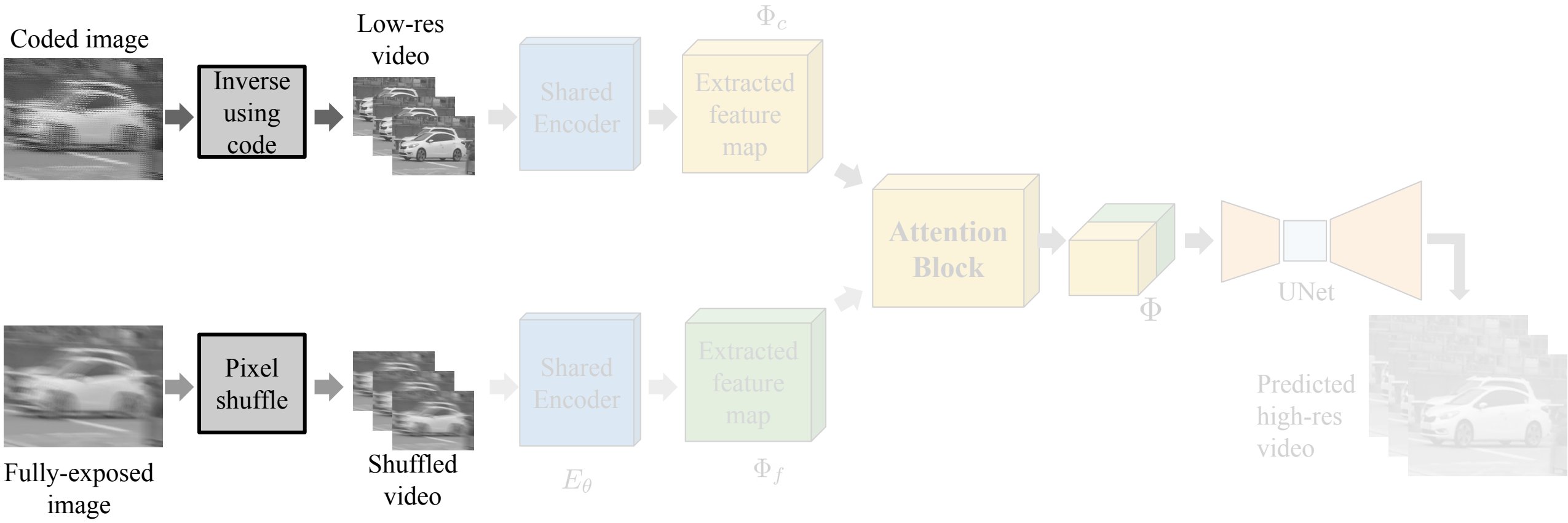
[Sarhangnejad et al. ‘19]

Coded-blurred video reconstruction



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

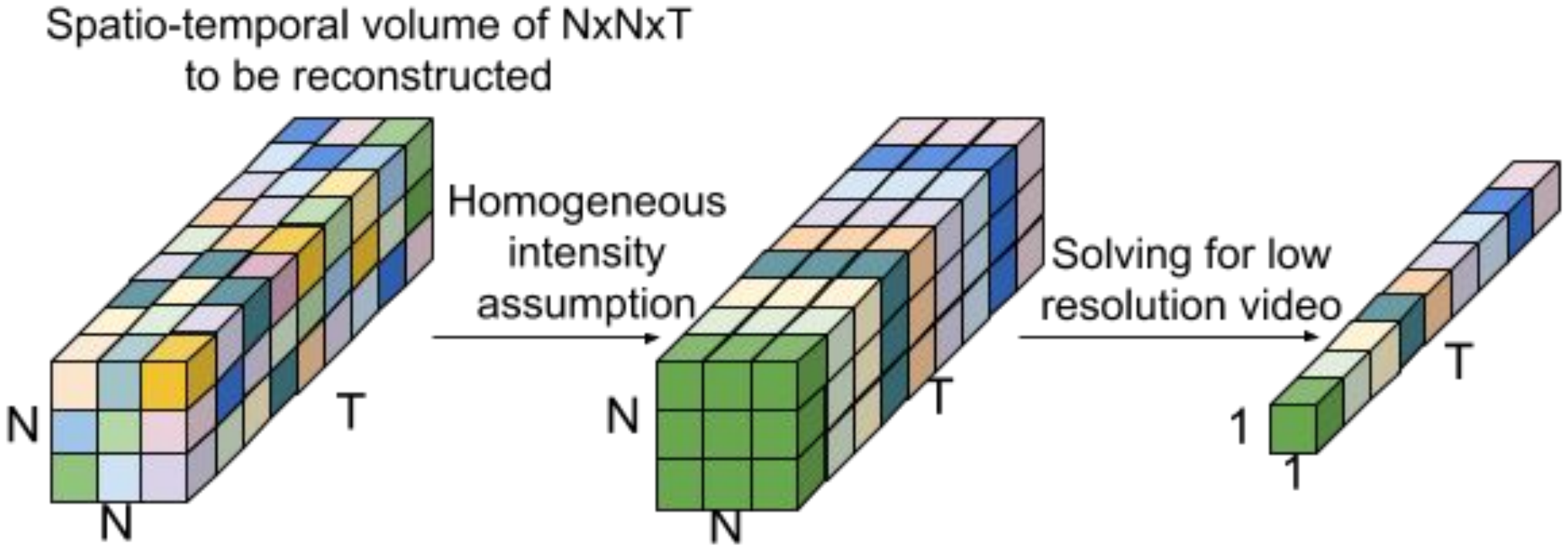
Video Reconstruction from Coded-Blurred Image Pair



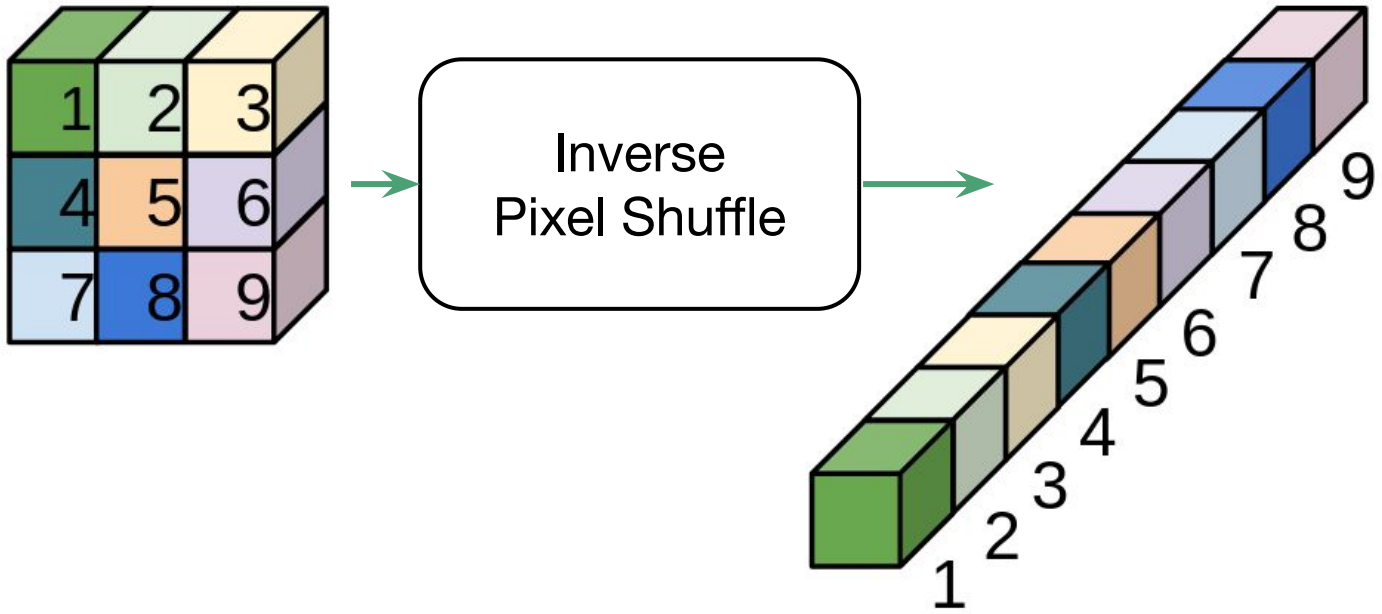
Extracting low-resolution videos from coded image

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

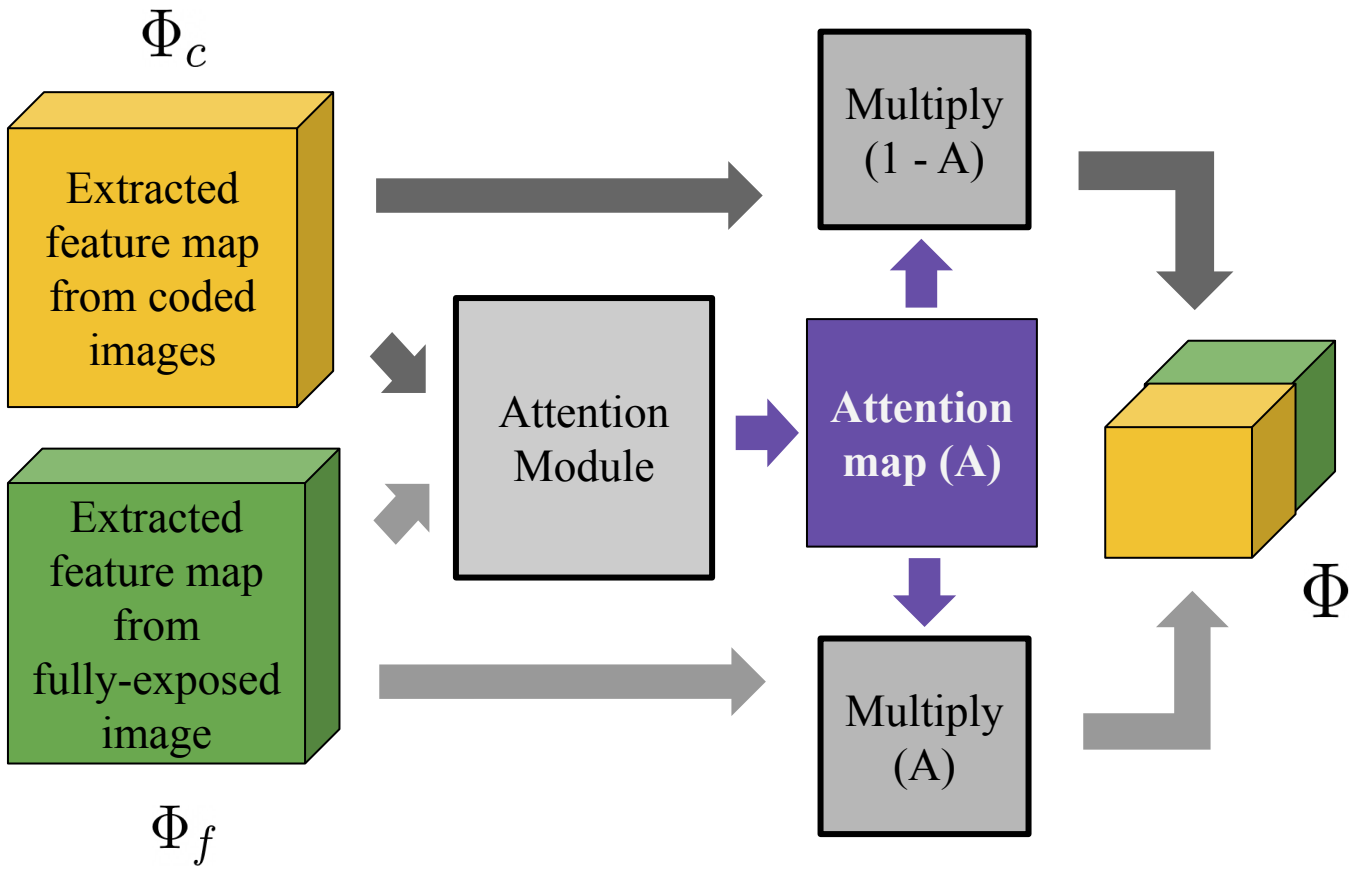


Extracting low-resolution video from blurred image



Rearranging the pixels of an image into a video

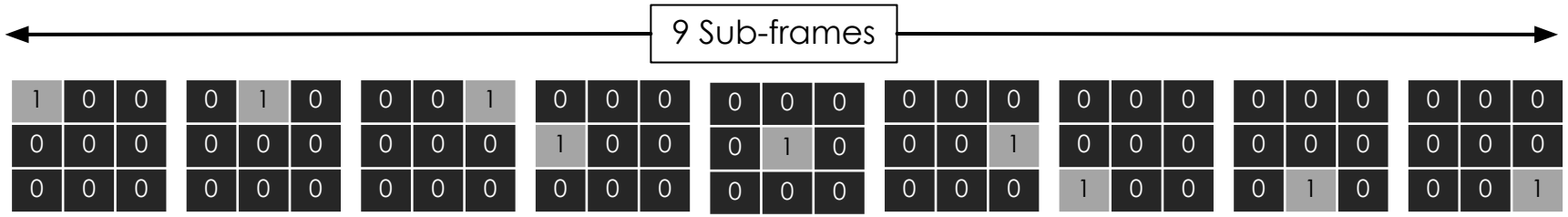
Attention Block: Spatio-temporal fusion of blurred-coded pair



Comparison of blurred vs. coded vs. coded-blurred video reconstruction

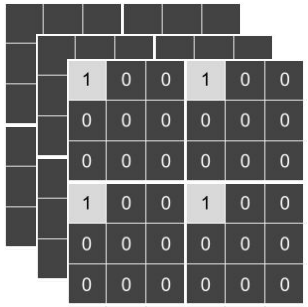
Exposure code used to obtain the coded images

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.

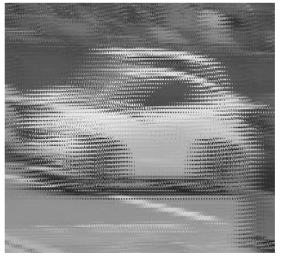


Ground truth frames

Pixel-wise multiply and average along time



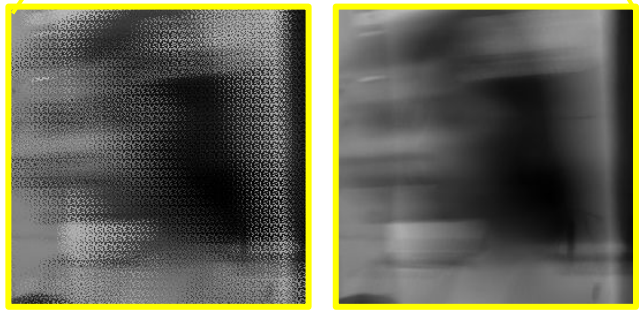
Repeated exposure code



Coded exposure image

Ground truth video

Input Images



Zoomed-in coded image

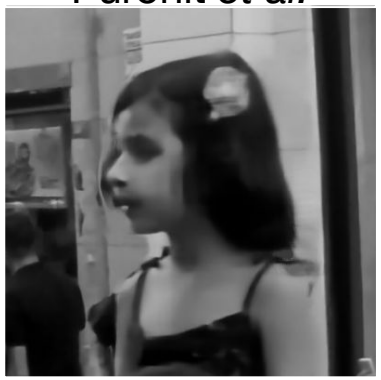
Zoomed-in blurred image

Blurred image as input

Jin *et al.*

PSNR: 26.23 dB
SSIM: 0.939

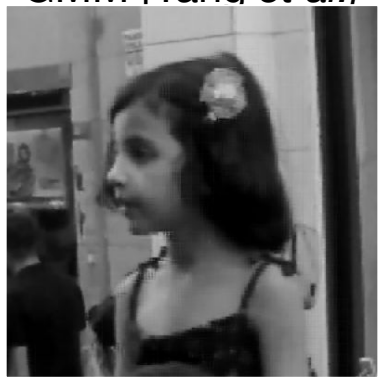
Purohit *et al.*



PSNR: 30.51 dB
SSIM: 0.967

Coded image as input

GMM [Yang *et al.*]



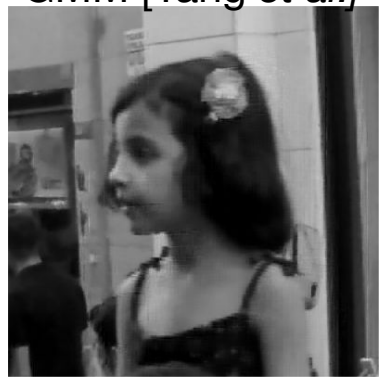
PSNR: 33.20 dB
SSIM: 0.973

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

Quantitative comparison

Input	**Blurred Image**			**Coded Image**		**Coded + blurred**	
Algorithm	Jin et al.	Purohit et al.	Ours	GMM [Yang et al.]	Ours	GMM [Yang et al.]	Ours
PSNR	22.89	23.48	23.86	30.27	32.52	32.39	**34.09**
SSIM	0.865	0.879	0.861	0.938	0.962	0.955	**0.971**

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