Saliency Guided Wavelet Compression for Low-Bitrate Image and Video Coding

Souptik Barua[1], Kaushik Mitra[2], and Ashok Veeraraghavan[1]

[1] Rice University, [2] Indian Institute of Technology, Madras

December 16, 2015

RICE®

DSP

MOTIVATION

Image/video coding so far has been studied keeping in mind
a particular end user in mind

Image/video coding so far has been studied keeping in mind a particular end user in mind

RICE

DSP

A new end user has emerged in the last 10 years!

A new end user has emerged in the last 10 years!

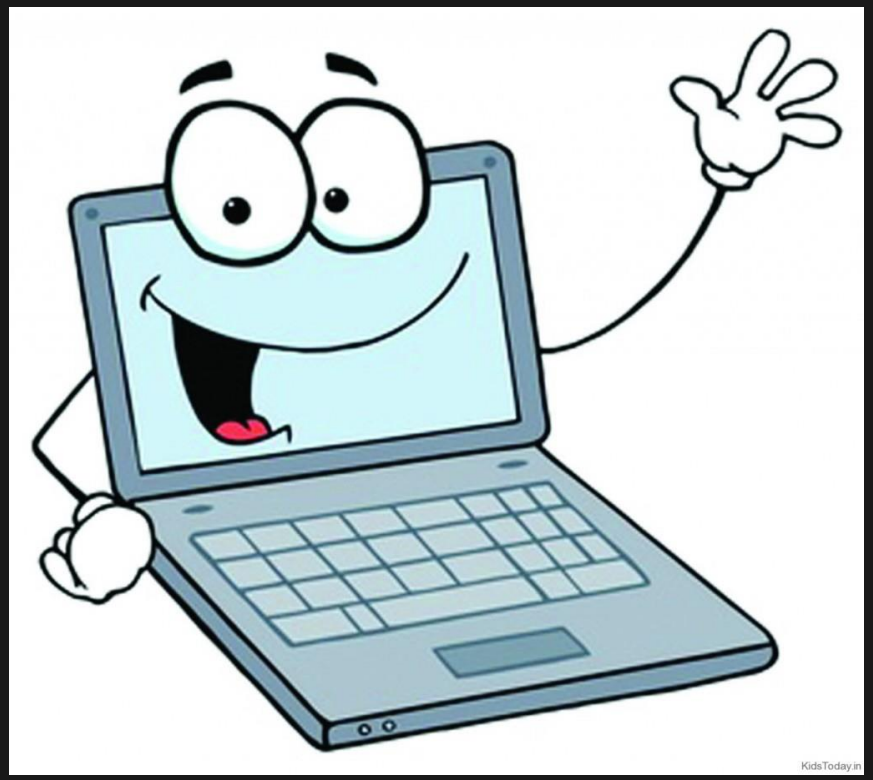
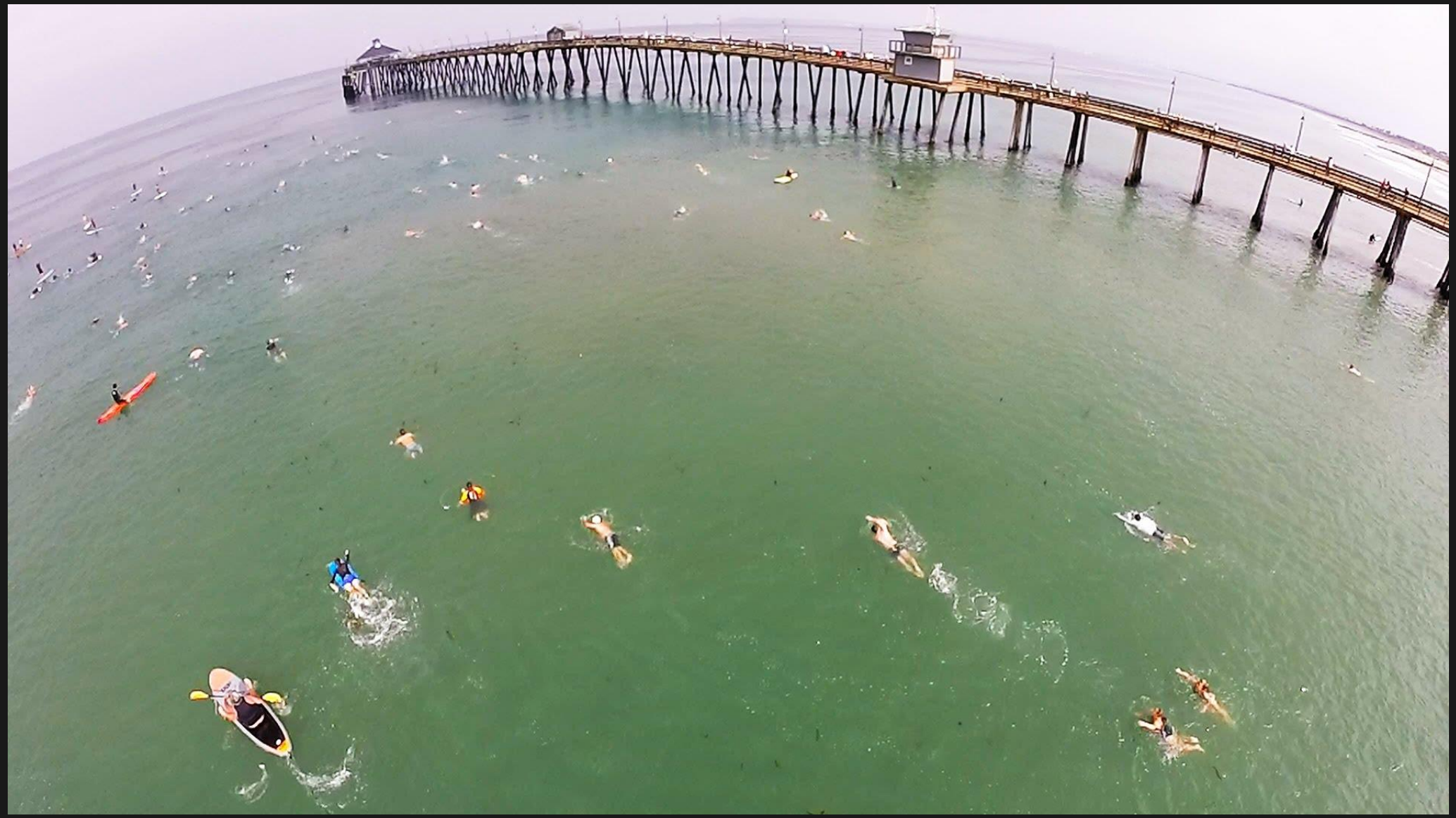


Image coding for Computer Vision tasks such as Object Detection?

Image coding for Computer Vision tasks such as Object Detection?

http://host.robots.ox.ac.uk:8080/pascal/VOC/voc2007/examples/person_06.jpg



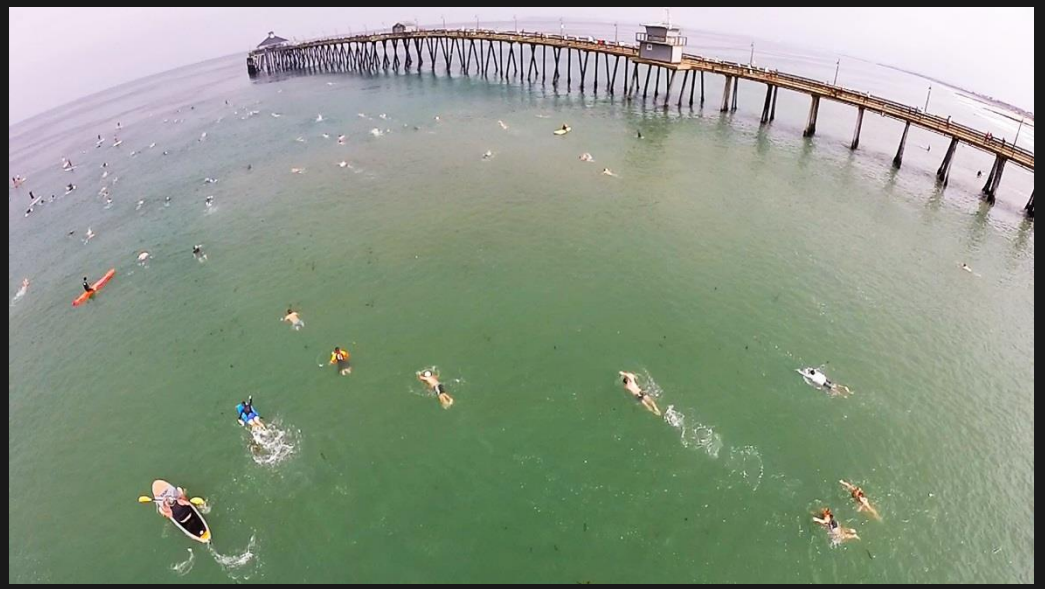
Scene

RICE

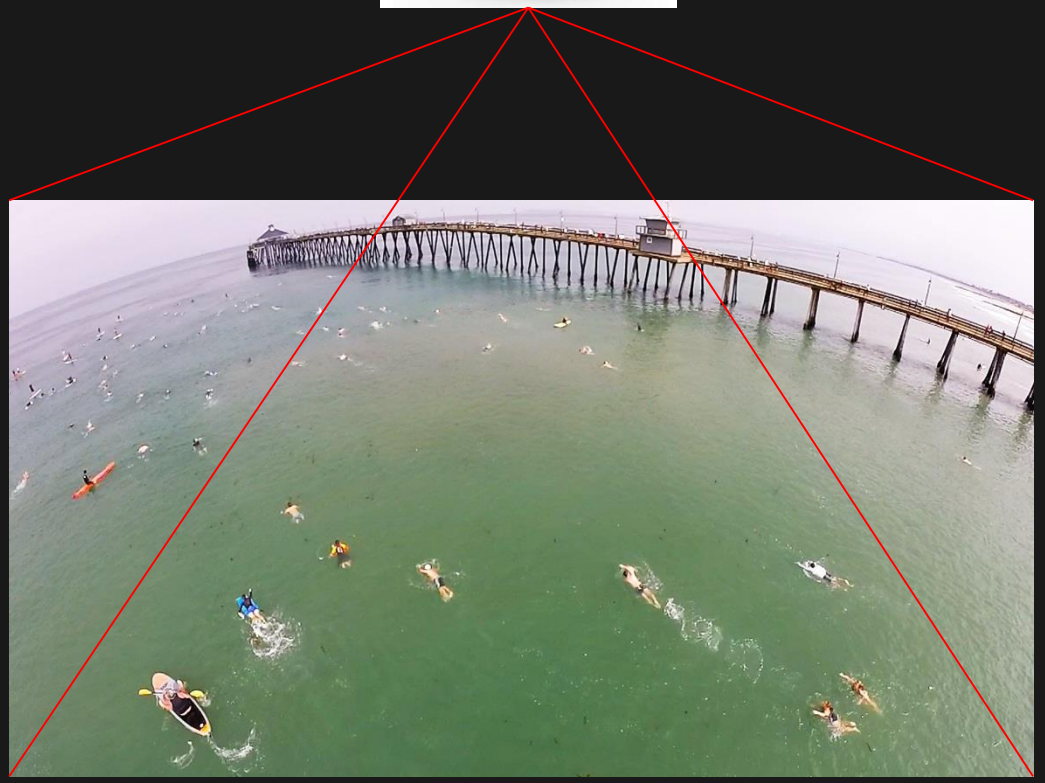
DSP

DJI Phantom drone

4K video capture



Scene



Scene

DJI Phantom drone

4K video capture

Issues

Scene

DJI Phantom drone

4K video capture

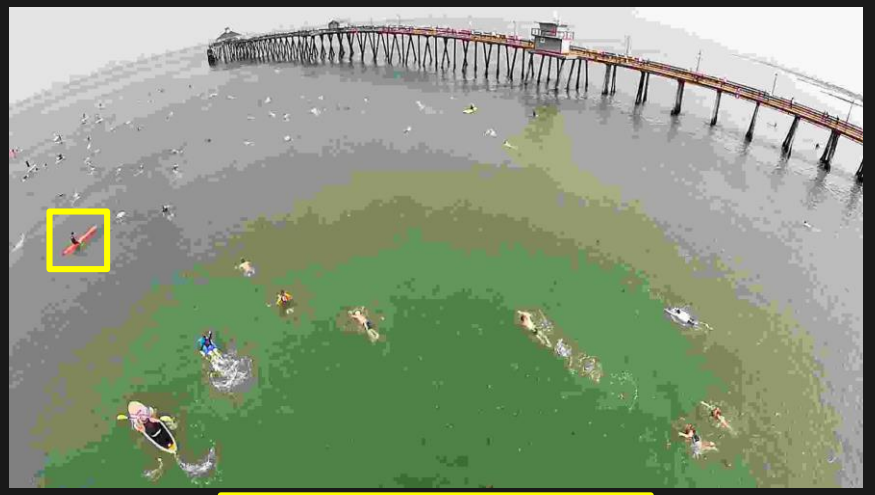
Bandwidth limited channel
(~1 Mbps*)

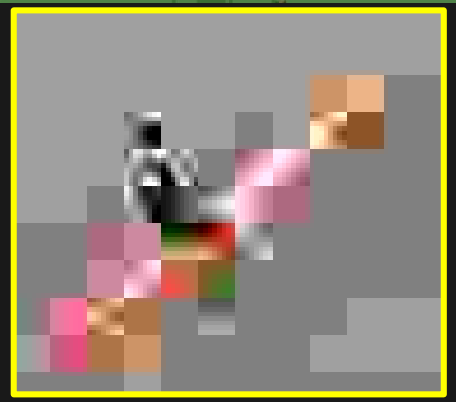
*https://en.wikipedia.org/wiki/Satellite_Internet_access

RICE

DSP

Bandwidth limited channel
(~1 Mbps)

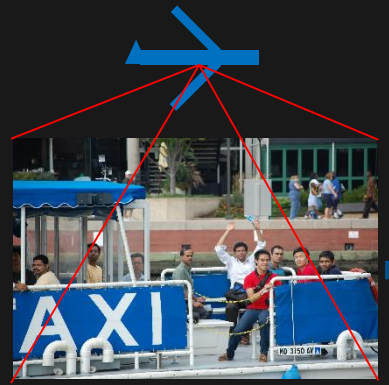




CONTRIBUTION

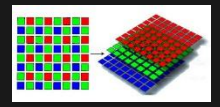
Let’s visit the JPEG pipeline!

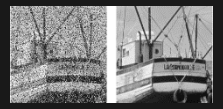
JPEG Pipeline



CMOS sensor

Image Processing





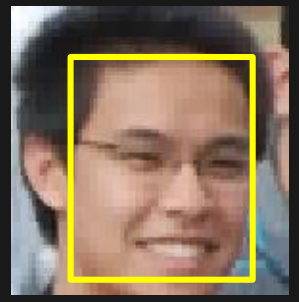
Compression

Transmit/store

Receive/load

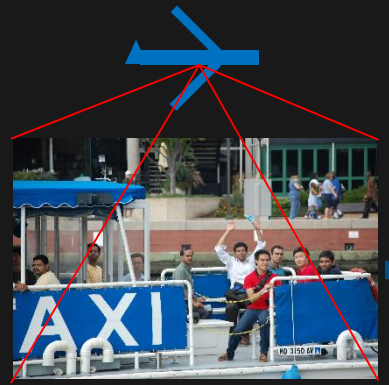
Decompression

Detection



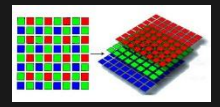
Lets remove some pieces, and add some!

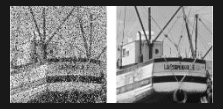
JPEG Pipeline



CMOS sensor

Image Processing





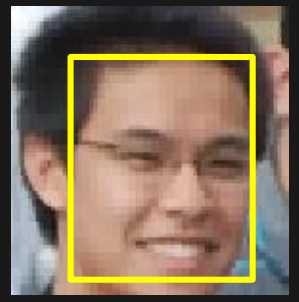
Compression

Transmit/store

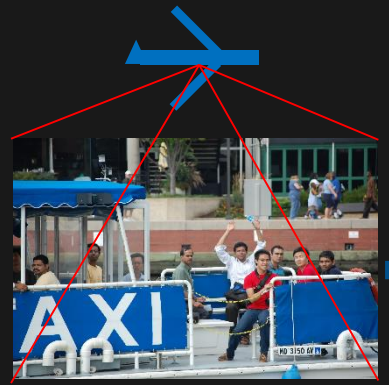
Receive/load

Decompression

Detection

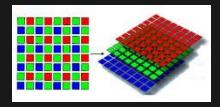


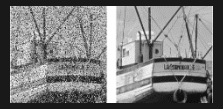
SBC Pipeline



CMOS sensor

Image Processing





Compression

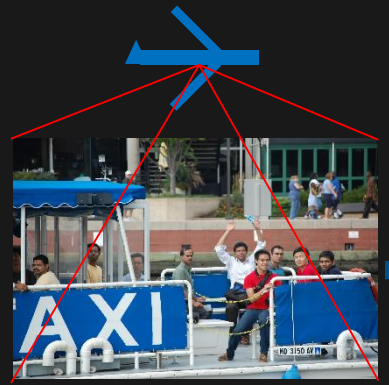
Transmit/store

Receive/load

Decompression

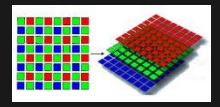
Detection

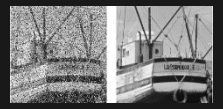
SBC Pipeline



CMOS sensor

Image Processing





Object Detection

Compression

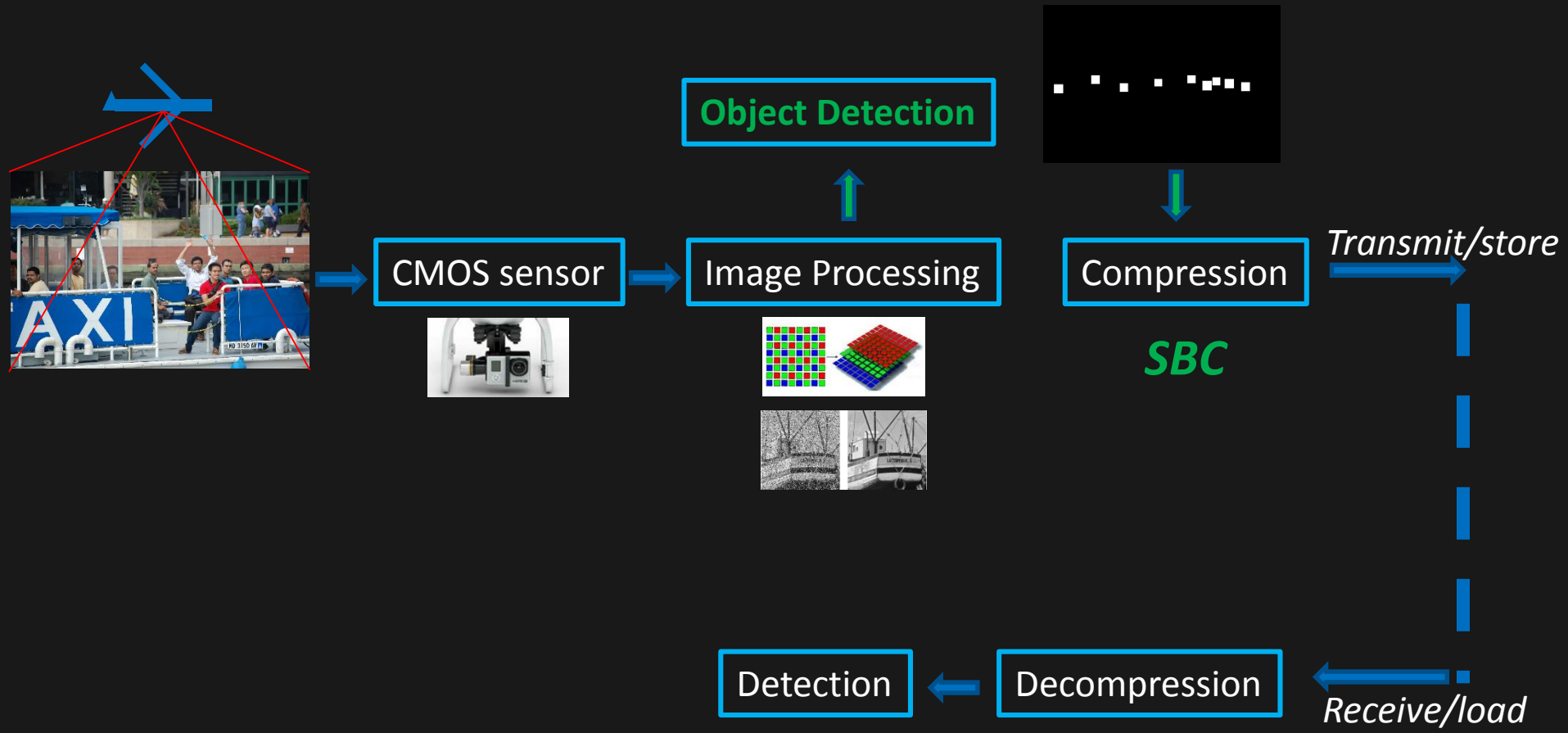
Transmit/store

Receive/load

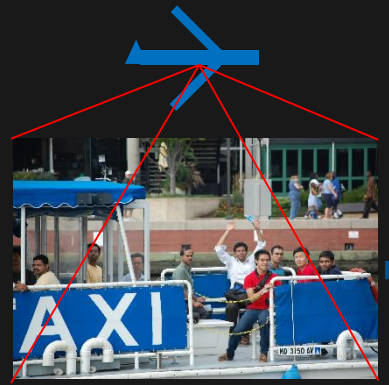
Decompression

Detection

SBC Pipeline

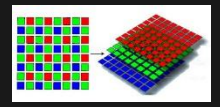


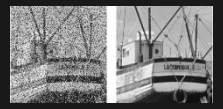
SBC Pipeline



CMOS sensor

Image Processing





Object Detection

Compression

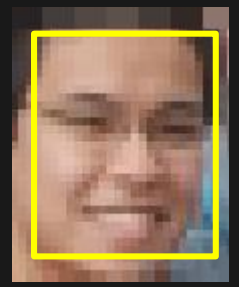
SBC

Transmit/store

Receive/load

Decompression

Detection



PROPOSED ALGORITHM

SBC: Related work

- JPEG-2000 ROI Encoder (2000)
- Bitplane-by-Bitplane Shift [Wang *et al* (2003)]
- Visual attention guided compression algorithms [Guo *et al* (2011), Shen *et al* (2013), Hadizadeh *et al*(2014)]

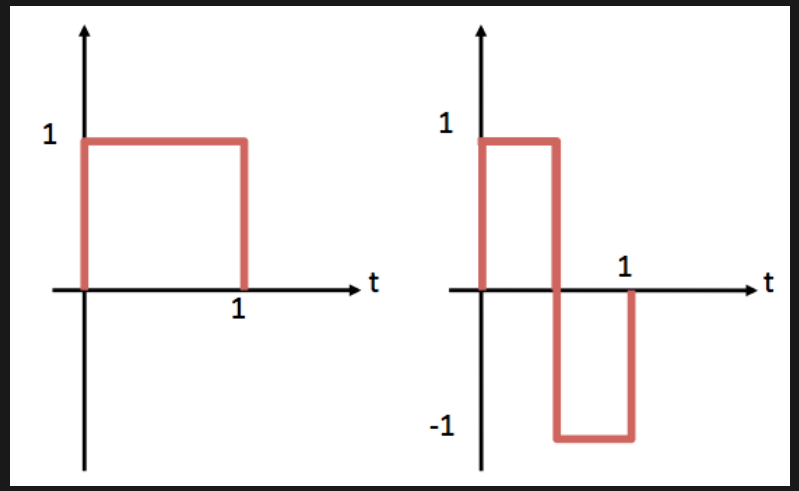
Limitations:

- Can handle only *two* levels of saliency
- Difficulty in *integrating* with JPEG-2000
- Underperforms when end application is *object detection*

Wavelet Transform

- Wavelets provide the localization property.
  - An object can be compactly represented by a few wavelet coefficients

- Haar Wavelets

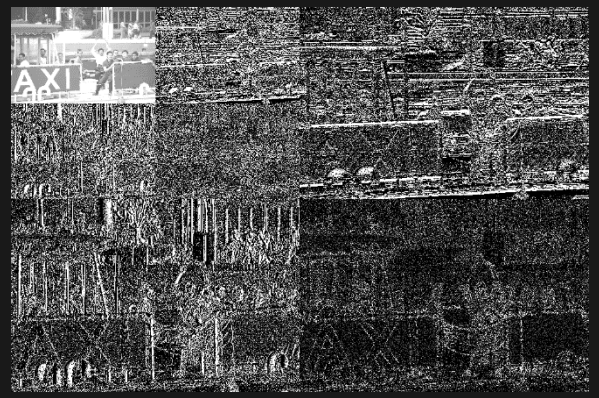


Wavelet Transform

Wavelet Transform

Wavelet transform
(level = 1)

Wavelet Transform



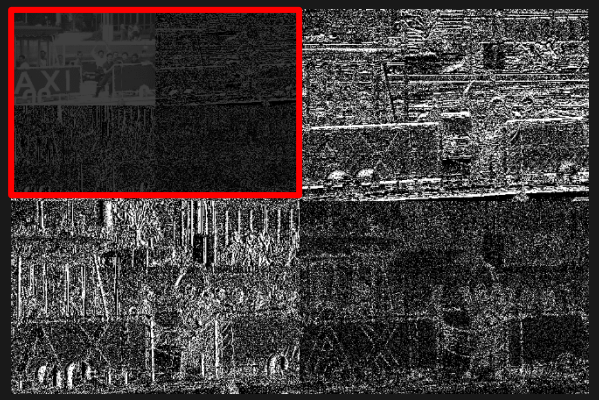
Wavelet transform
(level = 2)

Wavelet Transform

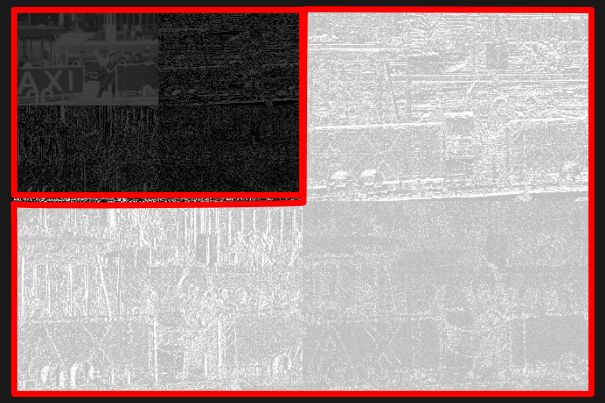
Wavelet transform
(level = 3)

How JPEG-2000 orders wavelet coefficients?

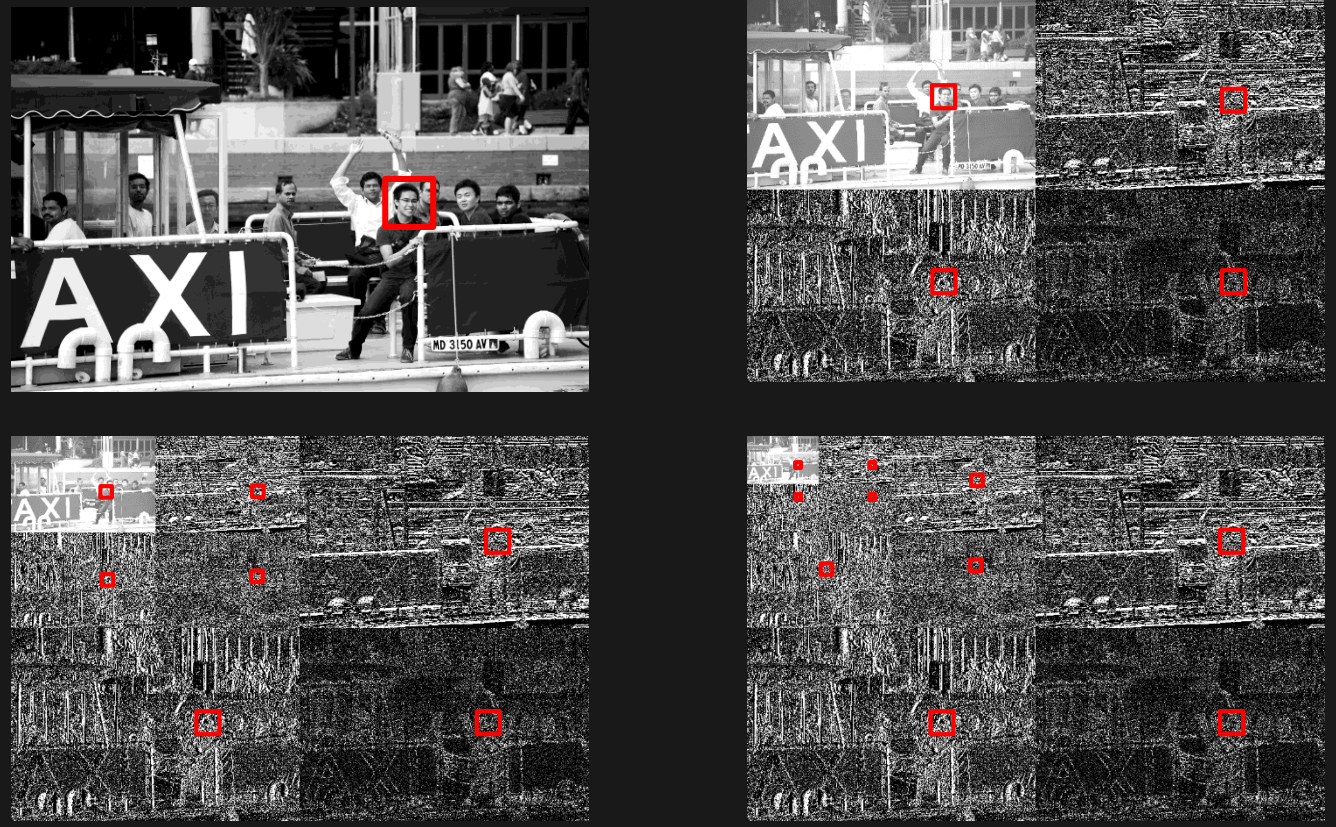

AXI
MD 3150 AV




AXI
MD 3150 AV



Wavelet Transform



SBC: Wavelet saliency computation

Original image

SBC: Wavelet saliency computation

Original image

SBC: Wavelet saliency computation

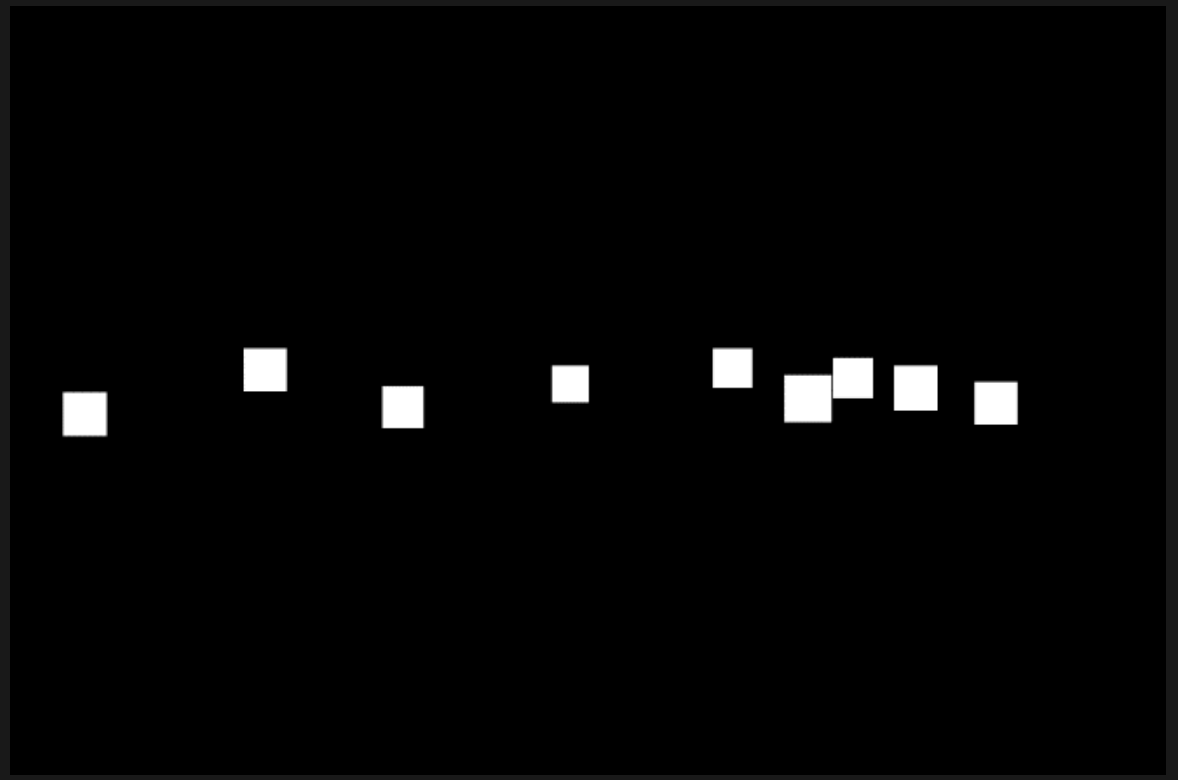


Image saliency map

SBC: Wavelet saliency computation

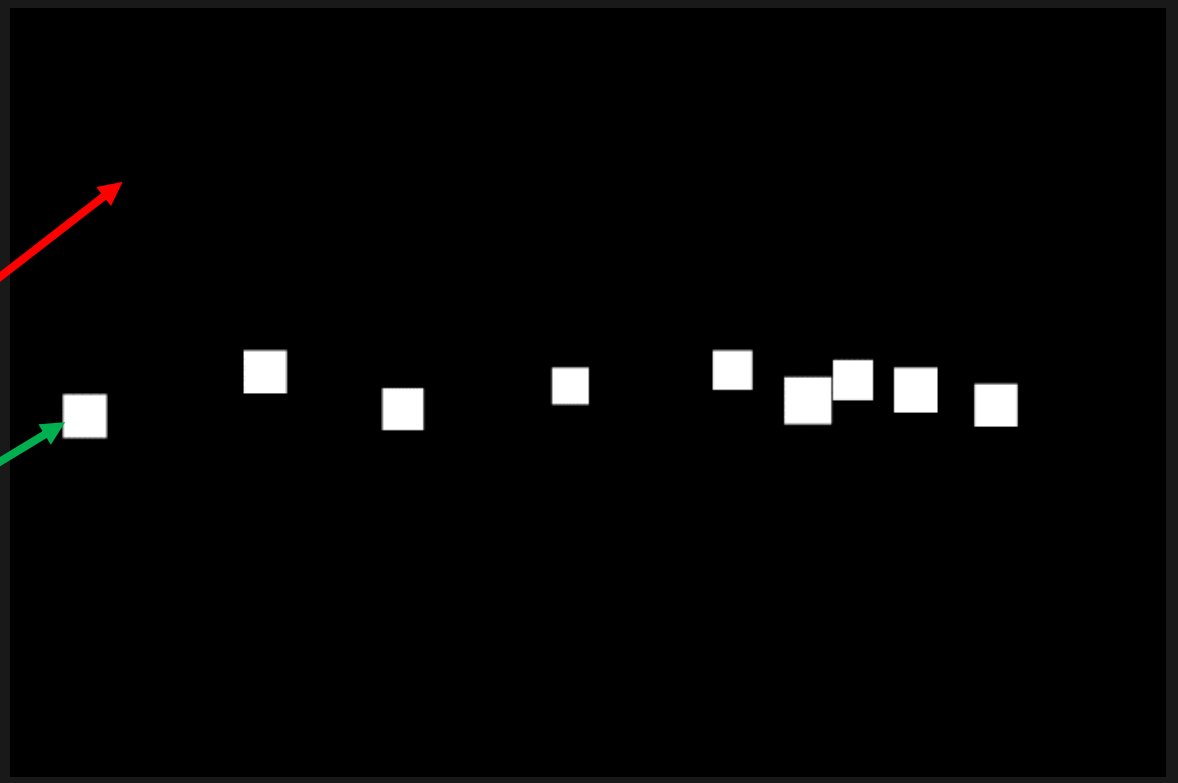


Image saliency map

No saliency;
Ordering as per wavelet level
(JPEG-2000)

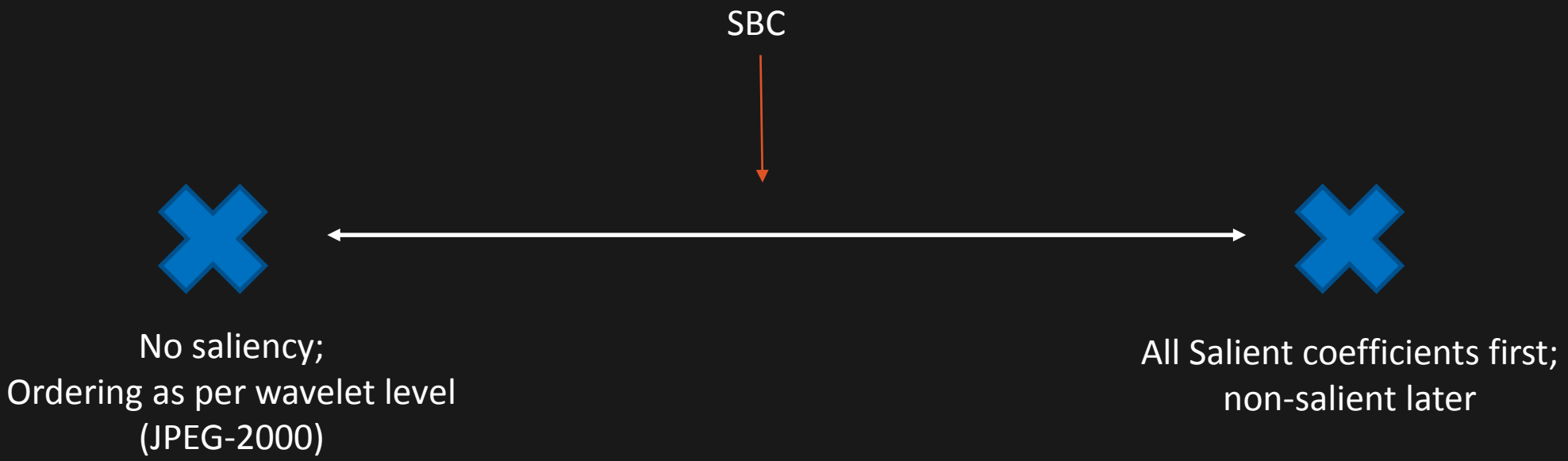
No saliency;
Ordering as per wavelet level
(JPEG-2000)

All Salient coefficients first;
non-salient later

RICE

DSP


SBC
No saliency;
Ordering as per wavelet level
(JPEG-2000)
All Salient coefficients first;
non-salient later



Flexibility in deciding relative importance of different objects in the image!

SBC: Wavelet saliency computation

Original image

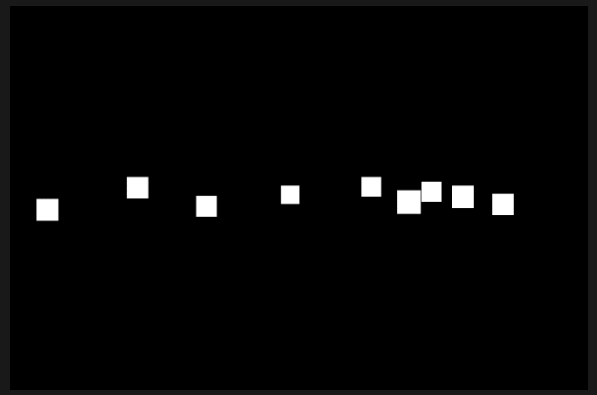


Image saliency map

SBC: Wavelet saliency computation

Original image

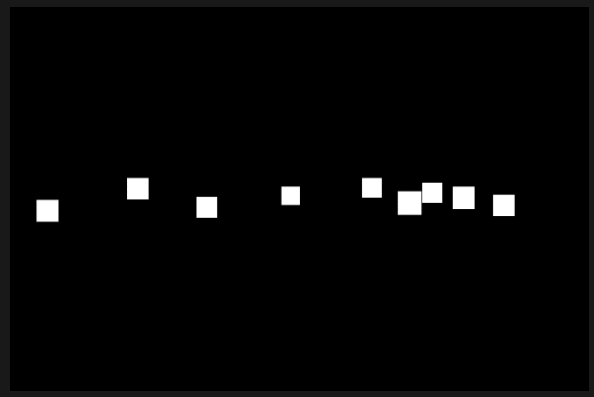
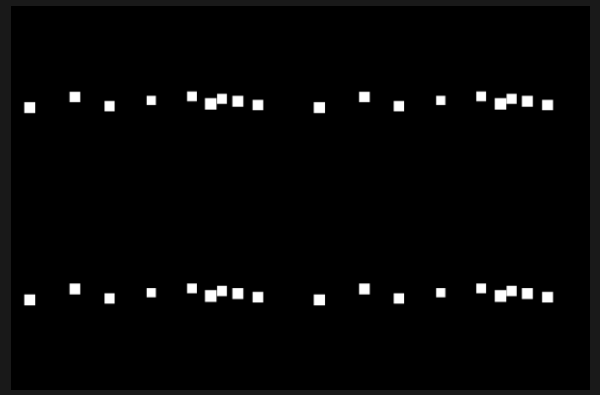


Image saliency map



Wavelet saliency map
Level 1

SBC: Wavelet saliency computation

Original image

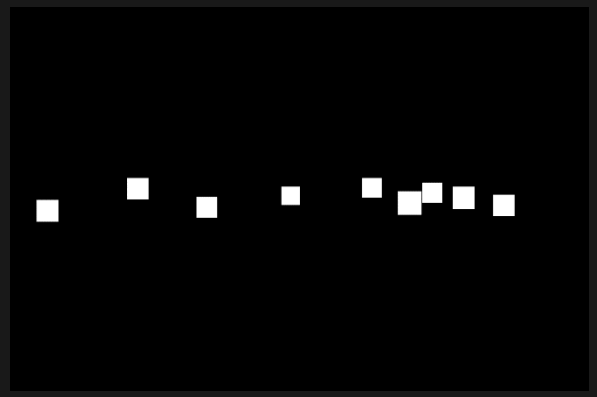
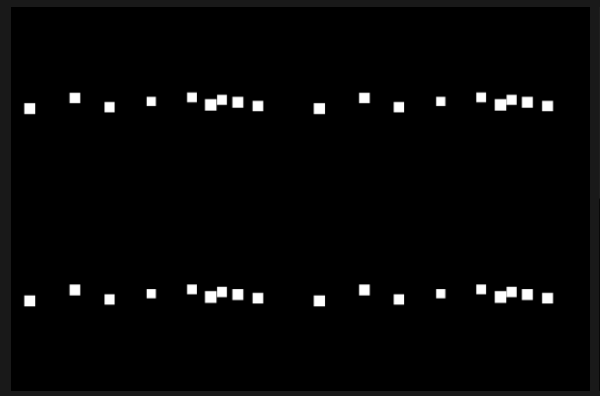


Image saliency map



Wavelet saliency map
Level 1

Wavelet saliency map
Level 2

SBC: Wavelet saliency computation

Original image

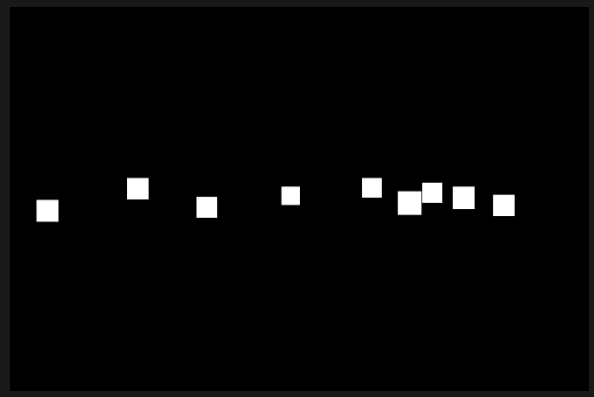
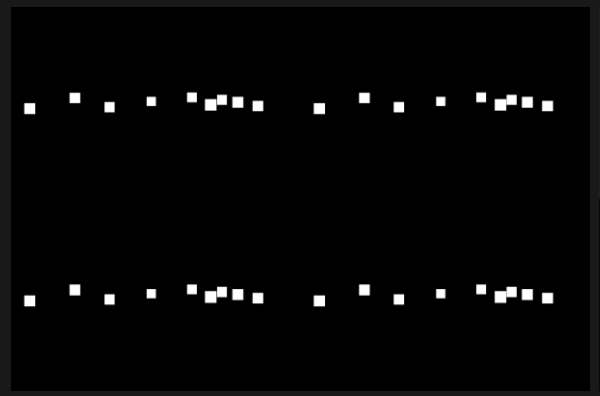


Image saliency map



Wavelet saliency map
Level 1

Wavelet saliency map
Level 2

Wavelet saliency map
Level 3

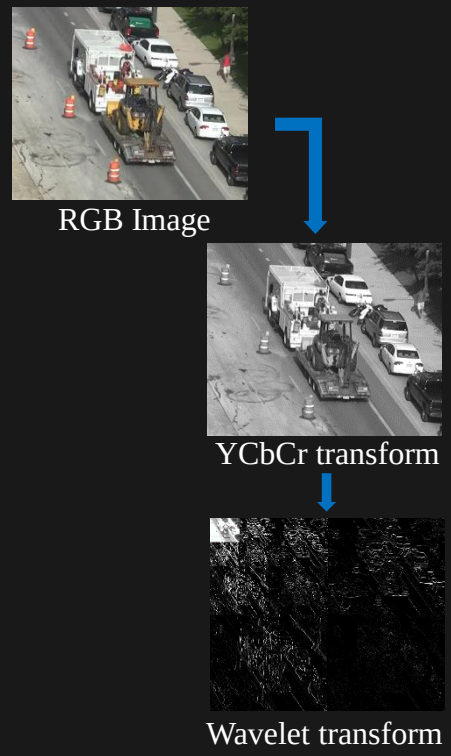
SBC: Algorithm Outline

RGB Image

ENCODER

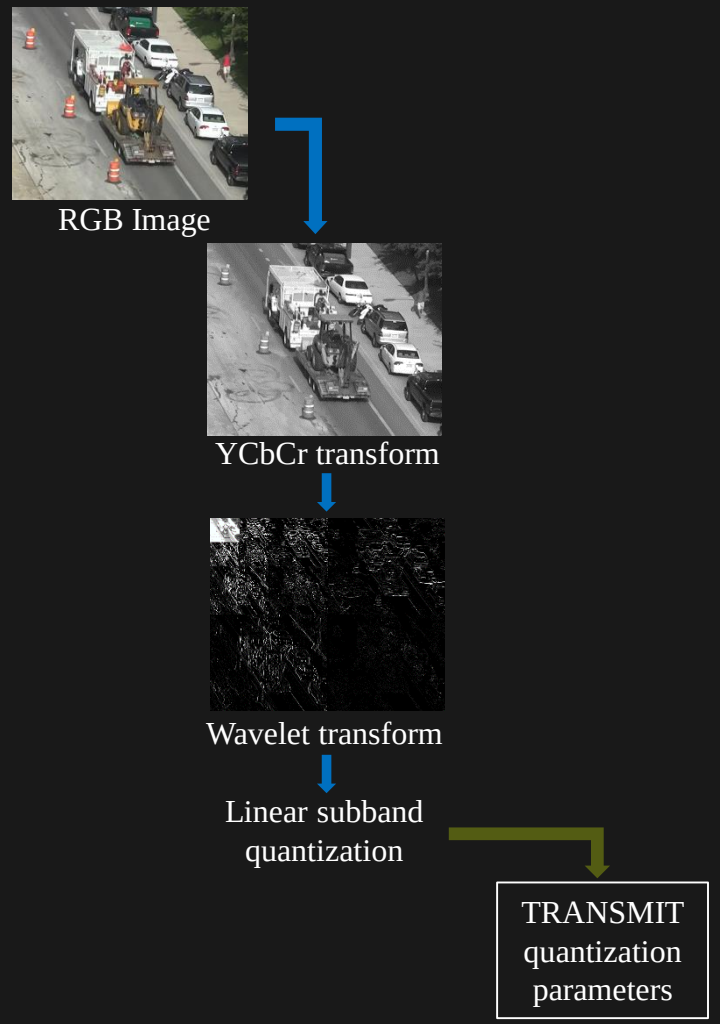
SBC: Algorithm Outline

ENCODER



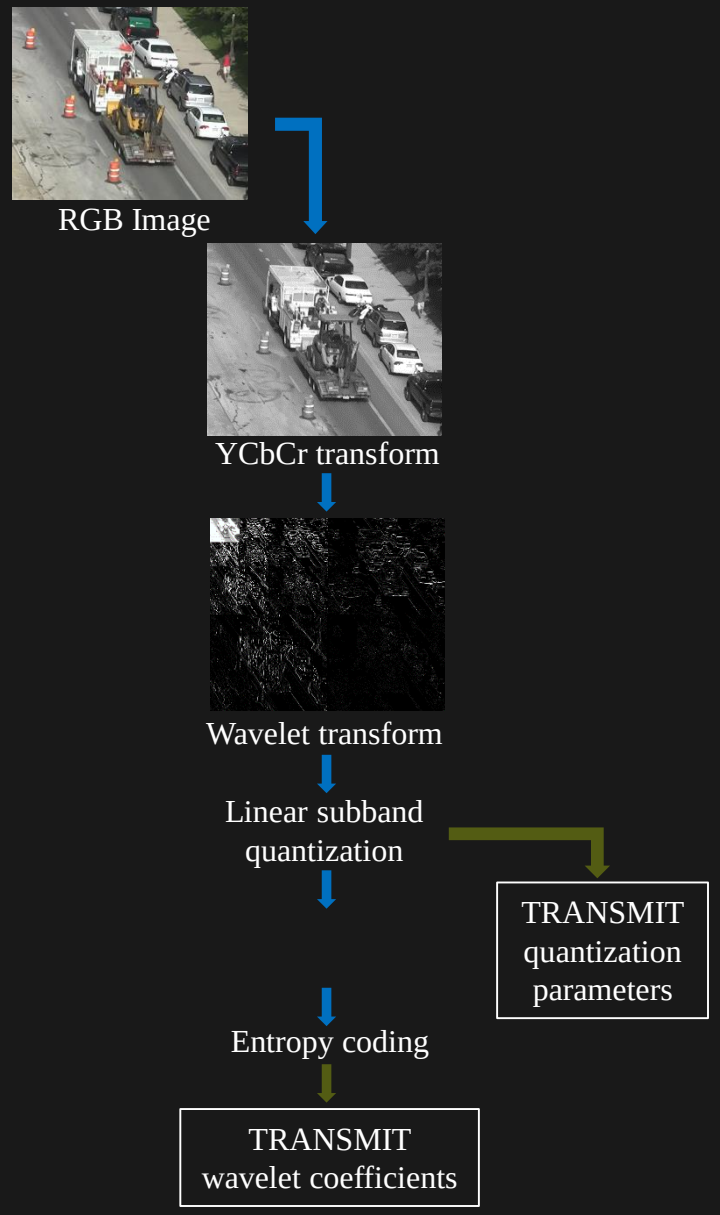
SBC: Algorithm Outline

ENCODER



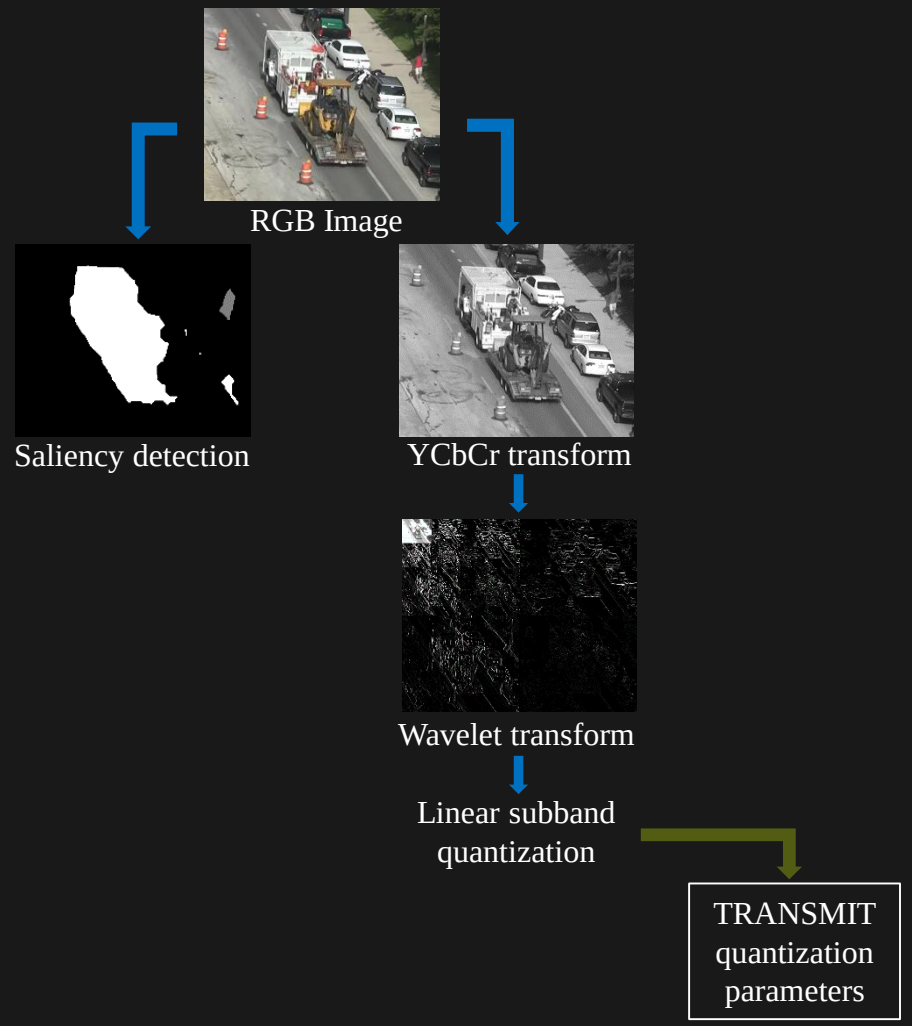
SBC: Algorithm Outline

ENCODER



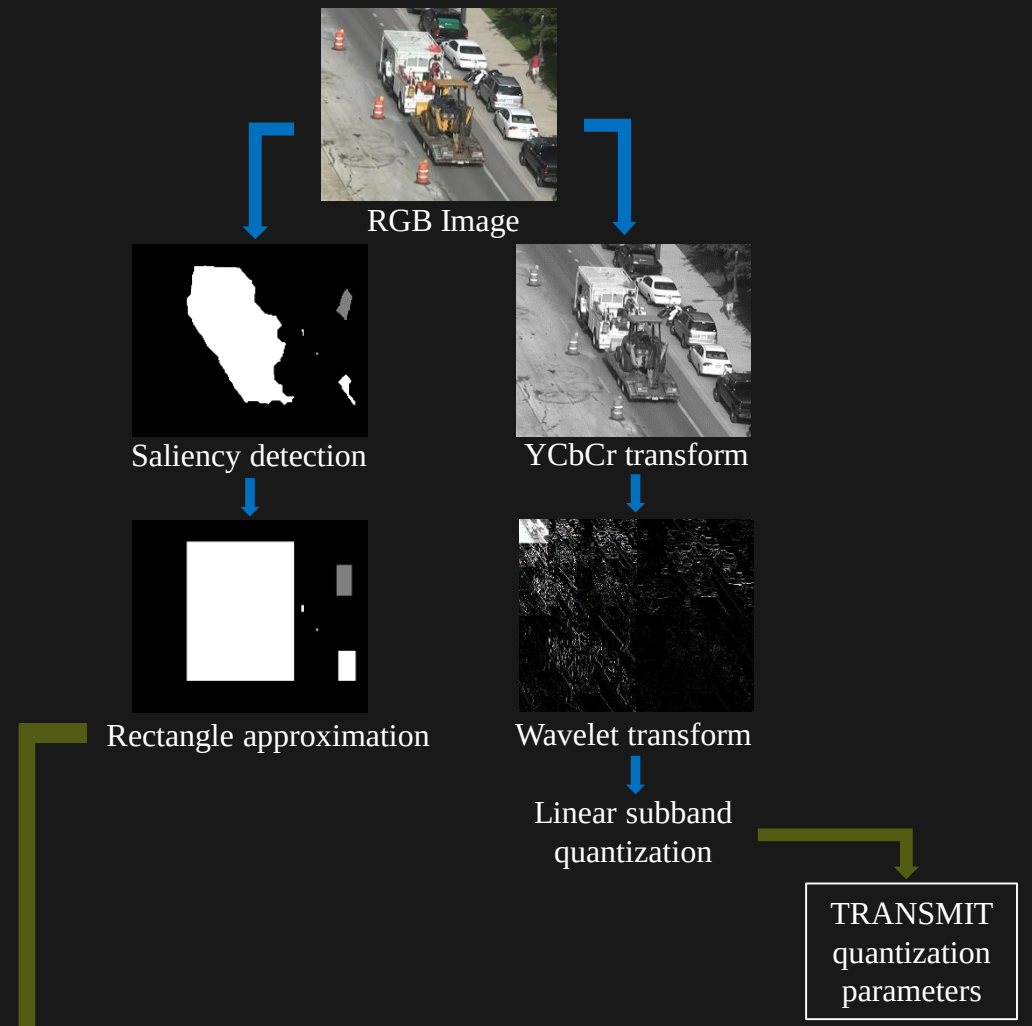
SBC: Algorithm Outline

ENCODER



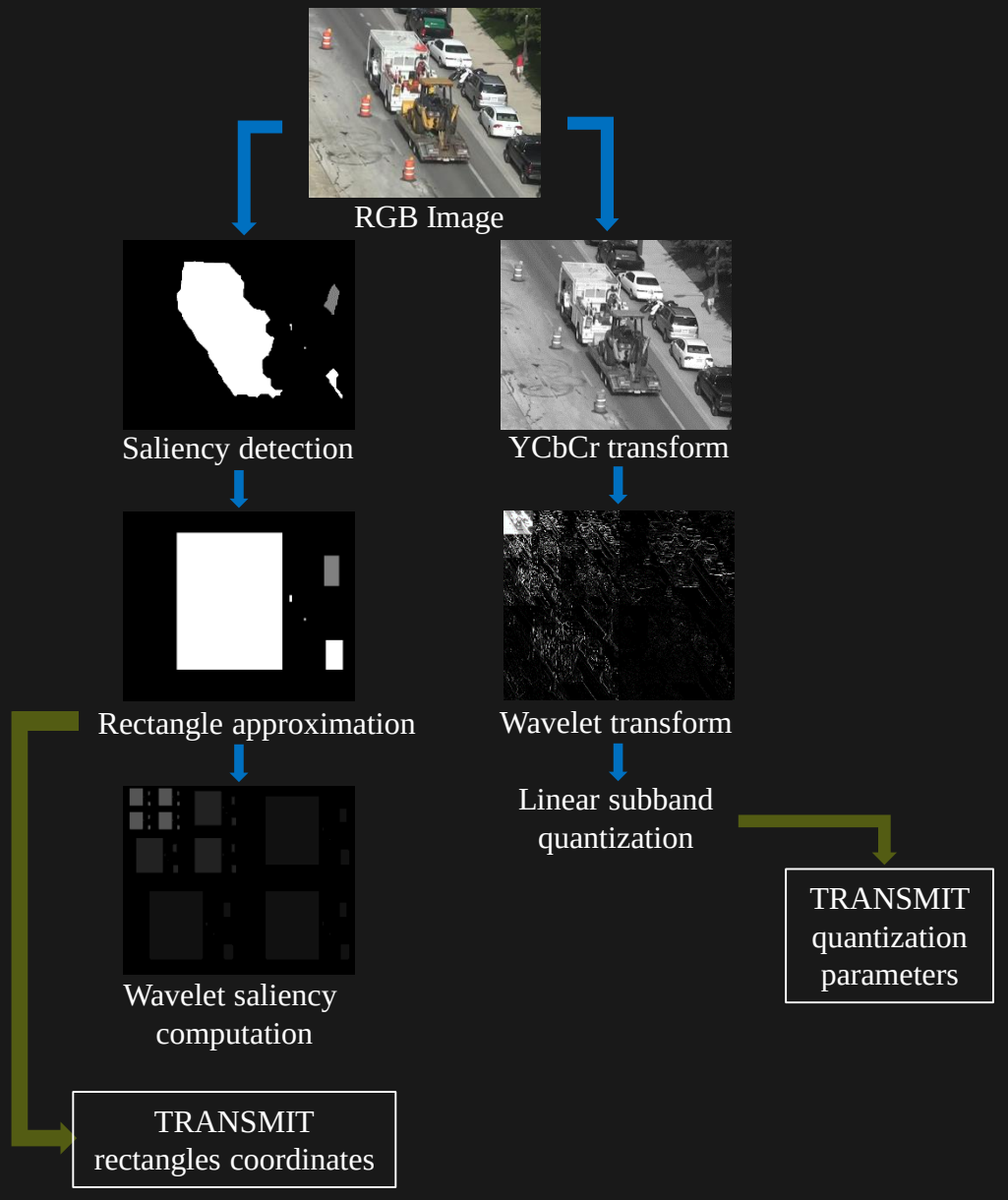
SBC: Algorithm Outline

ENCODER



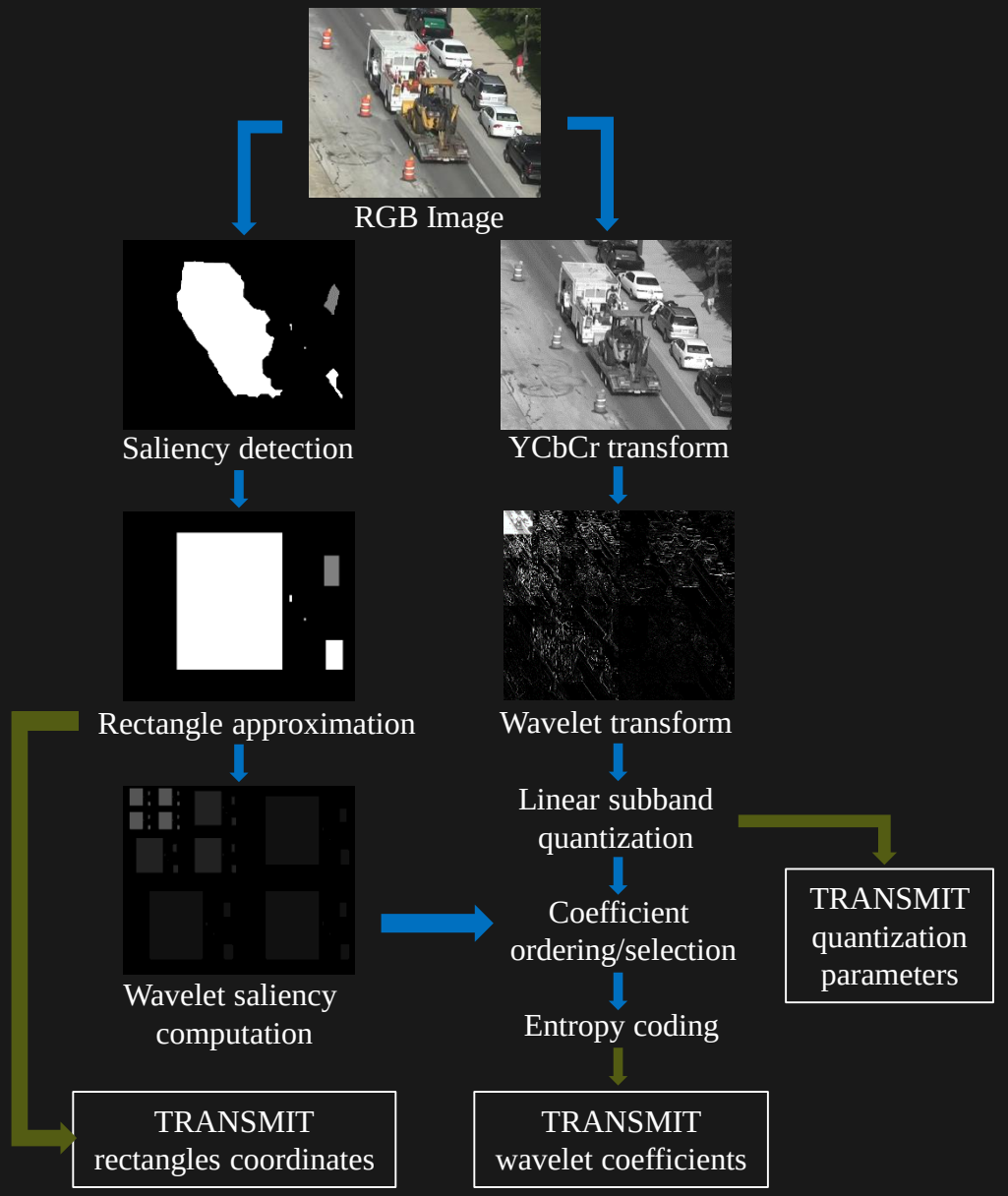
SBC: Algorithm Outline

ENCODER



SBC: Algorithm Outline

ENCODER

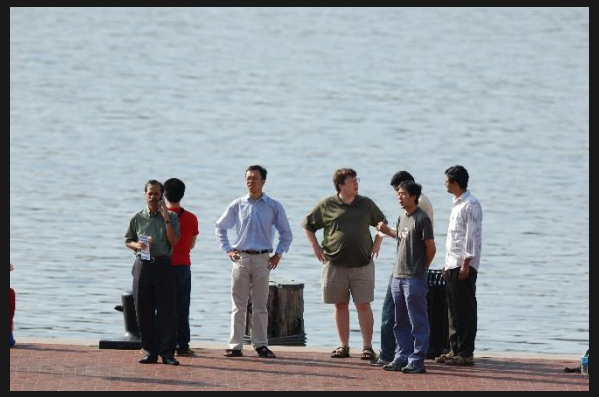


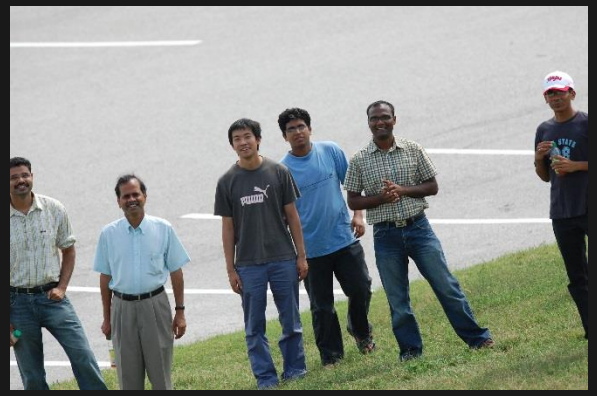
RESULTS: DETECTION PERFORMANCE

UMD Remote Faces dataset*

6MP RGB images of people at large distances

*Rama Chellappa, Jie Ni, and Vishal M Patel, “Remote identification of faces: Problems, prospects, and progress,” Pattern Recognition Letters, vol. 33, no. 14, pp. 1849–1859, 2012.





RICE

DSP

We'll apply OpenCV's face detector
to JPEG, JPEG-2000 and SBC compressed images

SBC: Face Detection performance

	JPEG		JPEG – 2000		SBC	
bpp	TDR (%)	FPR ($\times 10^{-4}$%)	TDR (%)	FPR ($\times 10^{-4}$%)	TDR (%)	FPR ($\times 10^{-4}$%)
0.02	-	-	84.8	2.5	**88.7**	1.5
0.04	37.2	0.2	86.0	2.5	**88.7**	1.6
0.06	59.5	0.5	88.0	2.8	**89.5**	2.1
0.08	82.9	0.9	88.0	2.8	**89.5**	2.3
0.10	**89.5**	1.6	89.5	2.4	89.5	2.3

Comparison of Face Detection performance for JPEG, JPEG-2000 and SBC using the popular Viola-Jones face detector

bpp: bits per pixel

RICE

DSP

SBC: Face Detection performance

	JPEG		JPEG – 2000		SBC	
bpp	TDR (%)	FPR ($\times 10^{-4}$%)	TDR (%)	FPR ($\times 10^{-4}$%)	TDR (%)	FPR ($\times 10^{-4}$%)
0.02	-	-	84.8	2.5	**88.7**	1.5
0.04	37.2	0.2	86.0	2.5	**88.7**	1.6
0.06	59.5	0.5	88.0	2.8	**89.5**	2.1
0.08	82.9	0.9	88.0	2.8	**89.5**	2.3
0.10	**89.5**	1.6	89.5	2.4	89.5	2.3

Comparison of Face Detection performance for JPEG, JPEG-2000 and SBC using the popular Viola-Jones face detector

TDR: True Detection Rate (It's a face, and you say face)

SBC: Face Detection performance

	$JPEG$		$JPEG-2000$		SBC	
bpp	TDR (%)	FPR ($\times 10^{-4}$%)	TDR (%)	FPR ($\times 10^{-4}$%)	TDR (%)	FPR ($\times 10^{-4}$%)
0.02	-	-	84.8	2.5	**88.7**	1.5
0.04	37.2	0.2	86.0	2.5	**88.7**	1.6
0.06	59.5	0.5	88.0	2.8	**89.5**	2.1
0.08	82.9	0.9	88.0	2.8	**89.5**	2.3
0.10	**89.5**	1.6	89.5	2.4	89.5	2.3

Comparison of Face Detection performance for JPEG, JPEG-2000 and SBC using the popular Viola-Jones face detector

FPR: False Positive Rate (It's NOT a face, and you say face)

SBC: Face Detection performance

	$JPEG$		$JPEG-2000$		SBC	
bpp	TDR (%)	FPR ($\times 10^{-4}$%)	TDR (%)	FPR ($\times 10^{-4}$%)	TDR (%)	FPR ($\times 10^{-4}$%)
0.02	-	-	84.8	2.5	**88.7**	1.5
0.04	37.2	0.2	86.0	2.5	**88.7**	1.6
0.06	59.5	0.5	88.0	2.8	**89.5**	2.1
0.08	82.9	0.9	88.0	2.8	**89.5**	2.3
0.10	**89.5**	1.6	89.5	2.4	89.5	2.3

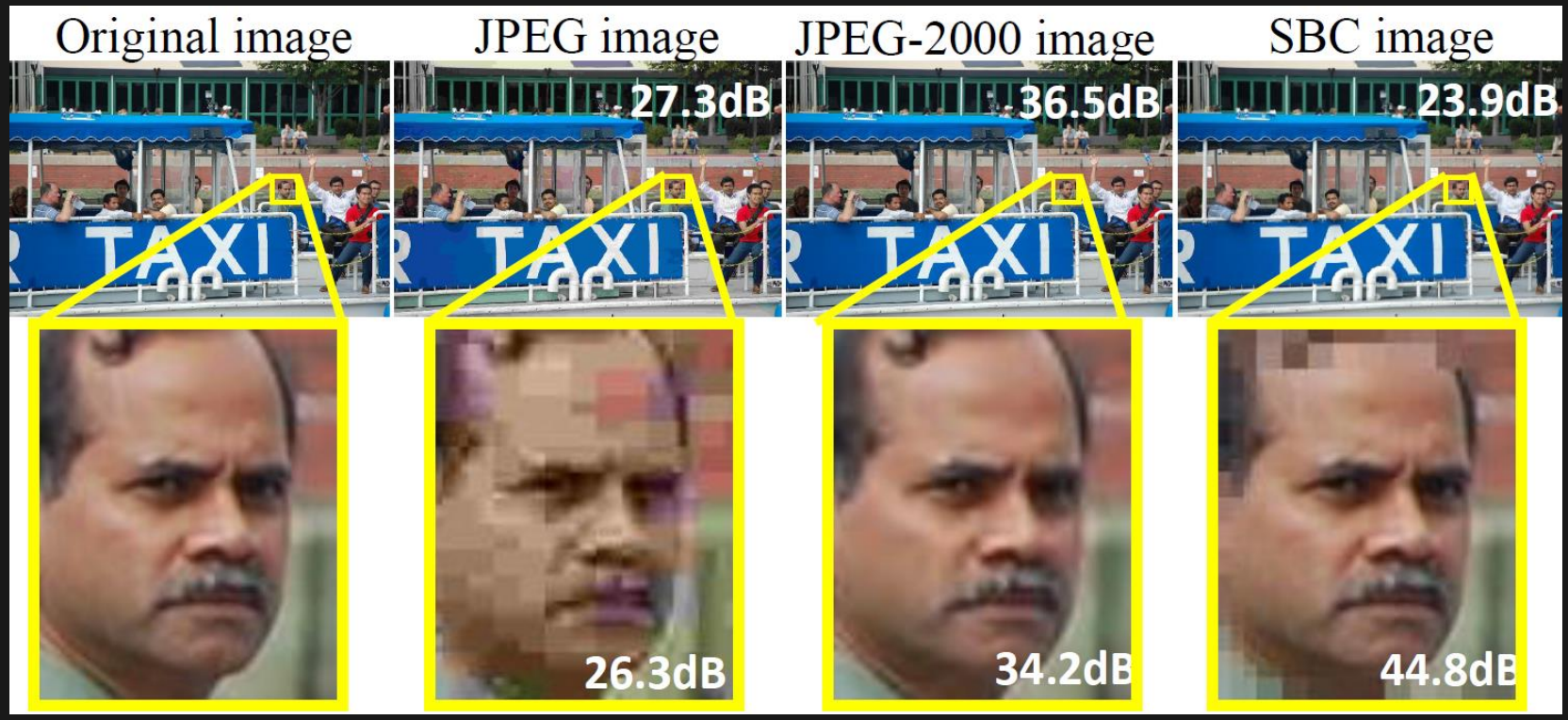
Comparison of Face Detection performance for JPEG, JPEG-2000 and SBC using the popular Viola-Jones face detector

RICE

DSP

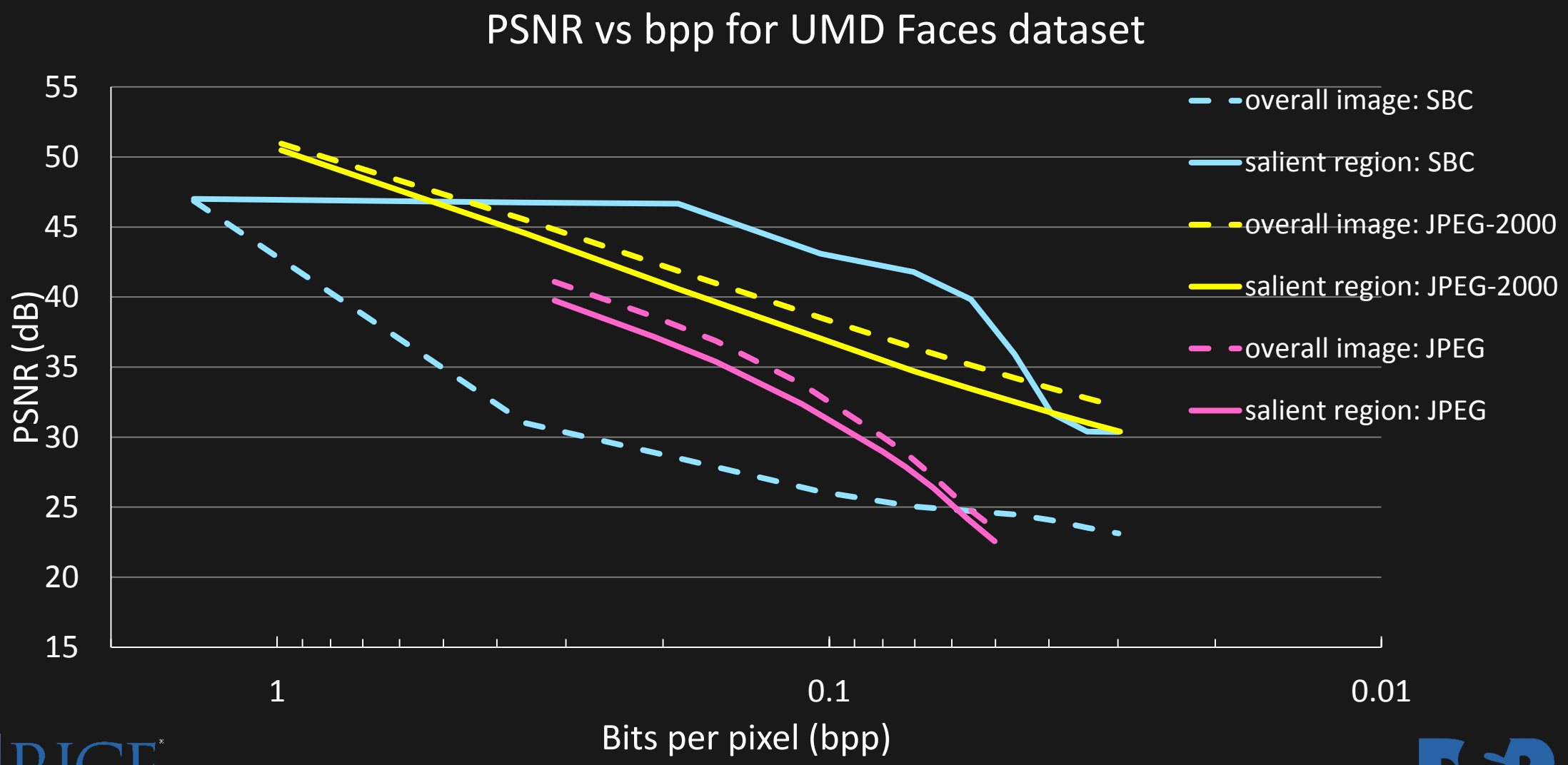
RESULTS: COMPRESSION PERFORMANCE

SBC: Image Compression performance

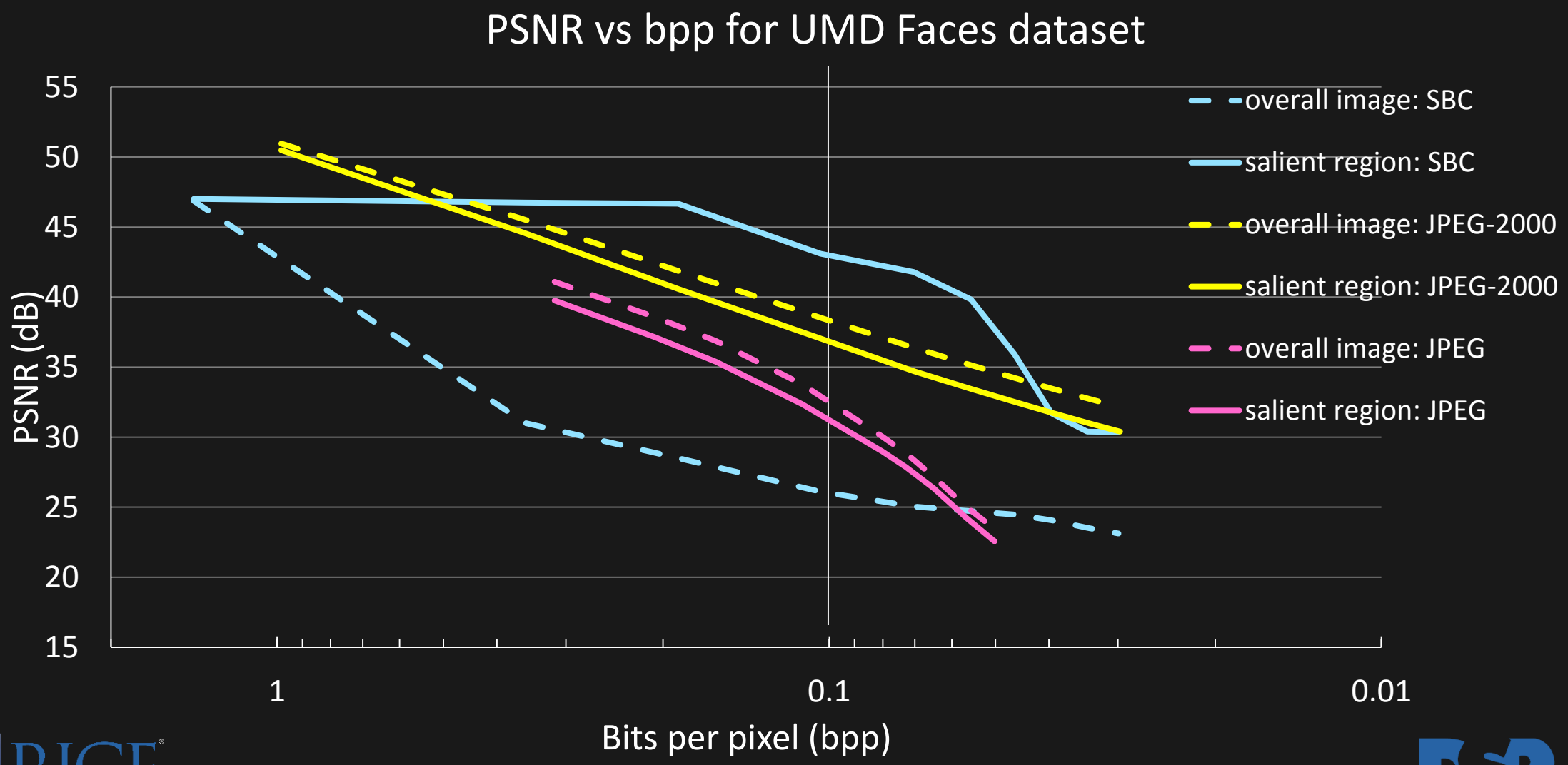


0.07 bpp

SBC: Image Compression performance



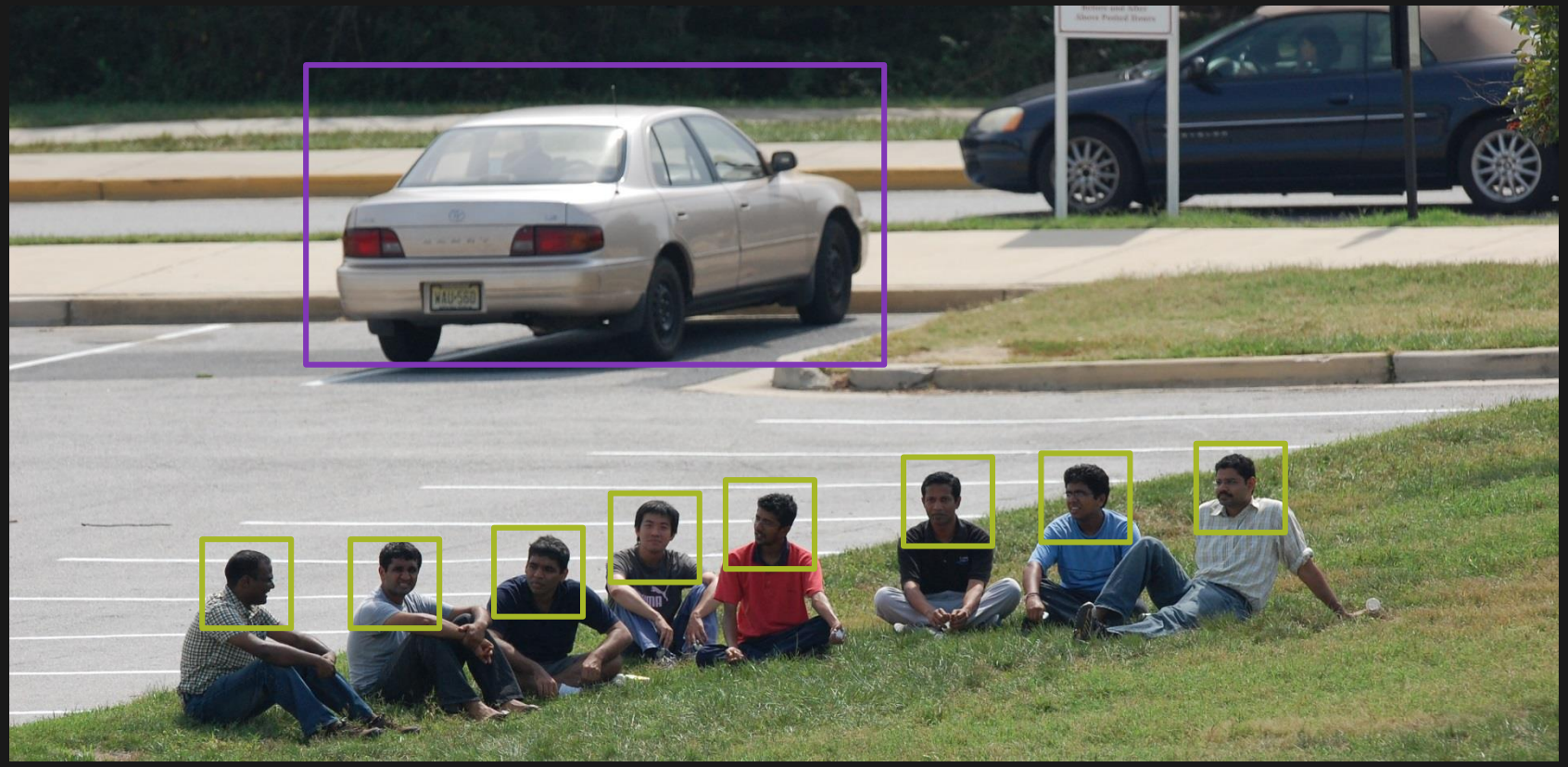
SBC: Image Compression performance



SBC: Multi-level Saliency performance

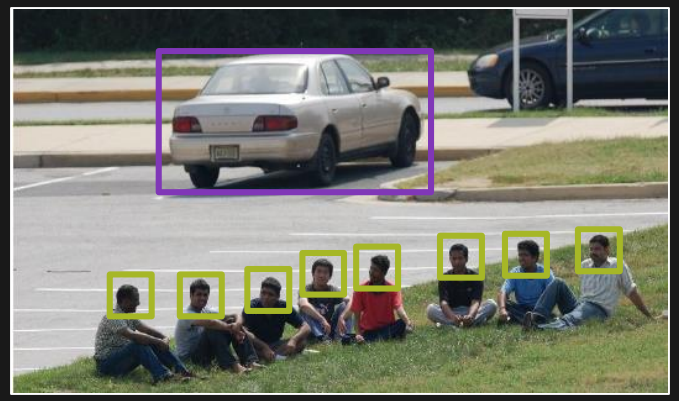
Original Image

SBC: Multi-level Saliency performance



Original Image

SBC: Multi-level Saliency performance

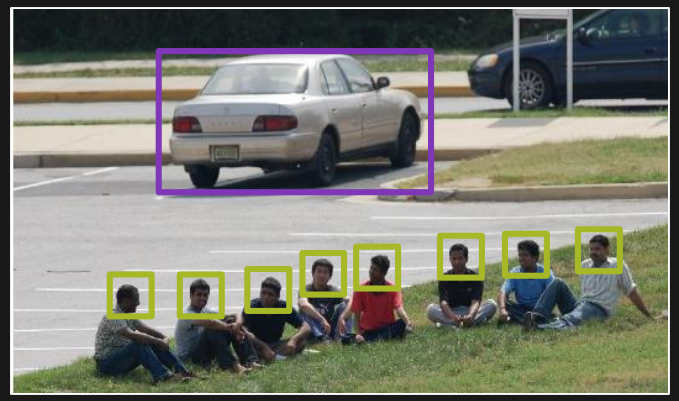


Original Image

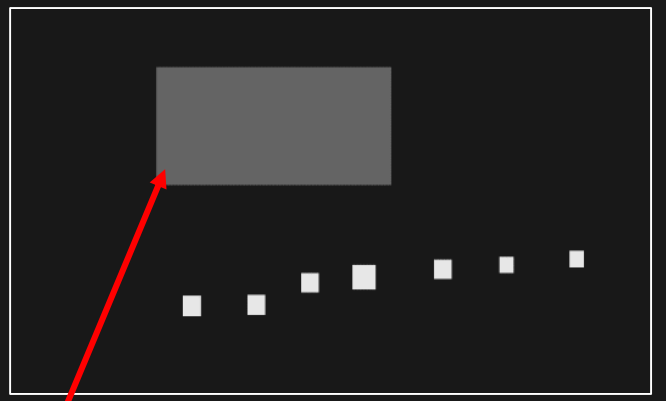
Saliency Map

High saliency *(saliency value = 17)*

SBC: Multi-level Saliency performance



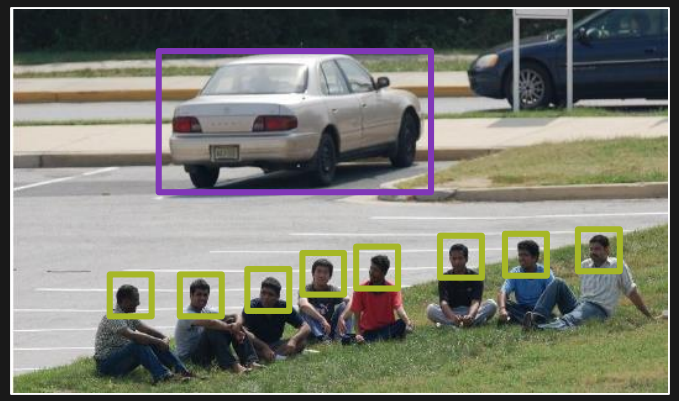
Original Image



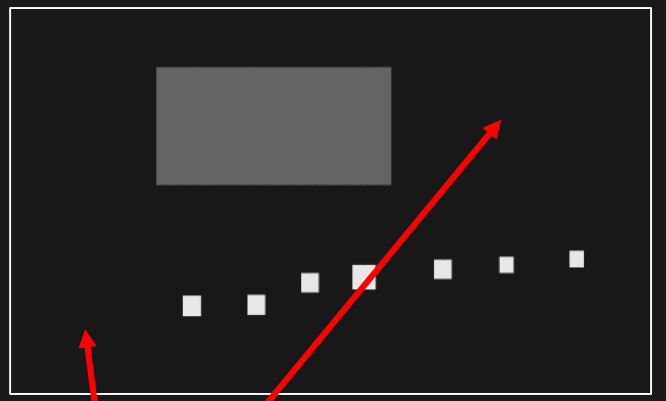
Saliency Map

Medium saliency *(saliency value = 5)*

SBC: Multi-level Saliency performance



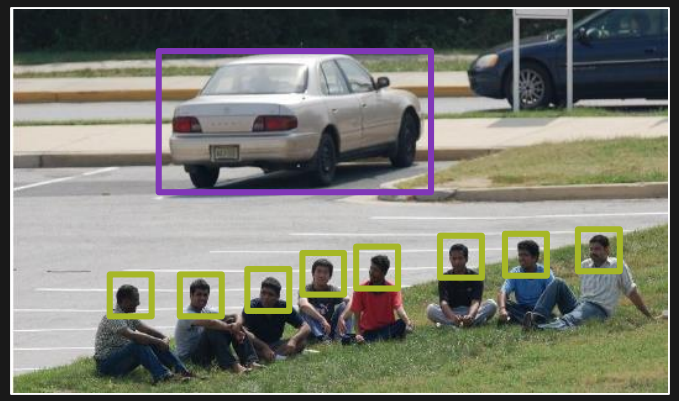
Original Image



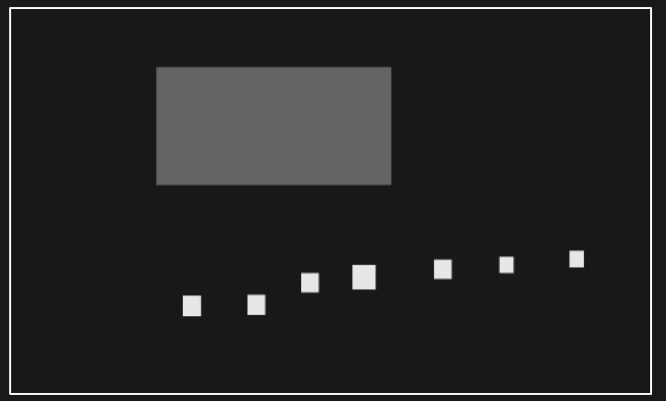
Saliency Map

Not salient (saliency value = 1)

SBC: Multi-level Saliency performance



Original Image



Saliency Map

SBC Image (0.07bpp)

SBC: Multi-level Saliency performance

Original

Reconstructed

VIRAT Video Dataset*

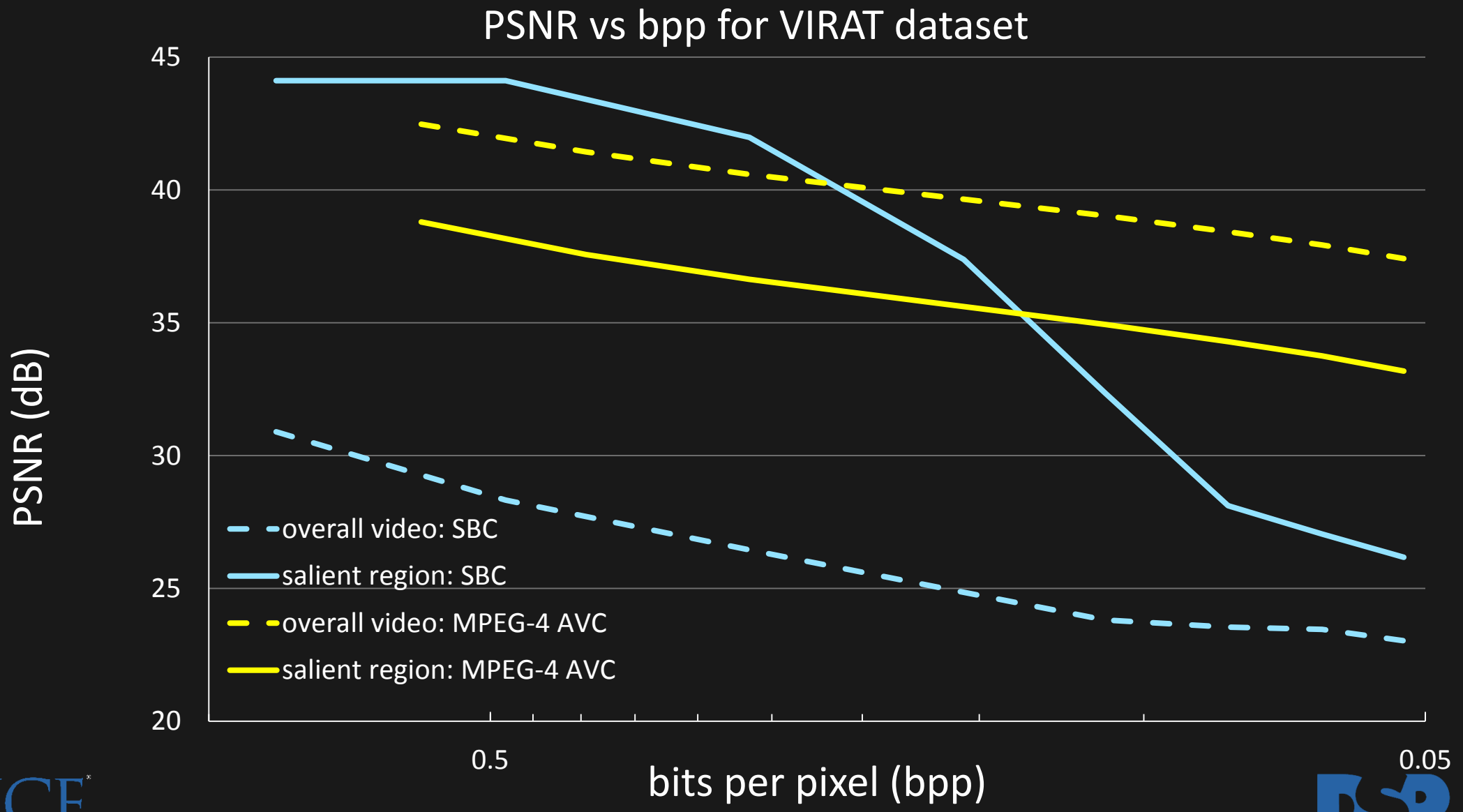
720p/1080p HD surveillance videos

*A Large-scale Benchmark Dataset for Event Recognition in Surveillance Video“by Sangmin Oh, Anthony Hoogs, Amitha Perera and others in Proceedings of IEEE Comptuer Vision and Pattern Recognition (CVPR), 2011

RICE

DSP

SBC: Video compression performance



FUTURE WORK

- Use different wavelets such as Daubechies
- Quantify the improved runtime performance of the SBC object detection pipeline
- Investigate the possibility of using motion estimation for the same object over different frames [Chien *et al* 2008]

CONCLUSION

- A saliency guided wavelet compression scheme for images/videos; tailored towards the object detection task in ultra-low bitrate scenarios
  - Detect objects in the raw captured frames
  - Compress object regions less compared to non-object regions

- Introduced the concept of wavelet saliency map: a flexible way of ordering wavelet transform coefficients

- Better face detection performance compared to JPEG/JPEG-2000

- Better image compression performance compared to JPEG/JPEG-2000

- Video compression and tracking performance at best comparable with MPEG-4 AVC