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

Souptik Barua\textsuperscript{1}, Kaushik Mitra\textsuperscript{2}, and Ashok Veeraraghavan\textsuperscript{1}

\textsuperscript{1} Rice University, \textsuperscript{2} Indian Institute of Technology, Madras

December 16, 2015
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.
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

DJI Phantom drone
4K video capture
DJI Phantom drone
4K video capture

Scene
Bandwidth limited channel (~1 Mbps*)

*https://en.wikipedia.org/wiki/Satellite_Internet_access
Bandwidth limited channel
(\sim 1 \text{ Mbps})
CONTRIBUTION
Let’s visit the JPEG pipeline!
Let's remove some pieces, and add some!
JPEG Pipeline

CMOS sensor → Image Processing → Compression

Transmit/store

Detection → Decompression

Receive/load
SBC Pipeline

CMOS sensor → Image Processing → Compression

Detection ← Decompression

Transmit/store

Receive/load
SBC Pipeline

CMOS sensor → Image Processing → Compression

Detection

Decompression

Receive/load

Transmit/store

Object Detection
SBC Pipeline

CMOS sensor → Image Processing → Compression

Object Detection

Transmit/store

Detection → Decompression

Receive/load
PROPOSED ALGORITHM
SBC: Related work

- Bitplane-by-Bitplane Shift [Wang et al (2003)]

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?
Wavelet Transform
SBC: Wavelet saliency computation
SBC: Wavelet saliency computation
SBC: Wavelet saliency computation

Image saliency map
SBC: Wavelet saliency computation

Image saliency map

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

SBC

All Salient coefficients first; non-salient later
Flexibility in deciding relative importance of different objects in the image!

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

All Salient coefficients first; non-salient later
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
SBC: Algorithm Outline

ENCODER

RGB Image
SBC: Algorithm Outline

ENCODER

RGB Image → YCbCr transform → Wavelet transform
SBC: Algorithm Outline

ENCODER

- RGB Image
- YCbCr transform
- Wavelet transform
- Linear subband quantization
- TRANSMIT quantization parameters
SBC: Algorithm Outline

ENCODER

RGB Image → YCbCr transform → Wavelet transform → Linear subband quantization → Entropy coding → TRANSMIT quantization parameters → TRANSMIT wavelet coefficients
SBC: Algorithm Outline

- RGB Image
- Saliency detection
- YCbCr transform
- Wavelet transform
- Linear subband quantization
- TRANSMIT quantization parameters

ENCODER
SBC: Algorithm Outline

ENCODER

- RGB Image
  - Saliency detection
  - YCbCr transform
  - Wavelet transform
  - Linear subband quantization
  - TRANSMIT quantization parameters
  - TRANSMIT rectangles coordinates
SBC: Algorithm Outline

ENCODER

- Saliency detection
- YCbCr transform
- Rectangle approximation
- Wavelet transform
- Linear subband quantization
- Wavelet saliency computation
- TRANSMIT quantization parameters
- TRANSMIT rectangles coordinates
SBC: Algorithm Outline

- RGB Image
- Saliency detection
- Rectangle approximation
- Wavelet transform
- Linear subband quantization
- Coefficient ordering/selection
- Wavelet saliency computation
- Entropy coding
- TRANSMIT rectangles coordinates
- TRANSMIT wavelet coefficients

ENCODER

TRANSMIT quantization parameters
RESULTS: DETECTION PERFORMANCE
UMD Remote Faces dataset*

6MP RGB images of people at large distances

We’ll apply OpenCV’s face detector to JPEG, JPEG-2000 and SBC compressed images
## SBC: Face Detection performance

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

<table>
<thead>
<tr>
<th>bpp (bpp)</th>
<th>TDR (%)</th>
<th>FPR ($\times 10^{-4}$%)</th>
<th>TDR (%)</th>
<th>FPR ($\times 10^{-4}$%)</th>
<th>TDR (%)</th>
<th>FPR ($\times 10^{-4}$%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.02</td>
<td>-</td>
<td>-</td>
<td>84.8</td>
<td>2.5</td>
<td>88.7</td>
<td>1.5</td>
</tr>
<tr>
<td>0.04</td>
<td>37.2</td>
<td>0.2</td>
<td>86.0</td>
<td>2.5</td>
<td>88.7</td>
<td>1.6</td>
</tr>
<tr>
<td>0.06</td>
<td>59.5</td>
<td>0.5</td>
<td>88.0</td>
<td>2.8</td>
<td>89.5</td>
<td>2.1</td>
</tr>
<tr>
<td>0.08</td>
<td>82.9</td>
<td>0.9</td>
<td>88.0</td>
<td>2.8</td>
<td>89.5</td>
<td>2.3</td>
</tr>
<tr>
<td>0.10</td>
<td><strong>89.5</strong></td>
<td>1.6</td>
<td>89.5</td>
<td>2.4</td>
<td>89.5</td>
<td>2.3</td>
</tr>
</tbody>
</table>

 bpp: bits per pixel
### SBC: Face Detection performance

<table>
<thead>
<tr>
<th>bpp</th>
<th>TDR (%)</th>
<th>FPR ($\times 10^{-4}$%)</th>
<th>TDR (%)</th>
<th>FPR ($\times 10^{-4}$%)</th>
<th>TDR (%)</th>
<th>FPR ($\times 10^{-4}$%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.02</td>
<td>-</td>
<td>-</td>
<td>84.8</td>
<td>2.5</td>
<td>88.7</td>
<td>1.5</td>
</tr>
<tr>
<td>0.04</td>
<td>37.2</td>
<td>0.2</td>
<td>86.0</td>
<td>2.5</td>
<td>88.7</td>
<td>1.6</td>
</tr>
<tr>
<td>0.06</td>
<td>59.5</td>
<td>0.5</td>
<td>88.0</td>
<td>2.8</td>
<td>89.5</td>
<td>2.1</td>
</tr>
<tr>
<td>0.08</td>
<td>82.9</td>
<td>0.9</td>
<td>88.0</td>
<td>2.8</td>
<td>89.5</td>
<td>2.3</td>
</tr>
<tr>
<td>0.10</td>
<td><strong>89.5</strong></td>
<td><strong>1.6</strong></td>
<td><strong>89.5</strong></td>
<td><strong>2.4</strong></td>
<td><strong>89.5</strong></td>
<td><strong>2.3</strong></td>
</tr>
</tbody>
</table>

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

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

<table>
<thead>
<tr>
<th></th>
<th>JPEG</th>
<th>JPEG - 2000</th>
<th>SBC</th>
</tr>
</thead>
<tbody>
<tr>
<td>bpp</td>
<td>TDR (%)</td>
<td>FPR ($\times 10^{-4} %$)</td>
<td>TDR (%)</td>
</tr>
<tr>
<td>0.02</td>
<td>-</td>
<td>-</td>
<td>84.8</td>
</tr>
<tr>
<td>0.04</td>
<td>37.2</td>
<td>0.2</td>
<td>86.0</td>
</tr>
<tr>
<td>0.06</td>
<td>59.5</td>
<td>0.5</td>
<td>88.0</td>
</tr>
<tr>
<td>0.08</td>
<td>82.9</td>
<td>0.9</td>
<td>88.0</td>
</tr>
<tr>
<td>0.10</td>
<td>89.5</td>
<td>1.6</td>
<td>89.5</td>
</tr>
</tbody>
</table>

**FPR: False Positive Rate** (It’s NOT a face, and you say face)
### SBC: Face Detection Performance

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

<table>
<thead>
<tr>
<th>bpp</th>
<th>JPEG TDR (%)</th>
<th>JPEG FPR ( \times 10^{-4}% )</th>
<th>JPEG – 2000 TDR (%)</th>
<th>JPEG – 2000 FPR ( \times 10^{-4}% )</th>
<th>SBC TDR (%)</th>
<th>SBC FPR ( \times 10^{-4}% )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.02</td>
<td>-</td>
<td>-</td>
<td>84.8</td>
<td>2.5</td>
<td>88.7</td>
<td>1.5</td>
</tr>
<tr>
<td>0.04</td>
<td>37.2</td>
<td>0.2</td>
<td>86.0</td>
<td>2.5</td>
<td>88.7</td>
<td>1.6</td>
</tr>
<tr>
<td>0.06</td>
<td>59.5</td>
<td>0.5</td>
<td>88.0</td>
<td>2.8</td>
<td>89.5</td>
<td>2.1</td>
</tr>
<tr>
<td>0.08</td>
<td>82.9</td>
<td>0.9</td>
<td>88.0</td>
<td>2.8</td>
<td>89.5</td>
<td>2.3</td>
</tr>
<tr>
<td>0.10</td>
<td><strong>89.5</strong></td>
<td>1.6</td>
<td>89.5</td>
<td>2.4</td>
<td><strong>89.5</strong></td>
<td>2.3</td>
</tr>
</tbody>
</table>
RESULTS: COMPRESSION PERFORMANCE
SBC: Image Compression performance

Original image

JPEG image 27.3dB

JPEG-2000 image 36.5dB

SBC image 23.9dB

0.07 bpp
SBC: Image Compression performance

PSNR vs bpp for UMD Faces dataset

- **Overall image:** SBC
- **Salient region:** SBC
- **Overall image:** JPEG-2000
- **Salient region:** JPEG-2000
- **Overall image:** JPEG
- **Salient region:** JPEG

PSNR (dB) vs Bits per pixel (bpp)
SBC: Image Compression performance

PSNR vs bpp for UMD Faces dataset

- Overall image: SBC
- Salient region: SBC
- Overall image: JPEG-2000
- Salient region: JPEG-2000
- Overall image: JPEG
- Salient region: JPEG
SBC: Multi-level Saliency performance
SBC: Multi-level Saliency performance
SBC: Multi-level Saliency performance

Original Image

Saliency Map

High saliency (saliency value = 17)
SBC: Multi-level Saliency performance

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

41.9 dB

31.7 dB

26.2 dB
VIRAT Video Dataset*

720p/1080p HD surveillance videos

SBC: Video compression performance

PSNR vs bpp for VIRAT dataset

- Overall video: SBC
- Salient region: SBC
- Overall video: MPEG-4 AVC
- Salient region: MPEG-4 AVC
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