Human-centric Compression

Are humans the best lossy image compressors?

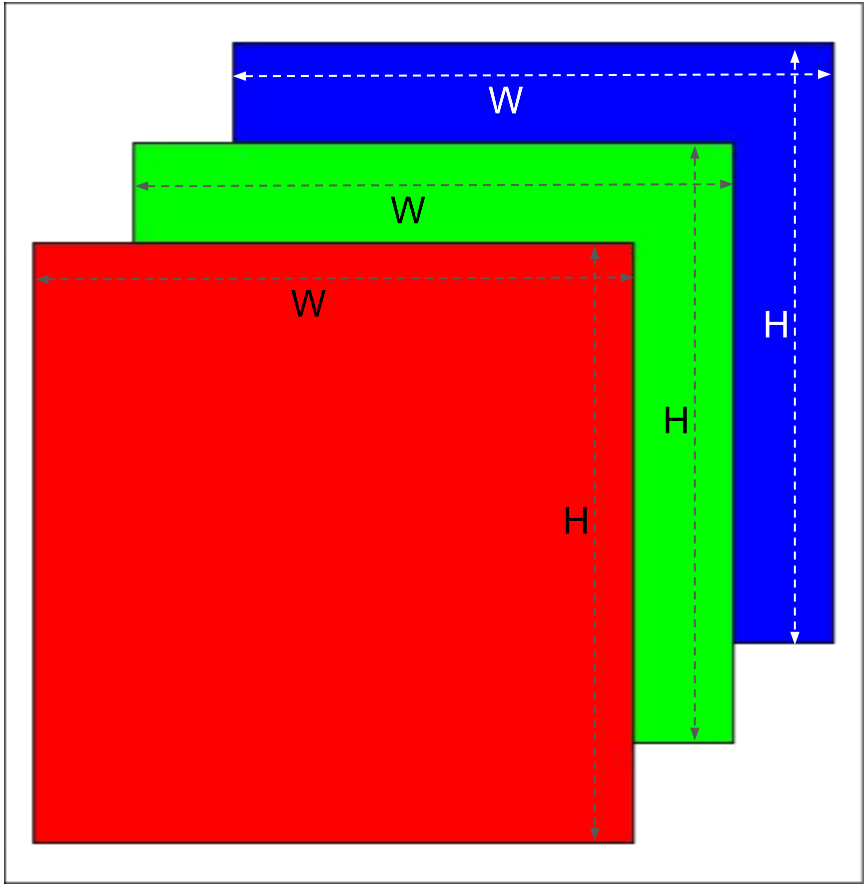
By Soham Mukherjee (Monta Vista), Sean Yang (St. Francis), and Ashu Bhown (Palo Alto)

Outline

- ➢ Introduction
- ➢ Traditional Compressors
- ➢ Human Compression explained
- ➢ Process
- ➢ Testing methodology
- ➢ Results
- ➢ Conclusions

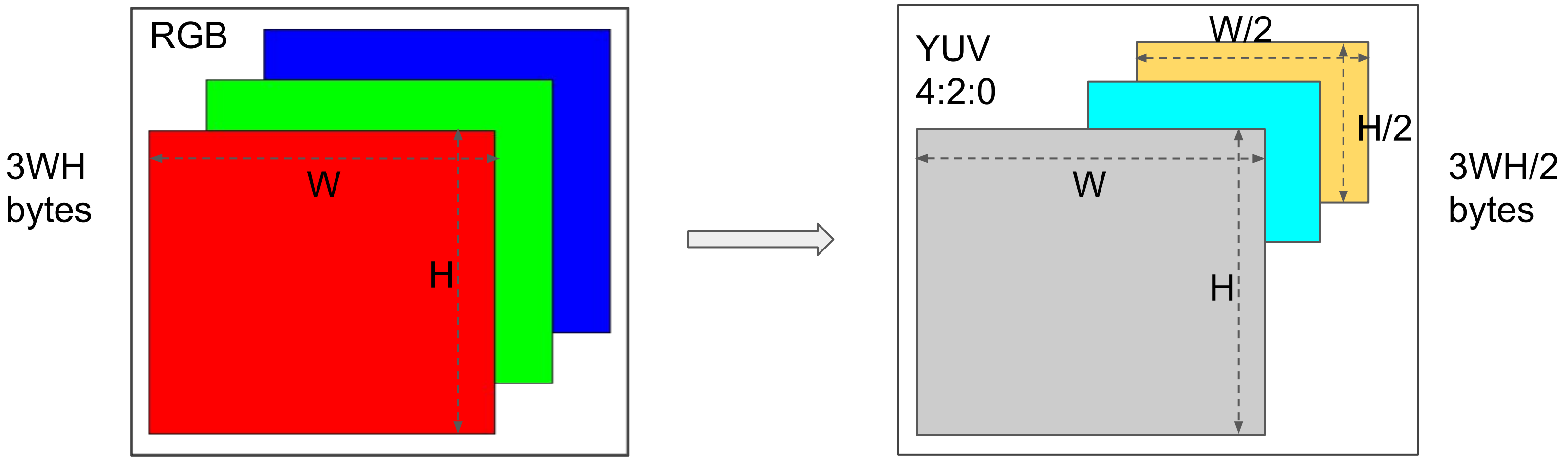
Introduction

- ➢ A digital color image is typically represented as three channels: red, green, and blue, each of same resolution WxH as the image.
- ➢ Each pixel value in each channel is represented using 8-bits (1 byte)
  - ○ 3 bytes per pixel
- ➢ Total of 3*W*H bytes per image uncompressed



Introduction

- Most modern compression schemes work in Y-U-V colorspace, with 4:2:0 color sampling
- Convert RGB to YUV
  - Y represents the luminance (the brightness) and U and V are the chrominance (color) components
  - Downsample U and V channels by factor of 2 in each dimension.



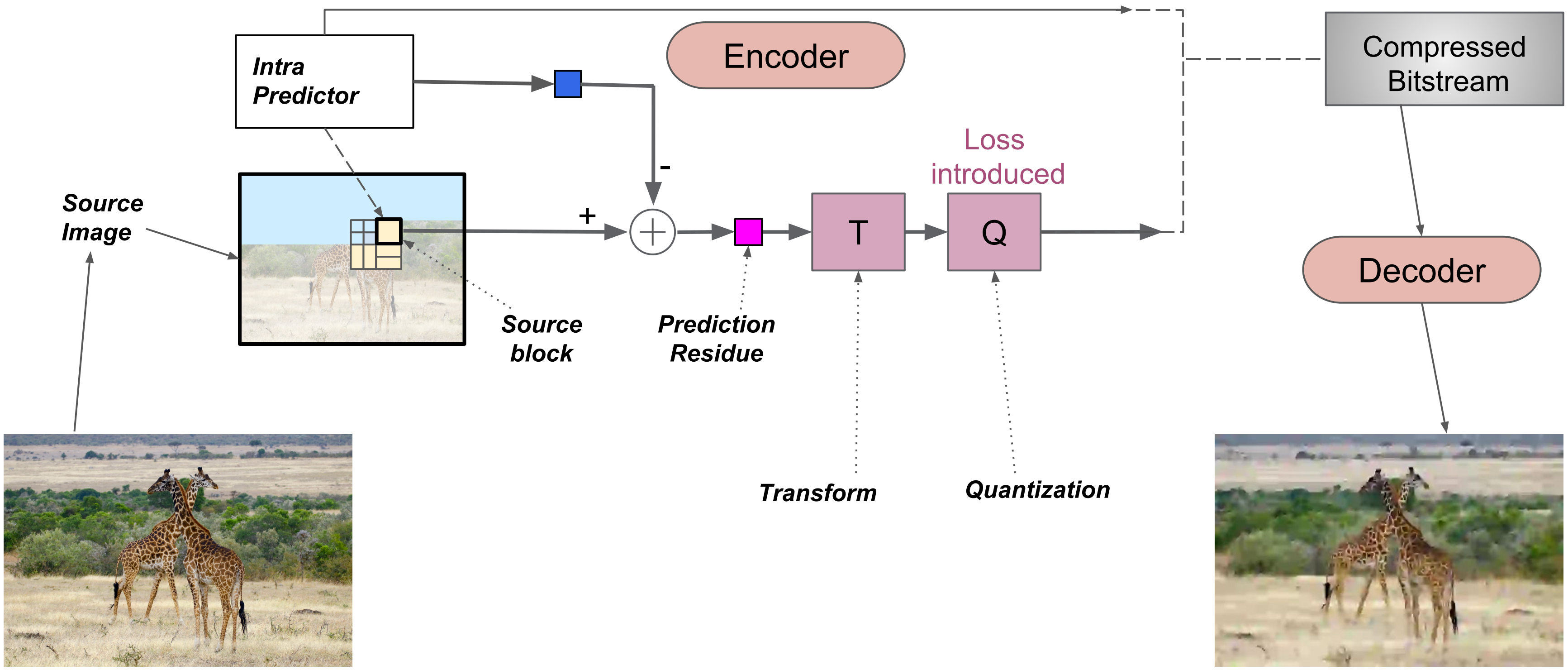
Introduction

- Explosion in digital images generated
  - High quality image capture devices ubiquitous
  - Example: 12 mega-pixel camera on iPhone X
    - Total of 36 MB per image (RGB) or 18 MB per image (YUV 4:2:0)
    - Sharing a photo album with just 100 pictures takes 1.8 GB data to be transmitted
- For ease of storage and sharing, compression is essential

Lossy vs Lossless

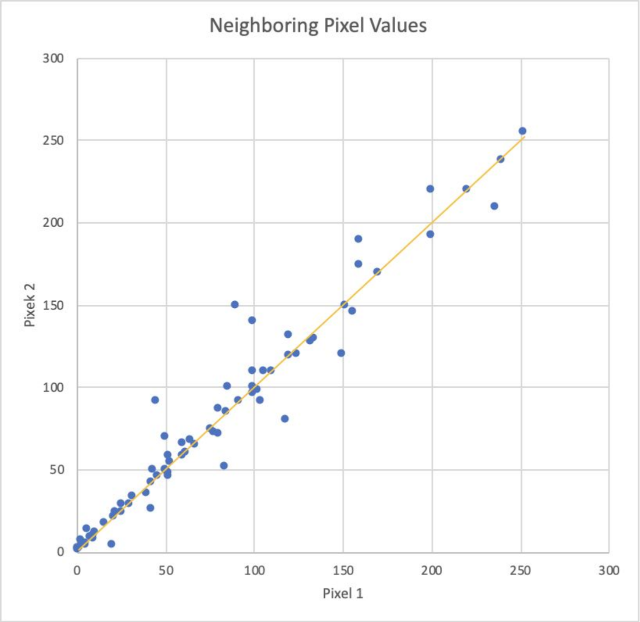
- Lossless compression would give us about 2:1 compression on an average - not enough
- Some loss must be tolerated
  - Especially for everyday sharing, as long as the image conveys the same information
  - Speed of transmission more important than getting an exact replica

Traditional Modern Compressor



Transform & Quantization

- Transform
  - Generally pixel neighboring one another will have similar values
  - Because of this we can rotate the graph such that a majority of the values resides on an axis
- Quantization
  - Rounds off the new pixel values on the rotated axis



Traditional Lossy Compressors

- 1992: JPEG
  - Joint Photographic Experts Group
  - Transform Encoding
- 2000: JPEG 2000
  - improved compression encoding method, but never made it mainstream due to compatibility issues
- 2010: WebP
  - Lossy algorithm by Google
  - Entropy Encoding
    - predicts the color of a pixel by looking at the surrounding fragments
    - reduces the size that traditional lossy compression algorithms could by an average of 25%

Traditional Lossy Compressors Flaws

- At very low bit-rates, the reconstruction is not able to represent the original image closely enough
- Compression Artifacts- distortions of the image
  - Staircase noise (aliasing) along curving edges
  - Blockiness
  - Posterization
- Generation Loss- repeatedly compressing and decompressing the file will cause it to progressively lose quality

Overall Goals

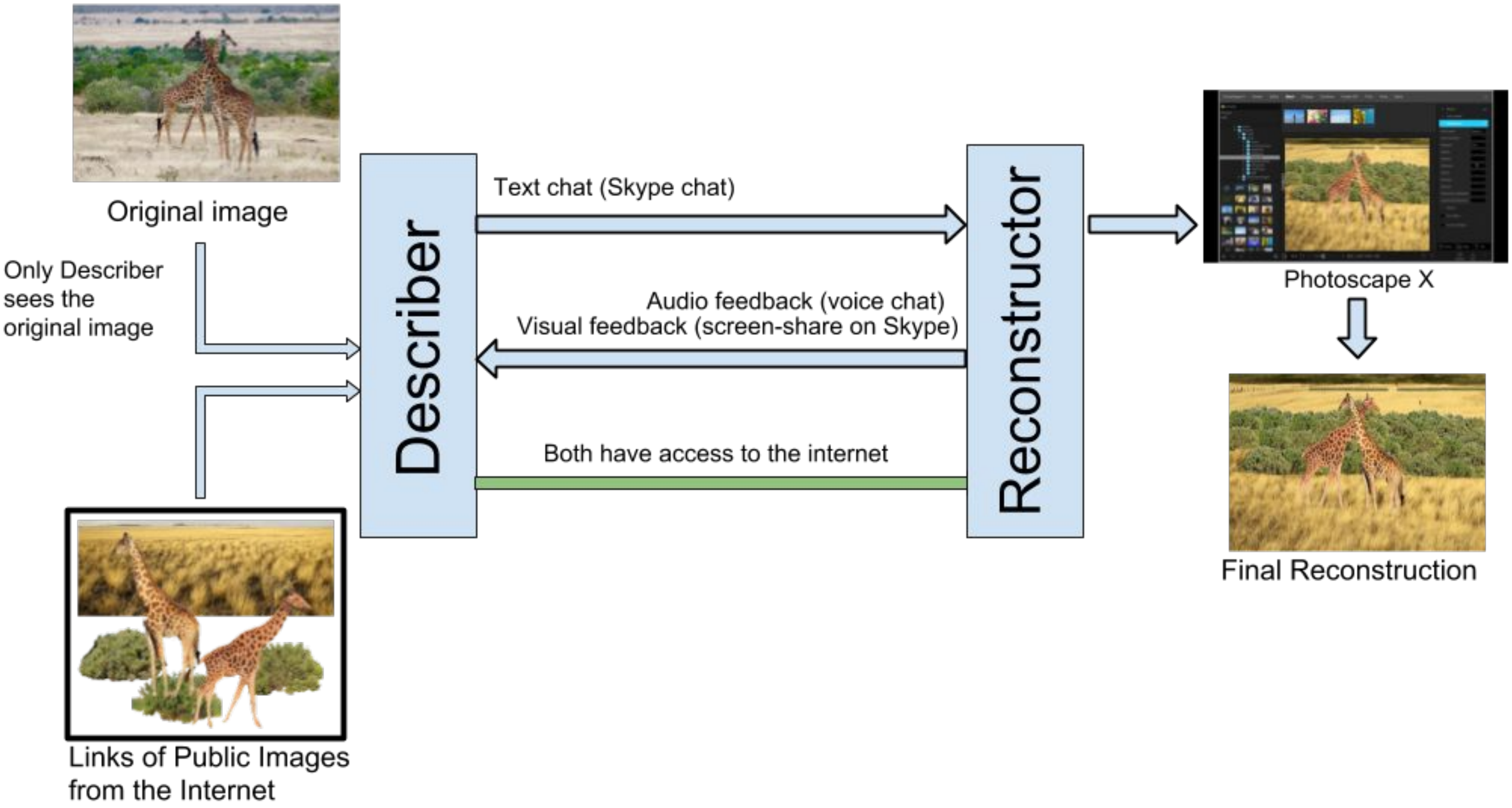
- ➢ To provide a more human centric approach to image compression that could be eventually implemented by neural nets
- ➢ To fully utilize the public resource of images already available on the Internet
- ➢ Question: Can we create more efficient image reconstructions by preserving only what humans perceive as important at low bit rates?
    - High level descriptions of parts of images rather than pixels
    - Using the English Language rather than encoding pixels

Human Compression Explained

- Our setup involves two distinct roles, referred to as the “describer” and the “reconstructor” respectively
- In short, the describer takes images and sends solely text-based information to the reconstructor, who attempts to recreate the image using any tools necessary

Experiment Set-up Using Skype

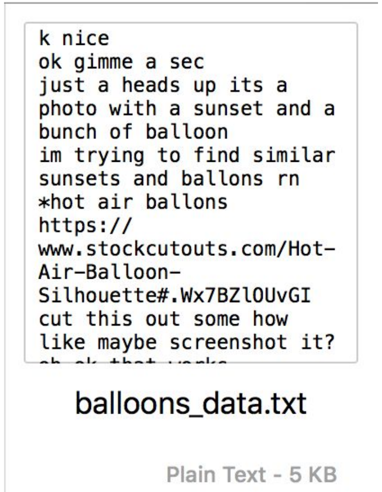
- ➢ Text Commands (Describer —> Reconstructor)
  - ○ The describer is only allowed to send messages to the reconstructor through the built-in Skype text chat.
  - ○ The describer must turn off their outgoing audio/video to avoid inadvertently leaking any information to the reconstructor.
- ➢ Feedback (Reconstructor —> Describer)
  - ○ The reconstructor may talk to the describer through audio/video/text chat.
  - ○ The reconstructor may share their partial reconstruction with the describer in real-time, by using the screen-share feature of Skype.


Original image
Only Describer sees the original image
Describer
Text chat (Skype chat)
Audio feedback (voice chat)
Visual feedback (screen-share on Skype)
Both have access to the internet
Reconstructor
Photoscape X
Final Reconstruction
Links of Public Images from the Internet

End of Process

When the reconstruction has been completed by the reconstructor to the level of describer’s satisfaction, the compression experiment is stopped.

- The transcript is processed by removing timestamps and compressing it using the bzip2 [16] compressor (an open source single file compressor program).
- The bzip2 encoded Skype transcript represents the final compressed representation of the input image.



The quality of image reconstruction can now be compared to that of a standard lossy image compressor, as described in the next section.

Reconstructor Stages

Skype Chat Excerpts

https://www.worldwildlife.org/habitats/grasslands

Try transformations
elongate the fence bit
only focus on the vertical...

there's a line of shrubbery that goes across the middle third of the image...
that's the largest bush in the pic
so keep the others sizes equal to or smaller that that and make it look continuous
and make sure to make the bushes smaller as you work your way up so that there's a sense of depth...

there will be a line of tiny shrubs along that line...
the line itself starts about a quarter from the left...

try and make the grass look less tall on the bottom...

when you're done with that take a look at these
https://public-media.smithsonianmag.com/filer/32/f2/32f24473-b380-43f5-94df-da0e58644439/16301090250_acf80be87f_o.jpg
https://img.purch.com/w/192/aHR0cDovL3d3dy5saXZlc2NpZW5jZS5jb20vaW1hZ2VzL2kvMDAwLzA2OC8wOTQvaTMwMC9naXJhZmZlLmpwZz8xNDA1MDA4NDQy
sure
while you're editing that giraffe
its spots are too dark
make it look like the other giragge…

make the right one bigger than the left
make the heads level
wait back
put the left one where it was before
good
now move the right giraffe to the left so that their necks cross
good
move them both to the center
make them both taller as well
their heads should be above the middle line of shrubs…

there's a ridgeline in the back
of very dense shrubs
but let's try something
I want you to place a shrub on the very top of the image
and stretch it from left to right…
it should be less green make it look hazier if that makes sense…

Final reconstruction

Original image

Testing methodology

Evaluating the quality of the reconstruction by the human compressors and WebP

1. Human compression: The given input image is compressed by the humans using the procedure described earlier. The size (in bytes) of the compressed representation of the image is recorded.

2. WebP compression: Next, we use the WebP compressor to lossily compress the input image to have a similar size as the human compression text representation.

3. Quality evaluation: Finally, we compare the quality of the WebP and human compressed images using human scorers on the Mechanical Turk platform.

Reaching Out To The Public

- We compare the quality of compressed images using human scorers (workers) on Amazon Mechanical Turk, a platform for conducting large scale surveys
- For each image, we display the original image and the human reconstruction and ask the workers to rate the reconstruction on a discrete scale of 1 to 10
- To capture the effects of human perception, the scale represents a general "level of satisfaction" with the reconstruction rather than a specific metric like accuracy
- We perform identical experiments for the WebP reconstructions. For every experiment, we collect 100 survey responses and obtain summary statistics.

What a worker would see:

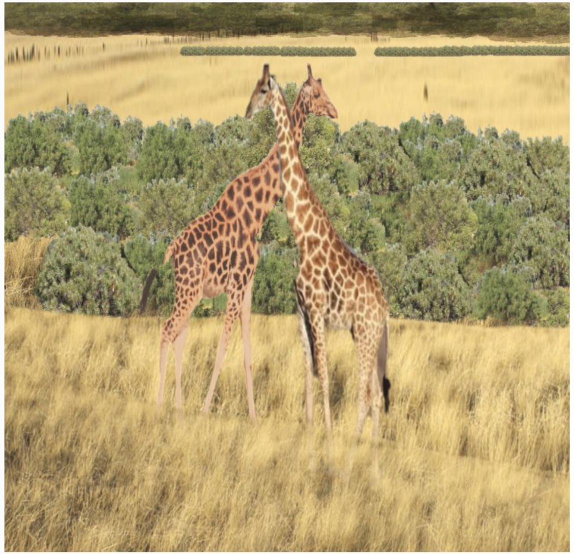
Instructions

The second image is a reconstruction of the first image.

- Compare the two images and rate your level of satisfaction from the reconstruction using the scale below (1=completely unsatisfied, 10=completely satisfied).

Original Image:

Image Reconstruction:



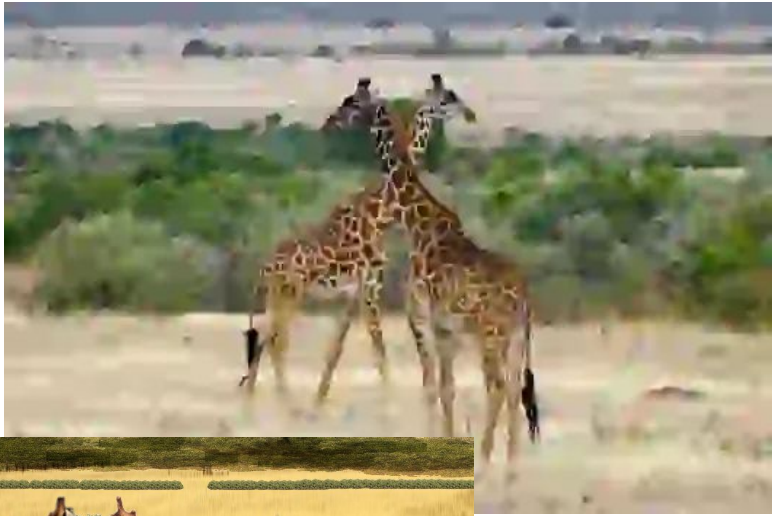
Level of Satisfaction:

- (•) 1 (completely unsatisfied) ( ) 2 ( ) 3 ( ) 4
- ( ) 5 ( ) 6 ( ) 7 ( ) 8 ( ) 9
- ( ) 10 (completely satisfied)

Selected Visual Results

Original

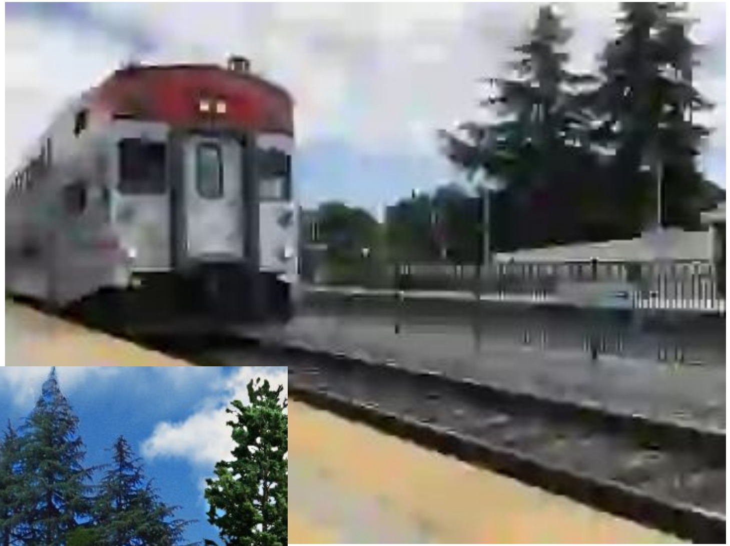
WebP



Human Compressed

Selected Visual Results

WebP



Original

Human Compressed

Selected Visual Results

Original

WebP

Human Compressed

Selected Visual Results

WebP

Original

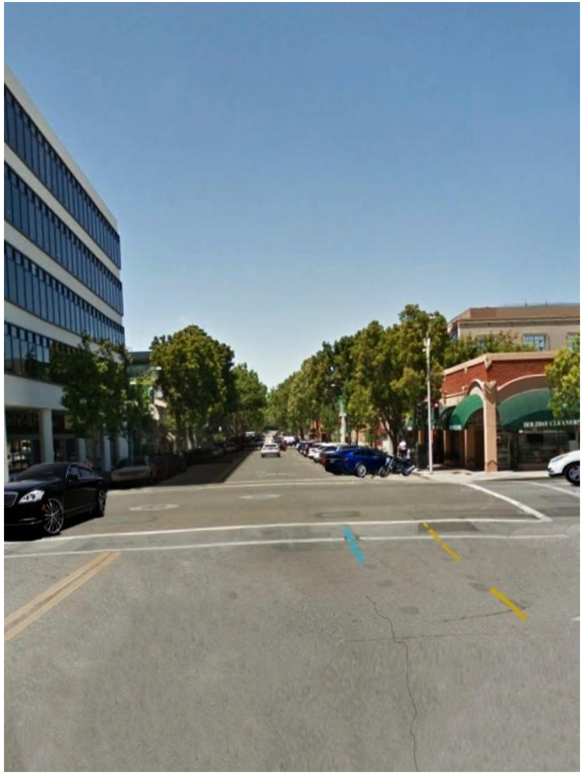
Human Compressed

Selected Visual Results

Original

WebP

Human Compressed



Selected Visual Results

Original

WebP

Human Compressed

Selected Visual Results

Original

WebP

Human Compressed

Results

➢ Mturk scores for Human and WebP reconstruction

Image	Original size (KB)	Compressed chat size (KB)	WebP size (KB)	Mean score		Median score	
				Human	WebP	Human	WebP
arch	1119	3.805	3.840	4.04	**5.1**	3	**5**
balloon	92	1.951	2.036	**6.22**	5.45	**7**	6
beachbridge	3263	4.604	4.676	**4.34**	3.92	4	4
eiffeltower	2245	4.363	4.394	**5.98**	5.77	6	6
face	1885	2.649	2.762	2.95	**5.47**	3	**6**
fire	4270	2.407	2.454	**6.74**	5.09	**7**	5
giraffe	5256	3.107	3.144	**6.28**	4.48	**7**	4
guitarman	1648	2.713	2.730	**4.88**	4.07	**5**	4
intersection	3751	3.157	3.238	**6.8**	4.15	**7**	4
rockwall	4205	6.613	6.674	4.41	**4.85**	4	**5**
sunsetlake	1505	4.077	4.088	**5.08**	4.82	5	5
train	3445	1.948	2.024	**6.85**	3.62	**7**	3
wolfsketch	1914	0.869	0.922	**8.25**	3.46	**9**	3

Conclusions

- Not a practical compression scheme, but ....
- Our experiment shows that human centric compression can be more powerful than traditional compression at very low bit rate
- Effective utilization of semantically and structurally similar images can dramatically improve compression ratio
  - Most public compressors do not take advantage of this rich public resource
  - Shows room for growth for traditional compression
- The human compression framework is useful as an exploratory tool, but not practical due to its labor-intensive nature.

Work Of The Future

- Limitations of our process
  - our drawing/editing skills
  - our avoidance of sophisticated software for image editing
  - the difficulty of manually searching for similar images
  - the inefficiency of the English language
- Neural network based models may be natural candidates for alleviating these problems and could eventually performance even better than that we have shown in this work