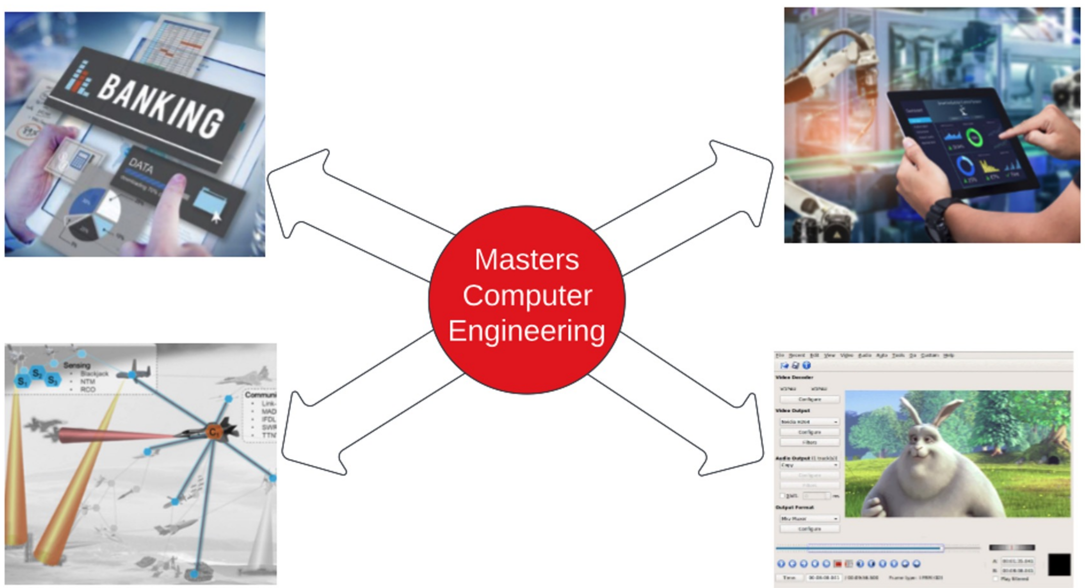
Evolution of Video Processing

Scott Labrozzi
VP Core Media Technology / Engineering Fellow
Disney Streaming

Where do I start …

NC STATE UNIVERSITY

1991



Where has this led me …

Co-Founder

Co-Founder

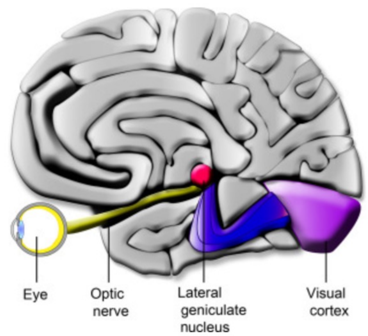
Acquired

Acquired/Became

- Co-Founding two Start-up Companies
- 30 Patents Issued to Date
- Developed and Co-Authored two Industry Standards
- Honored with Two Technical Emmys
- Loving every minute of it …

Having started in video processing, I never imagining that decades later I would still be working in and transforming in the domain

One-Slide Primer on Video Compression



- Understand human vision system (HVS)
  - We are more sensitive to black/white than color
  - Contrast, Noise, Frequency, Texture, Brightness, …
- Correlations
  - Data
  - Spatial
  - Temporal
- Predictive encoding
- Rate Control
  - To what degree loss is introduced to meet target bitrate

1992 …

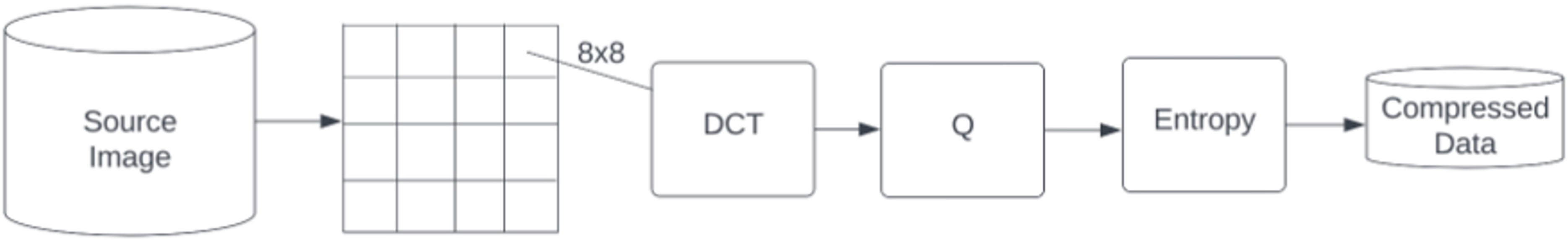
- **J**oint **P**hotographic **E**xperts **G**roup
  - Founded in late 1980s
  - Focus: Standard for Image-Based Compression
  - **JPEG** Published in 1992

- Spatial Redundancy
  - 20 year-old Transform from Nasir Ahmed, Discrete Cosine Transform (the **DCT**)
  - Spatial -> Frequency domain

- Exploits HVS’s varying sensitivities in the frequency domain
  - Quantization allows removal, with loss, spatial redundancy

- Statistical Characteristics
  - Quantized DCT Coefficients coded compactly based using Entropy encoding
  - Huffman Coding

JPEG High Level Block Diagram



In summary, the encoding tools we have for JPEG are

- DCT Transform
- Quantization
- Entropy (Huffman) Encoding

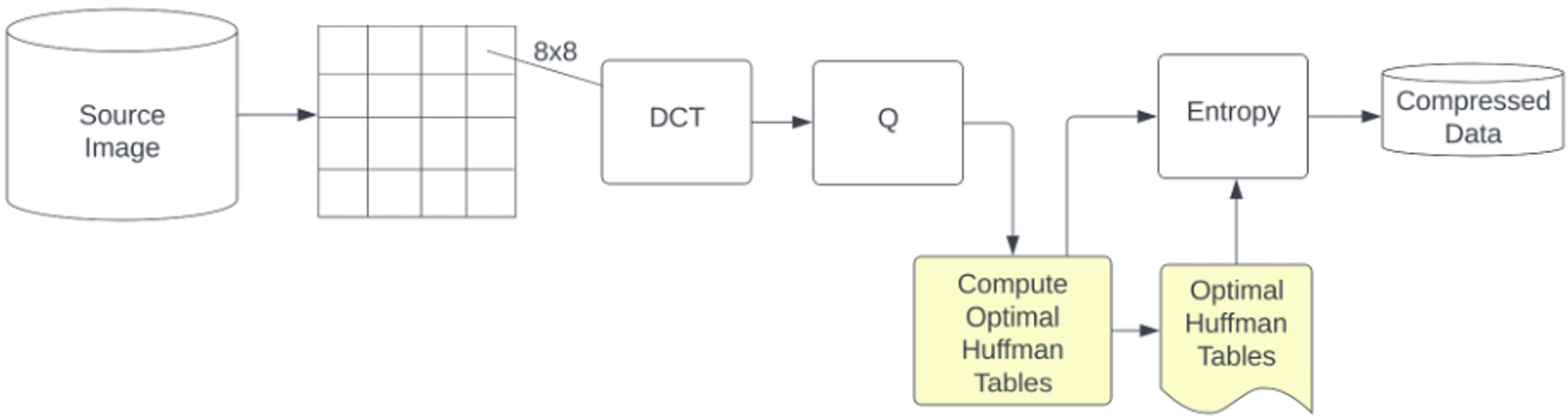
@Sun, My First Task: Make it Better, More Efficient!

Compression is an Art

- A Codec specification describes
  - A bitstream syntax
  - How *exactly* to Decompress from that syntax
  - (In)directly the set of tools (techniques) that can be used to produce the syntax
- A Codec specification does NOT describe
  - How to use the techniques, to achieve the syntax
  - Choices made to preserve vs lose, Rate control
- Compression is an Art
  - Prose vs Poetry
  - (A) -> Compression ART -> (B)
- Decompression is exact
  - (B) -> Decompress Specification -> (C)

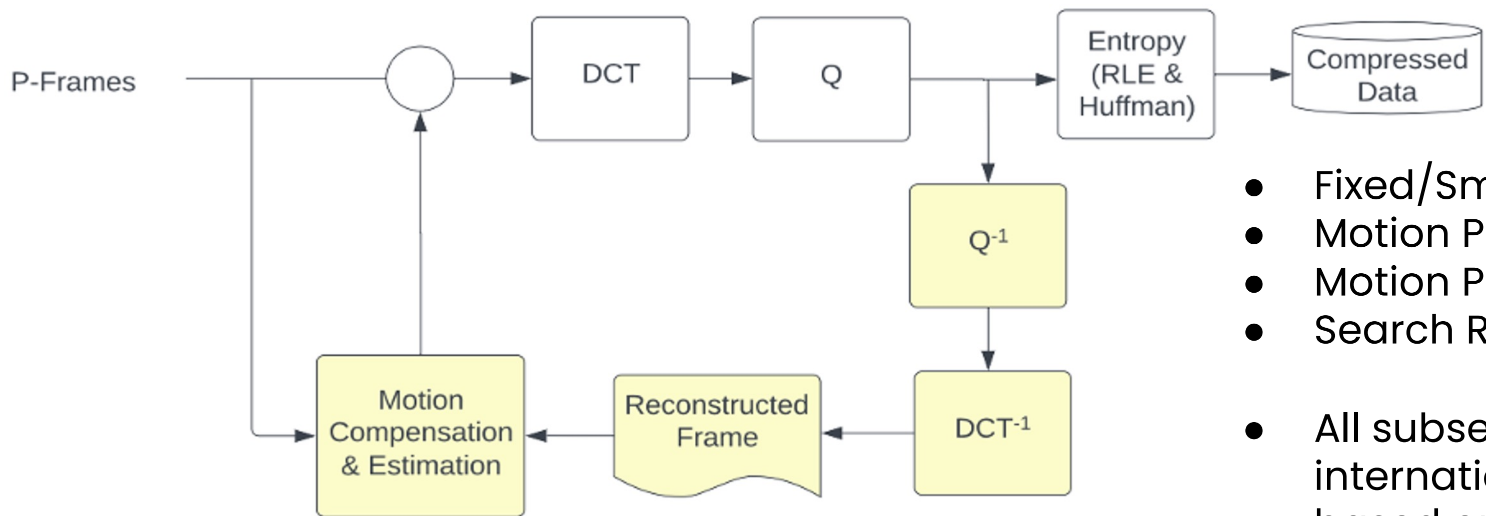
Optimal Huffman Encoding

- Entropy encoding exploits statistical characteristics of data
- JPEG provides a set of pre-defined Huffman tables
- Allows custom tables embedded in the output
- Create optimal Huffman tables given the quantized characteristics of a frame



Early Video Standard ...

- Video is sequences of frames .. could use MJPEG but exploits only spatial domain
- 1980s...
  - A standards study group formed to establish a standard for digital transmission of voice, video and data over the public telephone network for video conferencing
  - Future deployment of a transmission technology ISDN
  - H.261 published in 1988, not fully vetted until 1993



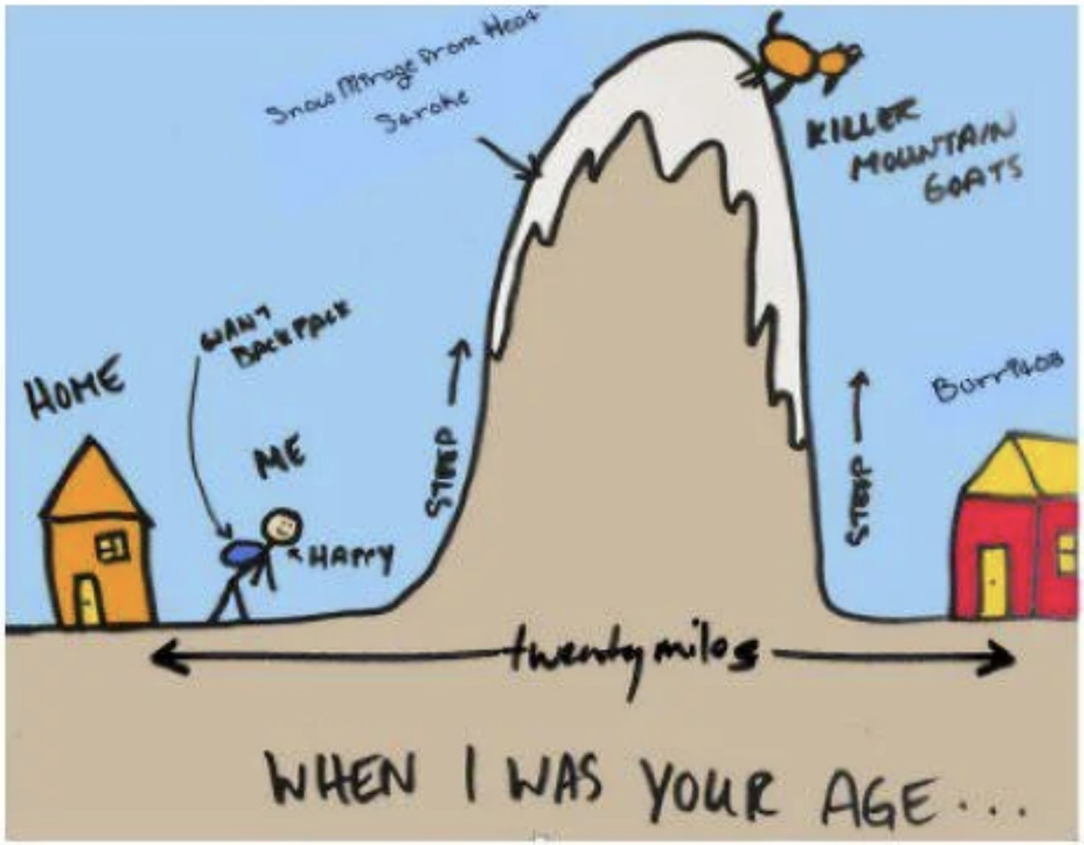
- Fixed/Small Resolutions
- Motion Prediction Forward
- Motion Prediction 1 Frame
- Search Range Limited

- All subsequent international standards based on H.261

Video .. Storage

- Applications / Use-Cases Drove Evolution of Video Codecs
- **M**otion **P**icture **E**xperts **G**roup
  - Formed in 1988 : Storage of Media on CD-ROMs (~ 1.5Mb/sec)
- **MPEG-1**
  - Similar to H.261
  - Didn't have resolution restrictions
    - Ideal operating point was ~ 352x240
  - By rearranging order frames are encoded
    - **B-Frames** : Bi-Directional Motion Compensation
  - Finally approved in 1992

You have it so Easy … When I Was Your Age …



By: Taylor Tam

Standard Definition (SD)
Analog-> Digital Television

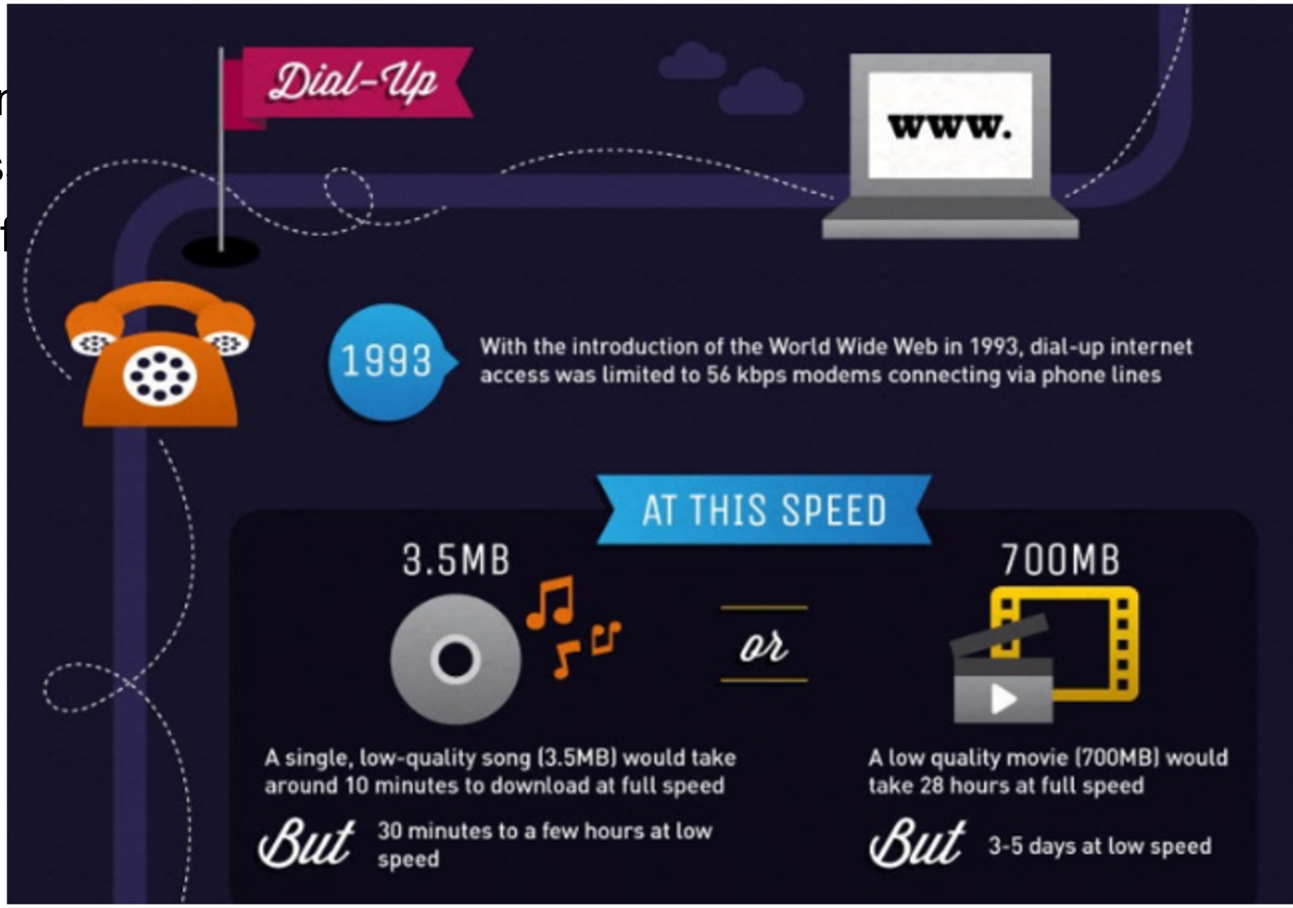
			Compared To (Smaller)		
	Width	Height	SD	HD	UHD
MPEG-1 Video	352	240	4x	24x	98x
SD Video	720	480		6x	24x
HD Video	1920	1080			4x
UHD Video	3820	2160			

- Use case: Broadcast Digital TV
  - Interlaced : a video *frame* is split into two temporally and spatially displaced *fields*.
- MPEG-2 / aka H.262 (1994)
  - Address interlacing television
  - Further compression techniques, such as adaptive quantization
  - MPEG-2 looks essentially like that of MPEG-1 and H.261.

A New Frontier …

This is your Parent’s Internet

- The i… eo
  acros…
- But o…



Proprietary Video / Streaming Formats

- The video standards to date were completely insufficient for the challenges of video on the internet

- The Rise of Proprietary Formats
  - Real Video (Real Networks)
  - Windows Media Video (Microsoft)
  - Flash Video (Adobe - Initially a version of H.263)

- First-Generation Streaming Formats

Osprey & What Could-have-Been

- Co-founded with 5 others from Sun Microsystems
- Early years
    - Focused on Video Encoding Hardware
- Late 90s
    - Transitioned to Video Capture
    - Partnered and became preferred solution for streaming
- Late 90s had this grand idea …
    - **A video hosting site that would allow consumers and businesses to upload and share videos**

The Dawn of High Definition (HD) TV

- In Parallel with Streaming on the Internet
  - Another Transformation in Broadcast TV : SD to HD
- Raleigh is Famous!
  - First Public HDTV Broadcast in US
    - 1996 by our own WRAL
- One thing to Demonstrate Technology …
  - Another thing entirely to deploy at scale
  - Challenges with High Definition
- Second Start-Up 2003
  - Co-Founding Inlet Technologies as their CTO
  - WRAL's parent company, Capital Broadcasting, played a critical role in supporting our efforts being our first angel investor

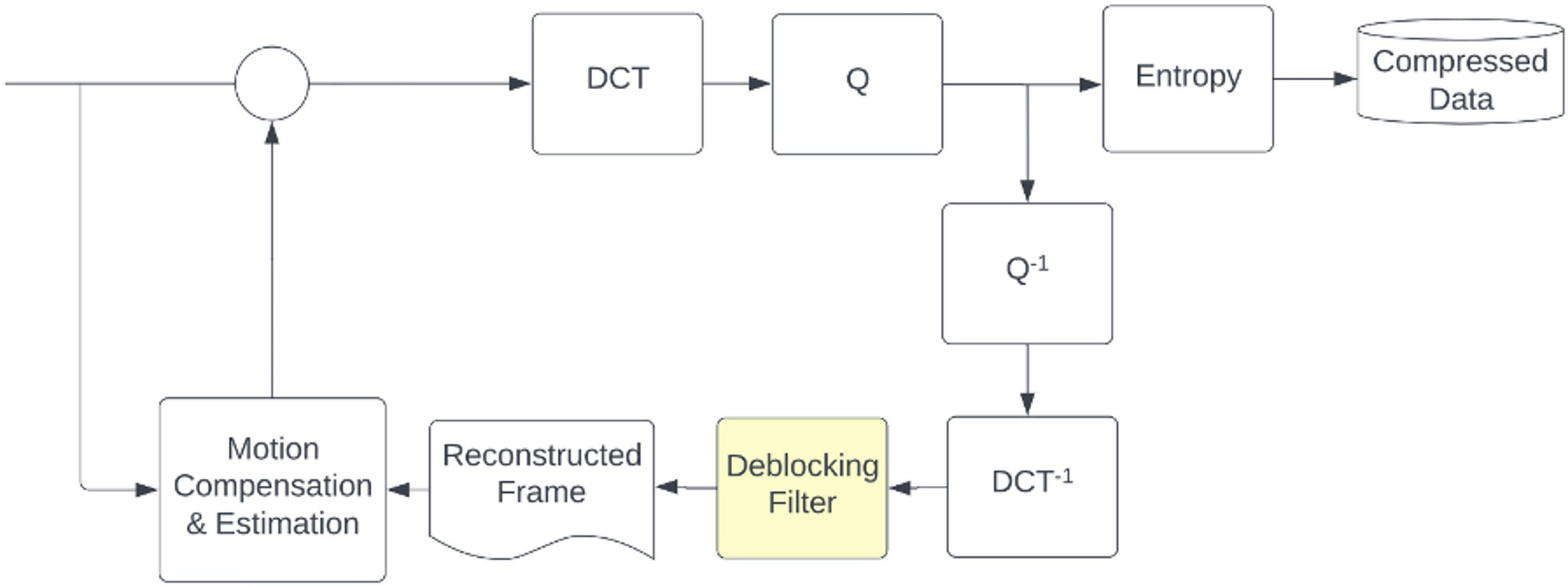
HD Compression for DVD

- Inlet's basis, much like MPEG-1, to address a storage problem
  - Two potential storage mediums coming to market at the time
    - HD-DVD Disc and Blu-ray Discs
  - MPEG-2's compression efficiencies were insufficient.

- Alternatives to MPEG-2
  - VC-1 : extension of the Windows Media Video format
  - H.264 (aka MPEG-4 Part10 aka AVC Advanced Video Codec)
    - Published in 2003

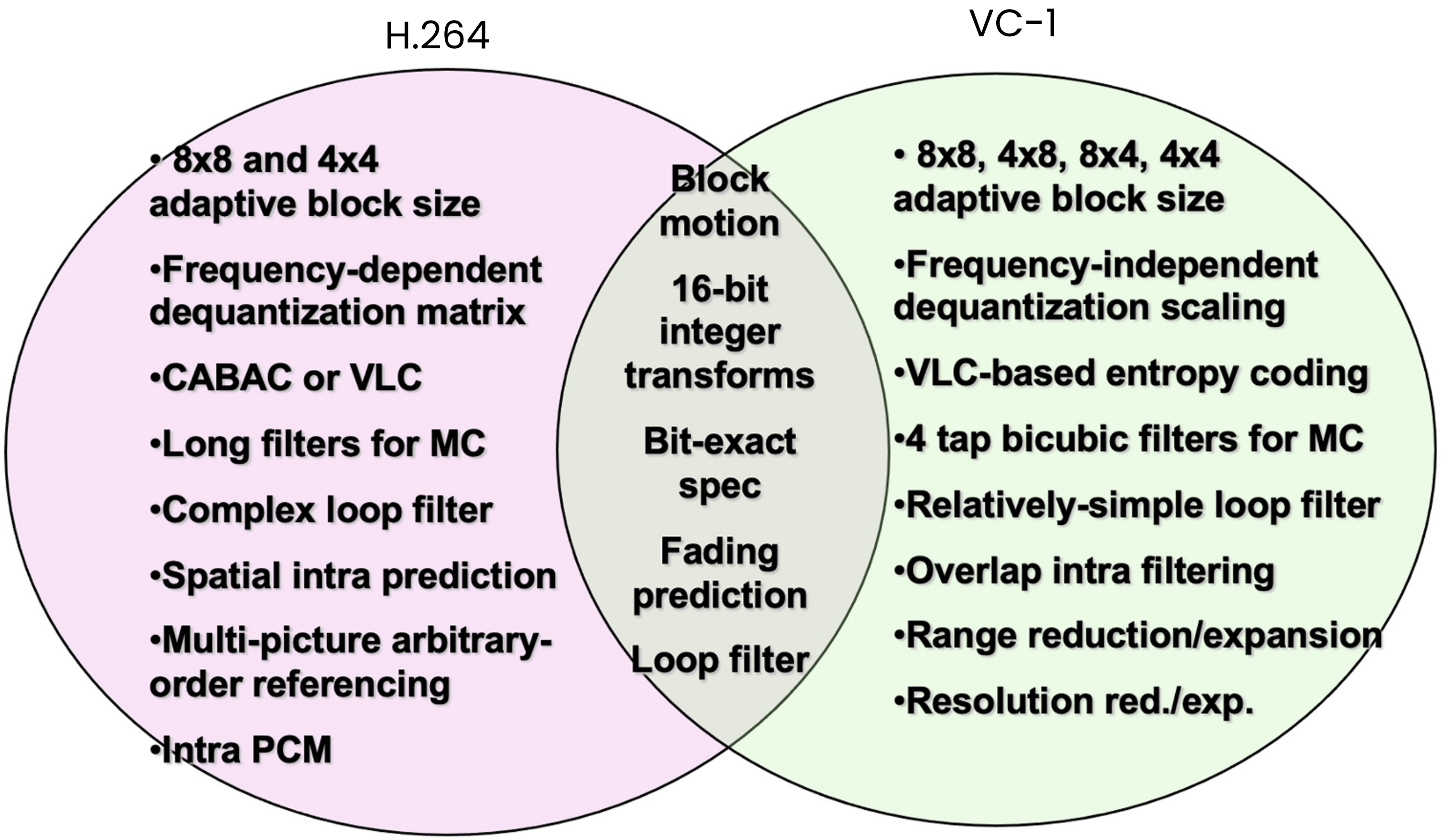
- VC-1 and AVC similar to MPEG-2
  - Both introduced an interesting new encoding tool – that of an in-loop deblocking filter

VC-1 / H.264



DCT
Q
Entropy
Compressed Data
Q-1
DCT-1
Deblocking Filter
Reconstructed Frame
Motion Compensation & Estimation

VC-1 / H.264 Improved and New Tools



Viability - Lesson on Compression Complexity

- Codecs are Asymmetric
    - Compression is highly more computationally complex than Decoding
    - Generational 6x-10x (Encode) vs < 2x (Decode)
- VC-1 < H.264 Complex in encoding tool set
    - Made VC-1 more viable to implement in 2003+ than H.264
        - Compute/Performance/Cost
- Understand technological viability and roadmaps
- At Inlet, as such, we chose to work on VC-1
    - Side note: 1st career patent on techniques to partition video compression across multiple hardware units

Shifting One's Strategy …

- Knowing when it's time
  - Market Dynamics
  - Technological Changes

- 2006, Inlet…
  - Pivoted / Focus on Streaming Delivery
  - Not Capture (Osprey) but *Software* Streaming Solutions

- **Fortuitous!!**
  - Led Inlet into the Advent of Adaptive Streaming

Adaptive Streaming

- Streaming before 2009
  - Source -> 1+ Bitrate Variants
  - Client Selected a Variant
  - Playback started after enough downloaded
  - Playback paused for rebuffering if network insufficient
- Adaptive Streaming
  - Source -> 1+ Bitrates
  - Encode is Segmented into short duration files
  - Client Hits Play
  - Playback starts quickly
  - Playback adapts across variants based on network conditions

Competition is Good …

- Several adaptive streaming formats came to market
    - Windows Media Smooth Streaming (based on VC-1)
    - Apple’s HTTP Live Streaming (based on AVC)
    - Adobe’s Adaptive Streaming (based on Flash Video)

- Inlet: developed a solution around all of them
    - Became the defacto solution for Adaptive Streaming

- Use strengths of one format to drive improvements in others

Adaptive Streaming's Viability for Traditional TV

- Consumer Internet Bandwidth Increases
  - Led to Adaptive Streaming's ~Viability as an **Alternative** to Traditional TV Providers

- Loosely
  - Much still to do behind the scenes
  - Was a named inventor on 15 patents over next 5 years

- Powered by MLBAM and Inlet/Cisco Live Adaptive Streaming Encoders
  - 2015 Sony Vue launched, combining live TV with traditional on-demand video
  - 2017 Hulu expanded their previously on-demand service with Live-TV

- Adaptive Streaming became an option **for** Traditional TV Providers
  - Was a primary author or two industry specifications / standards

The Dawn of Ultra High Definition (UHD) & the Codec Standards Schism

- In the background … Standards Org developing the next Gen Format
- H.265 / HEVC (High Efficiency Video Coding)
    - 2013 Approval
    - Again, based on similar concepts of predecessors
    - Many new tools, upwards of 50% savings for UHD resolution over AVC
    - Computational Complexity 6x+ Higher than AVC
- Egregious Licensing Terms / Uncertainty in Holders Shocked the Ecosystem
- 2015 Alliance for Open Media (AOM) Created
    - Cisco, Google, Amazon, Microsoft and others
    - Create royalty-free video standard alternatives to MPEG
    - AV1 Published in 2018
- Competition forced HEVC to change its licensing terms

Disney Streaming and the + of Disney+

- 2016 -> MLBAM -> Disney Streaming
  - Develop our own Live/VOD Adaptive Streaming Solution
  - No man's an Island ...

- The **+** in Disney+'s Product Name

- Walt Disney to his Imagineers: "Plus It"
  - Take a good idea and make it even better

- Team/I - passion to uphold the Quality associated with the Disney brand continually drives us to ensure the visual impact of our storytelling reaches our customers

- Two Technical Emmy's for technology developed for Disney+
  - *Driven by : how can we do better?*

Machine Learning Now / Future

- A Reflection : Final Semester, NCSU, 1991

- Now, Much Wiser …
  - Evolution of Video Compression Spanning 3+ Decades = Variations on a Theme
  - We've begun to exhaust potential
    - Explore new means for future gains
  - Exponential increase in complexity
    - Techniques happening faster than pace of computational progress

- ML is sought to be used from things such as decision making, in-loop filtering, etc.

Recent ML Centric COmpression Patent Filings

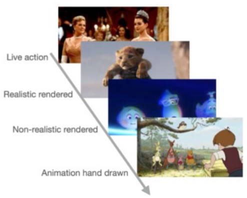
Systems and Methods for Image Compression at Multiple, Different Bitrates	Machine Learning Based Video Compression	Latent Space Constraints for Learned Video Compression	Codec Rate Distortion Compensating Down-sampler	Machine Learning Based Video Compression
Systems and Methods for Distortion Removal at Multiple Quality Levels	Machine Learning Based Video Compression	Normalizing Flow for Image Compression	Machine Learning Model Based Embedding for Adaptive Content Evaluation	Microdosing for Low Bitrate Video Compression
Content Adaptive Optimization for Neural Data Compression	Machine Learning Based Video Compression		Knowledge Distillation for GAN Based Video Codec	

ML Beyond Video Coding

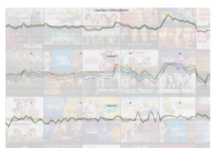
Content Characterization For Media Processing

Discern inter/intra content characteristics to drive multiple decisioning points in the media processing pipeline resulting in business and customer cost savings while maintaining or improving upon DSS's visual and delivered quality of experience.

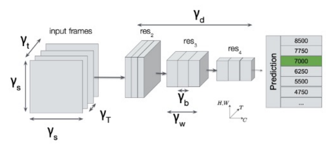
In the previous milestones, we developed a deep-learning based solution that can select the best encoding parameters according to subjective data gathered in partnership with Disney Streaming (DS). The main challenge has been that we cannot fully trust any objective quality metric and thus have to rely on scarce subjective data. This is particularly true in the case of film grain where current metrics favor overly smooth results that remove the original artistic intent (e.g., film grain). To overcome such issues, our solution uses a combination of contrastive learning and fine-tuning on a small set of subjective data. In this milestone, we are focusing on analyzing the performance of our solution and benchmarking it on novel hardware available to DS. We performed a number of the planned experiments already and identified certain bottlenecks that we plan to further investigate in the next steps within this milestone.



Content Genre Mapping



Encode Parameter Tuning



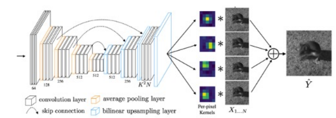
Base Stack Optimization

ML Pre-Processing

Develop a set of Advanced ML approaches to video pre-processing to improve visual quality while maintaining the studios creative intent through to the end user. Pre-processing tasks that are of specific interest in this project are downsampling and dithering. An expected outcome of pre-processing improvements will be cost savings and higher end user quality of experience due to compression efficiencies gained.

Down Scaling

We are experimenting with different machine learning architectures, color spaces, and perceptual losses to offer better content-adaptive downsampling. The current results shows advantages of using a learning-based approach for downsampling for some types of content when compared to state-of-the-art methods.



Neural Banding Detection and Mitigation

In the previous milestone, we focused on performance optimizations and compared different network architectures for our neural debanding solution. Moreover, we investigated how to extend it to perform neural dithering as a pre-processing step. Most of the performance evaluation on the previous milestone has focused on synthetically induced banding artifacts. In this milestone, we are focusing on testing and evaluating our neural debanding with in-the-wild compressed content. We are assessing both objective and subjective quality when compared to other state-of-the-art methods to analyze the advantages of our learned approach.

Bicubic (Pytorch)

Neural Network (L1)

A/V SYNC ANALYSIS

Project Description

The problem of determining, in a non-reference domain, whether audio is synchronized to video is highly desired in many production flows, both for file-based and real-time (live/linear) use-cases. The definitive cue for synchronization is lip-sync, although other visual and audio cues can be leveraged as well. Reference-based a/v sync solutions are also desired, but those only tell us if a test asset is in sync with the reference, not if the reference or test itself is in a/v sync. We desire to devise a ML data-driven model to automate the detection and any skew of sync between a/v. A skew value in live use cases (Hulu/E+/etc) could further be used in a feedback model to address a/v sync problems.

Other ML-based R&D Projects

Thank You & Q/A