Video quality metadata in compressed bitstreams

Ioannis Katsavounidis

Research Scientist

Video Infrastructure

Facebook

Outline

- Video content at Facebook
- Video quality measurement at Facebook
- Upload quality calculation
- Metadata in digital images
- Full-reference metrics as video quality metadata

OCULUS

Quest 2

Workplace

Workplace News

Supporting emergency services and government organizations during COVID-19

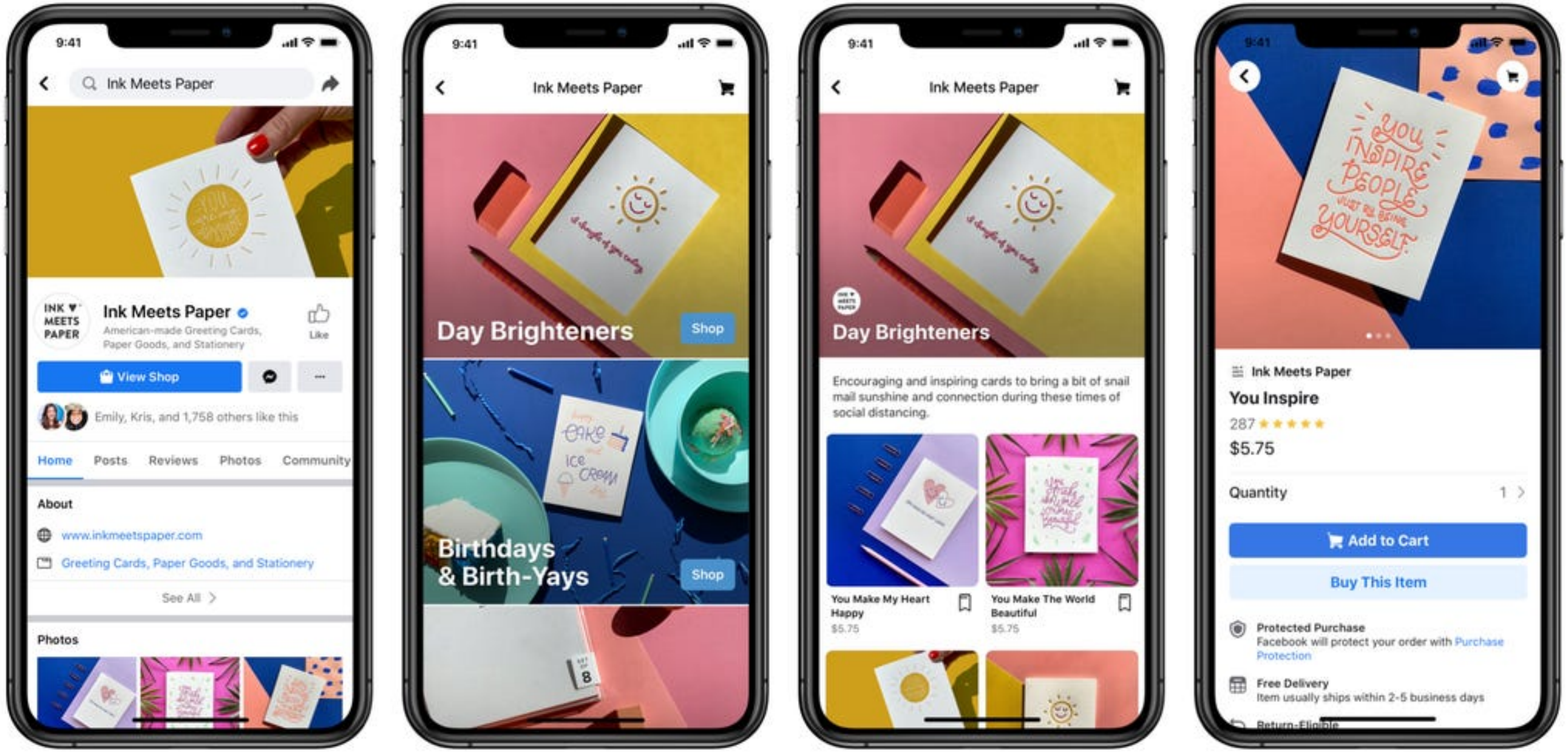
by Julien Codorniou

Today we signed with one of the world's most prestigious department stores and fashion houses, Libertie Globale, based in Paris. Our Excelsior and Decolletage range of silks and.... See more

Rooms2Live

Faith Communities

Shops



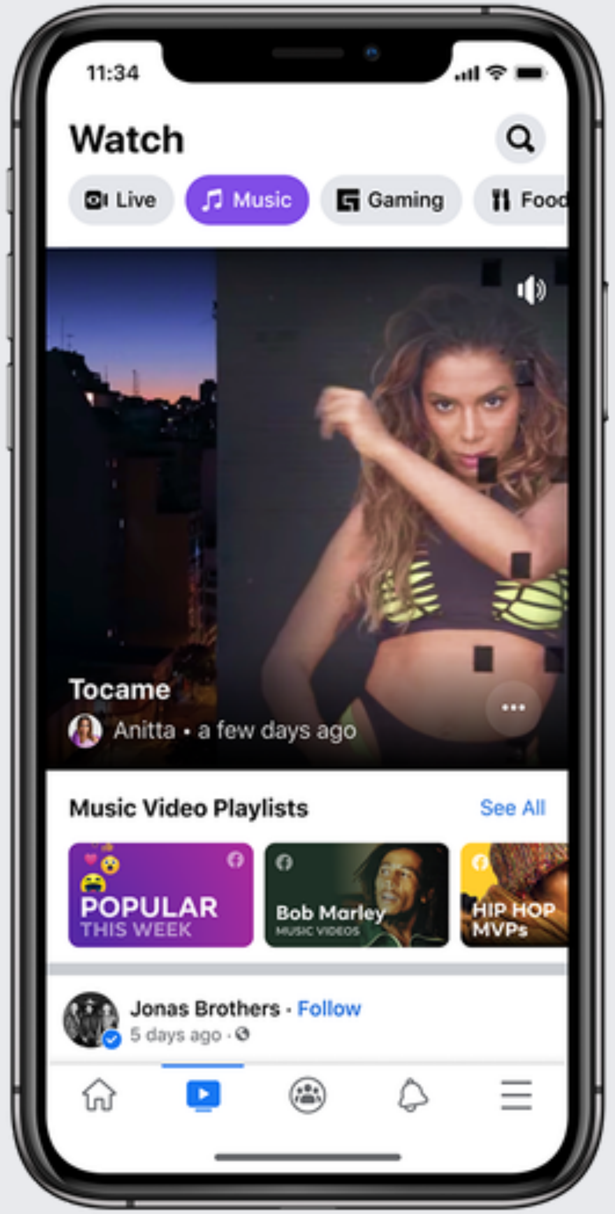
Emily, Kris, and 1,758 others like this

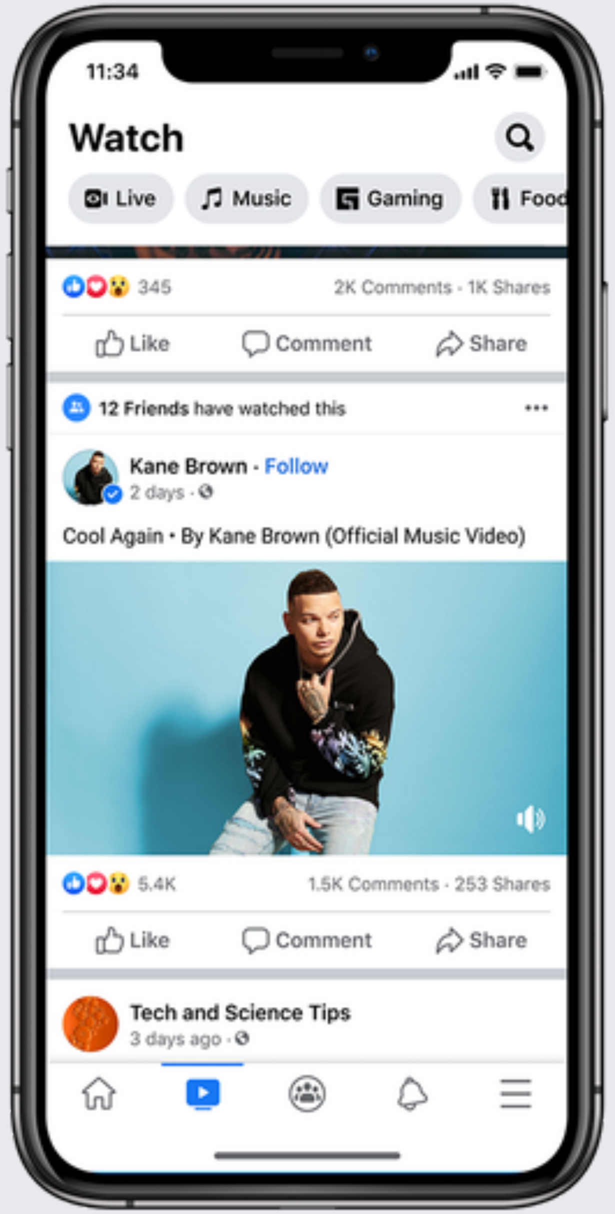
Encouraging and inspiring cards to bring a bit of snail mail sunshine and connection during these times of social distancing.

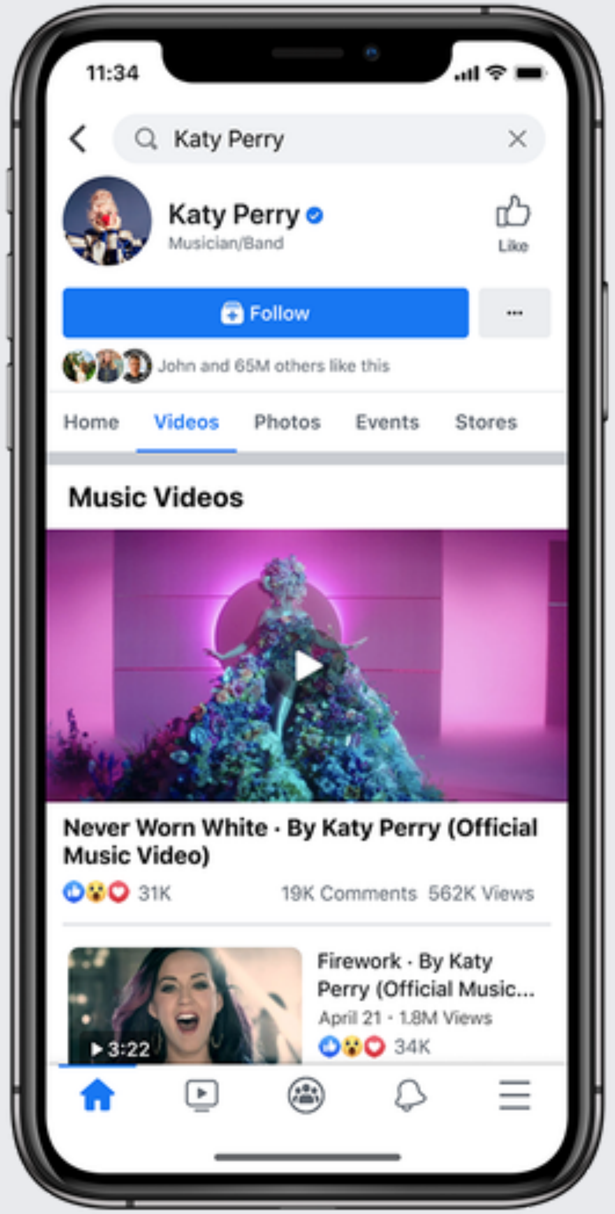
Facebook will protect your order with Purchase Protection

Item usually ships within 2-5 business days

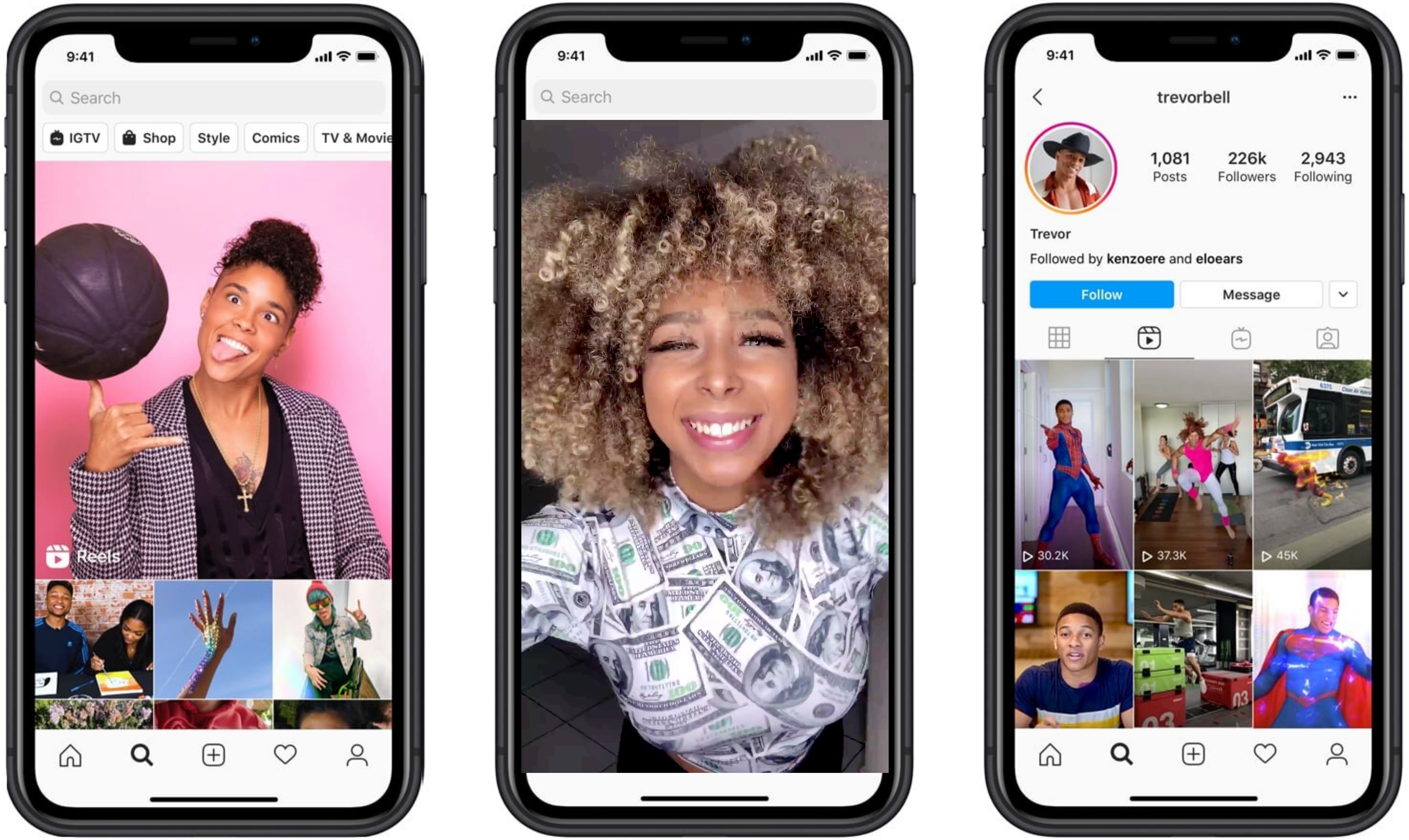
Premium Music Videos







Reels



trevorbell

1,081 Posts 226k Followers 2,943 Following

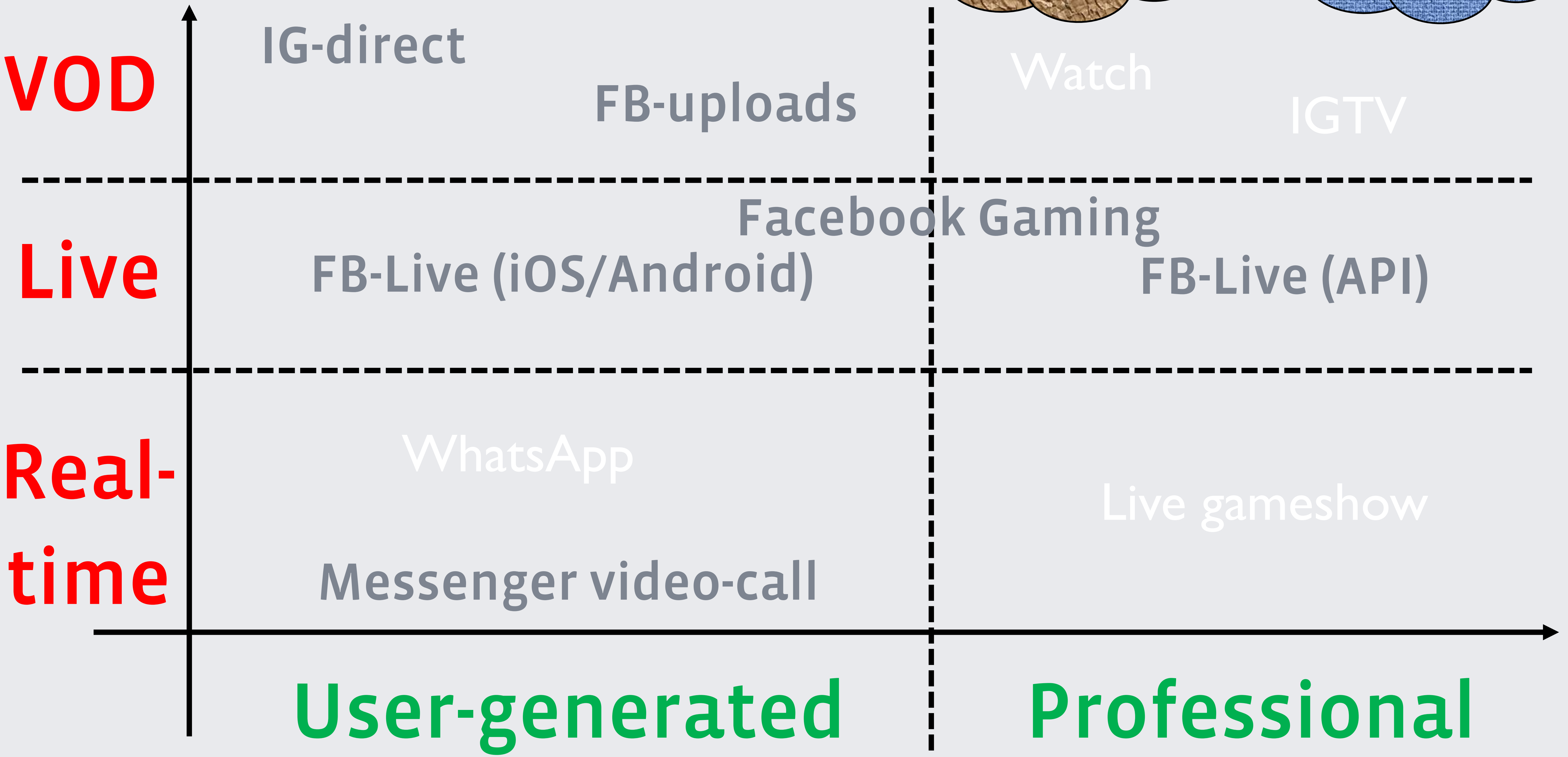
Trevor

Followed by **kenzoere** and **eloears**

Video content at FB

Portal

Oculus

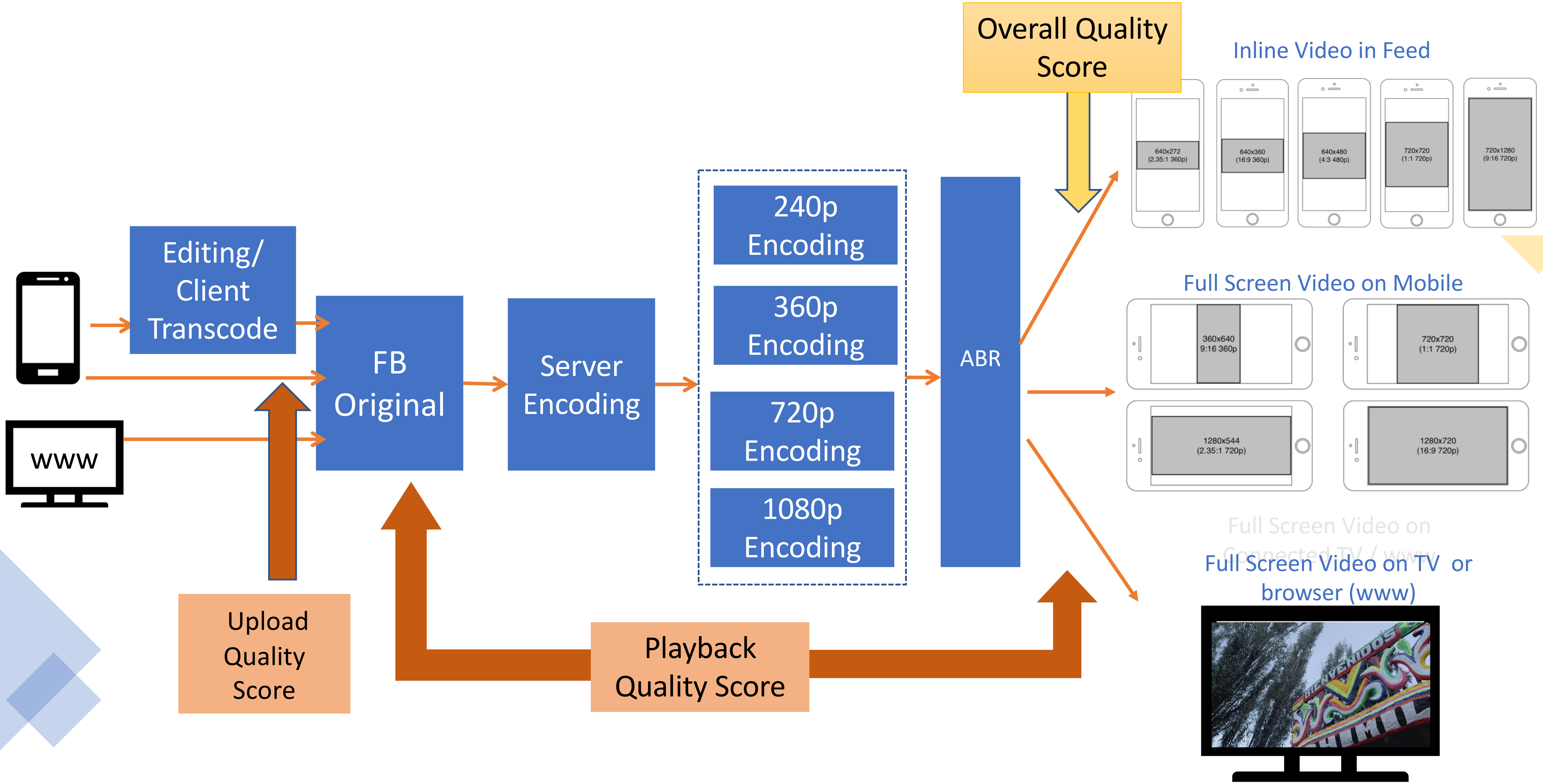


	User-generated	Professional
VOD	IG-direct FB-uploads	Watch IGTV
Live	Facebook Gaming FB-Live (iOS/Android)	FB-Live (API)
Real-time	WhatsApp Messenger video-call	Live gameshow

Challenge in Quality Assessment – Variation in Uploaded Video Quality

- **High quality ingested videos**
  - Curated content and some UGC
- Some **UGC can be really low quality**
  - In reshared UGC, source is already highly compressed
    - Downloading from WhatsApp/Messenger and uploading to FB
  - Client transcoding needed to upload reliably from poor connections (2G/3G)
    - High-quality source transcoded to low resolution.
- FB Products make it is **easy to edit/remix content prior to upload**
  - Memes often start with low-quality source and adds text/images on top.
  - Quality is in the “eye of the beholder”

Quality Metric (FB-MOS) Building Blocks



Overall Quality Score

Inline Video in Feed

Editing/ Client Transcode

FB Original

Server Encoding

240p Encoding

360p Encoding

720p Encoding

1080p Encoding

ABR

Full Screen Video on Mobile

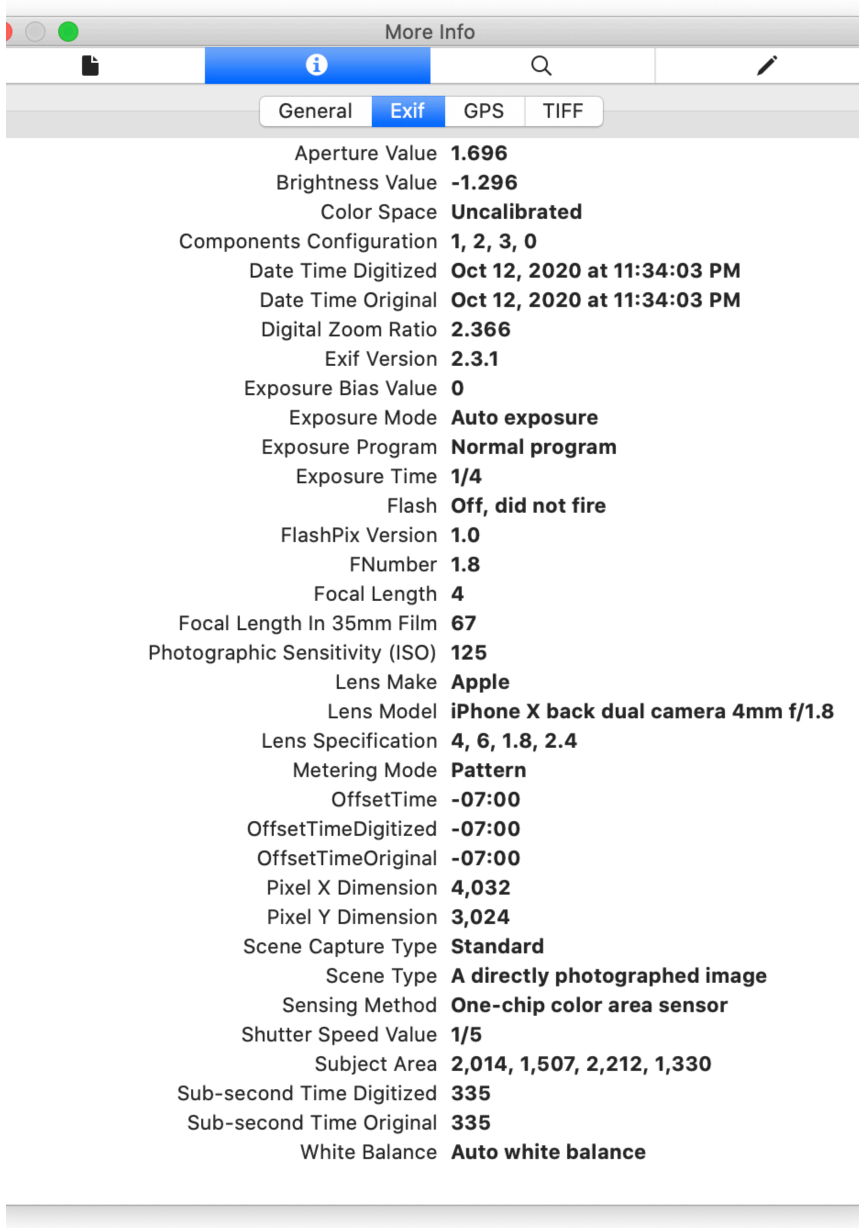
Full Screen Video on Connected TV / www

Full Screen Video on TV or browser (www)

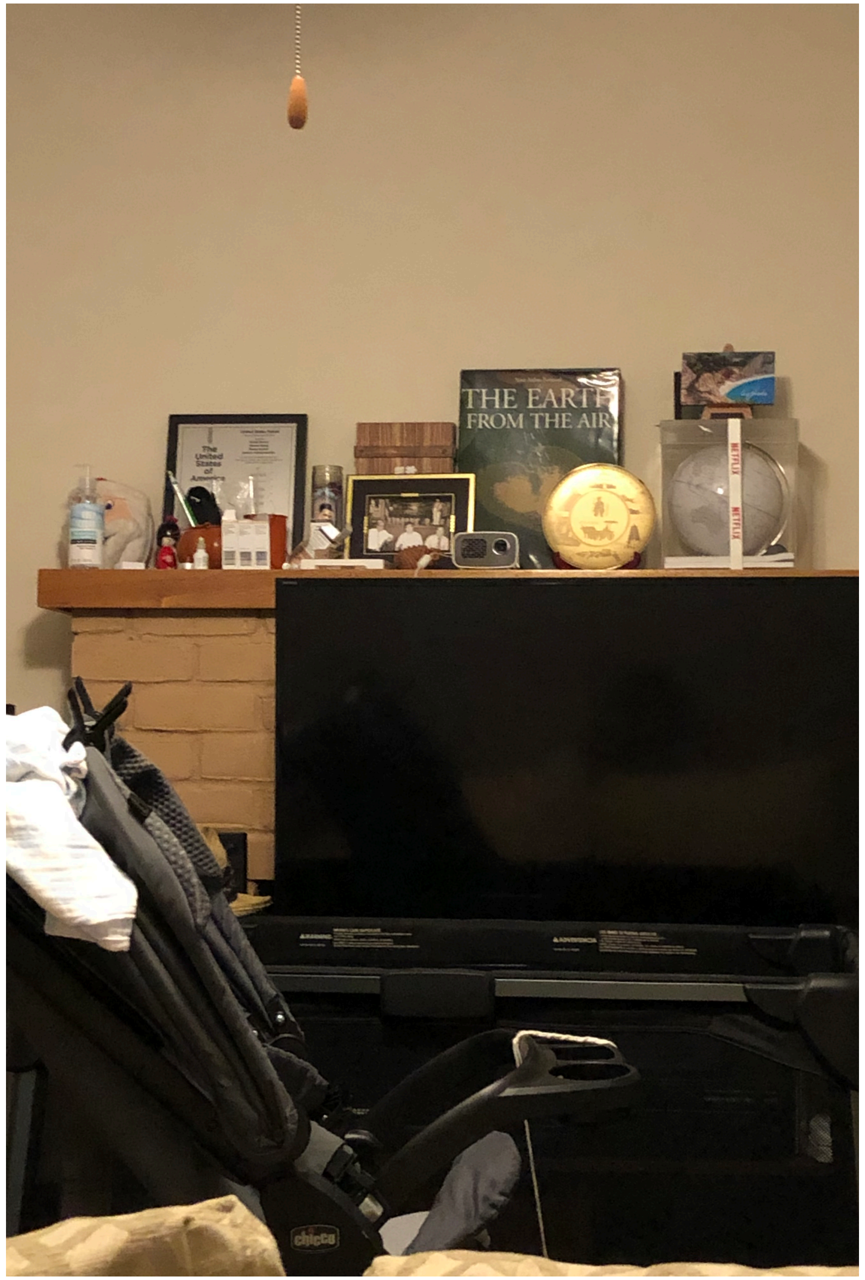
Upload Quality Score

Playback Quality Score

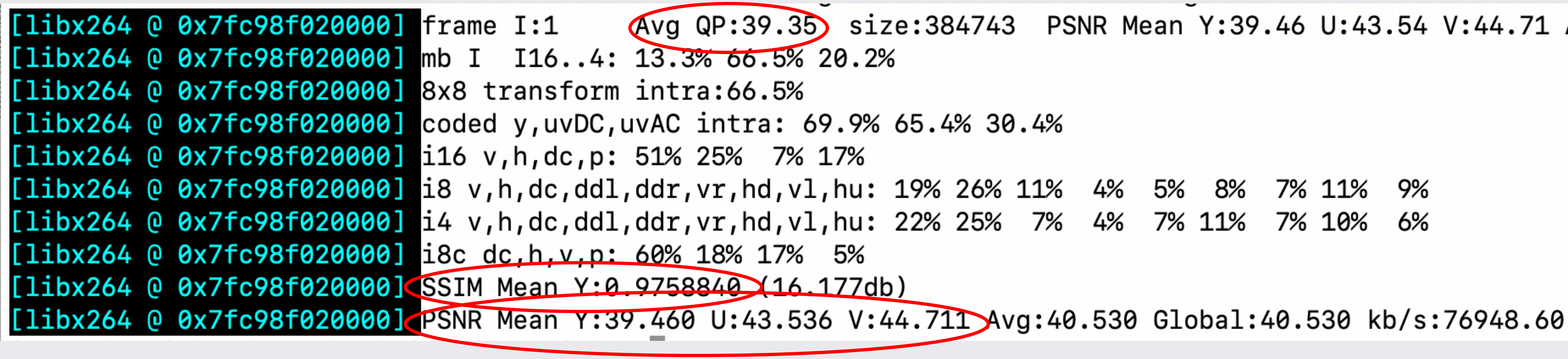
Image (JPEG) metadata: EXIF



Aperture Value	1.696
Brightness Value	-1.296
Color Space	Uncalibrated
Components Configuration	1, 2, 3, 0
Date Time Digitized	Oct 12, 2020 at 11:34:03 PM
Date Time Original	Oct 12, 2020 at 11:34:03 PM
Digital Zoom Ratio	2.366
Exif Version	2.3.1
Exposure Bias Value	0
Exposure Mode	Auto exposure
Exposure Program	Normal program
Exposure Time	1/4
Flash	Off, did not fire
FlashPix Version	1.0
FNumber	1.8
Focal Length	4
Focal Length In 35mm Film	67
Photographic Sensitivity (ISO)	125
Lens Make	Apple
Lens Model	iPhone X back dual camera 4mm f/1.8
Lens Specification	4, 6, 1.8, 2.4
Metering Mode	Pattern
OffsetTime	-07:00
OffsetTimeDigitized	-07:00
OffsetTimeOriginal	-07:00
Pixel X Dimension	4,032
Pixel Y Dimension	3,024
Scene Capture Type	Standard
Scene Type	A directly photographed image
Sensing Method	One-chip color area sensor
Shutter Speed Value	1/5
Subject Area	2,014, 1,507, 2,212, 1,330
Sub-second Time Digitized	335
Sub-second Time Original	335
White Balance	Auto white balance



Transcoding example (FFMPEG/x264)



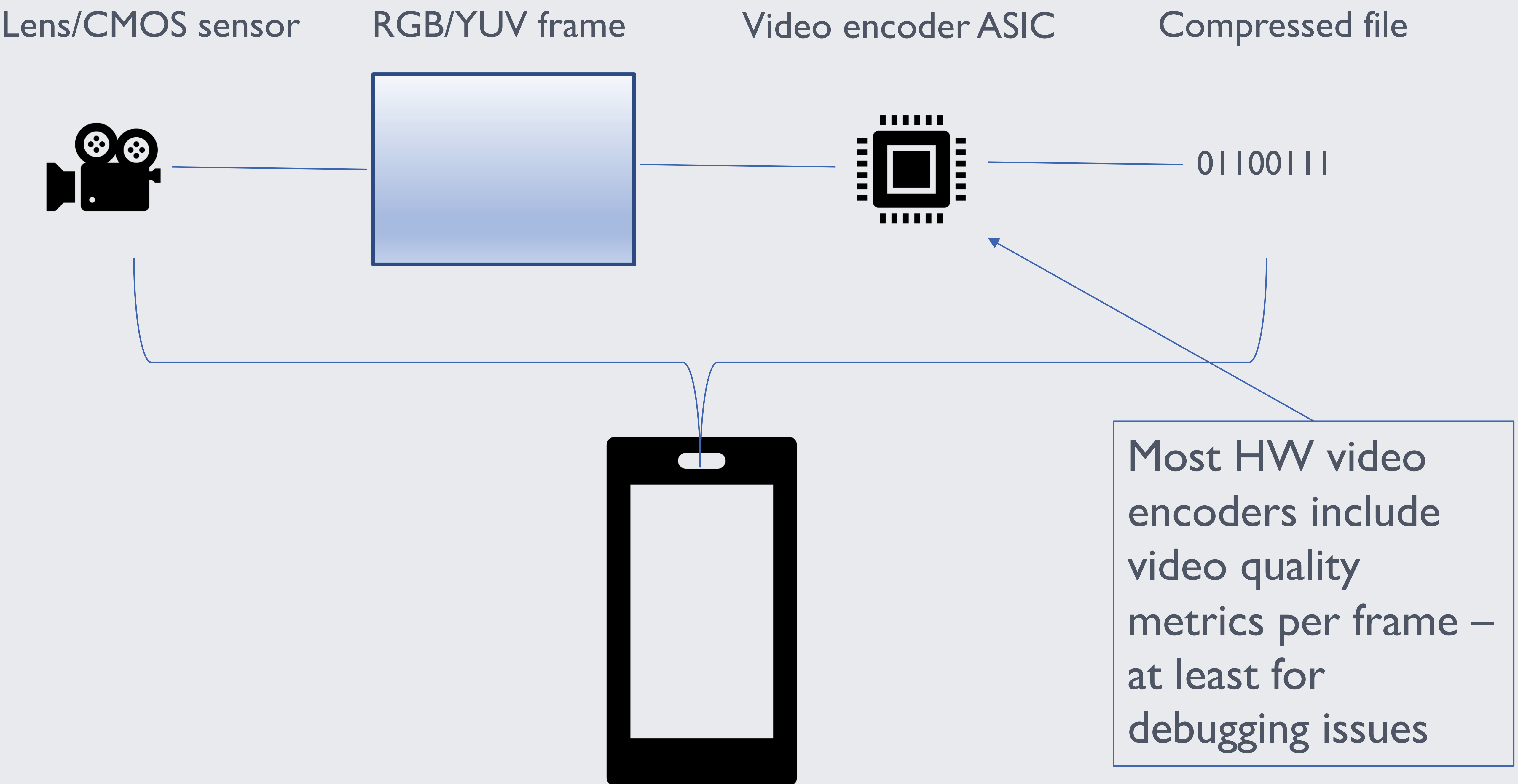
```
[libx264 @ 0x7fc98f020000] frame I:1     Avg QP:39.35  size:384743  PSNR Mean Y:39.46 U:43.54 V:44.71
[libx264 @ 0x7fc98f020000] mb I  I16..4: 13.3% 66.5% 20.2%
[libx264 @ 0x7fc98f020000] 8x8 transform intra:66.5%
[libx264 @ 0x7fc98f020000] coded y,uvDC,uvAC intra: 69.9% 65.4% 30.4%
[libx264 @ 0x7fc98f020000] i16 v,h,dc,p: 51% 25%  7% 17%
[libx264 @ 0x7fc98f020000] i8 v,h,dc,ddl,ddr,vr,hd,vl,hu: 19% 26% 11%  4%  5%  8%  7% 11%  9%
[libx264 @ 0x7fc98f020000] i4 v,h,dc,ddl,ddr,vr,hd,vl,hu: 22% 25%  7%  4%  7% 11%  7% 10%  6%
[libx264 @ 0x7fc98f020000] i8c dc,h,v,p: 60% 18% 17%  5%
[libx264 @ 0x7fc98f020000] SSIM Mean Y:0.9758840 (16.177db)
[libx264 @ 0x7fc98f020000] PSNR Mean Y:39.460 U:43.536 V:44.711 Avg:40.530 Global:40.530 kb/s:76948.60
```

Elementary video quality information about this encode is readily available

- Per frame average QP
- Per frame PSNR (Y/U/V)
- Per frame SSIM

At near-zero compute overhead

How about camera capture?



Lens/CMOS sensor

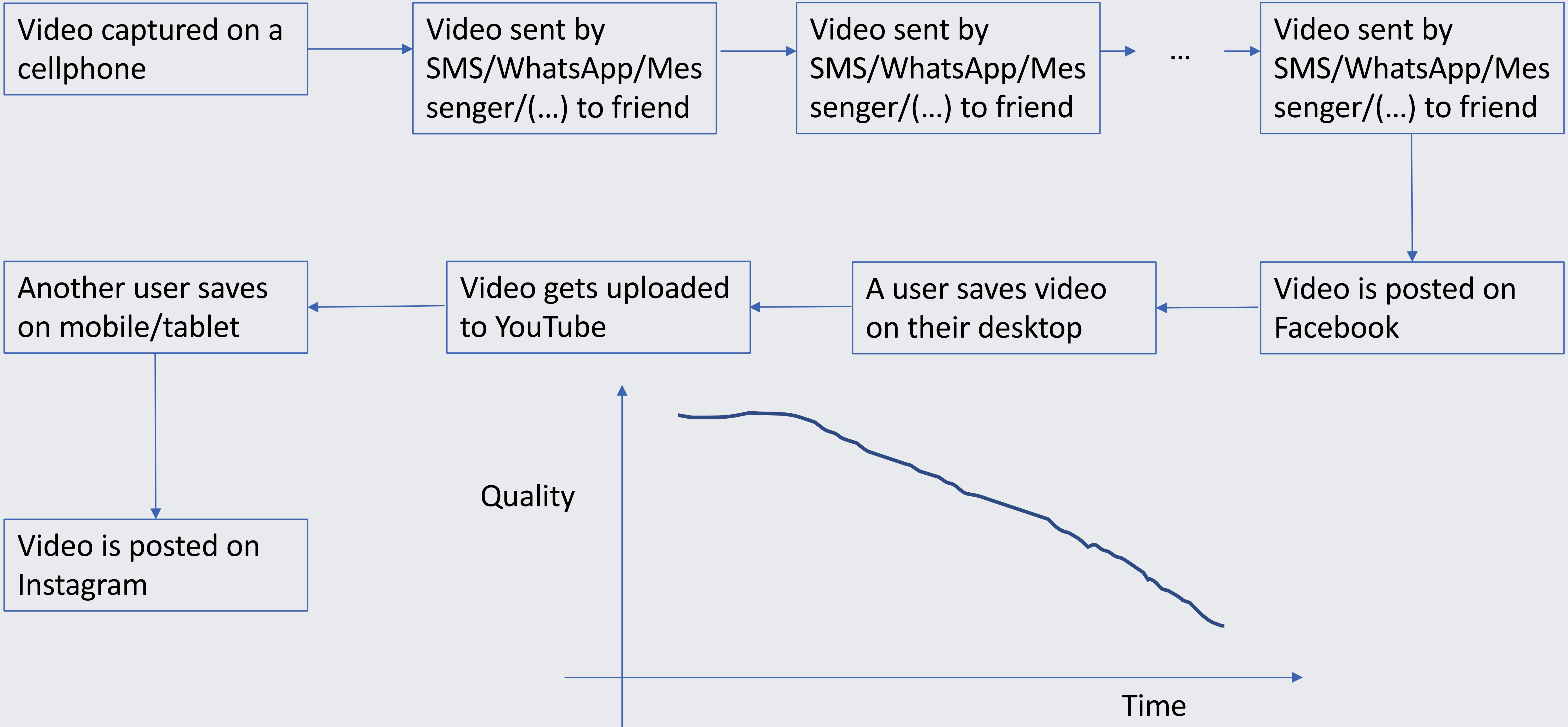
RGB/YUV frame

Video encoder ASIC

Compressed file

Most HW video encoders include video quality metrics per frame – at least for debugging issues

The life-cycle of a UGC video



Video captured on a cellphone

Video sent by SMS/WhatsApp/Messenger/(…) to friend

Video sent by SMS/WhatsApp/Messenger/(…) to friend

…

Video sent by SMS/WhatsApp/Messenger/(…) to friend

Video is posted on Facebook

A user saves video on their desktop

Video gets uploaded to YouTube

Another user saves on mobile/tablet

Video is posted on Instagram

Challenge

- Each transcoding pipeline estimates source video quality using no-reference metrics to determine best ingestion strategy
- During transcoding, full-reference quality metrics are generated to determine best encoding settings/ABR strategy
- Estimation errors propagate and accumulate when cascading multiple transcoding pipeline
- No-reference metrics require significant compute overhead

Existing proposals

- ISO/IEC 23001-10, MPEG Systems Technologies – Part 10: Carriage of timed metadata metrics of media in ISO base media file format
- ISO/IEC 23001-13, MPEG Systems Technologies – Part 13: Media orchestration
- ISO/IEC 13818-1:2015/AMD 6:2016 Carriage of Quality Metadata in MPEG2 Streams
- ISO/IEC 23009 Dynamic Adaptive Streaming over HTTP (DASH)

Existing proposals (cont’d)

- Video quality metrics covered by MPEG standards
  - PSNR
  - SSIM
  - MS-SSIM
  - VQM
  - PEVQ
  - MOS
  - FSIG

Existing proposals – pros and cons

- Good starting point, offering a system-level (container) mechanism to store per-frame quality metadata
- Primary use-case for MPEG proposal is to convey quality metadata to clients and facilitate delivery of video content through ABR algorithms
- Transcoding hasn’t been properly considered

What is missing

- More (newer) video quality metrics
  - VMAF
  - FB-MOS
- Multiple generations of full-reference metrics – cascade of transcoding steps
- Scaled (at different viewport resolutions) vs. non-scaled metrics
- Spatio-temporal aggregation methods
- Presence of video quality metadata in elementary video streams and system (container) formats

Our proposal – standard video quality metadata payload

- Video quality metric name (e.g. “SSIM”)
- Video quality metric version or model identifier (e.g. “v0.6.1”)
- Video quality raw score (e.g. “0.9256”)
- Video quality MOS score (e.g. “3.89”)
- 95% Confidence interval (e.g. “0.1” – this can be obtained from the statistical analysis of subjective data, as correlated with a given metric)
- Scaling method (e.g. “None”, for non-scaled or “Lanczos-5”)
- Temporal reference (e.g. “0-3”, when referring to the first 4 frames in a sequence)
- Aggregation method (e.g. “Arithmetic mean”)
- Generation index (e.g. “2”, if there were two prior encoding steps – perhaps an image sensor, and a first encoding)

Katsavounidis et al. “A case for embedding video quality metrics as metadata in compressed bitstreams”

When do we need no-reference video quality metrics?

- In the camera front-end, to estimate quality of raw input pixels
  - Although, camera metrics (aperture, ISO, speed) can help
- For legacy videos, i.e. those that don’t have video quality metadata
- For video broadcasting applications (transmission over noisy channels)
- For different (non-transcoding) image/video applications

Summary

- Full reference video quality metrics are readily available in most modern transcoding pipelines
- Including full-reference video quality metrics as metadata in compressed bitstreams takes very little space and provides a more accurate and “green” way of estimating source video quality
- Establishing a standard format to save such metadata at both elementary video bitstream level and system layer is crucial
- Both HW (device) makers and service providers have a lot to gain by offering such metadata in their compressed bitstreams