Monitoring video quality inside a network

Amy R. Reibman

AT&T Labs – Research
Florham Park, NJ

amy@research.att.com

Outline

- Measuring video quality (inside a network)
- Anatomy of packet loss impairments (PLI)
- Estimate MSE due to a PLI
- Predicting visibility of PLI
- Conclusions and challenges

Applications of video quality estimators

- Algorithm optimization
  - Automated in-the-loop assessment
- Product benchmarks
  - Vendor comparison to decide what product to buy
  - Product marketing to convince customer to give you $$
- System provisioning
  - Determine how many servers, how much bandwidth, etc.
- Content acquisition and delivery (and SLAs)
  - Enter into legal agreements with other parties
- Outage detection and troubleshooting

Measuring video quality inside the network

A video quality monitor for *inside the network* that is

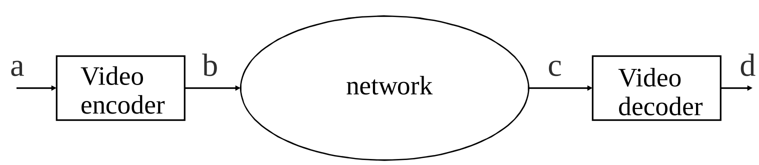
1. Real-time,
2. Per stream,
3. Scalable to many streams in network,
4. Measures only impact of network impairments,
5. Uses human perceptual properties, and
6. Accurate enough to answer the question:

   To what degree are *specific* network impairments affecting the quality of this *specific* video content?

Factors that affect video quality

- Video compression algorithm factors
  - Decoder concealment, packetization, GOP structure, …
- Network-specific factors
  - Delay, Delay variation, Bit-rate, Packet losses
- Network independent factors
  - Sequence
    - Content, amount of motion, amount of texture, spatial and temporal resolution
  - User
    - Eyesight, interest, experience, involvement, expectations
  - Environmental viewing conditions
    - Background and room lighting; display sensitivity, contrast, and characteristics; viewing distance

Where to measure?

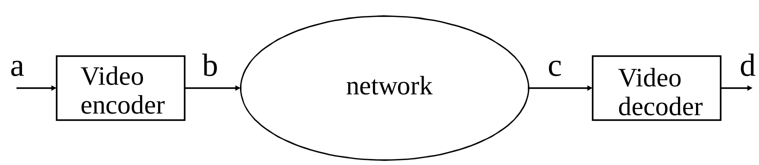


- In the network
  - If corporate network is managed by third party
  - Network operator does not have access to end-systems
  - For videos traversing multiple ISPs
- Between LAN and WAN, or at access/peering points between ISPs

What to measure?

- Not average network performance
  - Different ISPs,
  - Different bandwidth capacities,
  - Different time-varying loads
- Not only network-level measurements
  - Not all impairments produce same impact
  - Example: some packet losses are invisible, others are highly visible

What information can you gather?

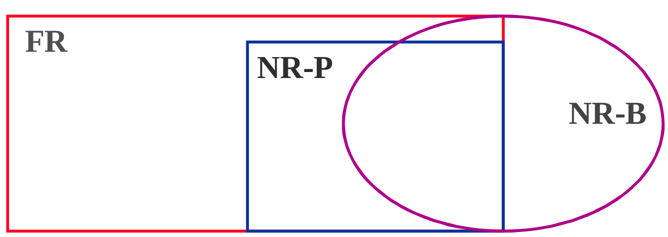


• Original video	X
• Encoding parameters	E(.)
• Complete encoded bitstream	E(X)
• Network impairments (losses, jitter)	L(.)
• Lossy bitstream	L(E(X))
• Decoder (concealment, buffer, jitter)	D(.)
• Decoded pixels	D(L(E(X)))

Constraints imposed by “inside the network”

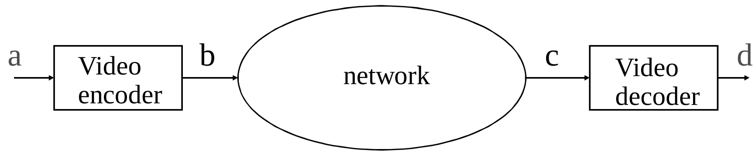
- Complexity, Scalability
  - If processing too complicated, can’t do for all streams
- Security, Proprietary algorithms
  - If encrypted content, can only process packet headers
- Structural constraints
  - Some data is unknowable (ex: environmental conditions)
  - Make reasonable assumptions about decoder (buffer handling, error concealment)
- Measurement point(s) location?
  - Miss impairments between measurement point and viewer
  - Not all measurements may be accurate

Categorizing image and video estimators



- Full and Reduced Reference (FR and RR)
  - Most available info; requires original and decoded pixels
- No-Reference Pixel-based methods (NR-P)
  - Requires decoded pixels: a decoder for each video stream
- No-Reference Bitstream-based methods (NR-B)
  - Processes packets containing bitstream, without decoder

Traditional FR video quality measurements



• Original video	X
• Encoding parameters	E(.)
• Complete encoded bitstream	E(X)
• Network impairments (losses, jitter)	L(.)
• Lossy bitstream	L(E(X))
• Decoder (concealment, buffer, jitter)	D(.)
• Decoded pixels	D(L(E(X)))

Why doesn't this solve our problem?

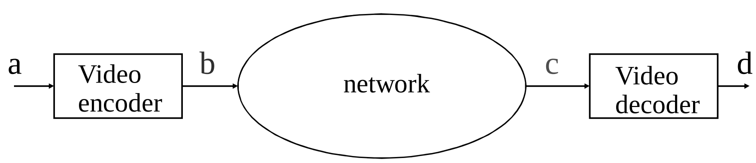
- Full-Reference: uses original and decoded video
  - Needs original video
  - Needs decoded video: a decoder for each stream in network
  - Cannot isolate impact of network impairments
  - Perceptual Full-Reference estimators are REALLY complicated!
  - Lots of parameter settings to get right

NR-Pixel methods for video quality

- No-Reference pixel QE: uses only decoded video
  - Still needs a decoder for each stream in network
  - Still cannot isolate impact of network impairments

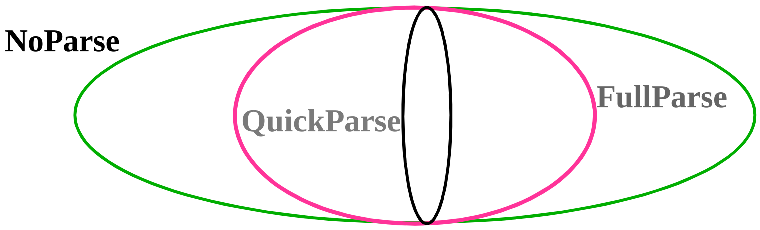
- Black-frame detection
- Video freezes
- Blockiness (Wu ‘97, Wang ‘00, …)
- Blurriness (Marziliano ‘02)
- Jerkiness (Pastrana-Vidal ‘05, Huynh-Thu ‘06)
- Ineffectiveness of error concealment (Yamada ‘07)
- Spatial Aliasing (Reibman ‘08)

No-reference Bitstream methods



• Original video	X
• Encoding parameters	E(.)
• Complete encoded bitstream	E(X)
• Network impairments (losses, jitter)	L(.)
• Lossy bitstream	L(E(X))
• Decoder (concealment, buffer, jitter)	D(.)
• Decoded pixels	D(L(E(X)))

NR-Bitstream methods for video quality

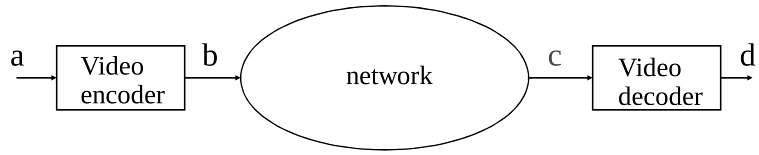


- FullParse – No *complete* decoding, but VLD
  - Mean, variance, spatial correlation, motion vectors
  - Location, extent, duration of losses
- QuickParse – “Easy-to-find” information only
  - Header information
  - Frame-level (or slice-level) summary information
- NoParse – Network-level stats only

ITU-T SG 12 standardization of QoS/QoE

- P.NAMS
  - Non-intrusive parametric model for quality assessment
  - Only packet-header information (IP through MPEG-2 TS)
  - Useful if payload is encrypted
  - Useful when processing capability is very limited
- P.NBAMS
  - Non-intrusive bitstream model for quality assessment
  - Allowed to use coded bitstream

Traditional network-based monitoring



• Original video	X
• Encoding parameters	E(.)
• Complete encoded bitstream	E(X)
• Network impairments (losses, jitter)	L(.)
• Lossy bitstream	L(E(X))
• Decoder (concealment, buffer, jitter)	D(.)
• Decoded pixels	D(L(E(X)))

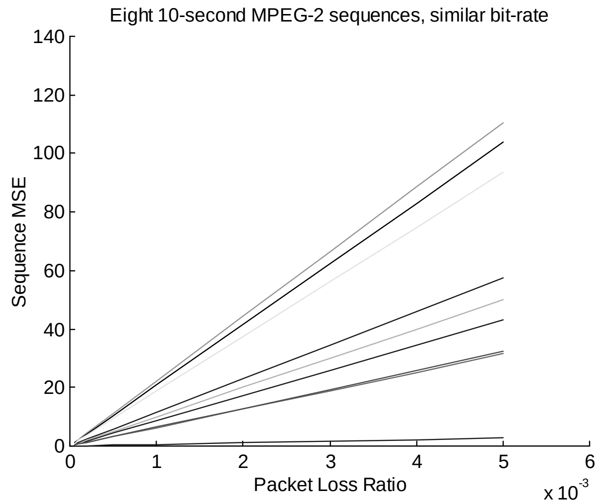
Why is PLR not enough?

- For MPEG-2, ***average*** MSE is linear with PLR

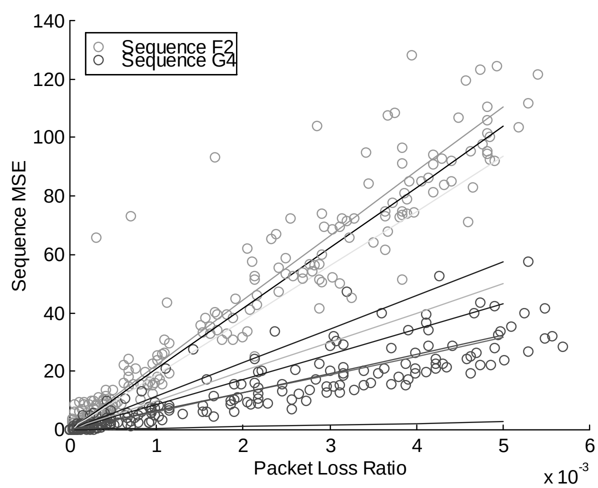
- What is the correct slope for a given bitstream?
- Depends on sequence-specific factors
  - Source content: motion, texture
- Depends on encoder-specific factors
  - Frequency of Intra information, bit-rate

- What is specific error for the given loss pattern?
- Depends on location of specific losses
  - Which frame type, individual spatial and temporal extent, scene change?

Influence of different content



Variation due to different losses



Quality assessment for networked video

- Compression effects
  - NR Estimation of MSE due to compression (Turaga '02, Ichigaya '04)
  - Motion-compensated edge artifacts (Leontaris '05)
- Packet loss effects
  - Estimate MSE (Reibman '02, Naccari '08)
  - Compute Mean Opinion Score (MOS) (Winkler 03, Liu 07, Lin 08)
  - Estimate visibility of individual packet losses (Kanumuri '04)
  - Estimate Mean Time Between Failures (Suresh '05)
- Timing effects (jitter)
  - Understand delivered video content in streaming scenario (Reibman '04, Gustafsson '08)

Estimating MSE due to packet loss

$$MSE = \frac{1}{N}\sum_n \sum_i (\hat{f}(n,i) - \tilde{f}(n,i))^2 = \frac{1}{N}\sum_n \sum_i e(n,i)^2$$

where $\hat{f}(n,i)$ is encoded value at pixel i frame n
and $\tilde{f}(n,i)$ is decoded value at pixel i frame n
and $e(n,i)$ is error for pixel i frame n

- What clues are in the bitstream to estimate MSE?
- Map unstructured problem into equivalent structured problem

Impact of network losses

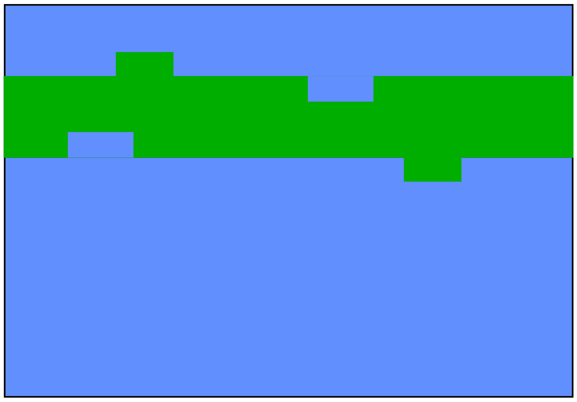
$\mathcal{M}_0$: set of macroblocks initially lost

Impact of network losses

$\mathcal{M}_0$: set of macroblocks initially lost

$e_0(n,i)$: initial magnitude of error

Impact of network losses

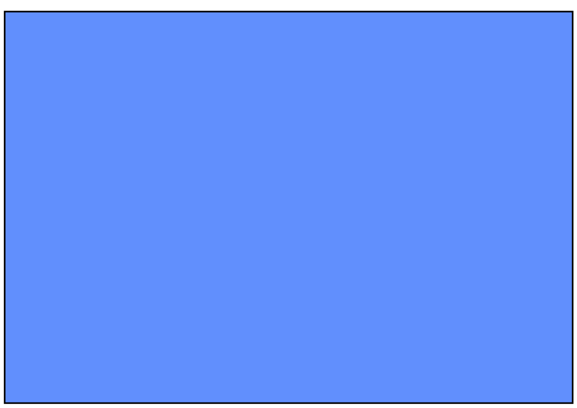


$\mathcal{M}_0$: set of macroblocks initially lost

$e_0(n,i)$: initial magnitude of error

ψ: prediction process (propagation of error; macroblock type and motion)

Impact of network losses



$\mathcal{M}_0$: set of macroblocks initially lost

$e_0(n,i)$: initial magnitude of error

ψ: prediction process (propagation of error; macroblock type and motion)

Characterization of the error

- Error is completely characterized by
  1. Which macroblocks are initially in error ($\mathcal{M}_0$)
  2. How large the initial error is in those macroblocks ($e_0(n,i)$)
  3. How the error propagates in space and time (ψ)

Characterization of the error

- Error is completely characterized by
  1. Which macroblocks are initially in error ($\mathcal{M}_0$)
     - Entire picture lost, 1 slice lost, 2 slices lost, …
  2. How large the initial error is in those macroblocks ($e_0(n,i)$)
     - Depends on source activity (still/moving)
     - Depends on encoder prediction
     - Depends on decoder’s concealment strategy
  3. How the error propagates in space and time (ψ)
     - Losses in B-frames only impact one frame
     - Received I-frame cleans out previous errors

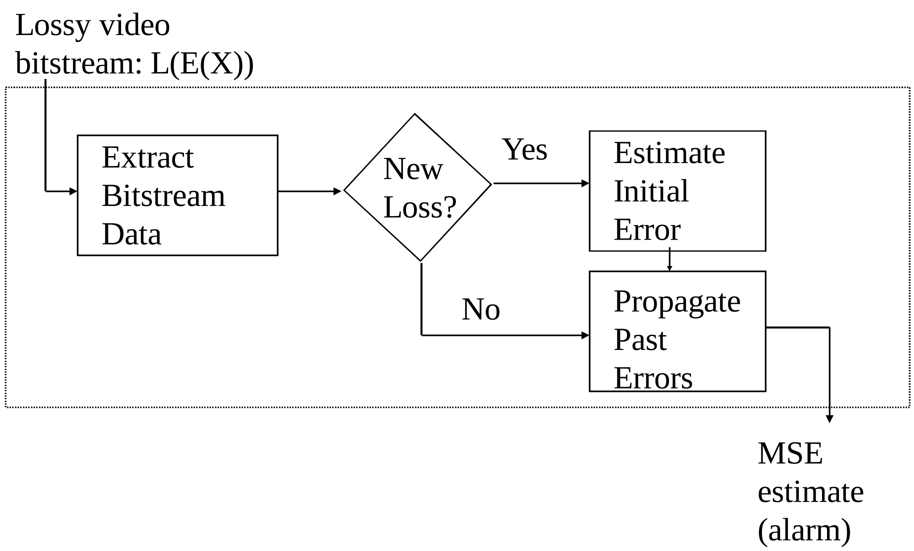
Characterization of the error: in the network

- Error is completely characterized by
  1. Which macroblocks are initially in error ($\mathcal{M}_0$)
     - Can be measured directly from lossy bitstream (NR-B)
     - Depends on compression, not on video content
  2. How large the initial error is in those macroblocks ($e_0(n,i)$)
     - Very hard to estimate accurately from lossy bitstream
     - Can be computed exactly given complete bitstream
  3. How the error propagates in space and time (ψ)
     - Characterized by motion vectors, macroblock types
     - Can be extracted exactly from the lossy bitstream (NR-B)

Calculating MSE due to packet loss

- Encoder-based estimation of MSE
  – Uncertainty of loss location, $\mathcal{M}_0$
  – Exact knowledge of propagation, ψ
  – Exact knowledge of initial error, $e_0(n,i)$
- Bitstream-based estimation of MSE
  – Exact knowledge of location of losses, $\mathcal{M}_0$
  – Exact knowledge of propagation, ψ
  – Unknown initial error, $e_0(n,i)$

Estimating MSE from lossy bitstream, L(E(X))



Extracting bitstream data

- How deeply can you process the packets?

QuickParse: Extracts slice-level information only

- Frame type, slice location, slice bit-rate, slice quantizer
- Knows which macroblocks; knows when errors stop
- Approximates spatial spread of the error propagation

FullParse: macroblock-level *-- no complete decoding!*

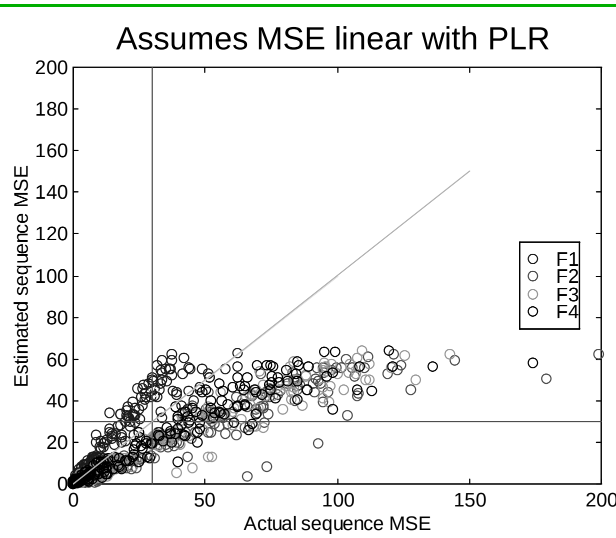
- Mean, variance, spatial correlation, motion vectors
- Knows exactly which macroblocks and how errors propagate

Performance comparison: data

- 225 sample packet loss traces
- 9 different PLR ranging from $5*10^{-5}$ to $5*10^{-3}$
- 25 sample traces per PLR

- 16 10-second MPEG-2 sequences
  - Wide range of sensitivity to packet loss
  - 8 sequences in training set; 8 sequences in test set

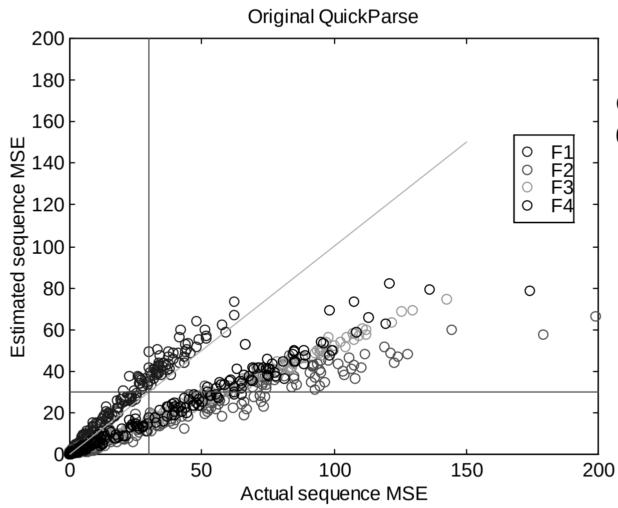
NoParse: Performance



Correlation 0.71

Probability of error: 9-14%

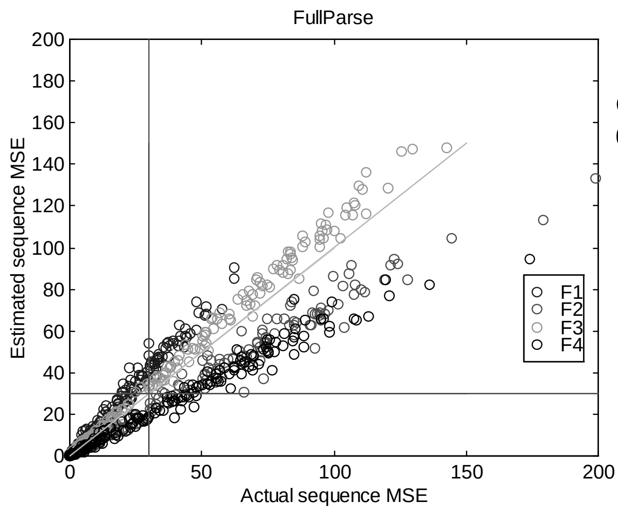
QuickParse: Performance



Correlation 0.79

Probability of error: 8-13%

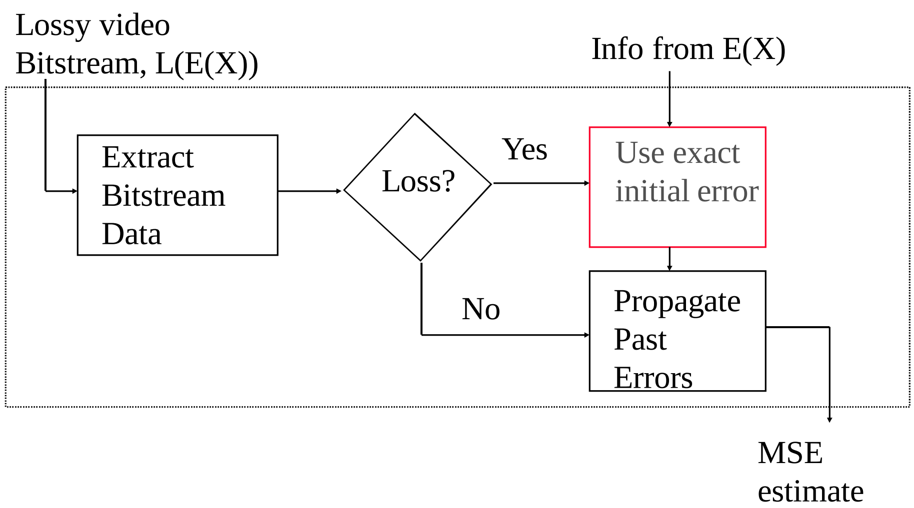
FullParse: Performance



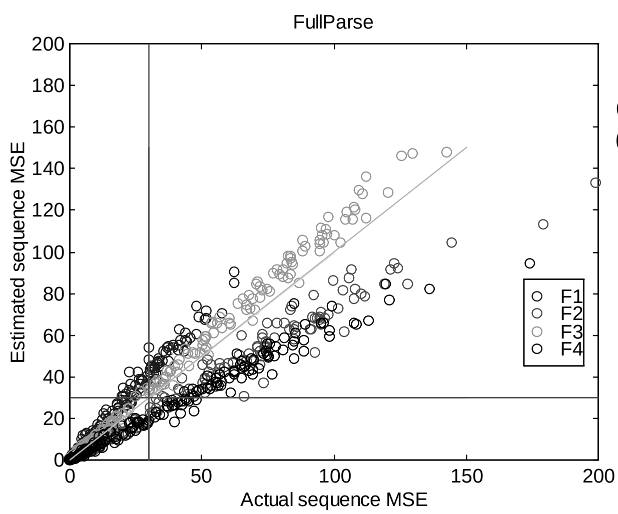
Correlation 0.95

Probability of error: 3-4%

Bounds: Estimating MSE from E(X)



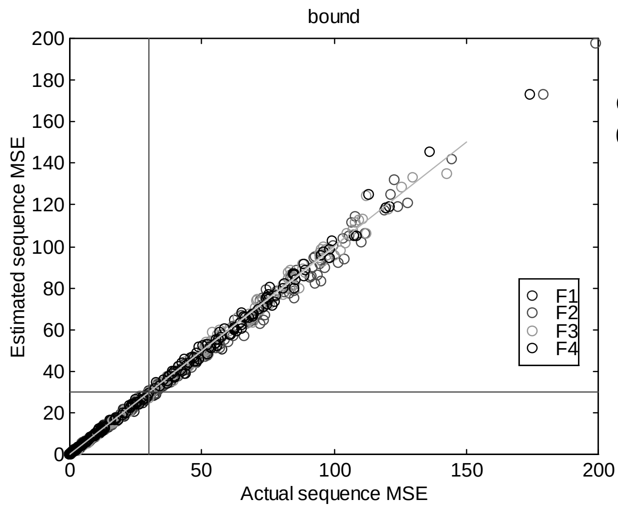
FullParse: Performance



Correlation 0.95

Probability of error: 3-4%

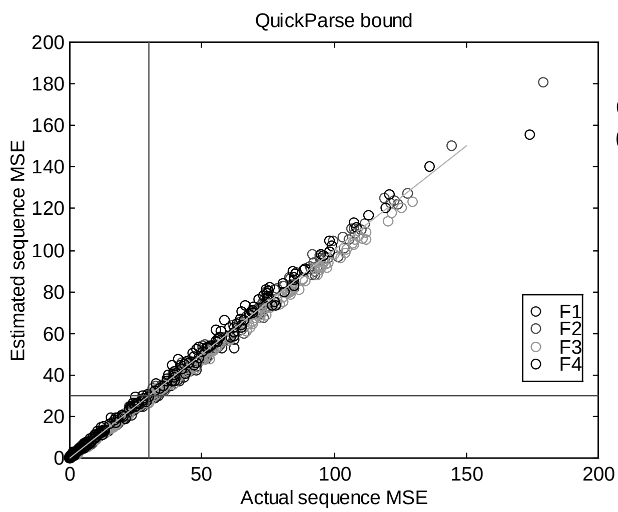
FullParse Bound: Performance



Correlation 0.998

Probability of error: 2%

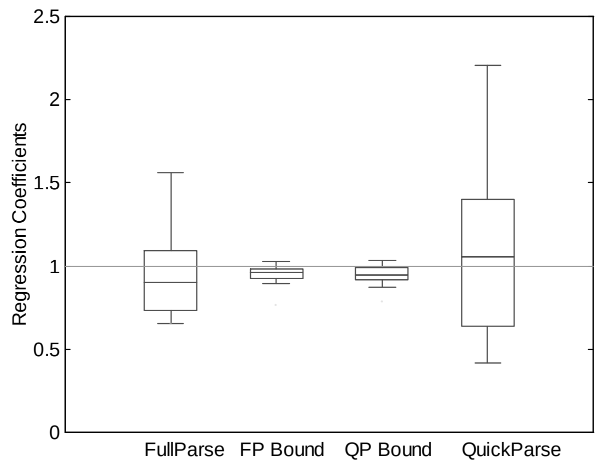
QuickParse bound: Performance



Correlation 0.995

Probability of error: 2%

Performance and bounds (16 seqs)



Observations: Broadcast MPEG-2 MSE

- QuickParse: Widely different slopes for different sequences
- FullParse: More accurate slopes, but room for improvement
- FullParse bound: Slopes consistently near one, but underestimated
- QuickParse bound: Nearly same as FullParse bound!

Inaccuracy of QuickParse is not due to simpler propagation, but to inaccurate estimate of initial error

- Reduce the complexity of FullParse
  - Estimate initial error with FP, propagate with QP

Outline

- Measuring video quality (inside a network)
- Anatomy of packet loss impairments (PLI)
- Estimate MSE due to a PLI
- Predicting visibility of PLI
  - NOT interested in quality given an average packet loss rate
  - Want to understand impact of each individual packet loss
- Conclusions and challenges

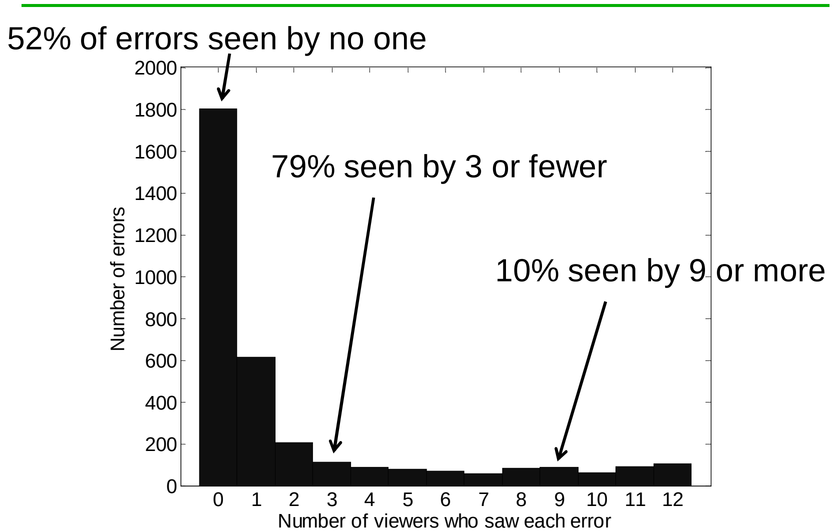
Visibility vs. quality

- Quality
  - How good is the video? How annoying are the artifacts?
  - Viewers provide MOS on a scale of 1—5
- Visibility
  - Did you see an artifact? What fraction of viewers saw artifact?
- Applications
  - High-quality video transport over a mostly reliable network
  - Design system so that less than some fraction of viewers will notice an impairment in the delivered video stream less than every (time period)?
  - Prioritization of packets to minimize visible impairments

Three Subjective DataSets

- Similar strategy (3455 isolated packet losses)
  - Measure each individual packet loss, NOT *average* quality
- Testing methodology
  - One packet loss every 4 seconds
  - Viewers are “immersed”, no audio, CRT display
  - “Press the space bar when you see an artifact”
  - 12 viewers for every PLI
- Wide range of parameters
  - Various compression standards (H.264, MPEG-2)
  - Different encoding parameters (Group of Picture, etc)
  - Different approaches for error concealment at decoder

Subjective test results: Ground truth



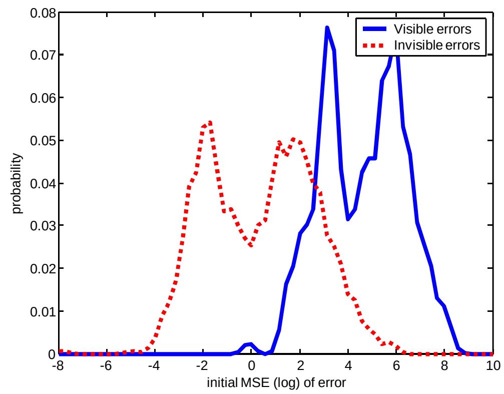
Visibility of packet loss impairments

- Depends on error itself
  - Size, spatial pattern, location, duration, amplitude
- Depends on decoded signal at location of error
  - New temporal edges (jerkiness), added horizontal edges, broken-up vertical edges
- Depends on encoded signal at location of error
  - Texture masking, luminance masking, motion masking may hide error
  - Motion tracking may enhance visibility in smoothly moving areas
  - This provides an implicit internal reference, even if not seen

Exploratory data analysis (EDA)

- Visibility as a function of one variable
  - Temporal duration: short one-frame errors are usually invisible
    - 1.5% of one-frame errors are seen by 75%+ of people
    - 63% of one-frame errors are seen by NO ONE!
  - Spatial extent: smaller errors more likely to be invisible
  - Motion: small motion losses typically invisible
  - Initial MSE: smaller errors more likely to be invisible
  - Scene motion: losses more likely to invisible with still camera

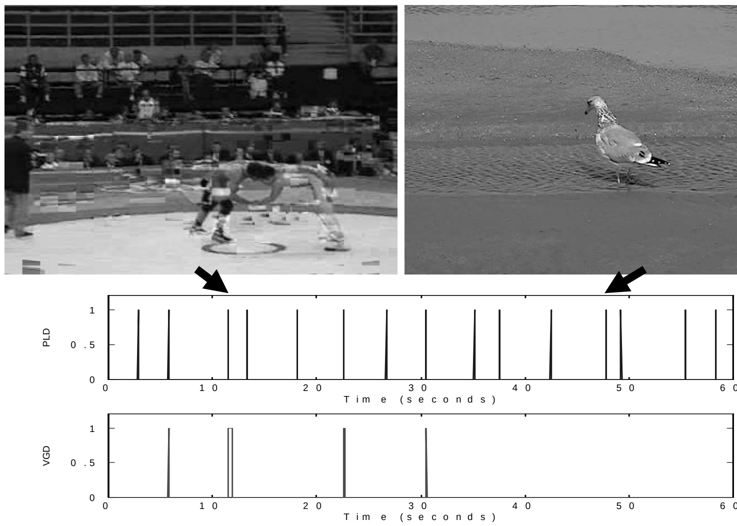
Initial MSE vs. visibility



Visual Glitch Detector for packet losses

- Always extract some information for all videos
  - Information about encoded signal
    - Local means and variances, motion, motion accuracy
  - Information about surrounding scene
    - Camera motion; Near a scene change?
- When there is a packet loss, extract:
  - Information about decoded signal
    - Extra edges possibly introduced
  - Information about error signal
    - Size, duration, initial MSE, initial SSIM
- Estimate visibility using logistic regression
  - Trained using subjective tests; Humans create “ground truth”

Visual Glitch Detector



Packet Losses Only

Visual Glitch Detector

Conclusions

- Many open problems in measuring video quality
- Characterizing impact of packet loss using $\mathcal{M}_0$, ψ, and $e_0(n,i)$ useful in many contexts related to video transport over networks
- Perceptual quality estimators can be very easy to implement
- Lots of room for improvement: No-Reference quality estimators that are effective
  - Across different image content and good enough for a legal contract

Thanks

- To all my immediate collaborators
- To E. Koutsofios for “lefty” graphics package
- To the community at large
- To all our subjective test participants
- To a patient audience

Collaborators

- Broadcast MPEG-2 with losses: MSE
  – Vinay Vaishampayan (AT&T)
  – Swamy Sermadevi (Cornell/Microsoft)
- Video streaming using Microsoft Media
  – Shubho Sen (AT&T)
  – Kobus van der Merwe (AT&T)
- Broadcast MPEG-2 with losses: Visibility
  – Sandeep Kanumuri (UCSD/DocomoUSA)
  – Vinay Vaishampayan (AT&T)
  – David Poole (AT&T)
  – Pamela Cosman (UCSD)

My journal papers on assessing quality

- A. R. Reibman, Y. Sermadevi and V. Vaishampayan, “Quality monitoring of video over a network", *IEEE Transactions on Multimedia*, vol. 6, no. 2, pp. 327-334, April 2004.
- S. Kanumuri, P. C. Cosman, A. R. Reibman, and V. Vaishampayan, “Modeling packet-loss visibility in MPEG-2 Video", *IEEE Transactions on Multimedia*, April 2006.
- A. Leontaris, P. C. Cosman, and A. R. Reibman, “Quality evaluation of motion-compensated edge artifacts in compressed video", *IEEE Transactions on Image Processing*, vol. 16, no. 4, pp. 943--956, April 2007.

My conference papers on assessing quality

- A. R. Reibman, Y. Sermadevi and V. Vaishampayan, “Quality monitoring of video over the Internet", *36th Asilomar Conference on Signals, Systems, and Computers*, vol. 2, pp. 1320-1324, Nov. 2002.
- A. R. Reibman and V. Vaishampayan, “Quality monitoring for compressed video subjected to packet loss", *IEEE International Conference on Multimedia and Expo (ICME'03)*, pp. I-17--20, vol. 1, July 2003.
- A. R. Reibman and V. Vaishampayan, ``Low-complexity quality monitoring of MPEG-2 video in a network", *IEEE International Conference on Image Processing (ICIP'03)*, pp. III-261--264, Sept. 2003.
- A. R. Reibman, S. Kanumuri, V. Vaishampayan, and P. C. Cosman, “Visibility of individual packet losses in MPEG-2 video", *IEEE International Conference on Image Processing (ICIP'04)*, pp. 171--174, Oct. 2004.
- S. Kanumuri, P. C. Cosman, and A. R. Reibman, “A generalized linear model for MPEG-2 packet-loss visibility", *Proc. International Workshop on Packet Video*, Dec. 2004.
- A. R. Reibman, S. Sen, and J. van der Merwe, “Network monitoring for video quality over IP", *Proc. Picture Coding Symposium*, Dec 2004.
- A. R. Reibman, S. Sen, and J. van der Merwe, “Analyzing the spatial quality of Internet streaming video", *First International Workshop on Video Processing and Quality Metrics*, Scottsdale, AZ, Jan. 2005. (http://enpub.fulton.asu.edu/resp/Upld/VPQM05/papers/226.pdf)
- A. R. Reibman, S. Sen, and J. van der Merwe, “Video quality estimation for Internet streaming", Fourteenth International World Wide Web Conference, Chiba Japan, May 2005.
- A. Leontaris and A. R. Reibman, “Comparison of blocking and blurring metrics for video compression", IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 585-588, March 2005.
- A. Leontaris, P. C. Cosman, and A. R. Reibman, “Measuring the added high frequency energy in compressed video", IEEE International Conference on Image Processing (ICIP'05), pp. 498-501, Sept. 2005.
- A. R. Reibman and T. Schaper, “Sub*jective performance evaluation for super-resolutio*n image enhancement", Second International Workshop on Video Processing and Quality Metrics, Scottsdale, Arizona, January 2006. *(http://enpub.fulton.asu.edu/resp/vpqm2006/papers06/303.pdf)*
- *A. R. Reib*man, R. Bell, and S. Gray, “Quality assessment for super-resolution image enhancement", IEEE International Conference on Image Processing, *October 2006.*
- *S. Kanumuri, S. G. Subramanian, P. C. Cosm*an, and A. R. Reibman, “Packet loss visibility in H.264 videos using a reduced reference method", IEEE International Conference on Ima*ge Processing, October 2006.*
- *A. R. Reibman and D. Poole, “Characteri*zing packet-loss impairments in compressed video", IEEE International Conference on Image Processing, Sept. 2007.
- A. R. Reibman and D. Poole, “Predicting packet-loss visibility using scene characteristic*s", Sixteenth International Packet Video W*orkshop, Nov. 2007.
- A. R. Reibman and S. Suthaharan, “A no-reference spatial aliasing measure for digital image resizing", IEEE International Conference on Ima*ge Processing, Oct. 2008.*
- *T.-L. Lin, P. C. Cosman*, and A. R. Reibman, “Perceptual impact of bursty versus isolated packet losses in H.264 compressed video", *IEEE International Conference on Image Processi*ng, Oct. 2008.