Quality versus Intelligibility: Evaluating the Coding Trade-Offs for American Sign Language Video

Frank Ciaramello, Jung Ko, and Sheila Hemami

Cornell University

CISS 2010
Princeton, NJ
Quality versus Intelligibility: Video Comparison

Encoded for Quality at 55 kbps

Encoded for Intelligibility at 55 kbps
Quality versus Intelligibility: Video Comparison

What do ASL users prefer?

“The extra clarity in the signer’s face makes it easier to understand.”

“The errors in the background are distracting and I prefer the video with the less clear signer but better background.”

Encoded for Quality at 55 kbps

Encoded for Intelligibility at 55 kbps
Outline

• Quality optimized video encoding
• Intelligibility optimized video encoding
• User controlled background quality
• Rate-distortion performance
• A quality-intelligibility convex hull
• Conclusions
H.264 Overview

- Motion/Intra Prediction
- Deblocking Filter
- Entropy Coder
- Coder
- T
- Q
- T⁻¹
- Q⁻¹
Rate and distortion for a macroblock $X$ depend on quantization step size ($Q$) and prediction mode ($p$).

$$\min_{p,Q} J(X, p, Q \mid \lambda) = D(X, p, Q) + \lambda R(X, p, Q)$$
A ‘Quality’ Optimized Coder

\[
\min_{p,Q} J(X,p,Q|\lambda) = D(X,p,Q) + \lambda R(X,p,Q)
\]

Distortion measured as \text{mean-squared-error}
An Intelligibility Optimized Coder

• Requires a distortion measure that quantifies ASL intelligibility.

• In ASL, information is communicated via facial expressions and hand shapes.

• Intelligibility is measured by the distortions in the linguistically relevant regions.
Computing Intelligibility

Intelligibility is measured by the distortions in linguistically relevant regions.

\[ D_{Intelli}(n) = \]
Computing Intelligibility

Intelligibility is measured by the distortions in linguistically relevant regions.

\[ D_{Intell}(n) = \alpha_F D_{FACE}(n) \]
Computing Intelligibility

Intelligibility is measured by the distortions in linguistically relevant regions.

\[ D_{Intell}(n) = \alpha_F D_{FACE}(n) + \alpha_H D_{HANDS}(n) \]
Computing Intelligibility

Intelligibility is measured by the distortions in linguistically relevant regions.

\[ D_{Intell}(n) = \alpha_F D_{FACE}(n) + \alpha_H D_{HANDS}(n) + \alpha_T D_{TORSO}(n) \]
Computing Intelligibility

Intelligibility is measured by the distortions in linguistically relevant regions.

\[ D_{Intell}(n) = \alpha_F D_{FACE}(n) + \alpha_H D_{HANDS}(n) + \alpha_T D_{TORSO}(n) + D_{Temporal}(n) \]

\[ D_{Temporal} \] quantifies time-varying distortions.
Computing Intelligibility

Intelligibility is measured by the distortions in linguistically relevant regions.

\[ D_{Intell}(n) = \alpha_F D_{FACE}(n) + \alpha_H D_{HANDS}(n) + \alpha_T D_{TORSO}(n) + D_{Temporal}(n) \]

\( \alpha \) controls the relative importance of each region.
An Intelligibility Optimized Coder

\[
\min_{p, Q} J(X, p, Q | \lambda) = D_{\text{Intell}}(X, p, Q) + \lambda R(X, p, Q)
\]
Quality versus Intelligibility
Coder Comparison

- Analyze the rate-distortion performance of each coder.
- Compute quality as:

\[ \text{PSNR} = 10 \log_{10} \frac{255^2}{\text{MSE}} \]

- Compute intelligibility as:

\[ \text{Intelligibility} = 10 \log_{10} \frac{C}{D_{\text{Intell}}} \]
Coder Comparison: Rate versus PSNR

- Quality Optimized Encoder
- Intelligibility Optimized Encoder
Coder Comparison: Rate versus Intelligibility

Intelligibility Optimized Encoder

Quality Optimized Encoder
Coder Comparison

Intelligibility Optimized Encoder

Quality Optimized Encoder

Intelligibility

PSNR (dB)

Bitrate (kbps)
Can we operate in here?
Outline

• Quality optimized video encoding
• Intelligibility optimized video encoding
• **User controlled background quality**
• Rate-distortion performance
• A quality-intelligibility convex hull
• Conclusions
User Controlled Background Quality

Allow user to increase background quality.

Background should never be less distorted than the signer.
User Controlled Background Quality

Implement encoding parameter, $\alpha_{\text{min}}$, that modifies region distortion weights.

$$D_{\text{Intell}} = \alpha_F D_{\text{FACE}} + \alpha_H D_{\text{HANDS}} + \alpha_T D_{\text{TORSO}}$$
User Controlled Background Quality

Implement encoding parameter, \( \alpha_{\text{min}} \), that modifies region distortion weights.

\( \alpha_{\text{min}} \) specifies the minimum allowed distortion weight.

\[
D_{\text{Intell}} = \alpha_F D_{\text{FACE}} + \alpha_H D_{\text{HANDS}} + \alpha_T D_{\text{TORSO}}
\]
User Controlled Background Quality

Implement encoding parameter, $\alpha_{\text{min}}$, that modifies region distortion weights.

$\alpha_{\text{min}}$ specifies the minimum allowed distortion weight.

$$D = \alpha_F D_{\text{FACE}} + \alpha_H D_{\text{HANDS}} + \alpha_{\text{min}} (D_{\text{TORSO}} + D_{\text{BG}})$$
Case 1: Intelligibility Optimized
Face > Hands > Torso > Background

Encoded at 55 kbps
Case 2: Nominal Background Weight
Face > Hands > Torso > Background

Encoded at 55 kbps

-Region Prioritized
-Region Not Prioritized

$\alpha_{\text{min}}$
Case 3: Face and Hands Prioritized

Face > Hands > Torso = Background

Encoded at 55 kbps

-Region Prioritized
-Region Not Prioritized

α_min

BG  Torso  Hands  Face
Case 4: Face Prioritized
Face > Hands = Torso = Background

Encoded at 55 kbps

- Region Prioritized
- Region Not Prioritized
Case 5: Quality Optimized Coder

Face = Hands = Torso = Background

Encoded at 55 kbps

-Region Prioritized
-Region Not Prioritized

$\alpha_{min}$

BG, Torso, Hands, Face
Outline

• Quality optimized video encoding
• Intelligibility optimized video encoding
• User controlled background quality
• Rate-distortion performance
• A quality-intelligibility convex hull
• Conclusions
Coder Comparison: Rate versus PSNR

PSNR (dB) vs. Bitrate (kbps)

- Quality Optimized Encoder
- Intelligibility Optimized Encoder
Coder Comparison: Rate versus PSNR

Quality Optimized Encoder

Intelligibility Optimized Encoder

Increasing $\alpha_{\text{min}}$
Coder Comparison: Rate versus Intelligibility

Intelligibility Optimized Encoder

Quality Optimized Encoder

Intelligibility

Very Easy

Easy

Neutral

Difficult

Very Difficult

Bitrate (kbps)
Coder Comparison:
Rate versus Intelligibility

Intelligibility Optimized Encoder

Quality Optimized Encoder

Increasing $\alpha_{\text{min}}$
Can we operate in here?
Quality-Intelligibility Trade-Offs

Where should we operate?
Outline

• Quality optimized video encoding
• Intelligibility optimized video encoding
• User controlled background quality
• Rate-distortion performance
• A quality-intelligibility convex hull
• Conclusions
Quality-Intelligibility Results

Intelligibility vs. PSNR (dB)

- Intelligibility Optimized Encoder
- Quality Optimized Encoder

Increasing Rate

Intelligibility
- Very Easy
- Easy
- Neutral
- Difficult

PSNR (dB)
Quality-Intelligibility Results

Intelligibility Optimized Encoder

Increasing Rate

Intelligibility

Very Easy

Easy

Neutral

Difficult

PSNR (dB)

16 18 20 22 24 26 28 30
Quality-Intelligibility Results

Increasing $\alpha_{\text{min}}$

Intelligibility Optimized Encoder

Quality Optimized Encoder

PSNR (dB)

Intelligibility

Very Easy
Easy
Neutral
Difficult
Nominal Background Weight Increases PSNR

PSNR (dB)

Intelligibility

Very Easy

Easy

Neutral

Difficult

PSNR Increase: 2dB-7dB

Nominal Background Weight Increases PSNR

PSNR (dB)

Intelligibility

Very Easy

Easy

Neutral

Difficult

PSNR Increase: 2dB-7dB

Nominal Background Weight Increases PSNR

PSNR (dB)

Intelligibility

Very Easy

Easy

Neutral

Difficult

PSNR Increase: 2dB-7dB
Prioritizing Face Increases Intelligibility

Intelligibility Increase: ~0.1
Summary

• Developed encoding parameter that controls priority given to multiple regions of interest.

• Parameter provides convex combination of intelligibility optimized and quality optimized encoders.

• Increased PSNR by up to 7dB with little decrease in intelligibility.