

ANALYSIS OF CODING TOOLS AND IMPROVEMENT OF TEXT READABILITY FOR SCREEN CONTENT

Holger Meuel

Julia Schmidt

Marco Munderloh

Jörn Ostermann

Coding of Text in Video Coding

- ▶ Neglection of high frequencies in state-of-the-art video coding
- ▶ Lots of high frequencies contained in text
- ▶ Therefore degradation of text

Transform based coding of text content (left: lossy, right: lossless)

Three Laws of Robotics

From Wikipedia, the free encyclopedia

The **Three Laws of Robotics** (often shortened to **The Three Laws** or **Three Laws**) are a set of rules devised although they were first introduced in a now-calling stories. The Three Laws are:

Properties of Screen Content (SC)

Application scenarios

- ▶ Office applications (e.g. text/spreadsheet processing)
- ▶ Text insertions into natural video (news tickers etc.)
- ▶ Streaming services, online gaming, video conferencing

Properties of letters and symbols

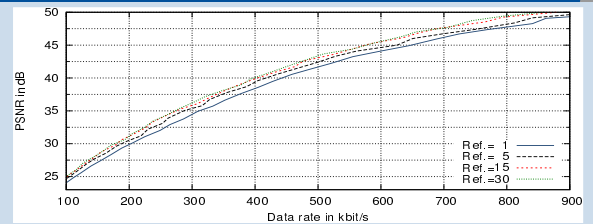
- ▶ Sharp edges introduced by letters and symbols
- ▶ Translational movement during scrolling and window movement

JCT-VC Test Sequence
Basketball Drive Text
(1920x1080)

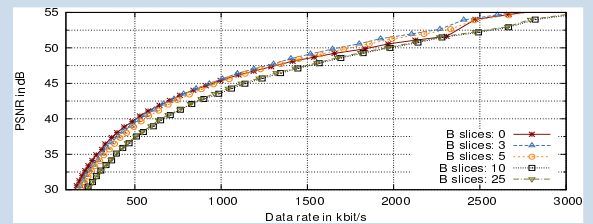


Existing Coding Tools in AVC & their Appropriateness for Screen Content Coding (SCC)

- ▶ Data rate ratios I/P & I/B:
 - ▶ for natural video: 20–1000 (I/P), 50–2000 (I/B)
 - ▶ for screen content: 500–10000 (I/P), 500–100000 (I/B)
 - ▶ Smaller movement in sequence → higher ratio
- ▶ Distance of Ref. Frames:
 - ▶ SC sequences typically contain very slow movements
 - ▶ Spreading reference slices over time as wide as possible is beneficial for high coding efficiency
- ▶ Hierarch. B slices:
 - ▶ Little difference between frames for slowly changing content
 - ▶ No additional information in hierarchical B slices
 - ▶ Recommendation to dismiss reference B Slices for SCC
- ▶ Number of B slices:
 - ▶ Larger temporal distance between reference slices
 - ▶ Increasing of residuals of P slices
 - ▶ Optimal results with 3 B Slices
 - ▶ Disable B slices completely for small movements
- ▶ Adaptive Quantisation Parameters (QPs):
 - ▶ QP changes are expensive
 - ▶ Fixed QP coding often better
- ▶ Resolution of Motion Vectors (MV):
 - ▶ Same as for natural video
- ▶ Spatial and Temporal Direct Mode (DM):
 - ▶ Use of Spatial/Temporal Direct Mode stays same in SC/camera captured sequences
 - ▶ 95–98% of DM coded blocks are better coded spatially



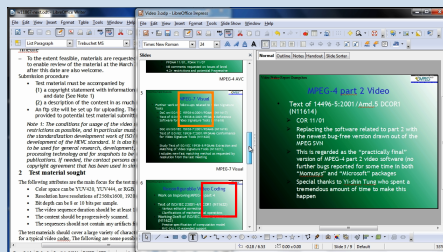
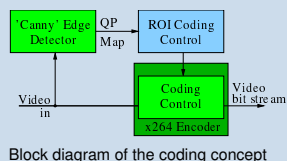
Coding performance for different number of reference slices (RD)



Coding performance for different numbers of B slices (RD diagram)

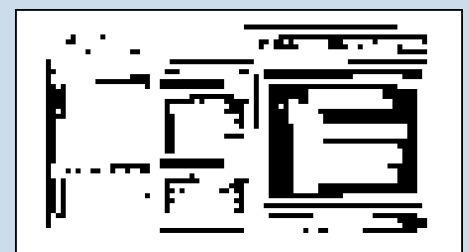
Improvement of Text Readability

- ▶ Requirements for text detection in SCC:
 - ▶ Runtime efficient separation of text/background
 - ▶ High detection rate → *Canny Edge Detector*
 - ▶ No adequate quality evaluation possible for SCC with PSNR (>45dB) → Subjective evaluation
- ▶ Advantage: Usage of standard coder with externally provided QP map
- ▶ Experimental results on the right



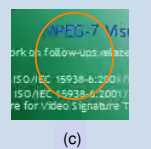
(a)

Frame 44 of the JCT-VC test sequence
Screen Capture Slide Event (1280x720)

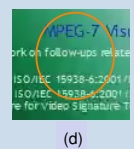


(b)

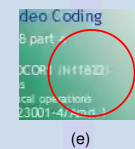
- (a) Original frame
- (b) Edges detected by Canny Edge Detector, white=text, black = coding with higher QP
- (c), (e) Frame after encoding with unmodified x264
- (d), (f) Frame after encoding/decoding using QP map



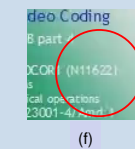
(c)



(d)



(e)



(f)