Analysis of Affine Motion Compensated Prediction and its Application in Aerial Video Coding

Holger Meuel

Institut für Informationsverarbeitung
Leibniz Universität Hannover

August 5^{th}, 2019

Video is everywhere!

Digital television: DVB
(T/T2, S/S2, C/C2)

Internet video

Video on demand

Aerial video

Mobile video

Video conferencing

Surveillance video

The Bit Rate Problem

- Data rate of one full HD sequence (1920 × 1080, 4:2:0): 622 Mbit/s
- More data rate needed for …
  - Higher resolutions, 4K, 8K, HDR, …
  - Multi-view video, e. g. 3D, 360°, …

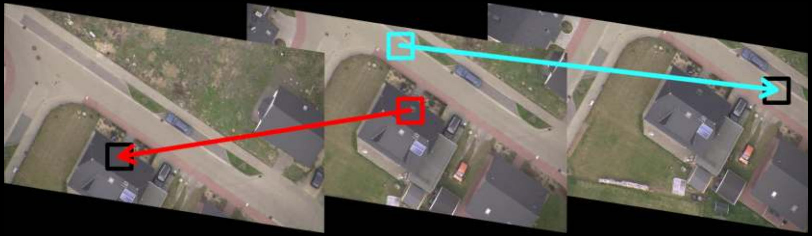
- For comparison:
  - Capacity of one Blu-ray (dual layer): ≈ 10 min video (HD)
  - Current broadband internet (DSL/ADSL/VDSL): 16–100 Mbit/s
  - Current mobile network (LTE Advanced): 500–1200 Mbit/s (shared!)

No economic storage and transmission of uncompressed video data!

Hybrid Video Coding for Data Compression

- Redundancy removal:
  1. **motion-compensated prediction**
  2. **entropy coding**
- Irrelevance removal: **transform & quantization**

Bit rates of compressed HD sequence ⇒ AVC: 10–12 Mbit/s / HEVC: 5–6 Mbit/s



Motion compensated prediction: blocks are predicted from similar image content

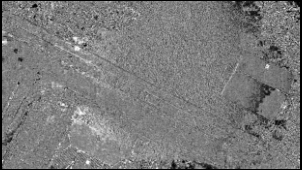
Motivation

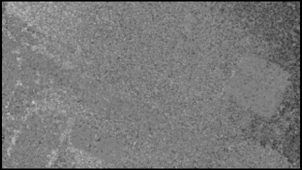
- Motion compensated prediction (MCP) as one key element in hybrid video coding
- High dependency between accuracy of motion estimation and prediction error
- Inaccurate motion estimation

  ⇒ High prediction error

  ⇒ High entropy

  ⇒ High bit rate





Aim of thesis:

Modeling of minimum required bit rate for encoding the prediction error as a function of the motion estimation accuracy using an **affine motion model**

Outline

Outline

Efficiency Analysis of Affine Motion Compensated Prediction

Efficiency Analysis of Simplified Affine Motion Compensated Prediction

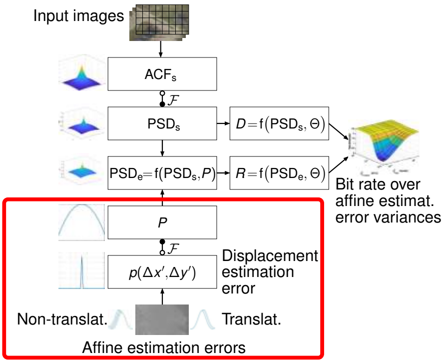
Experimental Results

Conclusion

Holger Meuel
meuel@tnt.uni-hannover.de

Overview: Bit Rate Derivation for Affine Estimation Errors

- Modeling of power spectral density (PSD) of signal
- Modeling of probability density function (pdf) $p(\Delta x', \Delta y')$ of displacement estimation error
- Derivation of $\mathrm{PSD_e}$ of displacement estimation error[1]
- Application of rate-distortion theory[2] $\Rightarrow$ bit rate



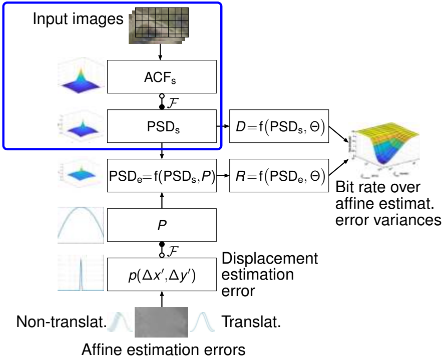
[1] Girod, "Efficiency of MoComp. Pred. for Hybrid Cod. of Video Seq.", Journ. on Sel. Areas in Comm., 1987

[2] Berger, "Rate Distortion Theory", Prentice-Hall, 1971

Holger Meuel
meuel@tnt.uni-hannover.de

Power Spectral Density (PSD) of the Signal

- Assumptions for video signal[3]
  - Isotropic autocorrelation function (ACF)
  - Exponentially decaying ACF
- Fitting of exponential parameter by measurements
  - JCT-VC test sequences
  - Aerial sequences
- Power spectral density PSD_s of video signal



[3] O'Neal et al., "Coding Isotropic Images", IEEE Transact. on Inform. Theory, 23(6):697–707, 1977

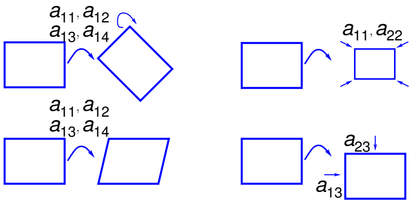
Affine Motion and Error Model

- Define affine motion model
- Derive errors from inaccurate affine motion estimation

Affine Motion Model

$$x' = a_{11} \cdot x + a_{12} \cdot y + a_{13}$$
$$y' = a_{21} \cdot x + a_{22} \cdot y + a_{23}$$

- $a_{11}, a_{12}, a_{21}, a_{22}$ non-translational parameters (rotation, scaling, shearing)
- a_{13} and a_{23} translational parameters



Holger Meuel
meuel@tnt.uni-hannover.de

Affine Motion Estimation

Perfect affine motion:

$$
\begin{aligned}
x' &= a_{11} \cdot x && + a_{12} \cdot y && + a_{13} \\
y' &= a_{21} \cdot x && + a_{22} \cdot y && + a_{23}
\end{aligned}
$$

- Inaccuracies introduced by affine motion parameter estimation (indicated by $\hat{\cdot}$)

$$
\begin{aligned}
\Delta x' = \hat{x}' - x' &= \underbrace{(\hat{a}_{11} - a_{11})}_{e_{11}} \cdot x && + \underbrace{(\hat{a}_{12} - a_{12})}_{e_{12}} \cdot y && + \underbrace{(\hat{a}_{13} - a_{13})}_{e_{13}} \\
\Delta y' = \hat{y}' - y' &= \underbrace{(\hat{a}_{21} - a_{21})}_{e_{21}} \cdot x && + \underbrace{(\hat{a}_{22} - a_{22})}_{e_{22}} \cdot y && + \underbrace{(\hat{a}_{23} - a_{23})}_{e_{23}}
\end{aligned}
$$

Holger Meuel
meuel@tnt.uni-hannover.de

Affine Error Model

Displacement estimation error:

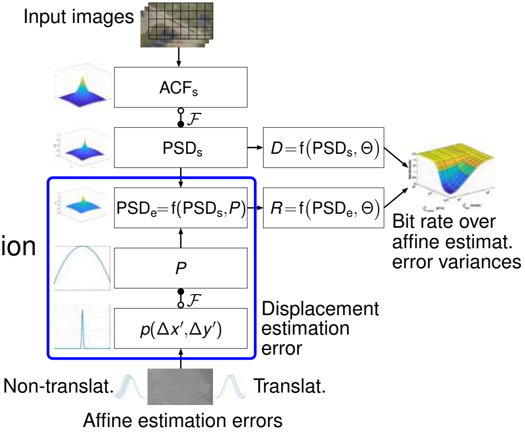
$$\begin{aligned}\Delta x' &= e_{11} \cdot x + e_{12} \cdot y + e_{13} \\ \Delta y' &= e_{21} \cdot x + e_{22} \cdot y + e_{23}\end{aligned}$$

Conclusions for displacement estimation errors

- $\Delta x'$, $\Delta y'$ can be described by affine model
- $\Delta x'$, $\Delta y'$ depend on location

Power Spectral Density of Error Signal

- Statistical modeling of displacement estimation error
- Calculation of power spectral density of prediction error



Probability Density Function Derivation (pdf)

- Assumption: errors are random variables which follow zero-mean Gaussian distributions

⇒ Joint pdf for statistically independent errors:

$$p(e_{11}, \ldots, e_{23}) = p(e_{11}) \cdot \ \ldots \ \cdot p(e_{23})$$

- But wanted: probability density function $p(\Delta x', \Delta y')$ of displacement estimation errors $\Delta x'$, $\Delta y'$

Pdf of Displacement Estimation Error

$$p(\Delta x', \Delta y') = \frac{1}{2\pi\sigma_{\Delta x'}\sigma_{\Delta y'}} \cdot \exp\left(-\frac{\Delta x'^2}{2\sigma^2_{\Delta x'}}\right) \cdot \exp\left(-\frac{\Delta y'^2}{2\sigma^2_{\Delta y'}}\right)$$

$$\text{with } \sigma^2_{\Delta x'} = \sigma^2_{e_{11}} x^2 + \sigma^2_{e_{12}} y^2 + \sigma^2_{e_{13}}$$

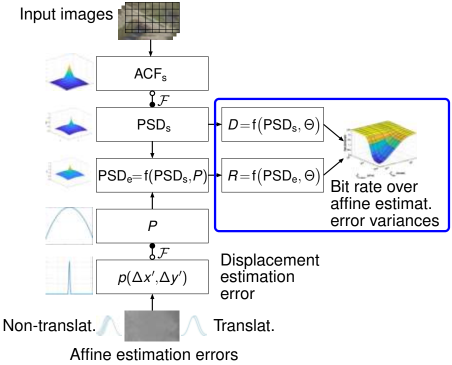
$$\text{and } \sigma^2_{\Delta y'} = \sigma^2_{e_{21}} x^2 + \sigma^2_{e_{22}} y^2 + \sigma^2_{e_{23}}$$

Conclusions

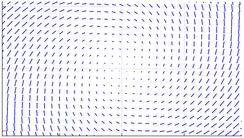
- Displacement estimation error pdf is a function of affinity estimation errors
- Pdf of the displacement estimation error is Gaussian distributed
- Variances $\sigma^2_{\Delta x'}$ and $\sigma^2_{\Delta y'}$ depend on location (x, y)

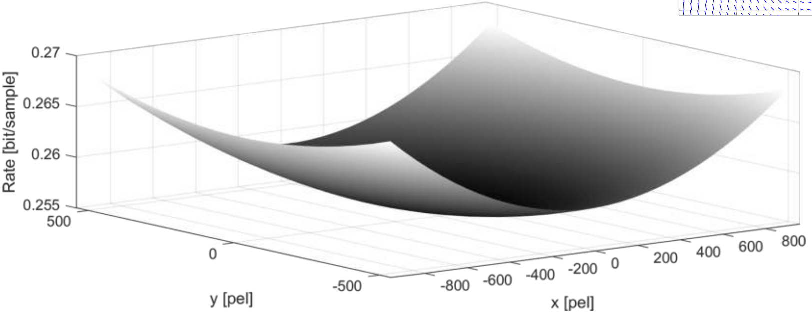
Rate-Distortion Theory

- Definition of target distortion *D* (30 dB SNR)
- Variation of generating function Θ to achieve predefined distortion *D*
- *One* Θ value corresponds to *one* distortion
- Calculation of rate *R* using same Θ value



Location-Dependent Bit Rate

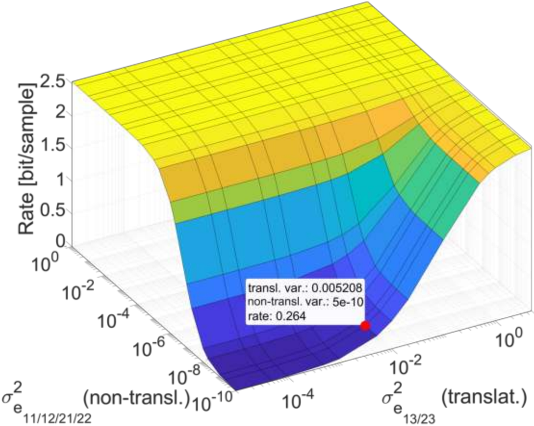




Equal location-dependent variances ($\sigma^2_{e_{11}} = \sigma^2_{e_{12}} = \sigma^2_{e_{21}} = \sigma^2_{e_{22}} = 5 \cdot 10^{-10}$) and equal location-independent, translational variances ($\sigma^2_{e_{13}} = \sigma^2_{e_{23}} = 0.0052$)

Holger Meuel
meuel@tnt.uni-hannover.de

Minimum Required Bit Rate for Prediction Error Coding



Left axis: location-dependent variances set equal;

right axis: location-independent, translational variances set equal

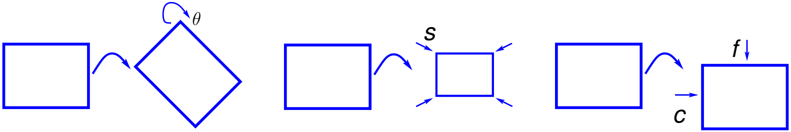
Outline

Motion Model

Physically motivated affine motion model:

$$x'_s = s \cdot \cos(\theta) \cdot x + s \cdot \sin(\theta) \cdot y + c$$
$$y'_s = -s \cdot \sin(\theta) \cdot x + s \cdot \cos(\theta) \cdot y + f$$

- c and f translational parameters
- Non-translational parameters
  - Rotation angle θ
  - Scaling factor s



Simplified Affine Motion Model

Model as used by JVET and in the JEM software[4]:

$$
\begin{aligned}
x'_{\mathrm{s}} &= (a+1)\cdot x && + b\cdot y && + c \\
y'_{\mathrm{s}} &= -b\cdot x && + (a+1)\cdot y && + f
\end{aligned}
$$

---

[4] Li et al.: “An Efficient Four-Parameter Affine Motion Model for Video Coding,” IEEE Transact. on Circuits and Systems for Video Technology, vol. PP, no. 99, pp. 1–1, 2017

Simplified Affine Motion Estimation

Estimated motion:

$$\hat{x}'_s = (\hat{a} + 1) \cdot x + \hat{b} \cdot y + \hat{c}$$
$$\hat{y}'_s = -\hat{b} \cdot x + (\hat{a} + 1) \cdot y + \hat{f}$$

- Inaccuracies introduced by simplified affine motion parameter estimation (indicated by $\hat{\cdot}$)

$$\Delta x'_s = \hat{x}'_s - x'_s = \underbrace{(\hat{a} - a)}_{e_a} \cdot x + \underbrace{(\hat{b} - b)}_{e_b} \cdot y + \underbrace{(\hat{c} - c)}_{e_c}$$
$$\Delta y'_s = \hat{y}'_s - y'_s = \underbrace{(-\hat{b} + b)}_{-e_b} \cdot x + \underbrace{(\hat{a} - a)}_{e_a} \cdot y + \underbrace{(\hat{f} - f)}_{e_f}$$

Holger Meuel
meuel@tnt.uni-hannover.de

Simplified Affine Error Model

Displacement estimation error:

$$\Delta x'_s = e_a \cdot x + e_b \cdot y + e_c$$
$$\Delta y'_s = -e_b \cdot x + e_a \cdot y + e_f$$

- Statistically independent error terms
- Statistical modeling of simplified affine estimation errors by their probability density functions (pdfs)

Conclusions for displacement estimation errors

- $\Delta x'_s$, $\Delta y'_s$ describable by simplified affine model
- $\Delta x'_s$, $\Delta y'_s$ depend on location again
- $\Delta x'_s$, $\Delta y'_s$ are interdependent

Pdf of the Displacement Estimation Error

$$p\left(\Delta x_s', \Delta y_s'\right) = \frac{1}{2\pi\sigma_{\Delta x_s'}\sigma_{\Delta y_s'}\sqrt{1-\rho^2}}$$

$$\cdot \exp\left(-\frac{1}{2\left(1-\rho^2\right)}\left[\frac{\Delta x_s'^2}{\sigma^2_{\Delta x_s'}} + \frac{\Delta y_s'^2}{\sigma^2_{\Delta y_s'}} - \frac{2\rho\cdot\Delta x_s'\cdot\Delta y_s'}{\sigma_{\Delta x_s'}\cdot\sigma_{\Delta y_s'}}\right]\right)$$

with
$$\sigma^2_{\Delta x_s'} = M\cdot\sqrt{\left(\sigma^2_{e_a}y^2+\sigma^2_{e_b}x^2+\sigma^2_{e_f}\right)\cdot\left(1-\rho^2\right)}$$

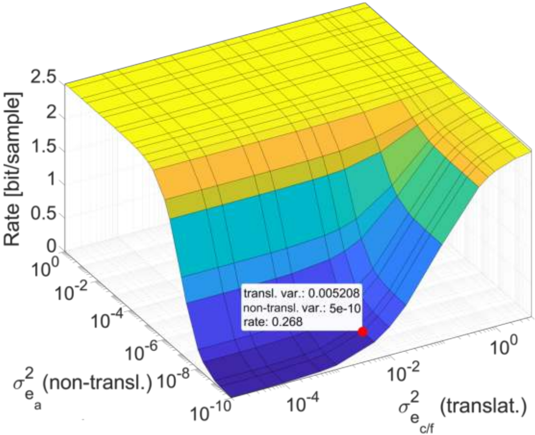
$$\sigma^2_{\Delta y_s'} = M\cdot\sqrt{\left(\sigma^2_{e_a}x^2+\sigma^2_{e_b}y^2+\sigma^2_{e_c}\right)\cdot\left(1-\rho^2\right)}$$

$$M = \left(\left(x^2+y^2\right)^2\sigma^2_{e_b} + y^2\sigma^2_{e_c} + x^2\sigma^2_{e_f}\right)\sigma^2_{e_a}$$

$$\rho = \frac{\left(\sigma^2_{e_a}xy - \sigma^2_{e_b}xy\right)}{\sqrt{\sigma^2_{e_a}y^2+\sigma^2_{e_b}x^2+\sigma^2_{e_f}}\cdot\sqrt{\sigma^2_{e_a}x^2+\sigma^2_{e_b}y^2+\sigma^2_{e_c}}}$$

- Bivariate zero-mean Gaussian distribution with correlation coefficient ρ
- Variances $\sigma^2_{\Delta x_s'}$, $\sigma^2_{\Delta y_s'}$ depend on locations x, y

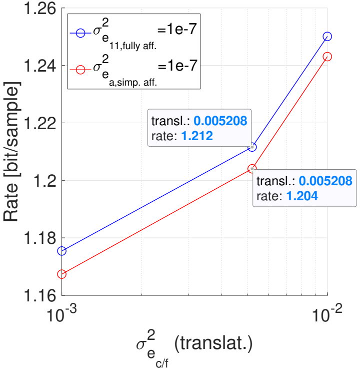
Minimum Required Bit Rate for Prediction Error Coding



Left axis: location-dependent variances in realistic ratio $\sigma^2_{e_b} = 2\sigma^2_{e_a}$;
right axis: location-independent, translational variances equal $\sigma^2_{e_c} = \sigma^2_{e_f}$

Comparison between Fully and Simplified Affine Model

- If motion in scene can be covered by both models, i. e. no shearing contained
  - Only 4 instead of 6 parameters for simplified model
  - ⇒ Smaller total estimation error
  - ⇒ Slightly smaller bit rates for practical applications



Location-dependent variances in ratio $\sigma^2_{e_b} = 2\sigma^2_{e_a}$

Holger Meuel
meuel@tnt.uni-hannover.de

Outline

Holger Meuel
meuel@tnt.uni-hannover.de

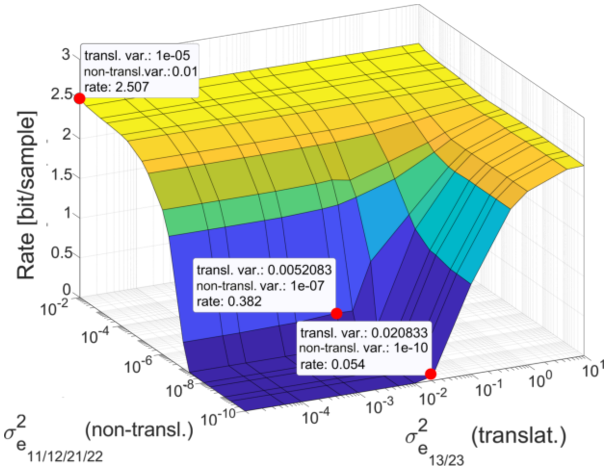
Outline

Experimental Setup

- Video signal with artificially introduced motion of specific variances
- Most-trivial motion estimation always predicting “no motion”

⇒ Introduced motion becomes exactly prediction error

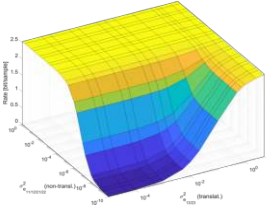
Measured Prediction Error Bit Rates



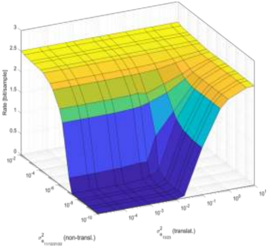
Holger Meuel
meuel@tnt.uni-hannover.de

Comparison between Model and Measurement

- Qualitatively perfect match between theory and measurement
- Slight overestimation of bit rates by model (2.53 vs. 2.507 bit/sample at maximum)
- Distinctive lower plateau in experimental data
- Measurements confirm supremum as predicted by the model



Theory (fully aff.)



Measurement

Holger Meuel
meuel@tnt.uni-hannover.de

Distinct Affine Test Sequences[5]

ShieldsPart, frame 1

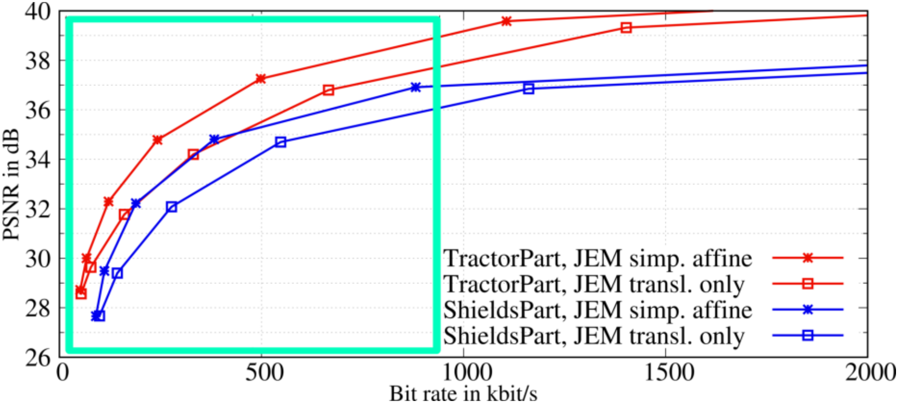
ShieldsPart, frame 100

TractorPart, frame 1

TractorPart, frame 100

[5] Li et al., “An Efficient Four-Parameter Affine Motion Model for Video Coding”

Operational Rate-Distortion Curves

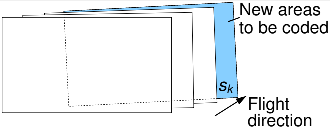


Outline

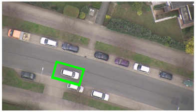
Holger Meuel
meuel@tnt.uni-hannover.de

Low Bit Rate Coding System for Aerial Video Sequences

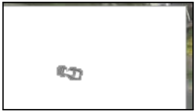
Idea: Exploit characteristics of aerial planar videos



I.) Detect region of interest (ROI): *new areas*



II.) Detect region of interest (ROI): *moving objects*



III.) Encode *only* areas containing *new areas* or *moving objects*

Reconstruct non-ROI areas by affine global motion compensation

Holger Meuel
meuel@tnt.uni-hannover.de

Test Sequences (TAVT)

350 m sequence

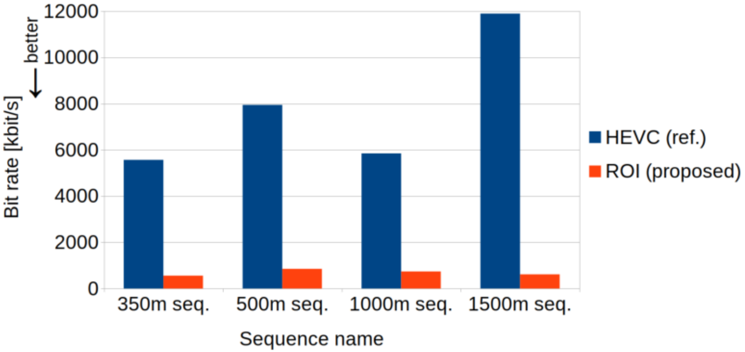
500 m sequence

1000 m sequence

1500 m sequence

Evaluation of the Aerial Coding System

- Common peak signal-to-noise ratio (PSNR) as quality criterion
- Evaluation in ROI areas only, PSNR $\approx 38\,\text{dB}$
- Sequences from TAVT



Holger Meuel
meuel@tnt.uni-hannover.de

Subjective Evaluation of Aerial Coding System

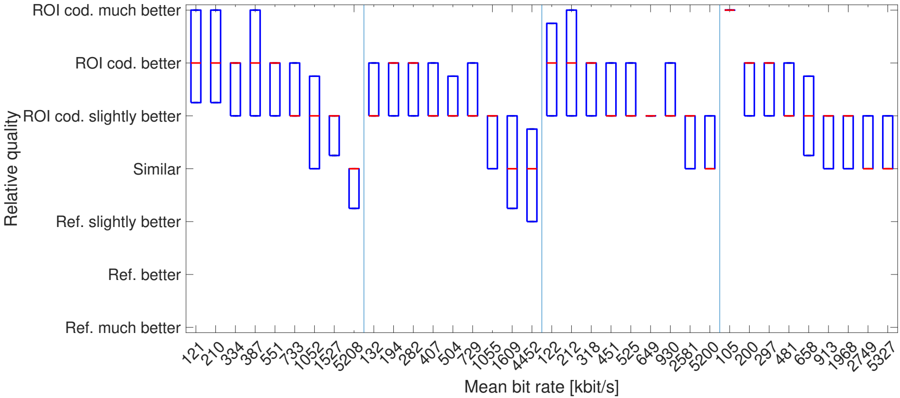
Original

HEVC 150 kbit/s

ROI HEVC 150 kbit/s

(Video comparison @150 kbit/s)

Subjective Tests of Aerial Video Coding System



Result from 27 test subjects judging *sharpness of details*;
test conditions according to ITU-T Rec. P.913

Outline

Efficiency Analysis of Affine Motion Compensated Prediction

Efficiency Analysis of Simplified Affine Motion Compensated Prediction

Experimental Results

Conclusion

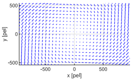
Conclusion

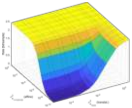
Analysis of affine motion compensated prediction

- **Modeling of displacement estimation error as function of affine motion estimation accuracy**
- Video signal characteristics & rate-distortion theory

⇒ Minimum required bit rate for prediction error coding

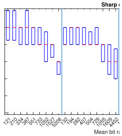
- **Mathematical modeling of bit rate estimation for simplified and fully affine motion compensated prediction in video coding**





Low bit rate aerial video coding system

- Exploiting affine global motion in aerial videos
- Encoding of *new areas* and *moving objects* only
- Superior quality compared to standardized video codecs



Thank you for your attention.