Application of the Rate-Distortion Theory for Affine Motion Compensated Prediction in Video Coding

Holger Meuel

Institut für Informationsverarbeitung
Leibniz Universität Hannover, Germany

June 19th, 2019

Motivation

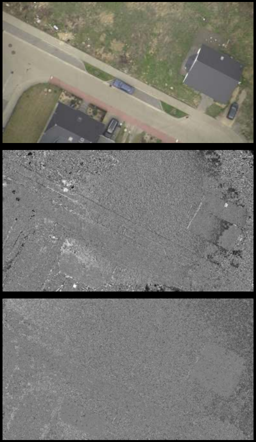
- Motion compensated (MC) prediction as one key element in hybrid video coding
- High dependency between accuracy of motion estimation (ME) and prediction error (PE)
- Inaccurate motion estimation

  $\Rightarrow$ High prediction error

  $\Rightarrow$ High entropy $\Rightarrow$ High bit rate

Goal:

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



Original aerial frame (top), "bad" MC/high PE (middle), "good" MC/small PE (bottom)

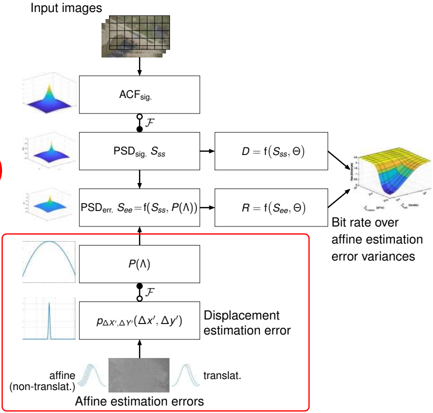
Outline

Outline

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) $p_{\Delta X', \Delta Y'}(\Delta x', \Delta y')$ of displacement estimation error
- Derivation of PSD of displacement estimation error $S_{ee}(\Lambda)$[1]
- Application of rate-distortion theory $\Rightarrow$ bit rate[2]



[1] Bernd Girod, "The Efficiency of Motion-Compensating Prediction for Hybrid Coding of Video Sequences," in IEEE Journal on Selected Areas in Communicat., vol. 5, no. 7, pp. 1140–1154, 1987

[2] Toby Berger, "Rate Distortion Theory: A Mathematical Basis for Data Compression", Prentice-Hall electrical eng. series, Prentice-Hall, 1971

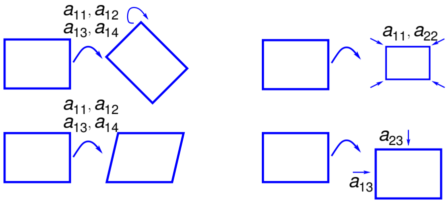
Outline

Motion Model

Affine motion model:

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

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



Holger Meuel
meuel@tnt.uni-hannover.de

Affine Motion Estimation

Estimated affine motion:

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

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

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

Affine Error Model

Displacement estimation error *in the frame*:

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

- Independent error terms e_{ij}, $i = \{1, 2\}$, $j = \{1, 2, 3\}$
- Statistical modeling of affine estimation errors by their probability density functions (pdfs)

Outline

Holger Meuel
meuel@tnt.uni-hannover.de

Probability Density Function Derivation

- Assumption: e_{ij} follow zero-mean Gaussian distributed pdfs

⇒ Joint pdf for independent e_{ij}:

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

- But wanted: probability density function $p_{\Delta X',\Delta Y'}(\Delta x', \Delta y')$ of displacement estimation errors $\Delta x'$, $\Delta y'$

Probability Density Function of the Displacement Estimation Error

With transformation theorem for pdfs:

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

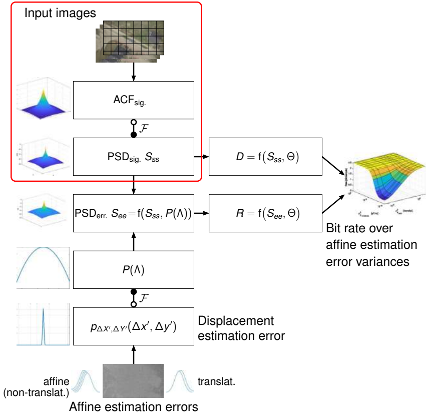
$$\text{with } \sigma^2_{\Delta x'} = \sigma^2_{e_{11}} x^2 + \sigma^2_{e_{12}} y^2 + \sigma^2_{e_{13}}$$
$$\text{and } \sigma^2_{\Delta y'} = \sigma^2_{e_{21}} x^2 + \sigma^2_{e_{22}} y^2 + \sigma^2_{e_{23}}$$

- Gaussian distributed pdf of the displacement estimation error
- Variances $\sigma^2_{\Delta x'}$ and $\sigma^2_{\Delta y'}$ depend on location x, y

Outline

Signal and Error Power Spectral Density Functions

- Model video signal
- Assumption of isotropic autocorrelation function
- Determination of power spectral density S_{ss} of video signal by Wiener–Khinchin theorem
- Calculation of power spectral density S_{ee} of displacement estimation error



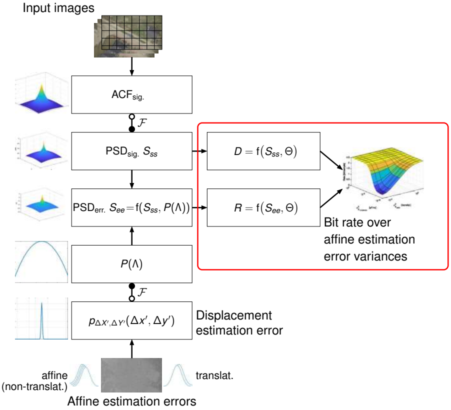
Holger Meuel
meuel@tnt.uni-hannover.de

Outline

Rate-Distortion Theory[3]

$$D = \frac{1}{4\pi^2} \iint_{\Lambda} \min\left[\Theta, S_{ss}(\Lambda)\right] \mathrm{d}\Lambda$$

$$R(D) = \frac{1}{8\pi^2} \iint_{\Lambda:\left(S_{ss}(\Lambda)>\Theta \text{ and } S_{ee}(\Lambda)>\Theta\right)} \log_2 \left[\frac{S_{ee}(\Lambda)}{\Theta}\right] \mathrm{d}\Lambda \text{ bit}$$

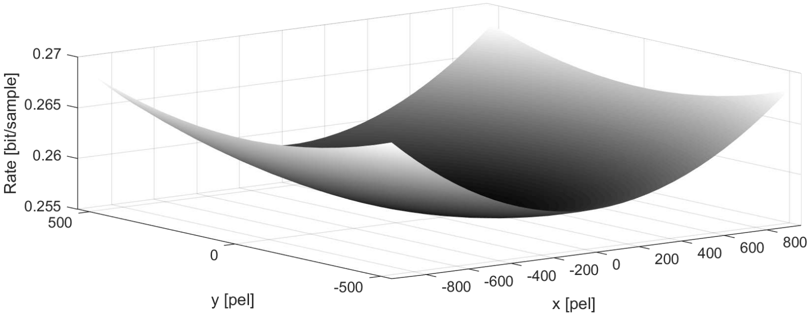


Θ: generating function varying distortion D and corresponding rate $R(D)$

[3] based on Toby Berger, "Rate Distortion Theory: A Mathematical Basis for Data Compression", Prentice-Hall electrical eng. series, Prentice-Hall, 1971

Outline

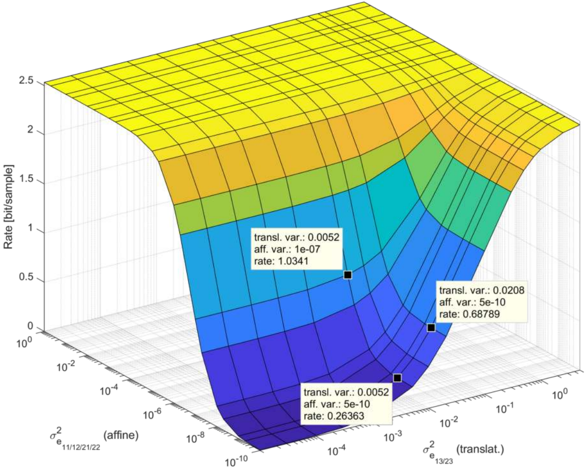
Location Dependent Bit Rate



Bit rate

Variances $\sigma^2_{e_{11}} = \sigma^2_{e_{12}} = \sigma^2_{e_{21}} = \sigma^2_{e_{22}} = 5 \cdot 10^{-10}$ and translational quarter-pel resolution ($\sigma^2_{e_{13}} = \sigma^2_{e_{23}} = 0.0052$), full HD resolution frame

Minimum Required Bit Rate for Prediction Error Coding



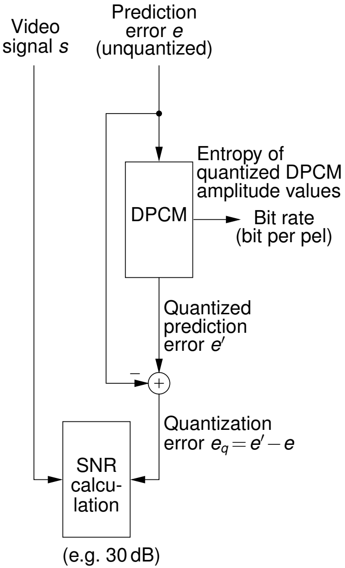
Distortion SNR $=$ 30 dB, $\sigma^2_{e_{11}} = \sigma^2_{e_{12}} = \sigma^2_{e_{21}} = \sigma^2_{e_{22}}$ and $\sigma^2_{e_{13}} = \sigma^2_{e_{23}}$, full HD resolution, isolines for translational quarter- (0.0052) and half-pel resolution marked

Outline

Experimental Setup

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

⇒ Introduced motion becomes exactly prediction error *e*

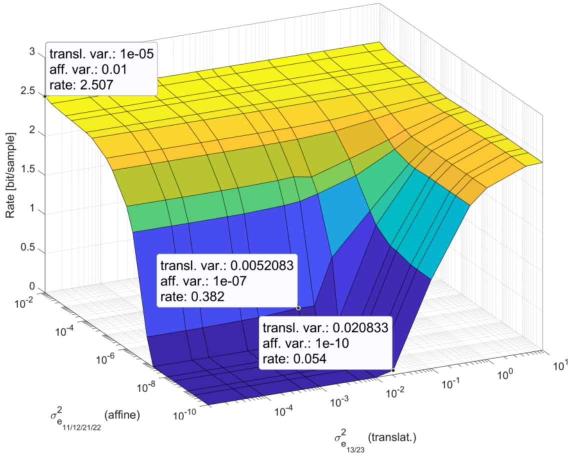


Experimental accomplishment:

Data rates of 30 randomly drawn, different motions for each combination of purely affine and translational variances averaged

Holger Meuel
meuel@tnt.uni-hannover.de

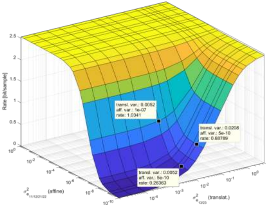
Measured Bit Rates for Encoding the Prediction Error



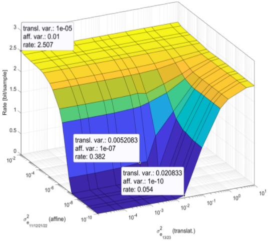
Measured bit rate for encoding the prediction error as a function of the motion estimation error variances, full HD resolution frame

Comparison between Theory and Experimental Data

- Qualitatively perfect match between theory and measurement
- Slight overestimation of bit rates by model (2.53 instead of 2.507 bit/sample at maximum)
- More pronounced lower plateau in experimental data due to interpolation filter



Theory



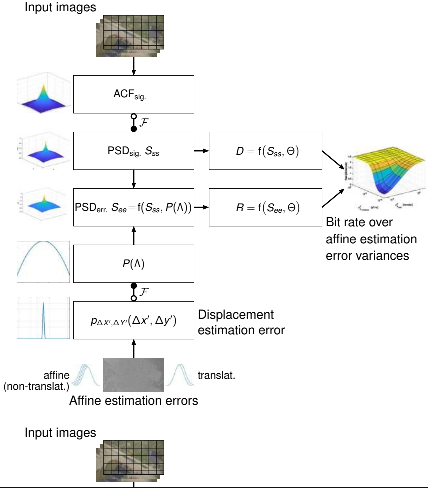
Measurement

Holger Meuel
meuel@tnt.uni-hannover.de

Real-World Application of the Model?

Consideration of simplified affine model as used in upcoming VVC

- Similar procedure, but:
- More complicated pdf of displacement estimation error
- JEM block size of 128×128



Distinct Affine Test Sequences[4]

ShieldsPart, frame 1

ShieldsPart, frame 100

TractorPart, frame 1

TractorPart, frame 100

[4] L. Li et al., "An Efficient Four-Parameter Affine Motion Model for Video Coding", IEEE Transact. on Circuits and Syst. for Video Tech., PP(99):1–1, 2017

Model vs. Real-World Measurements

- Block size: 128×128 pel as in JEM
- Translational quarter-pel, non-translational 1/16 pel accuracy

Sequence name	Model w/o signaling [bit/sample]	Model w/ signaling[5] [bit/sample]	Measured [bit/sample]	Remarks
ShieldsPart	0.398	0.5	0.71	Model approximates minimum bit rate 
TractorPart	0.058	0.07	0.012	Isotropic assumption violation, low-contrast signal, high amount of blur 

Conclusion:

Model provides valuable indications of the prediction error bit rate as function of affine motion estimation accuracy

[5] Sven Klomp, „Decoderseitige Bewegungsschätzung in der Videocodierung“, Fortschritt-Berichte VDI: Reihe 10, Informatik/Kommunik., 2012, ISBN 978-3-18-382010-8

Outline

Application of RD Theory for Affine MCP in Video Coding

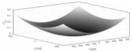
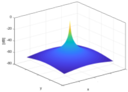
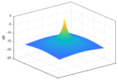
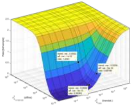
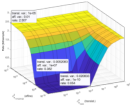
Model for affine motion compensation in video coding:

- Modeling of pdf of displacement estimation error $p_{\Delta X', \Delta Y'}(\Delta x', \Delta y')$
- Consideration of power spectral density of video signal
- Derivation of power spectral density of displacement estimation error
- Application of rate-distortion function

$\Rightarrow$ **Minimum bit rate for coding the prediction error**

Experimental verification:

- Confirmation of theoretical findings
- Application to simplified affine motion compensated prediction as employed in upcoming VVC

Holger Meuel
meuel@tnt.uni-hannover.de