Leibniz Universität Hannover

Rate-Distortion Theory for Affine (Global) Motion Compensation in Video Coding

Holger Meuel

Institut für Informationsverarbeitung
Leibniz Universität Hannover, Germany

Presentation of the ICIP 2018 paper at the SVCP, July 4th–6th, 2018

Motivation

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

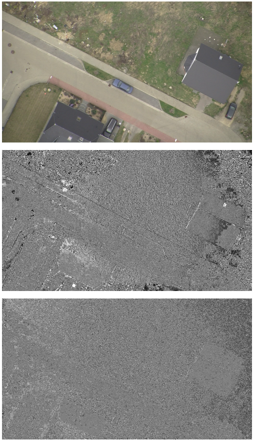
  $\Rightarrow$ High prediction error

  $\Rightarrow$ High entropy

  $\Rightarrow$ High bit rate

Goal:

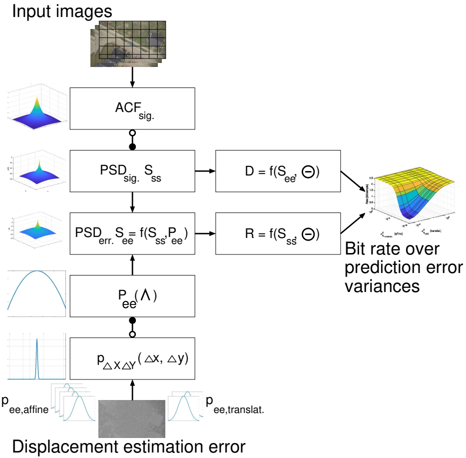
Model the prediction error bit rate as a function of the displacement estimation error for an **affine motion model**



Original aerial frame (top)

Overview[2]

- Model of power spectral density (P.S.D.) of signal
- Model of probability density function (p.d.f.) of displacement estimation error
- Derivation of P.S.D. of displacement estimation error $S_{ee}(\Lambda)$
- Application of rate-distortion theory ⇒ bit rate



[2] Modeling based on B. 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

Outline

Holger Meuel
meuel@tnt.uni-hannover.de

Outline

Affine Motion Model and Error Model

$$x = a_{11} \cdot x' + a_{12} \cdot y' + a_{13} \tag{1}$$

$$y = a_{21} \cdot x' + a_{22} \cdot y' + a_{23} \tag{2}$$

- a_{13} and a_{23} translational parameters
- $a_{11,12,21,22}$ "purely affine" parameters (rotation, scaling and shearing)
- Perturbation (indicated by $\hat{\cdot}$) by error terms e_{ij}, $i = \{1,2\}$, $j = \{1,2,3\}$ caused by inaccurate estimation

$$\Delta x = \hat{x} - x = \underbrace{(\hat{a}_{11} - a_{11})}_{e_{11}} \cdot x' + \underbrace{(\hat{a}_{12} - a_{12})}_{e_{12}} \cdot y' + \underbrace{(\hat{a}_{13} - a_{13})}_{e_{13}}$$

$$\Delta x = e_{11} \cdot x' + e_{12} \cdot y' + e_{13} \tag{3}$$

$$\Delta y = e_{21} \cdot x' + e_{22} \cdot y' + e_{23} \tag{4}$$

Holger Meuel
meuel@tnt.uni-hannover.de

Probability Density Function (p.d.f.) of the Displacement Estimation Error

Assumption: e_{ij} zero-mean Gaussian distributed with p.d.f.:

$$p(e_{ij}) = \frac{1}{\sqrt{2\pi\sigma_{e_{ij}}^2}} \cdot \exp\left(-\frac{e_{ij}^2}{2\sigma_{e_{ij}}^2}\right) \tag{5}$$

with $i = \{1,2\}$ and $j = \{1,2,3\}$

Joint p.d.f. for independent e_{ij}:

$$p_{E_{11},\ldots,E_{23}}(e_{11},\ldots,e_{23}) = p(e_{11}) \cdot \ \ldots \ \cdot p(e_{23}) \tag{6}$$

Holger Meuel
meuel@tnt.uni-hannover.de

Probability Density Functions Conversion

- Given now: joint p.d.f. $p_{E_{11},\ldots,E_{23}}(e_{11},\ldots,e_{23})$
- Wanted: p.d.f. $p_{\Delta X,\Delta Y}(\Delta x,\Delta y)$ of displacement estimation errors Δx, Δy

With transformation theorem for p.d.f.s:

$$p_{\Delta X,\Delta Y}(\Delta x,\Delta y)=\int_{\mathbb{R}^6} p_{E_{11},\ldots,E_{23}}(e_{11},\ldots,e_{23}) \cdot\delta\big(\Delta x-(x'e_{11}+y'e_{12}+e_{13})\big) \cdot\delta\big(\Delta y-(x'e_{21}+y'e_{22}+e_{23})\big)\,\mathrm{d}e_{11}\ldots\mathrm{d}e_{23} \qquad (7)$$

Holger Meuel
meuel@tnt.uni-hannover.de

Probability Density Function of the Displacement Estimation Error

$$p_{\Delta X,\Delta Y}(\Delta x, \Delta y) = \frac{1}{2\pi\sigma_{\Delta x}\sigma_{\Delta y}} \cdot \exp\left(-\frac{\Delta x^2}{2\sigma^2_{\Delta x}}\right) \cdot \exp\left(-\frac{\Delta y^2}{2\sigma^2_{\Delta y}}\right) \quad (8)$$

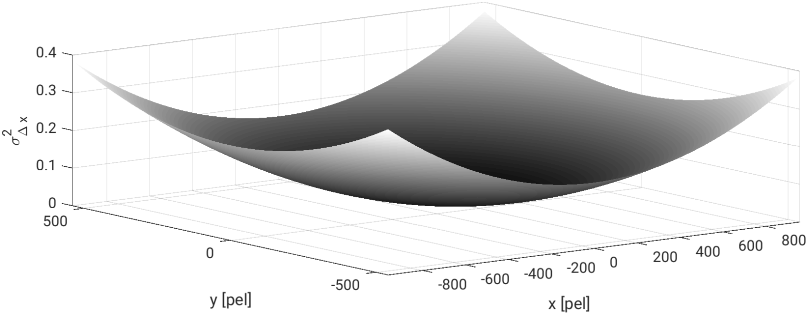
$$\text{with } \sigma^2_{\Delta x} = \sigma^2_{e_{11}} x'^2 + \sigma^2_{e_{12}} y'^2 + \sigma^2_{e_{13}} \quad (9)$$

$$\text{and } \sigma^2_{\Delta y} = \sigma^2_{e_{21}} x'^2 + \sigma^2_{e_{22}} y'^2 + \sigma^2_{e_{23}} \quad (10)$$

Variances $\sigma^2_{\Delta x}$ and $\sigma^2_{\Delta y}$ depend on locations x', y'!

Holger Meuel
meuel@tnt.uni-hannover.de

Location Dependent Variance $\sigma^2_{\Delta x}$ of Gaussian Distributed Displacement Estimation Error p.d.f.s for a Full HD Image



$\sigma^2_{e_{11}} = 2.3\text{e}-7$, $\sigma^2_{e_{12}} = 4.6\text{e}-7$ (measured from TAVT[3]) and $\sigma^2_{e_{13}} = 0.04$ (like Girod[2])

[3] *TNT Aerial Video Testset (TAVT)*, Institut für Informationsverarbeitung (TNT), Leibniz Universität Hannover, 2014, online: `https://www.tnt.uni-hannover.de/project/TAVT/`

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

Outline

Holger Meuel
meuel@tnt.uni-hannover.de

Signal and Error Power Spectral Density Functions

- Assumption of isotropic autocorrelation function[4]:

$$R_{ss}(\Delta x, \Delta y) = E\left[s(x,y) \cdot s(x - \Delta x,\ y - \Delta y)\right]$$
$$:= \exp\left(-\alpha\sqrt{\Delta x^2 + \Delta y^2}\right) \qquad (11)$$

- Determination of power spectral density of the video signal (Wiener–Khinchin theorem):

$$S_{ss}(\Lambda) = \mathcal{F}\left(R_{SS}(\Delta x, \Delta y)\right) \qquad (12)$$

- Power spectral density of displacement estimation error[2]:

$$S_{ee}(\Lambda) = 2\, S_{ss}(\Lambda)\left[1 - \mathrm{Re}(P(\Lambda)\right] + \Theta \qquad (13)$$

---

[2] B. 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, Aug 1987

[4] J. O’Neal and T. Natarajan, “Coding Isotropic Images”, IEEE Transact. on Inform. Theory, vol. 23, no. 6, pp. 697–707, Nov. 1977

Outline

Holger Meuel
meuel@tnt.uni-hannover.de

Rate-Distortion Theory[5]

$$D = \frac{1}{4\pi^2} \iint\limits_{\Lambda} \min \left[\Theta, S_{ss}(\Lambda)\right] \mathrm{d}\Lambda \,, \tag{14}$$

$$R(D) = \frac{1}{8\pi^2} \iint\limits_{\substack{\Lambda:\left(S_{ss}(\Lambda)>\Theta \right.\\ \left.\text{and } S_{ee}(\Lambda)>\Theta\right)}} \log_2 \left[\frac{S_{ee}(\Lambda)}{\Theta}\right] \mathrm{d}\Lambda \text{ bit} \tag{15}$$

with Θ being a parameter that generates the function $R(D)$ by taking on all positive real values

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

Outline

Holger Meuel
meuel@tnt.uni-hannover.de

Pixel Correlations

Sequence	Corr. ρ_x	Corr. ρ_y
Values from Girod	0.928	0.934
BasketballDrive* (HD)	0.9782	0.9488
BQTerrace* (HD)	0.9680	0.9659
Cactus* (HD)	0.9741	0.9812
Kimono* (HD)	0.9883	0.9900
ParkScene* (HD)	0.9634	0.9518
Mean of HD sequences*	**0.9744**	**0.9677**

Measured horizontal and vertical correlations between adjacent pixels for typical test sequences (*: 100 frames each).

Holger Meuel
meuel@tnt.uni-hannover.de

Variances of Affine Transformation Parameters

	$\sigma^2_{e_{11}}$	$\sigma^2_{e_{12}}$	$\sigma^2_{e_{21}}$	$\sigma^2_{e_{22}}$	mean $(\sigma^2_{e_{11}}, \sigma^2_{e_{22}})$	mean $(\sigma^2_{e_{12}}, \sigma^2_{e_{21}})$
350m seq.	2.03e−7	6.03e−7	6.59e−7	2.24e−7	**2.13e−7**	**6.31e−7**
500m seq.	1.94e−7	5.09e−7	3.63e−7	1.94e−7	**1.94e−7**	**4.35e−7**
1000m seq.	1.74e−7	4.05e−7	4.13e−7	2.12e−7	**1.93e−7**	**4.09e−7**
1500m seq.	3.19e−7	3.80e−7	3.69e−7	3.46e−7	**3.33e−7**	**3.75e−7**
Mean	**2.23e−7**	**4.74e−7**	**4.51e−7**	**2.44e−7**	**2.33e−7**	**4.63e−7**

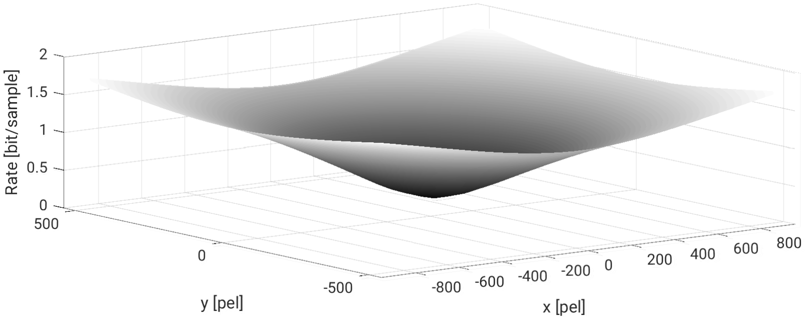
Measured variances $\sigma^2_{e_{ij}}$ of affine transformation parameters of aerial videos from the TAVT data set[3].

Note: $\sigma^2_{e_{11}}$ and $\sigma^2_{e_{22}}$ as well as $\sigma^2_{e_{12}}$ and $\sigma^2_{e_{21}}$ are pairwise similar.

[3] *TNT Aerial Video Testset (TAVT)*, Institut für Informationsverarbeitung (TNT), Leibniz Universität Hannover, 2014, online: `https://www.tnt.uni-hannover.de/project/TAVT/`

Holger Meuel
meuel@tnt.uni-hannover.de

Simulated Location Dependent Bit Rate

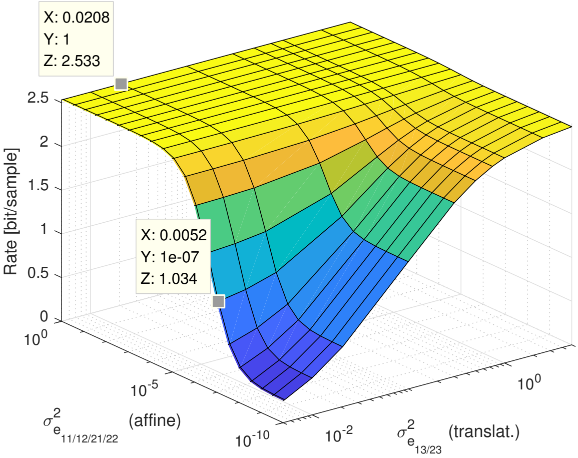


Bit rate

Simulation for Gaussian distributed displacement estimation error p.d.f.s for full HD image and variances $\sigma^2_{e_{11}} = \sigma^2_{e_{22}} = 2.3\text{e}-7$, $\sigma^2_{e_{12}} = \sigma^2_{e_{21}} = 4.6\text{e}-7$.

Holger Meuel
meuel@tnt.uni-hannover.de

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.
Datatips indicate isolines for translational quarter- (0.0052) and half-pel resolution.

Outline

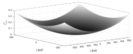
Holger Meuel
meuel@tnt.uni-hannover.de

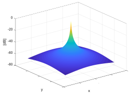
Summary: RD Theory for Affine MCP in Video Coding

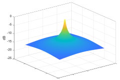
Model for affine motion compensation in video coding:

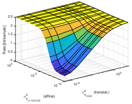
- Model the displacement estimation error as a function of the motion estimation accuracy:
- Model the p.d.f. of displacement estimation error $p_{\Delta X,\Delta Y}(\Delta x, \Delta y)$
- Model the P.S.D. of video signal S_{ss}
- Derivation of power spectral density of displacement estimation error S_{ee}
- Application of rate-distortion function

$\Rightarrow$ **Model for minimum required bit rate for prediction error coding**









Holger Meuel
meuel@tnt.uni-hannover.de