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

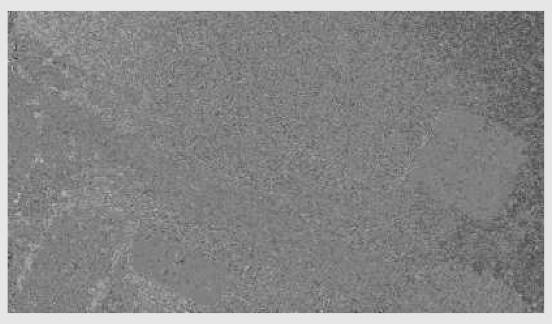
Holger Meuel · Stephan Ferenz · Yiqun Liu · Jörn Ostermann

Scenario and Goal

- 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
  - ⇒ High prediction error
  - ⇒ High entropy
  - ⇒ High bit rate

Original aerial frame (TAVT data set [1])

Bad estimation ⇒ high prediction error



Good estimation ⇒ low prediction error

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

Probability Density Function of Displacement Estimation Error

- Affine motion model:

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

- Error model:

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

- Probability density function (pdf) of the estimation error:

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

- Joint pdf for independent e_{ij}: $p_{E_{11},\ldots,E_{23}}(e_{11},\ldots,e_{23}) = p(e_{11})\cdot\ \ldots\ \cdot p(e_{23})$
- With transformation theorem for pdfs:

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

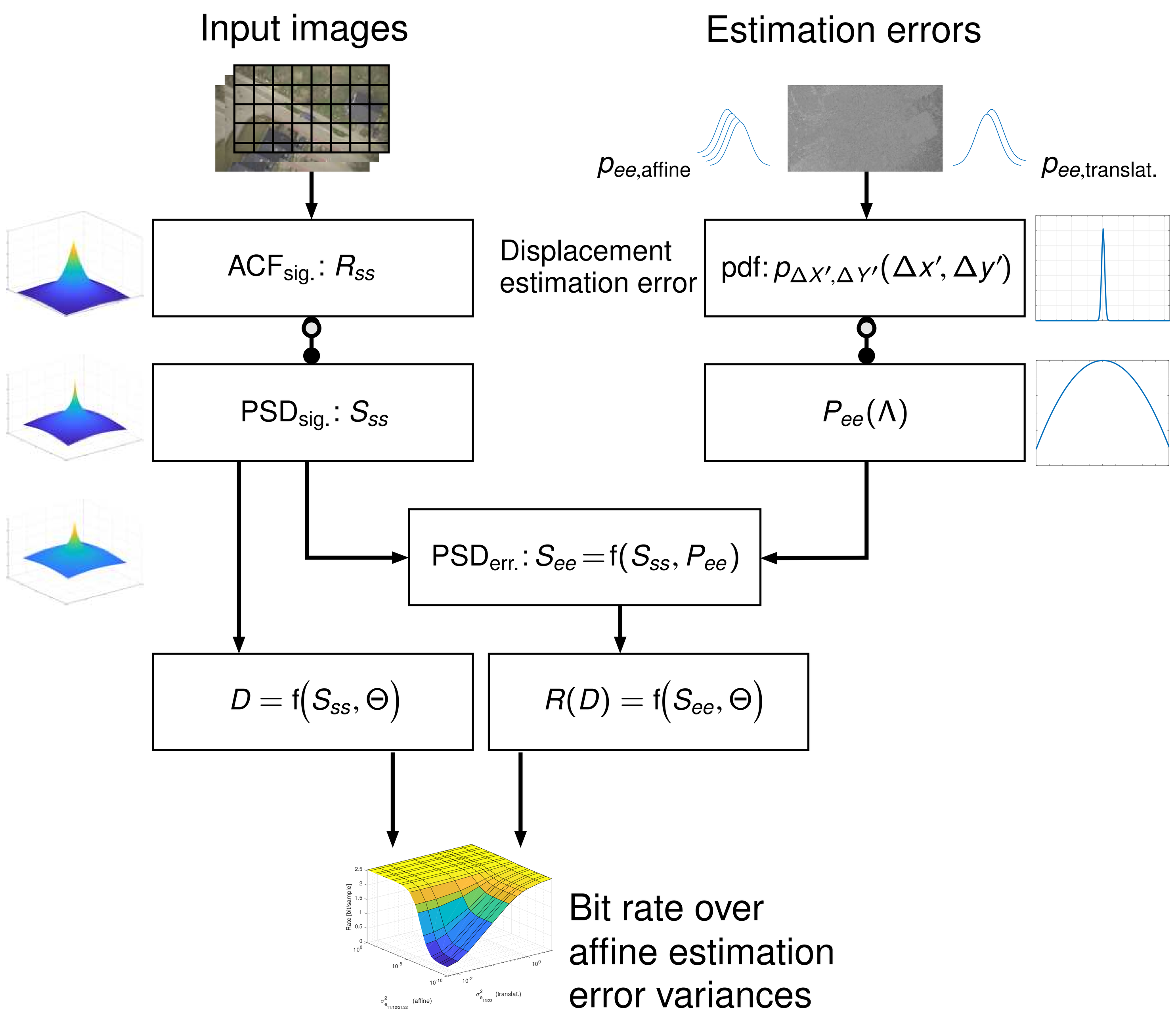
Pdf of displacement estimation error:

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

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

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

- $P_{ee}(\Lambda)$ is the Fourier transform of $p_{\Delta X,\Delta Y}(\Delta x,\Delta y)$!



Flow diagram of the analysis

Rate-Distortion Analysis

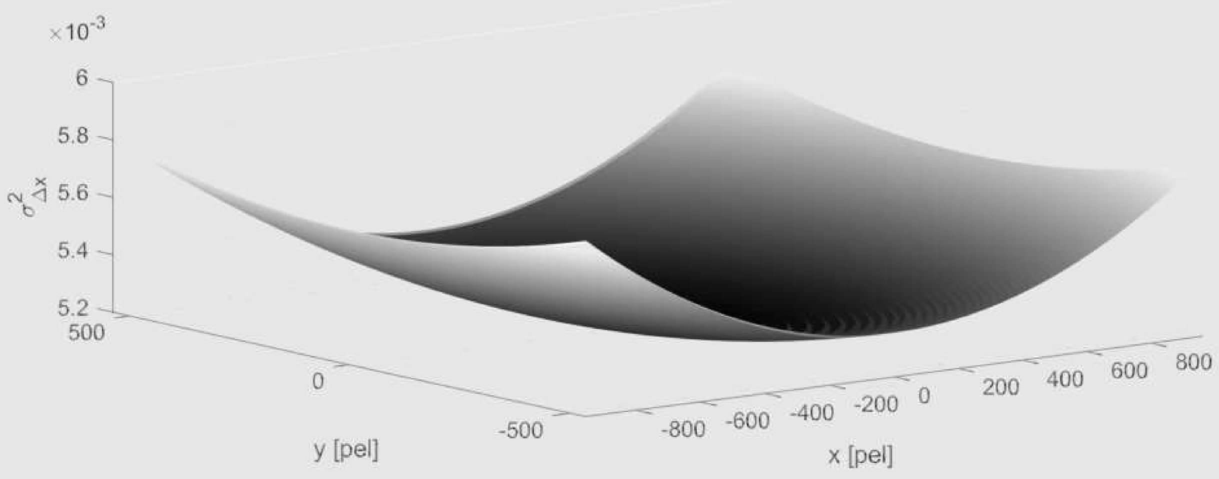
- Assumption for signal autocorrelation function:

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

- Measured correlations for HD sequences: $\alpha_x = 0.9744$, $\alpha_y = 0.9677$
- Measured affine estimation error variances:

$\sigma^2_{e_{11}}$	$\sigma^2_{e_{12}}$	$\sigma^2_{e_{13}}$	$\sigma^2_{e_{21}}$	$\sigma^2_{e_{22}}$	$\sigma^2_{e_{23}}$
$3.27\cdot10^{-10}$	$6.73\cdot10^{-10}$	$3.06\cdot10^{-5}$	$6.61\cdot10^{-10}$	$3.19\cdot10^{-10}$	$2.83\cdot10^{-5}$

Mean of purely affine variances $\sigma^2_{e_{11/12/21/22}} \approx 5e{-10}$



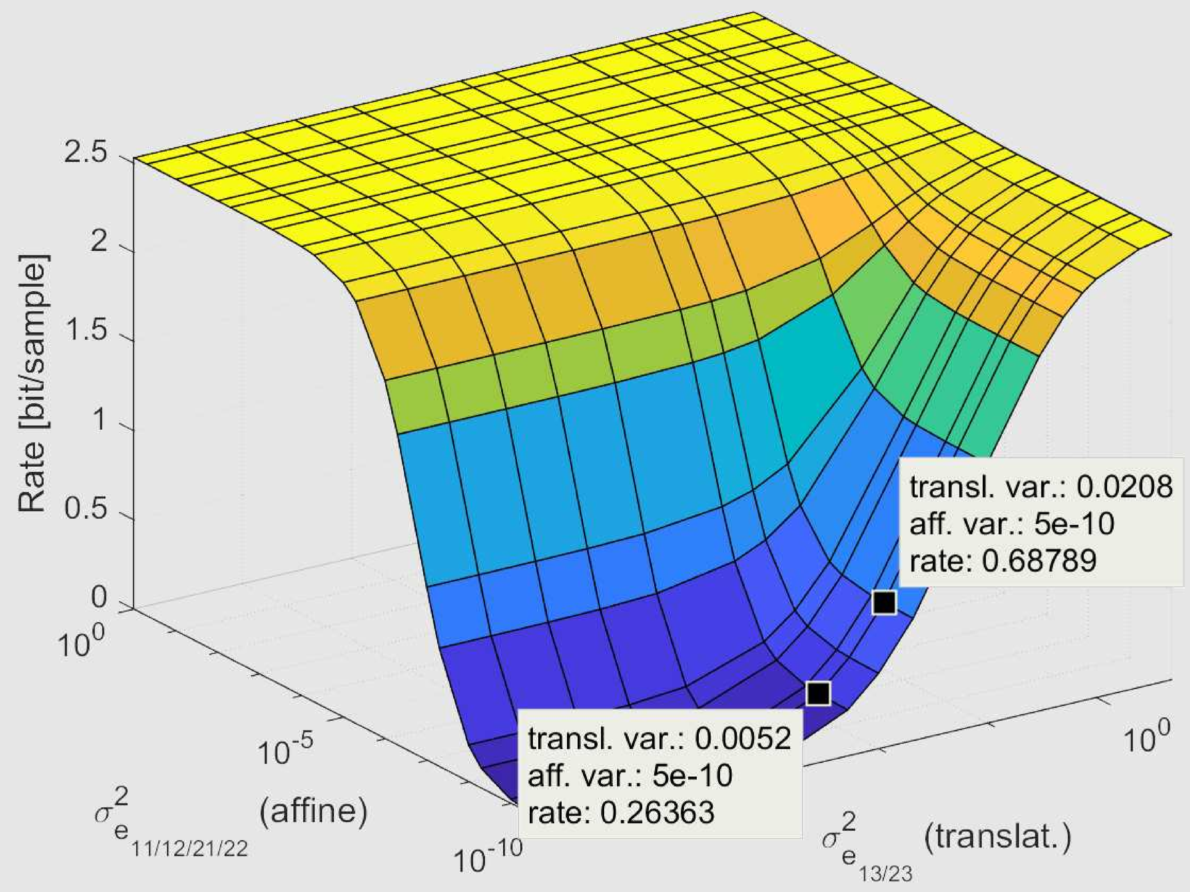
Location dependent variance $\sigma^2_{\Delta x}$ of Gaussian distributed displacement estimation error pdf for a full HD resolution image and $\sigma^2_{e_{11/12}} = 5e{-10}$ and translational quarter-pel resolution ($\sigma^2_{e_{13}} = 0.0052$).

- Power spectral density (PSD) $S_{ss}(\Lambda)$ of the signal is the Fourier transform of R_{ss} (Wiener-Khinchin theorem)
  - $\Lambda = (\omega_x, \omega_y)$ 2D spatial frequencies
- PSD of the displacement estimation error [2]:

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

  - Θ: parameter that generates the function $R(D)$ by taking on all positive real values
- Rate-distortion theory:

$$D = \frac{1}{4\pi^2}\iint_{\Lambda}\min\big[\Theta, S_{ss}(\Lambda)\big]\,d\Lambda$$
$$R(D) = \frac{1}{8\pi^2}\iint_{\Lambda:\left(S_{ss}(\Lambda)>\Theta \text{ and } S_{ee}(\Lambda)>\Theta\right)}\log_2\left[\frac{S_{ee}(\Lambda)}{\Theta}\right]d\Lambda\ \text{bit}$$



Minimum required bit rate for prediction error coding for a distortion of 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: isolines for translational $\frac{1}{4}$- (lower datatip) and $\frac{1}{2}$-pel resolution.

Summary

- Rate-distortion analysis of *affine* motion compensated prediction
- Model valid for block-based and global motion compensated prediction
- Much smaller bit rates achievable for accurate state-of-the-art motion estimators!

Main contribution: Derivation of the pdf of the location dependent displacement estimation error $p_{\Delta X,\Delta Y}(\Delta x,\Delta y|x',y')$!

Reference: [1] *TNT Aerial Video Testset* (TAVT), 2010–2014, URL: https://www.tnt.uni-hannover.de/project/TNT_Aerial_Video_Testset/
[2] B. Girod, "The Efficiency of Motion-Compensating Prediction for Hybrid Coding of Video Sequences," in IEEE Journal on Sel. Areas in Communicat., vol. 5, no. 7, pp. 1140–1154, Aug. 1987

Institut für Informationsverarbeitung
Holger Meuel
meuel@tnt.uni-hannover.de
https://www.tnt.uni-hannover.de/staff/meuel
Leibniz Universität Hannover