Advanced Video Processing for Future Multimedia Communication Systems

André Kaup

Chair of Multimedia Communications and Signal Processing
Friedrich-Alexander University Erlangen-Nürnberg

FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Future Multimedia Communication Systems

Trend in video to make communication more immersive by

- higher image resolution and quality
- stereo and multi-view systems

Data rate of uncompressed UHD video:

3840 x 2160 x 24 bit/pel x 30 frames/s

= **6 Gbit/s** or **746 MB/s**

One CD every second

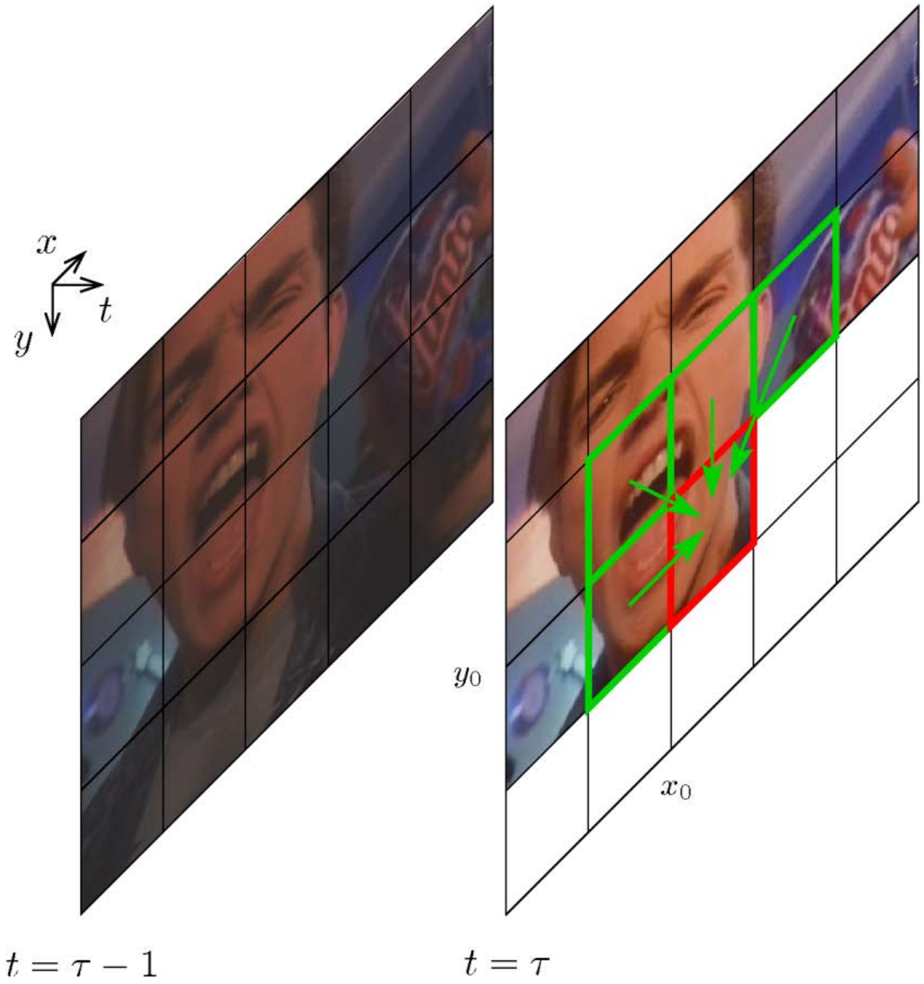
One DVD every 6 seconds

One Blu-ray disc every 36 seconds

⇨ **Conclusion:** compression is necessary and higher resolutions are more challenging with respect to efficiency and picture quality

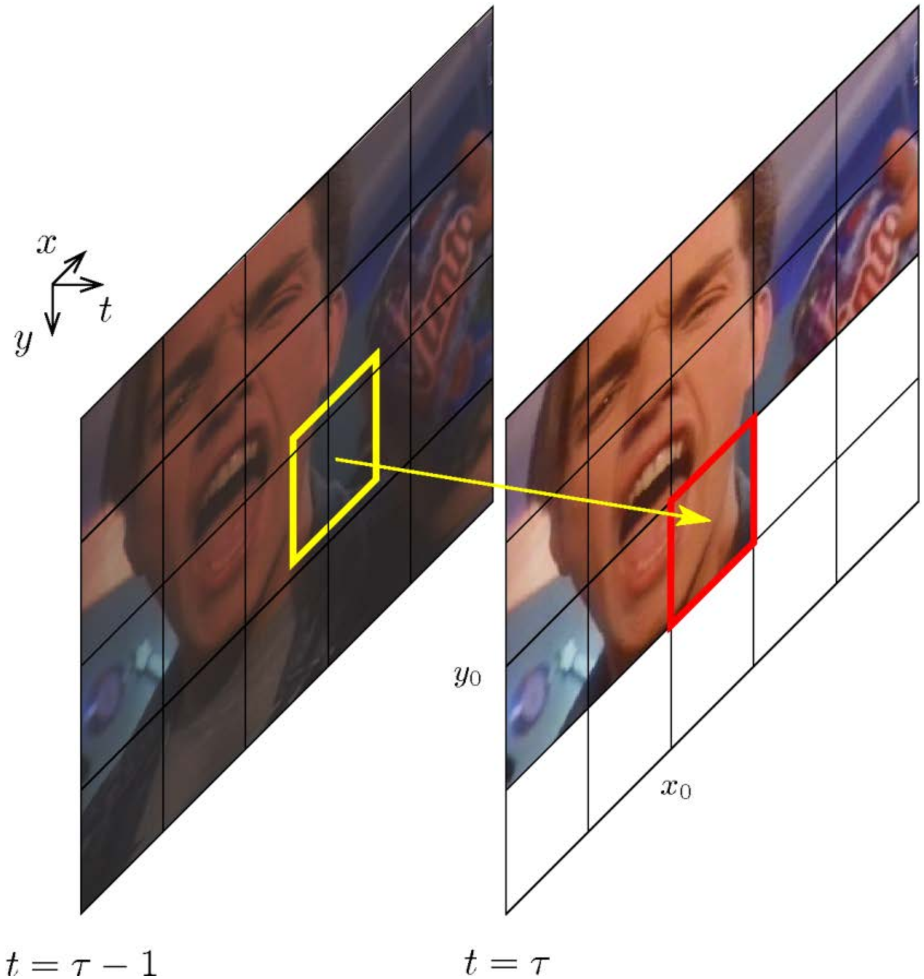
Intraframe Prediction

- **Assumption:**
  images have (locally) high spatial correlation
- **Intraframe prediction**
  uses already decoded image blocks in causal neighborhood
- **Best prediction mode**
  according to optimization constraint is transmitted as side information



Interframe Prediction

- **Assumption:** video has (locally) high temporal correlation
- **Interframe predicion** using motion compensated reference
- **Multiple** reference images for long term prediction
- **Motion vector** is transmitted as side information

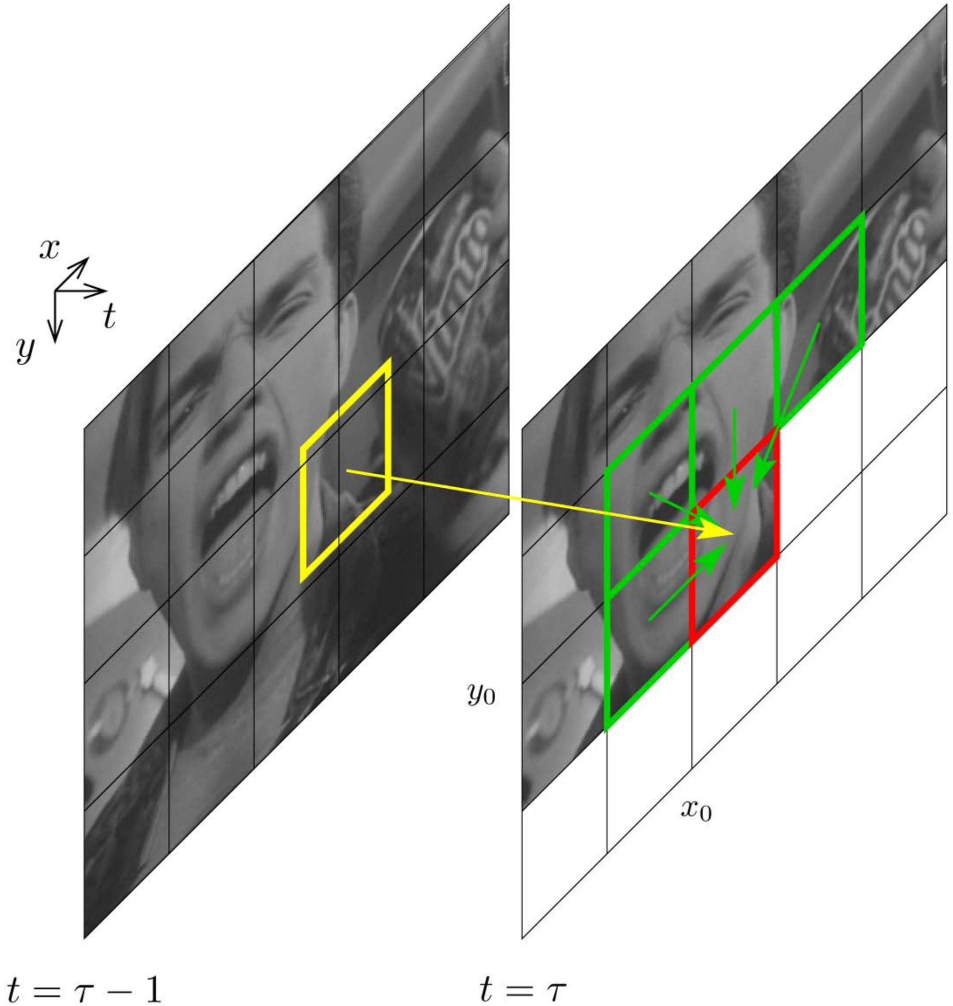


Overview

- Advancing compression efficiency
  - **Spatiotemporal prediction**
  - In-loop video denoising
  - Measurement and prediction of energy consumption
- Improving displaying quality
  - Scalable lossless compression
  - Multi-view video using super-resolution
  - Random sampling and reconstruction
- Conclusions and outlook

Spatiotemporal Prediction

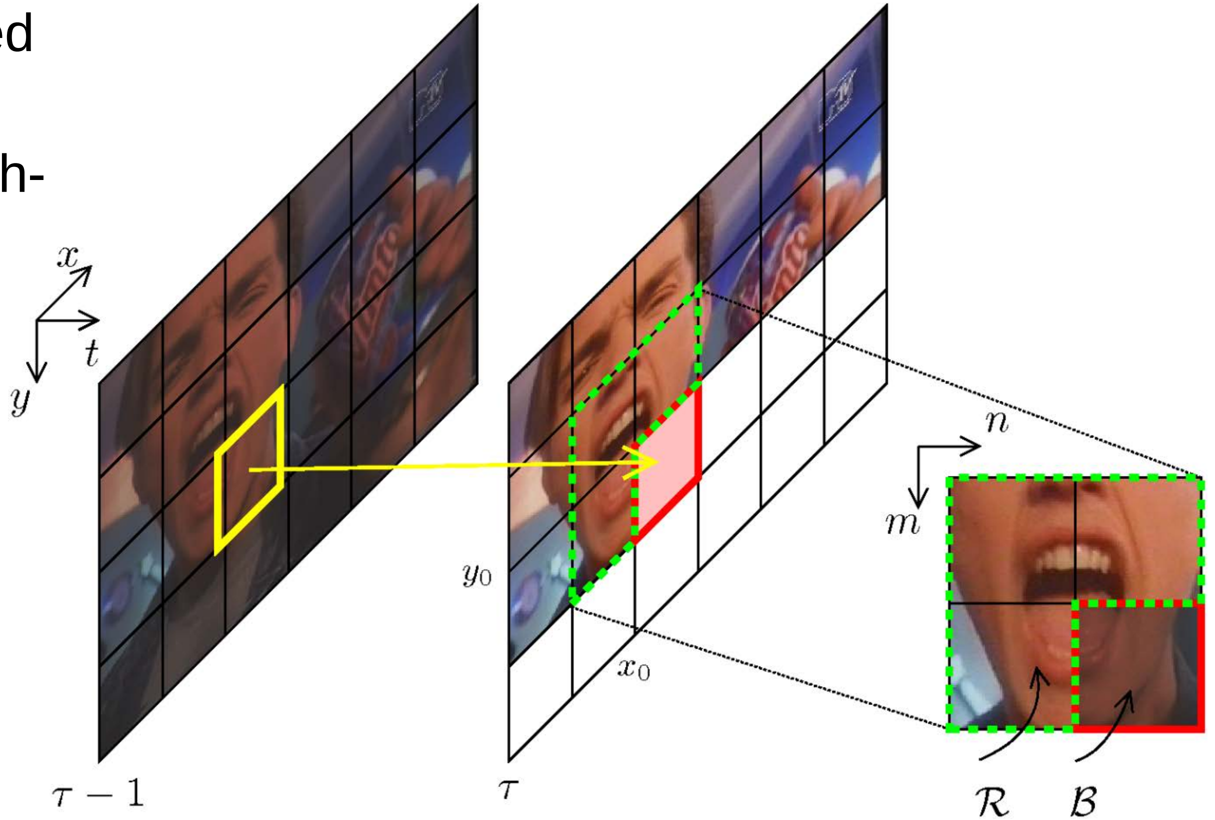
- **State-of-the-art:** switching between interframe und intraframe prediction modes
- **Decision** taken using rate-distortion optimization
- **Extended approach:** joint spatiotemporal prediction as additional mode



Non-Local Means Refined Prediction

Processing area $\mathcal{L} = \mathcal{B} \cup \mathcal{R}$ is regarded for prediction

- Motion compensated block $\mathcal{B}$
- Reconstructed neighboring blocks $\mathcal{R}$



Non-Local Means Refined Prediction

Task: Recover original signal $s\left[m,n\right]$ from given signal samples $\widetilde{s}\left[m,n\right]$ in area $\mathcal{L} = \mathcal{B} \cup \mathcal{R}$

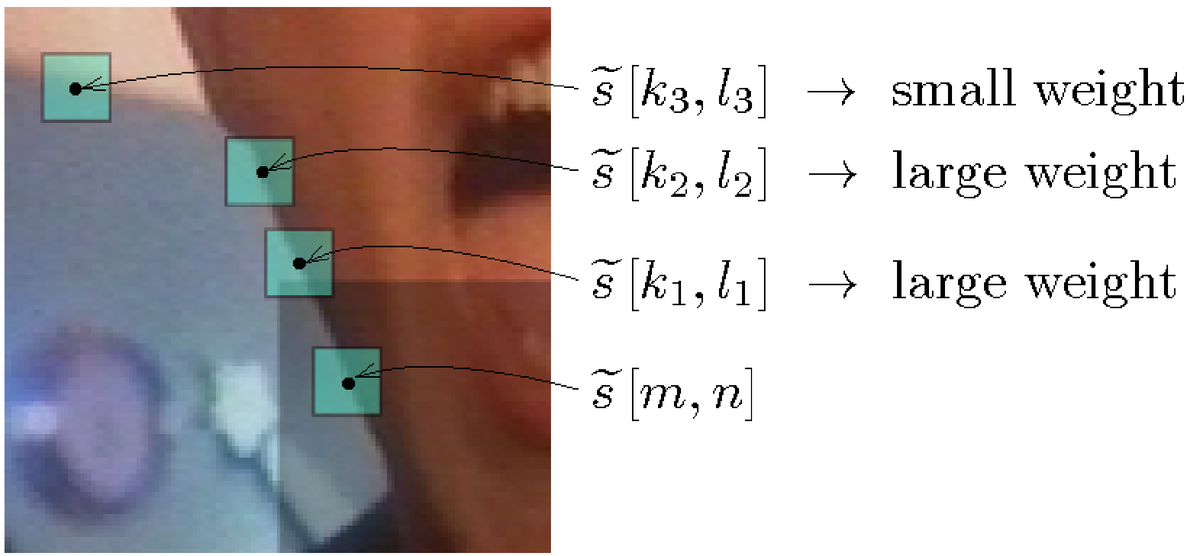
Non-local means: Estimate refined samples $\hat{s}\left[m,n\right]$ in $\mathcal{B}$ using weighted non-local average filter

$$\hat{s}\left[m,n\right] = \frac{\sum\limits_{(k,l)\in\mathcal{L}} \widetilde{s}\left[k,l\right] w_{(m,n)}\left[k,l\right]}{\sum\limits_{(k,l)\in\mathcal{L}} w_{(m,n)}\left[k,l\right]}$$

Non-Local Means Refined Prediction

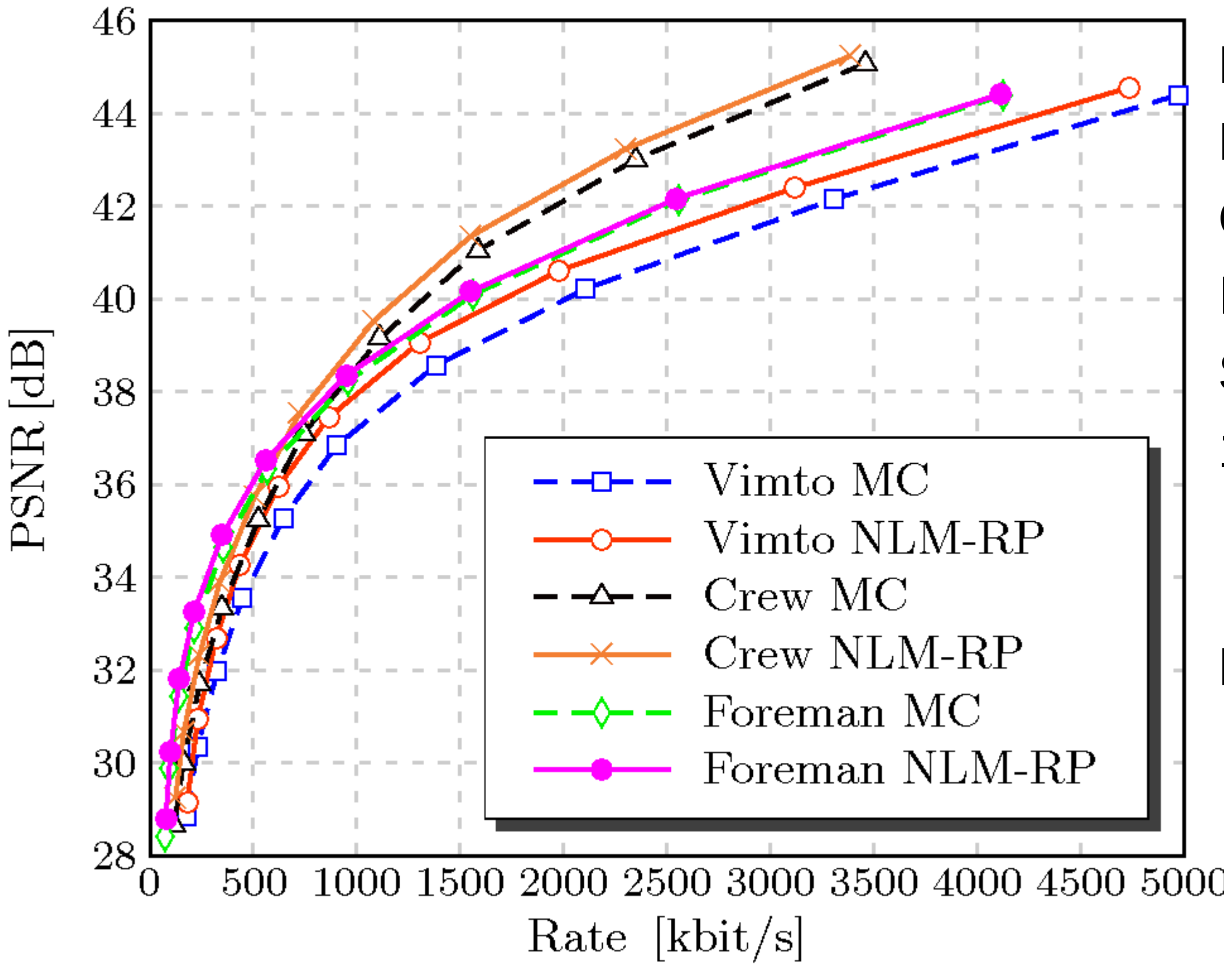
Example for weight calculation:

- Samples with similar neighborhood get a large weight
- Samples with dissimilar neighborhood get a small weight



Test Sequences

Simulation Results



H.264/AVC JM10.2

Baseline Profile, Level 2.0

CIF sequences

IPPP…, 100 frames

Search range: 16 sample

1 bit/block for signaling the new mode

NLM-RP parameters:

$h = 25$

$d_{\mathrm{m}} = 3$

[Seiler, Richter, Kaup, PCS 2010]

Simulation Results

Motion compensated prediction

QP34: 33.54 dB @ 447 kbit/s

Non-local means refined prediction

QP34: 34.25 dB @ 434 kbit/s

Simulation Results

Motion compensated prediction	Non-local means refined prediction

QP34: 33.54 dB @ 447 kbit/s

QP34: 34.25 dB @ 434 kbit/s

Overview

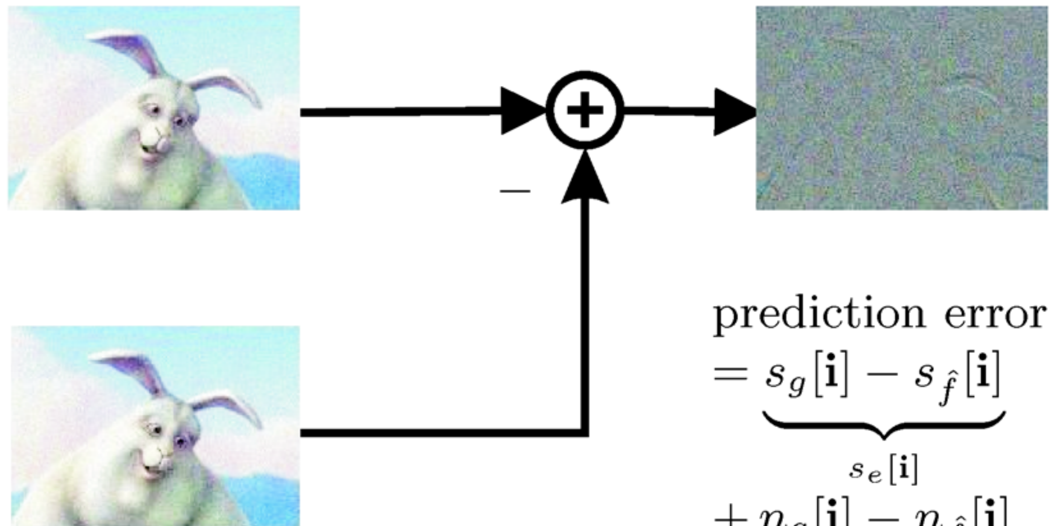
- Advancing compression efficiency
  - Spatiotemporal prediction
  - **In-loop video denoising**
  - Measurement and prediction of energy consumption
- Improving displaying quality
  - Scalable lossless compression
  - Multi-view video using super-resolution
  - Random sampling and reconstruction
- Conclusions and outlook

Prediction Error Signal

- **Problem**
  - Prediction error signal has more noise than the current frame itself

current frame $g[\mathbf{i}]$
$= s_g[\mathbf{i}] + n_g[\mathbf{i}]$

motion compensated
reference frame $\hat{f}[\mathbf{i}]$
$= s_{\hat{f}}[\mathbf{i}] + n_{\hat{f}}[\mathbf{i}]$



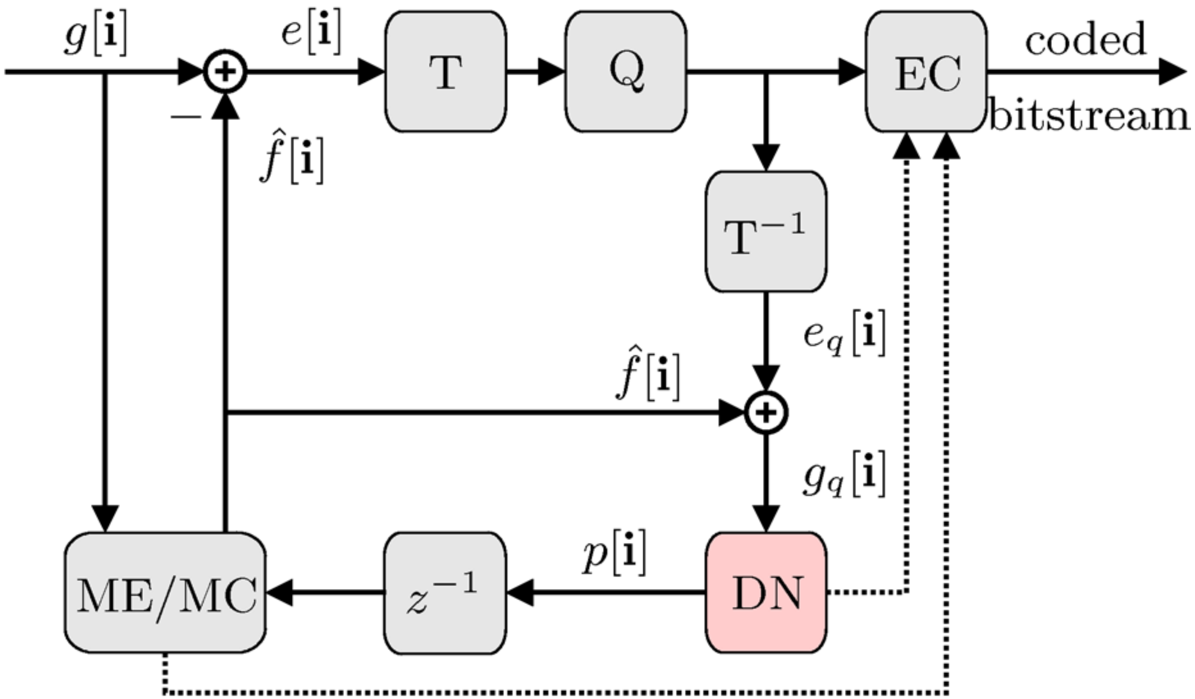
prediction error $e[\mathbf{i}]$

$$= \underbrace{s_g[\mathbf{i}] - s_{\hat{f}}[\mathbf{i}]}_{s_e[\mathbf{i}]} + \underbrace{n_g[\mathbf{i}] - n_{\hat{f}}[\mathbf{i}]}_{n_e[\mathbf{i}]}$$

- **Solution**
  - Remove noise from the predictor

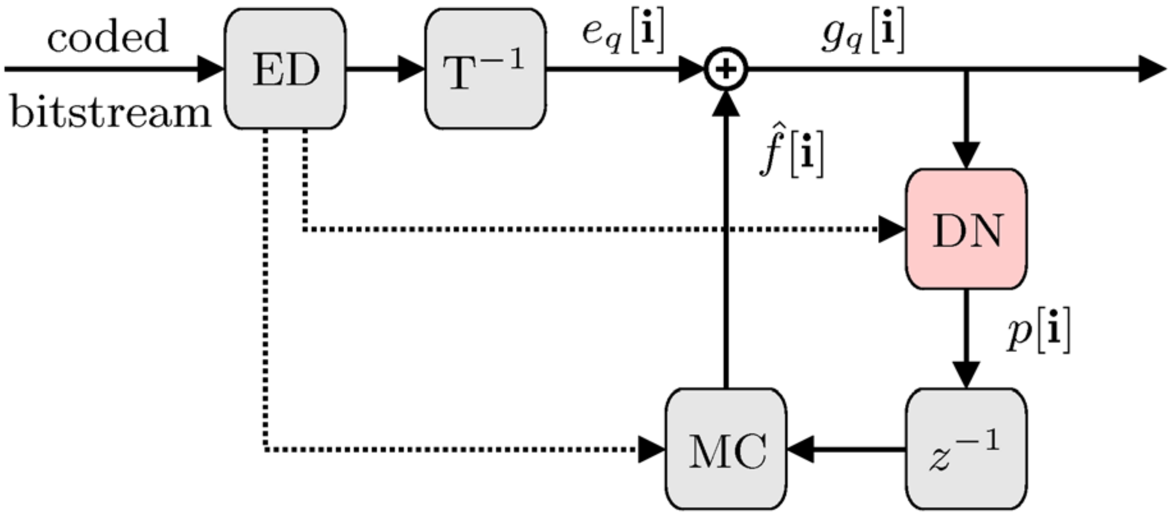
Inter-Frame Encoder with In-Loop Denoising

- Simplified block diagram of an inter-frame encoder **with in-loop denoising**



Inter-Frame Decoder with In-Loop Denoising

- Simplified block diagram of an inter-frame decoder **with in-loop denoising**

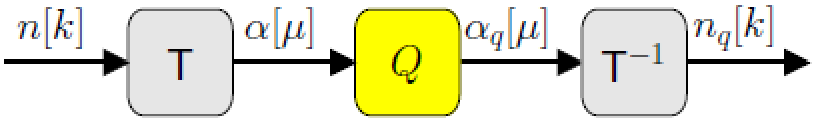


- Denoising is performed **after** displaying the decoded frames

Quantization of Noise

- Noise filtering works on **transformed** and **quantized** reference signal

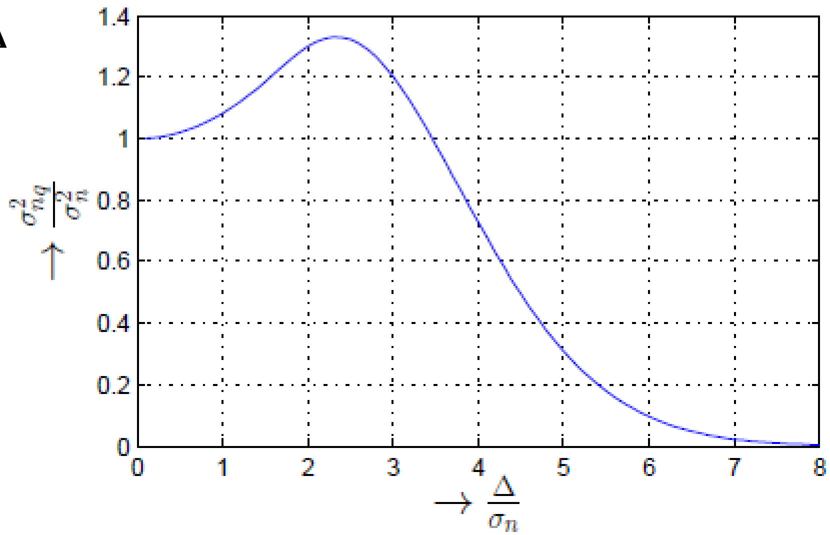
- Analytical model for Gaussian noise and perfect prediction



- **Observation:** Noise depends on Δ

$$\sigma_{n_q}^2(\Delta) = 2 \cdot \Delta^2 \sum_{\lambda=1}^{\infty} \lambda^2 w(\lambda)$$

$$w(\lambda) = \mathrm{erf}\left(\frac{\lambda\Delta + \frac{\Delta}{2}}{\sigma_n\sqrt{2}}\right) - \mathrm{erf}\left(\frac{\lambda\Delta - \frac{\Delta}{2}}{\sigma_n\sqrt{2}}\right)$$

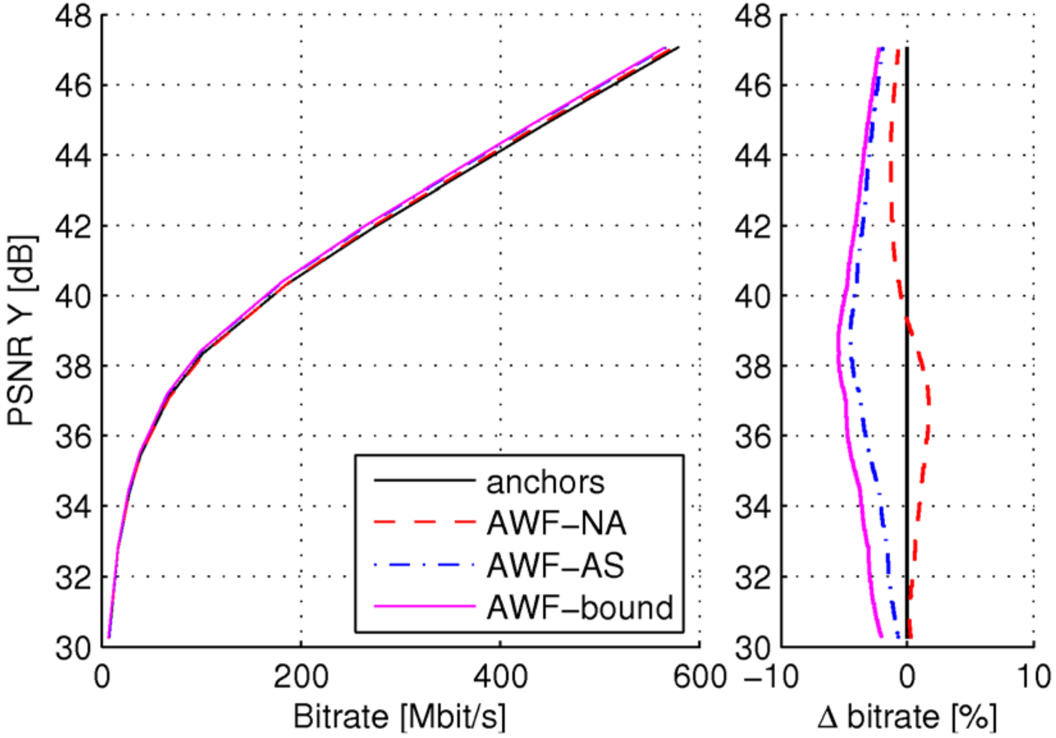


Simulation Conditions

- HEVC reference software HM-2.2
  - Coding of 100 frames
  - QP 2 {12 … 37}
  - Coding configurations: ldlc_P, ldhe_P, ldlc, ldhe
- SVT test sequences from ftp://vqeg.its.bldrdoc.gov/
  - Resolution of 3840x2160 pixels with 50 frames per second
  - Using a centrically cropped version of 2560x2160 pixels
- Denoising parameters
  - AWF: window of 3x3 pixels
  - Δ_L=1, Δ_H=3

Simulation Results for *ParkJoy* (ldlc)

- Estimated noise of the input sequence $\sigma_n \approx 1.8$



Overview

- Advancing compression efficiency
  - Spatiotemporal prediction
  - In-loop video denoising
  - **Measurement and prediction of energy consumption**
- Improving displaying quality
  - Scalable lossless compression
  - Multi-view video using super-resolution
  - Random sampling and reconstruction
- Conclusions and outlook

Energy Consumption of a Video Decoder

Battery constraint:

- Operating time of portable devices is limited by battery capacity
- Especially critical for HD and UHD content

Goal: Extend operating times by reducing the required decoding energy

Modeling the Decoding Energy

- Decoding energy estimated through **processing time**:

$$\hat{E}_{\mathrm{dec}} = P \cdot t_{\mathrm{dec}} + E_0$$

[Herglotz, Kaup, ISCAS 2015]

Mean power

Decoding time

Offset Energy

- Decoding energy estimated through **processor events**:

$$\hat{E}_{\mathrm{dec}} = \sum_{i=0}^{M} f(x_i)$$

M=3 events considered:

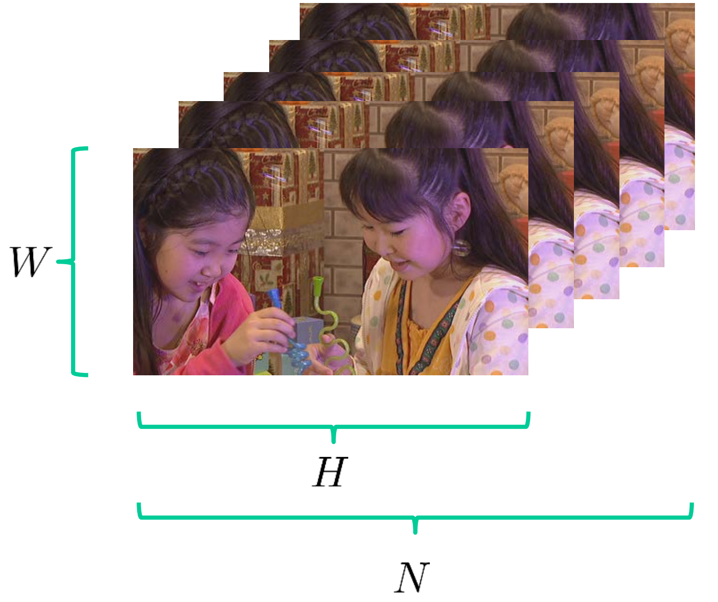
- Number of instruction fetches
- Level 1 data cache misses
- Hardware interrupts

Modeling the Decoding Energy – Cont'd

- Decoding energy estimated through **high-level features**:

$$\hat{E}_{\mathrm{dec}} = W \cdot H \cdot N \cdot \left(\alpha + \beta \cdot \frac{B}{W \cdot H \cdot N}^{\gamma} \right)$$

W: Frame width (pixels)
H: Frame height (pixels)
N: Number of frames
B: Bit stream file size
α, β, γ: System specific variables



[Herglotz, Kaup, EUSIPCO 2015]

Modeling the Decoding Energy – Cont'd

- Decoding energy estimated through **bit stream features**:

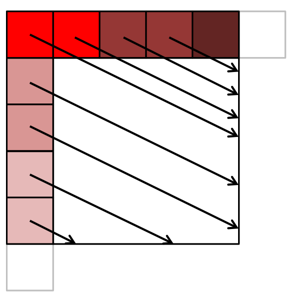
$$\hat{E}_{\text{dec}} = \sum_{\forall i} n_i \cdot e_i$$

[Herglotz, Kaup, ICIP 2014]

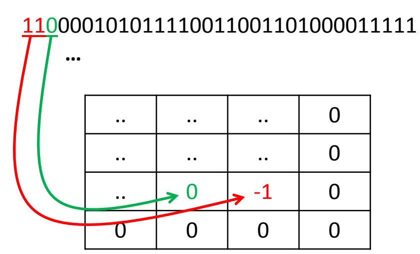
i: Feature index (up to 90)
n_i: Number of occurrences
e_i: Feature specific energy

- Examples:

Intra prediction (mode and block size)



Coefficient decoding



Estimation Accuracies

- **Test set:** 120 sequences, 16-40 frames, QP=10,32,45
- **Encoder configurations:** intra, low delay (P), random access,
- **Software:** HM-13.0, libde165, FFmpeg
- **Hardware:** Pandaboard, Beagleboard, FPGA

Estimation error: $$\varepsilon = \frac{\hat{E}_{\mathrm{dec}} - E_{\mathrm{dec}}}{E_{\mathrm{dec}}}$$

- **Mean absolute estimation errors:**

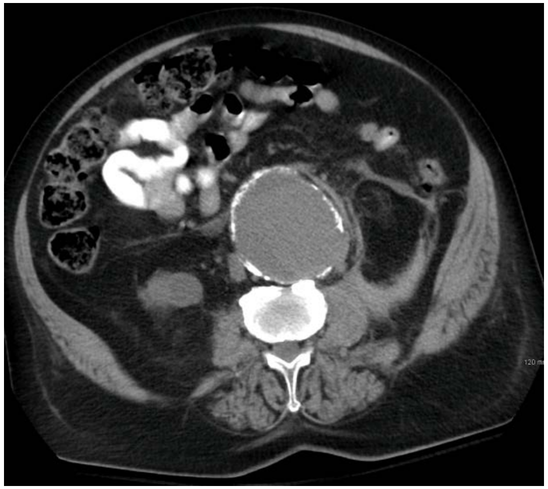
Model	$\overline{\varepsilon}$
Decoding time	3.20%
Processor events	3.24%
High-level features	9.36%
Bit stream features (90)	3.30%
Bit stream features (27)	5.46%

Overview

- Advancing compression efficiency
  - Spatiotemporal prediction
  - In-loop video denoising
  - Measurement and prediction of energy consumption
- Improving displaying quality
  - **Scalable lossless compression**
  - Multi-view video using super-resolution
  - Random sampling and reconstruction
- Conclusions and outlook

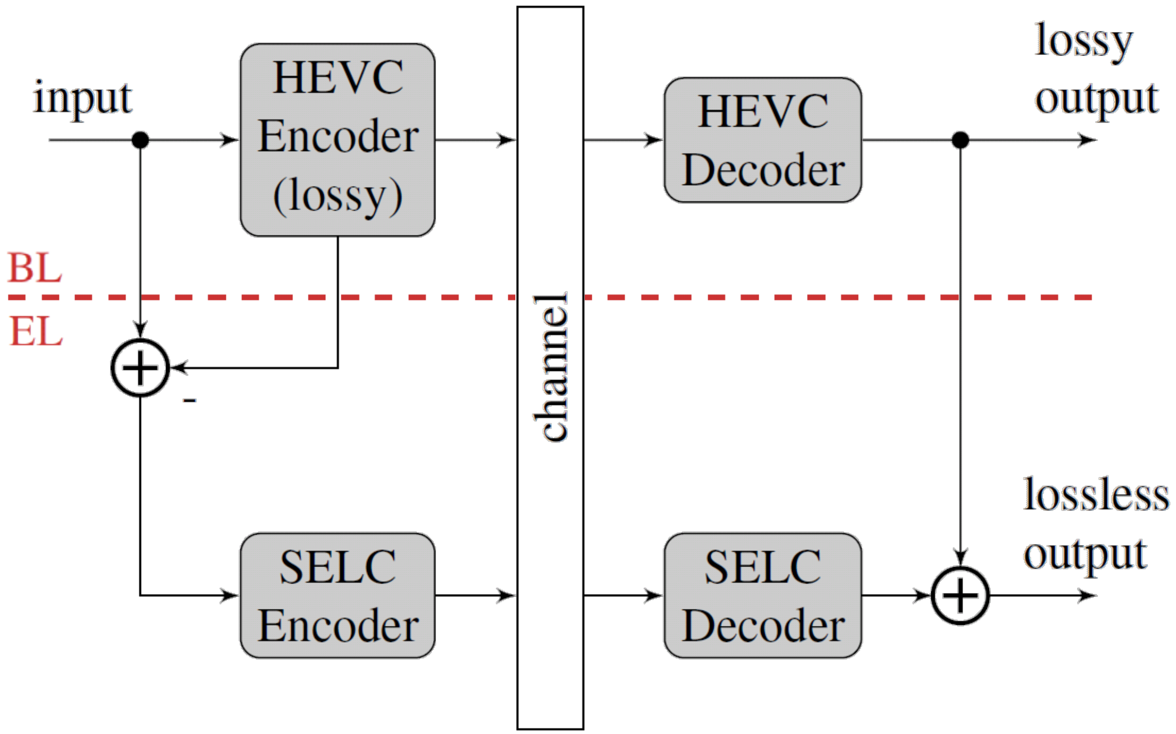
Motivation

- New video coding standard HEVC primarily targeting consumer applications with **lossy** compression
- Need for **lossless** compression in professional applications
  - Medical imaging (telemedicine)
  - Archiving (cinema)
  - …
- High bitrate ⇔ limited channel capacity
- **Scalable lossless coding** using two layers
  - Lossy base layer (BL)
  - Lossless enhancement layer (EL)

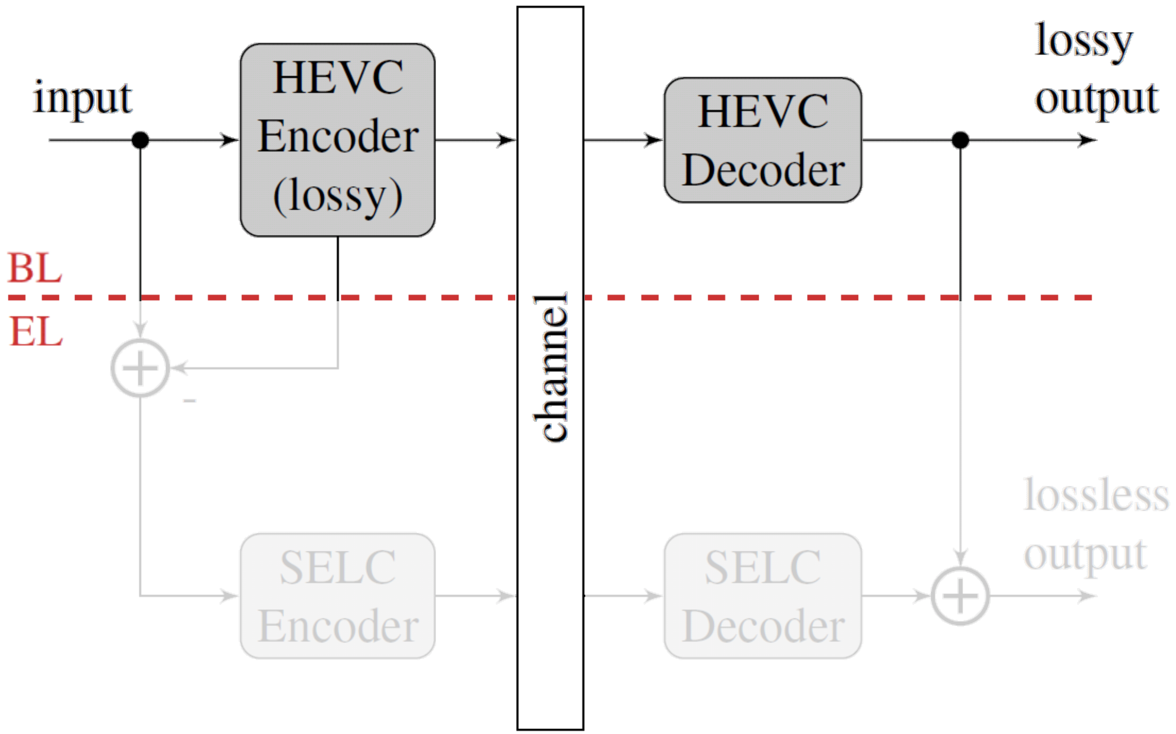


en.wikipedia.org/wiki/File:RupturedAAA.png

System Overview

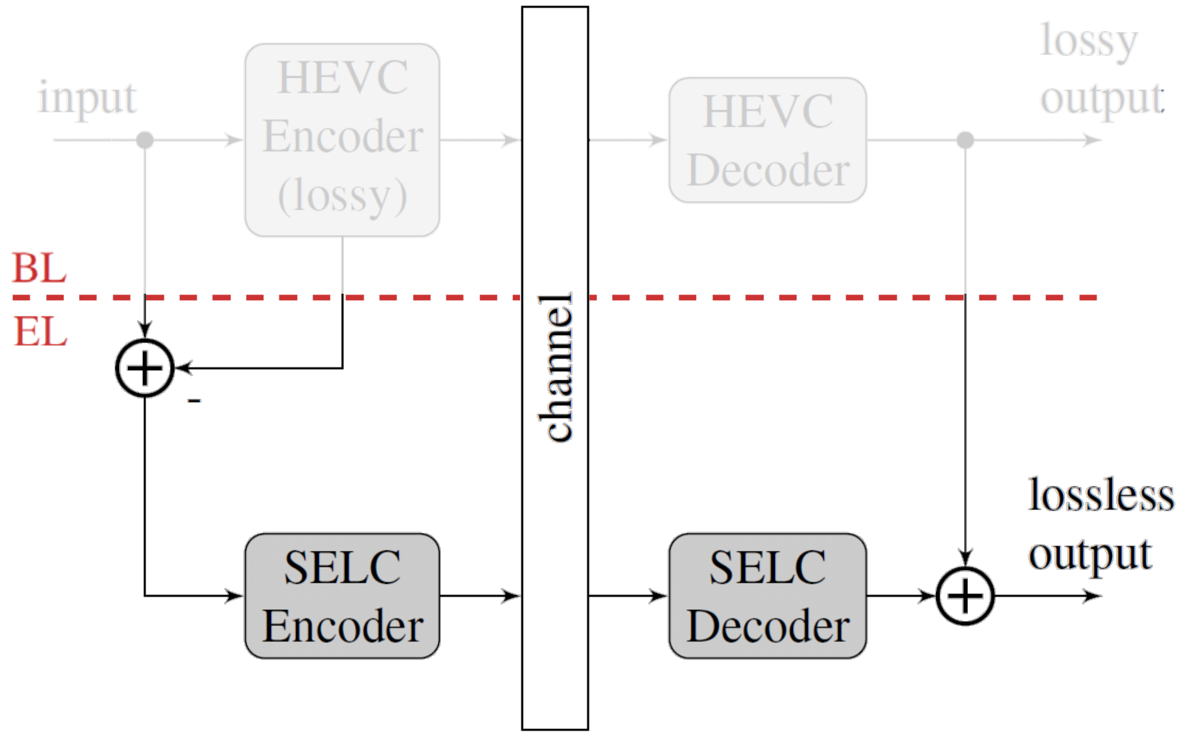


Base Layer



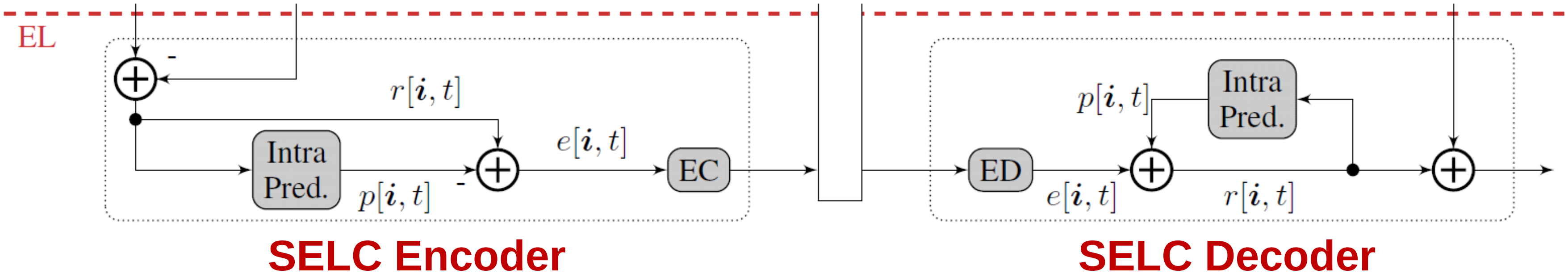
⇨ Lossy BL compression using HEVC

Enhancement Layer



⇨ Lossless EL coding using the proposed *__S__ample-based Weighted Prediction for __E__nhancement __L__ayer __C__oding* (SELC)

Enhancement Layer Coding



Intra prediction:

- Non-linear sample-based weighted prediction (SWP)
- Implemented using fast lockup tables

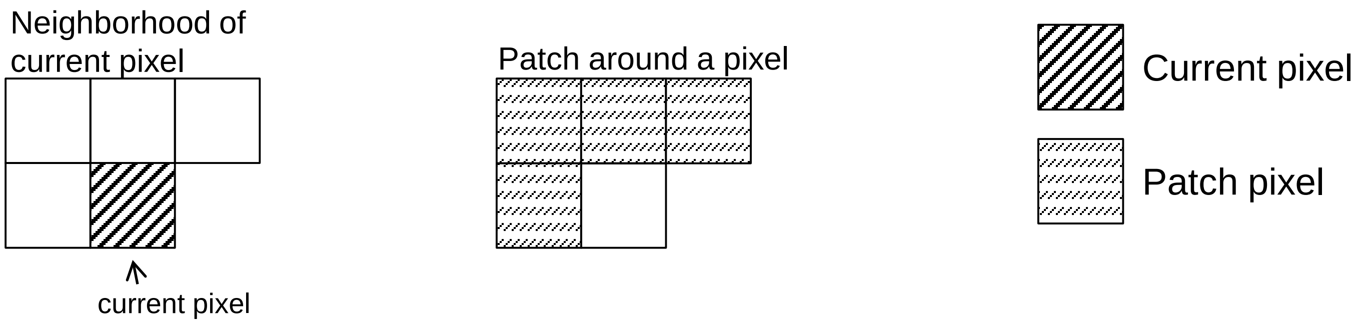
Entropy coding/decoding:

- Modified context-adaptive binary arithmetic coding (CABAC)

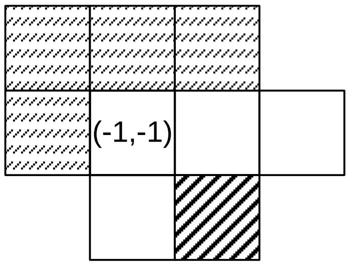
[Wige, Kaup, ICIP 2013]

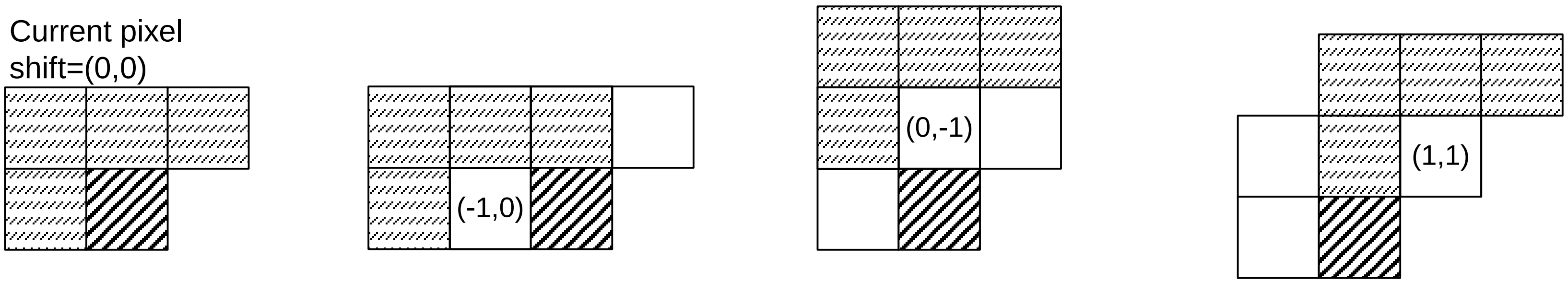
Intra Prediction (SWP) I

⇨ Four-pixel neighborhood and four-pixel patch



⇨ Patch around the current pixel is compared to the patches of the neighborhood pixels ...





[3] P. Amon et al., “RCE2: Sample-based weighted intra prediction for lossless coding,” document JCTVC-M0052, JCT-VC, Apr. 2013.

Experimental Results

Coding efficiency:

Relative bitrate differences[1] for EL coding compared to SHM-2.1

HM-11.0				SELC			
QP22	QP27	QP32	QP37	QP22	QP27	QP32	QP37
1.2%	1.0%	0.3%	0.8%	-2.6%	-4.7%	-6.5%	-7.3%

Runtime:

Relative runtime increase[2] for EL processing compared to BL processing only

	SHM-2.1				HM-11.0				SELC			
	QP22	QP27	QP32	QP37	QP22	QP27	QP32	QP37	QP22	QP27	QP32	QP37
Enc	25.3%	30.6%	34.9%	37.7%	18.5%	22.5%	25.5%	27.7%	0.6%	0.7%	0.9%	0.8%
Dec	244.4%	338.9%	443.6%	536.2%	260.1%	361.9%	451.4%	533.8%	202.8%	279.6%	334.3%	374.8%

[1]: average values w/o ElFuente [2]: average values for all sequences

Overview

- Advancing compression efficiency
  - Spatiotemporal prediction
  - In-loop video denoising
  - Measurement and prediction of energy consumption
- Improving displaying quality
  - Scalable lossless compression
  - **Multi-view video using super-resolution**
  - Random sampling and reconstruction
- Conclusions and outlook

Super-Resolution

- **Super-Resolution** (SR) is a key issue in image and video processing domain
- **Goal:** create reasonable high-frequency content for a low-resolution image or video sequence

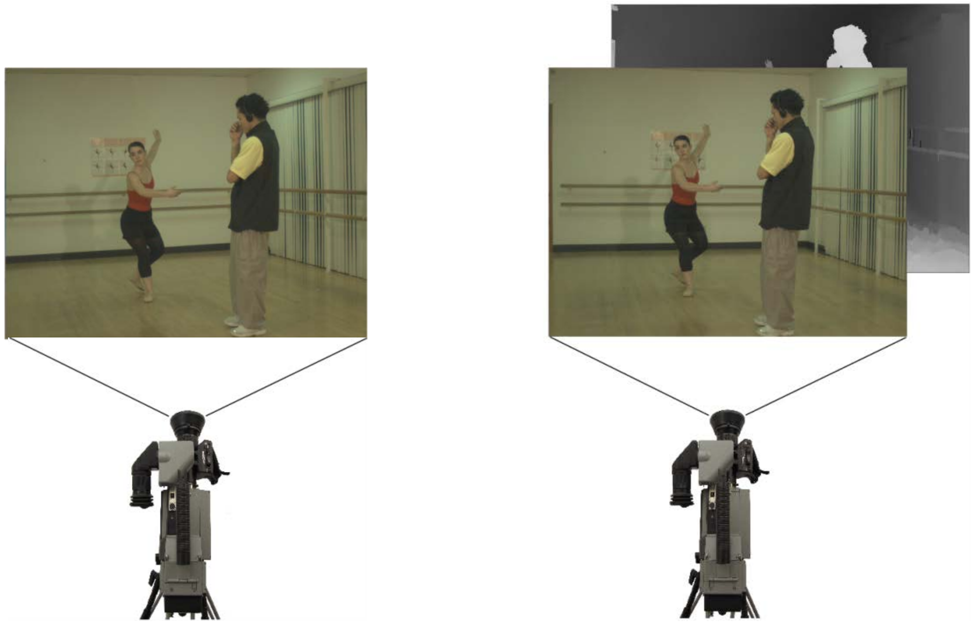
+

?

=

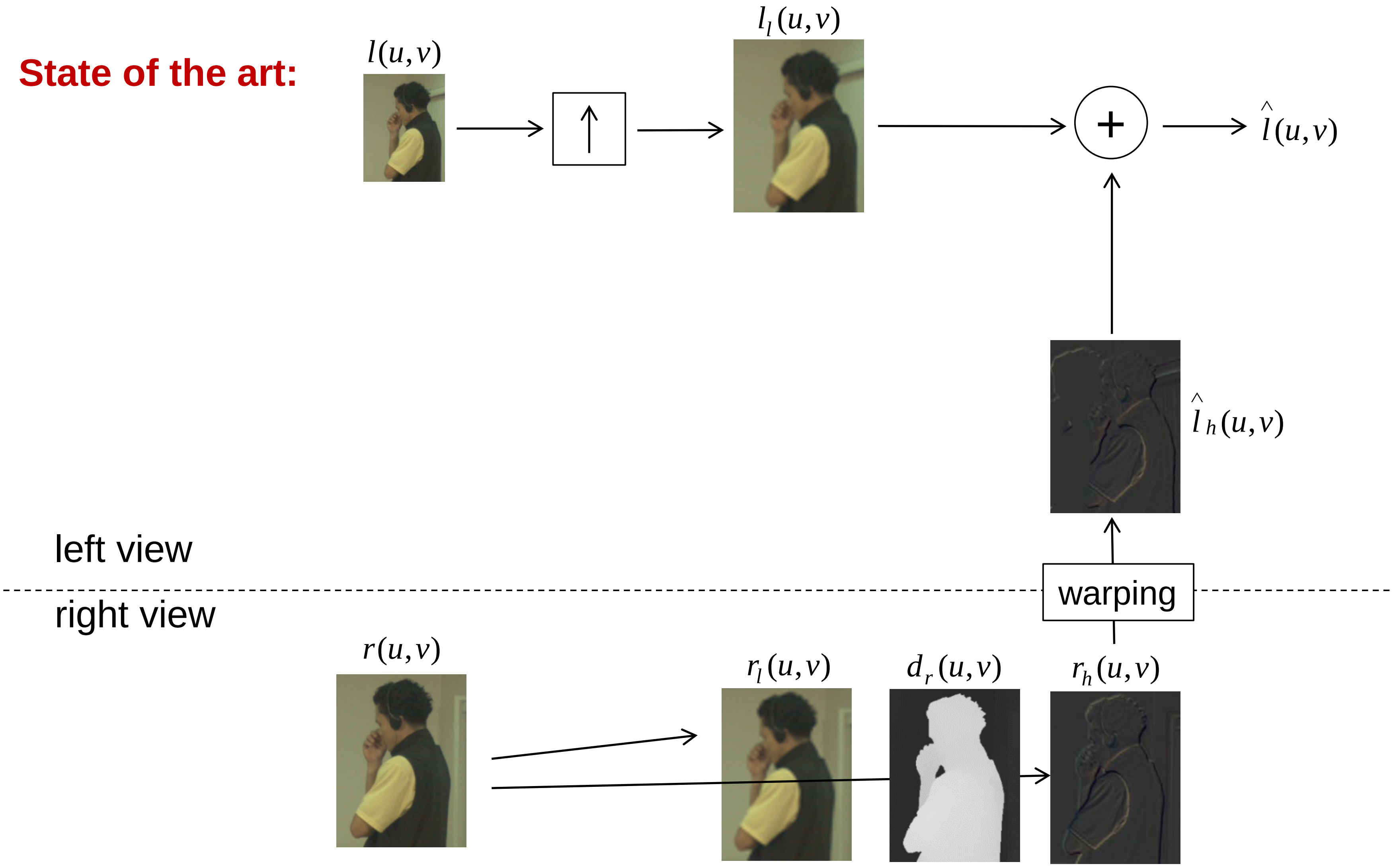
Motivation

- **Mixed-resolution** multi-view video plus depth format (MR-MVD)



- **Goal:** Usage of neighboring high-frequency content to refine low-resolution destination view

Super-Resolution Based on High-Frequency Synthesis



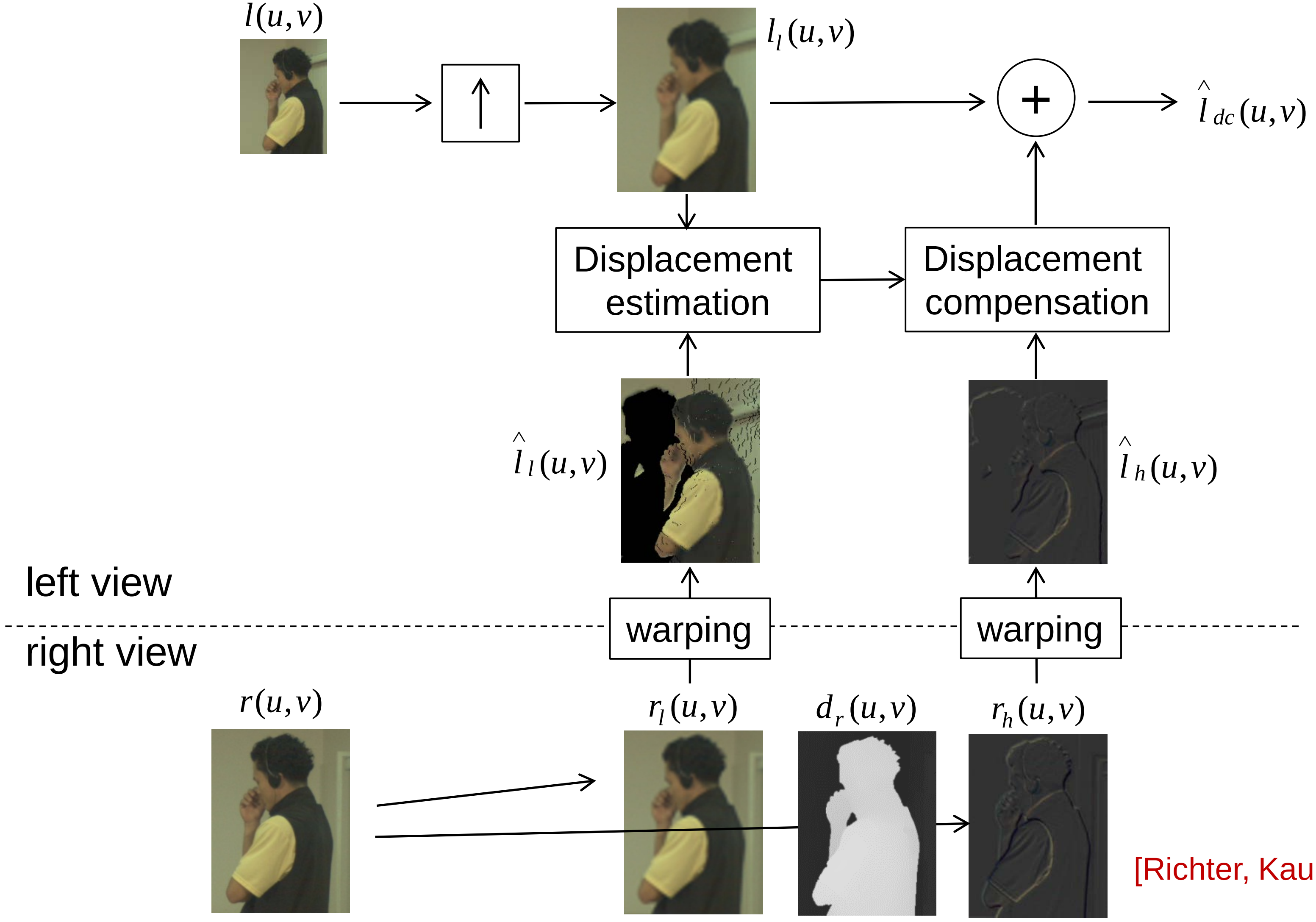
Super-Resolution Based on High-Frequency Synthesis

- Impact of additional depth inaccuracies on visual SR quality:

original | translation | scale | zoom

- Different depth distortion scenarios have different impact on SR quality
- **Goal:** Create an algorithm that is robust to each of those distortions

Displacement-Compensated Super-Resolution



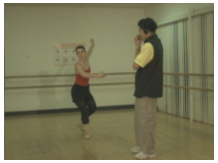
[Richter, Kaup, CSVT 2015]

Simulation Results

- **Translation:**
  - Shifting all depth entries 5 pixel positions to the top right.
- **Scaling:**
  - Limiting the 8 bit depth entries [0; 255] to [0; 127].
- **Zoom:**
  - Dropping 10% of rows and columns and resizing the cropped depth map via nearest neighbor interpolation.

Simulation Results

PSNR evaluation, $\downarrow 2$

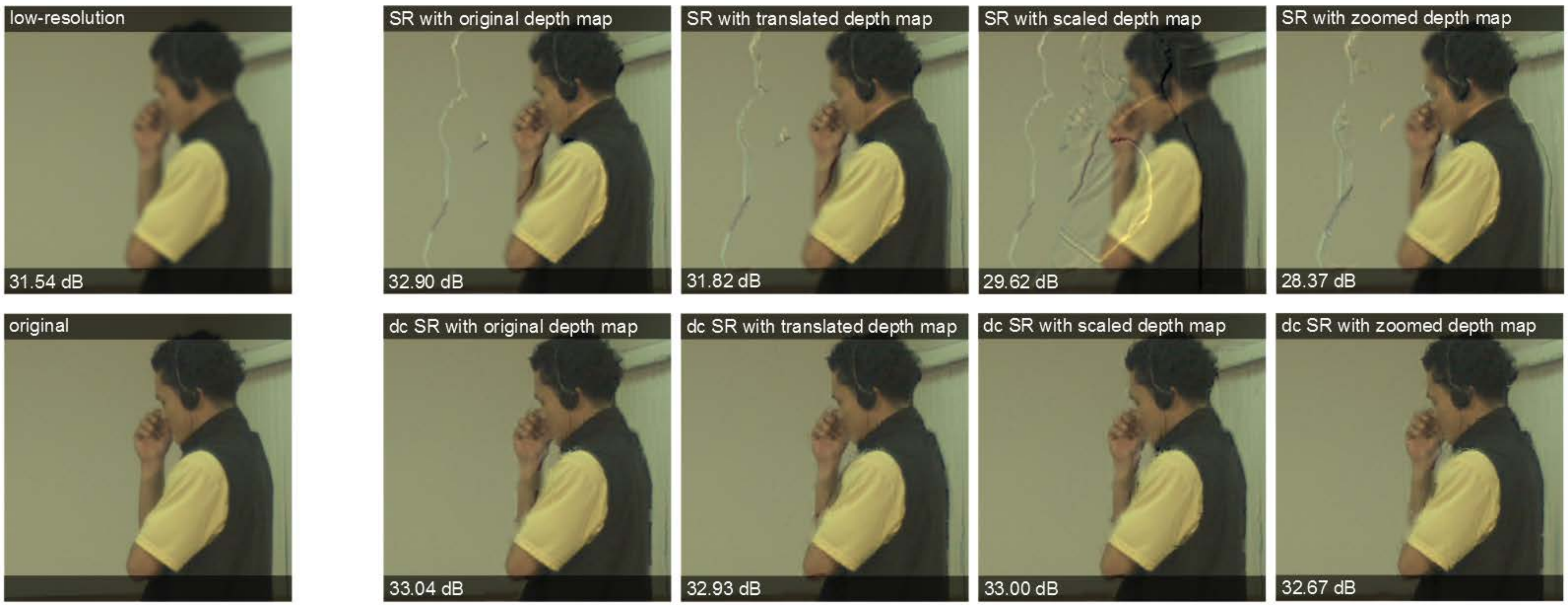



		Ballet	*Breakdancers*	*Cones*	*Teddy*	*Avg. gain*
	$l_l(u,v)$	36.97	**38.82**	33.10	33.71	
Original depth	$\hat{l}(u,v)$	36.68	37.38	34.63	35.09	
	$\hat{l}_{dc}(u,v)$	**38.01**	37.95	**34.71**	**35.43**	**0.58**
Translated depth	$\hat{l}(u,v)$	35.98	37.47	34.18	34.82	
	$\hat{l}_{dc}(u,v)$	**38.11**	38.04	**34.61**	**35.48**	**0.95**
Scaled depth	$\hat{l}(u,v)$	34.77	36.11	30.35	31.04	
	$\hat{l}_{dc}(u,v)$	**37.83**	38.01	**34.24**	**35.06**	**3.22**
Zoomed depth	$\hat{l}(u,v)$	34.16	36.86	32.23	33.16	
	$\hat{l}_{dc}(u,v)$	**37.80**	37.86	**34.09**	**35.07**	**2.10**

Simulation Results

Visual comparison: *ballet*

$l_l(u,v)$ $\hat{l}(u,v)$

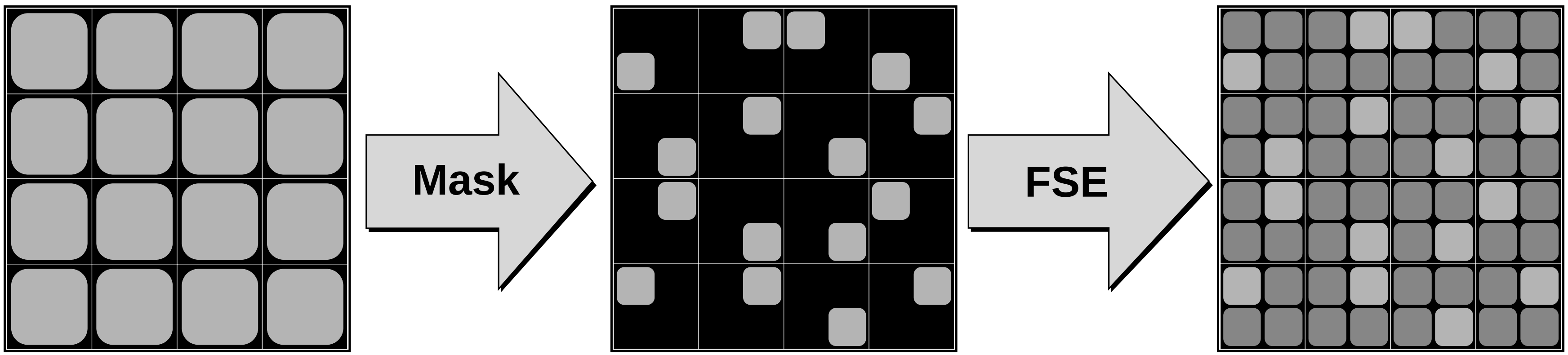


$l(u,v)$ $\hat{l}_{dc}(u,v)$

Overview

- Advancing compression efficiency
  - Spatiotemporal prediction
  - In-loop video denoising
  - Measurement and prediction of energy consumption
- Improving displaying quality
  - Scalable lossless compression
  - Multi-view video using super-resolution
  - **Random sampling and reconstruction**
- Conclusions and outlook

Example: ¼ Sampling



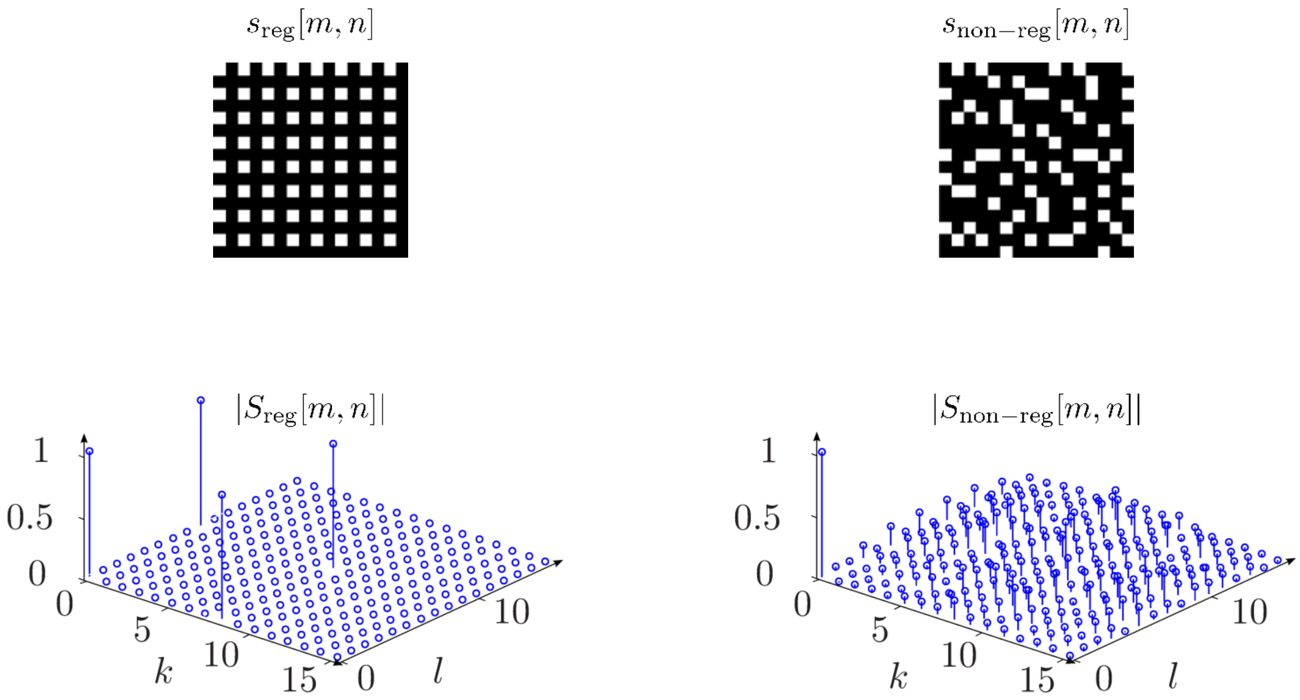
Low-resolution Sensor | Masked Sensor | High-resolution Image

Large pixel | Acquired pixel | Reconstructed pixel

[Schöberl, Seiler, Kaup, ICIP 2011]

Aliasing

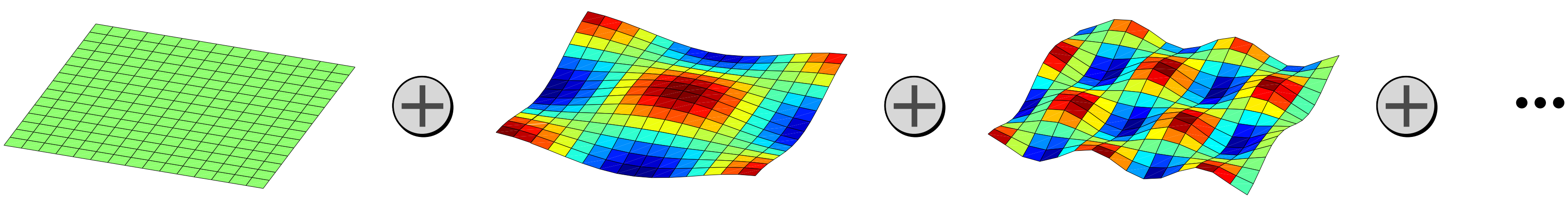
⇨ Regular versus non-regular sub-sampling



Frequency Selective Extrapolation

- ⇨ Sparse signal model generation as a weighted superposition of Fourier basis functions

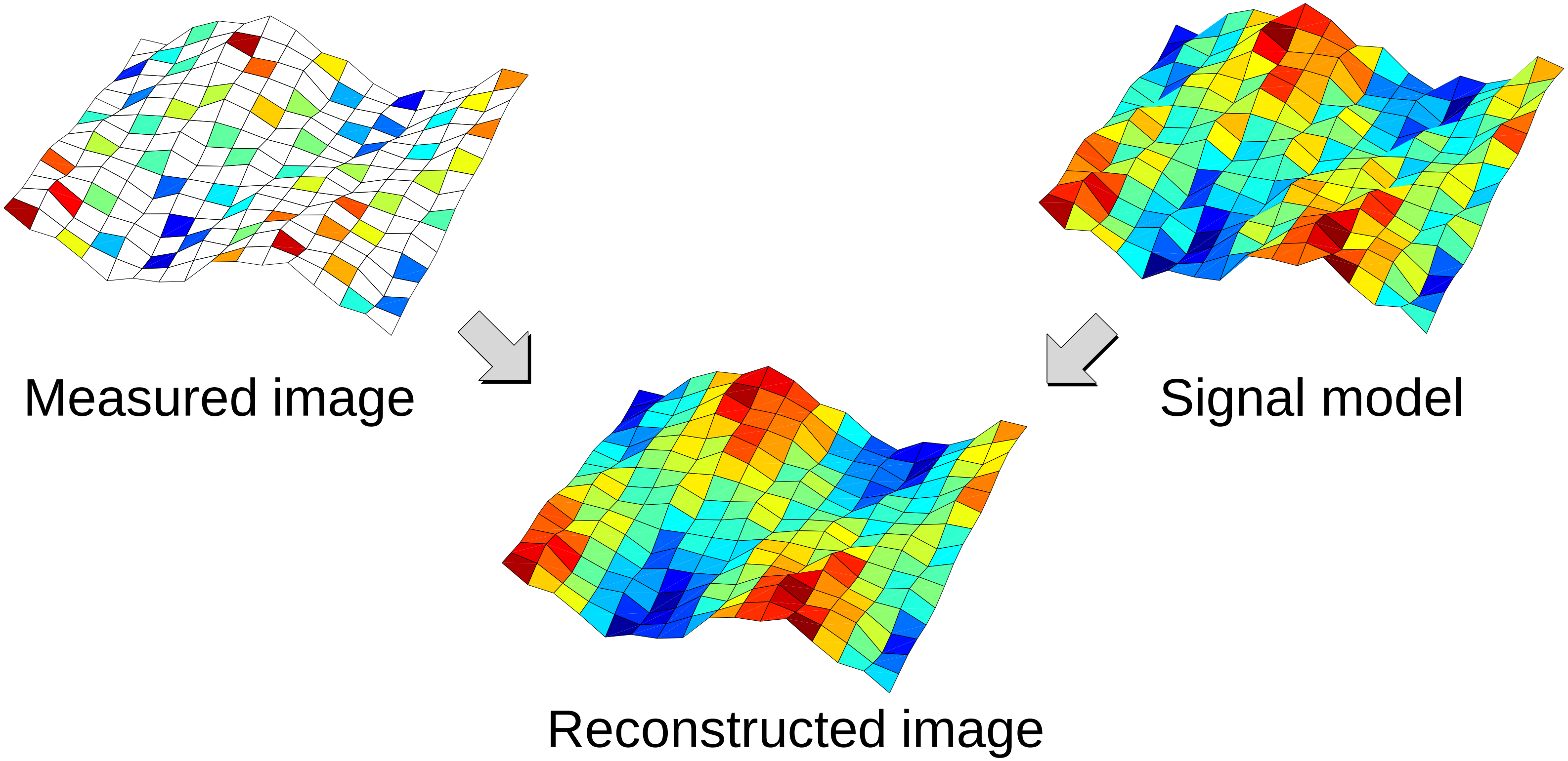
$$g[m,n] = \sum_{(k,l)\in\mathcal{K}} \hat{c}_{(k,l)}\varphi_{(k,l)}[m,n]$$



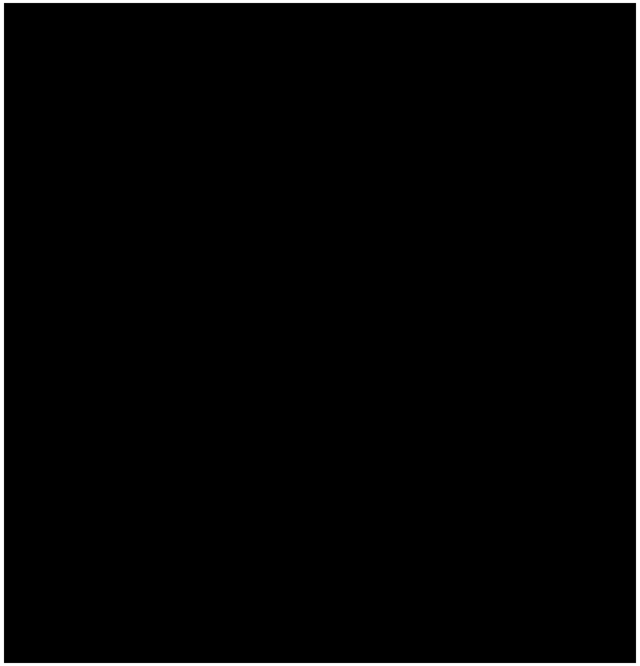
$\varphi_1[m,n]$ $\varphi_2[m,n]$ $\varphi_3[m,n]$

[Seiler, Kaup, SPL 2010]

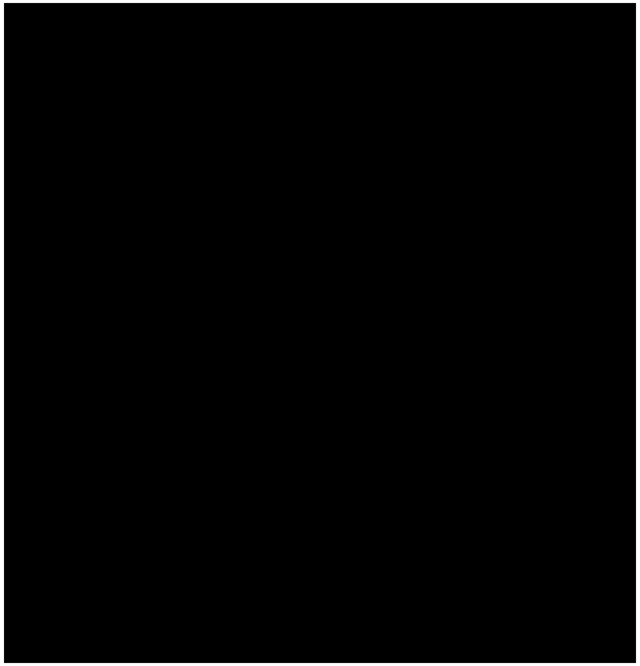
Frequency Selective Extrapolation



Reconstruction by Frequency Selective Extrapolation



Sampled image



Reconstructed image

Comparison

Low resolution image

Reconstructed image

Simulation Results on Image Data Base

Reconstruction algorithm	PSNR [dB] (KODAK)	PSNR [dB] (TECNICK)
Frequency Selective Extrapolation	**28.80**	**31.50**
Linear Interpolation	27.31	29.81
Steering Kernel Regression	27.55	30.30

[M. Jonscher, J. Seiler, T. Richter, M. Bätz, A. Kaup, ICIP 2014]

Summary and Conclusions

Future video communication systems will require more efficient compression and be more immersive

Efficient compression

- Video is a cube: ► Spatiotemporal prediction
- Noise might be significant: ► In-loop denoising
- Energy will play a role: ► Decoding energy measurement

Improved immersiveness

- Picture quality matters: ► Scalable lossless coding
- 3D is on the way: ► Super-resolution for multi-view
- Sampling revisited: ► Random pixel reconstruction

About the Future

"IT MAKES NO SENSE TO WORRY ABOUT THE FUTURE. BY THE TIME YOU GET THERE, IT'S THE PAST!"