Leibniz Universität Hannover

tnt

Extended Proceedings

4th Summer School on Video Compression and Processing (SVCP) 2018

4–6 July 2018

Hannover, Germany

https://svcp2018.tnt.uni-hannover.de

Contents

1 Program 3
2 Keynote 4
3 Corporate presentation 5
4 Sessions 6
5 Posters 9
6 Collection of presentation slides and posters 10

1 Program

Wednesday, 4 July 2018

09:00 – 13:00	Registration and reception
13:00 – 14:00	Lunch
14:00 – 14:15	Welcome remarks
14:15 – 16:15	Session 1
16:15 – 16:45	Coffee break
16:45 – 17:45	Keynote
18:30 – 22:00	BBQ & social activities

Thursday, 5 July 2018

09:00 – 10:30	Session 2
10:30 – 11:00	Coffee break
11:00 – 12:30	Session 3
13:00 – 14:00	Lunch
14:15 – 15:30	Posters
15:30 – 16:00	Coffee break
16:00 – 17:00	Session 4
17:00 – 18:00	Corporate presentation
18:30 – 22:00	Dinner & social activities

Friday, 6 July 2018

09:00 – 10:30	Session 5
10:30 – 11:00	Coffee break
11:00 – 12:30	Special session: research at TNT
12:30 – 12:45	Closing remarks
13:00 – 14:00	Lunch
14:00 – 16:00	*Meeting ITG Fachausschuss "Bildkommunikation und Bildverarbeitung"*

2 Keynote

De-hyping Neural Networks

BODO ROSENHAHN
LEIBNIZ UNIVERSITÄT HANNOVER, INSTITUT FÜR INFORMATIONSVERARBEITUNG

The progress in machine learning in the past decade, including examples such as IBM Watson and AlphaGo, strikes our society as ingenious and astounding, yet also frightening. Interestingly, the recent progress did not involve new conceptual breakthroughs. Instead, it was facilitated by further refinement of algorithms that already existed in the 70s and 80s. By implementing these refined algorithms on high-performance computers, and by training the corresponding statistical models with more data, impressive results are yielded. Media reports on machine learning algorithms that possess equal or super-human performance (regarding very specific tasks). Many research projects are driven by the interdisciplinary application of neural networks to advance the state-of-the-art.

In this talk I will give a brief introduction to the concepts behind neural networks and deep learning. I will de-mystify media news to ground neural networks as a useful and important, but also as a difficult tool. Machine learning research is not only about applying neural networks to something. Instead, it is also about understanding and nudging the data. The goal of machine learning research is to ensure that neural networks have a chance to learn something meaningful, thus providing a basis to solve tasks in signal processing, video coding and many other fields.

Bodo Rosenhahn studied Computer Science (minor subject Medicine) at the University of Kiel. He received the Dipl.-Inf. and Dr.-Ing. degrees from the University of Kiel in 1999 and 2003, respectively. From 10/2003 until 10/2005, he worked as postdoc at the University of Auckland (New Zealand), funded with a scholarship from the German Research Foundation (DFG). From 11/2005 to 08/2008 he worked as senior researcher at the Max Planck Institute for Informatics in Saarbrücken. Since 09/2008 he is Full Professor at Leibniz University Hannover, heading a group on automated image interpretation.

His works received several awards, including a DAGM Prize 2002, the Dr.-Ing. Siegfried Werth Prize 2003, the DAGM Main Prize 2005, the IVCNZ best student paper award, the DAGM Main Prize 2007, the Olympus Prize 2007, the ICPRAM best student paper award 2014, the CMC best student paper award 2014, the WACV Challenge Award 2015, the Günter Enderle Award (Eurographics) 2017 and the CVPR 2017 Multi-Object Tracking Challenge. In 2011, the European Commission awarded Bodo Rosenhahn with a EUR 1.43 million ERC Starting Grant and in 2013 with a POC Grant. He published more than 180 research papers, journal articles and book chapters, holds more than 10 patents and edited several books.

3 Corporate presentation

Real-time Video Conferencing

ALEXANDER GEHLERT
LOGMEIN GERMANY GMBH, DRESDEN

Online video conferencing solutions like GoToMeeting enable real-time communication and collaboration over the internet. A crucial building block for such applications is the video processing and coding chain, especially at the large-scale LogMeIn, as one of the world's top 10 SaaS companies, is operating on.

The presentation gives some details on how video processing and coding is done to enable the processing of multiple billions of video minutes per year.

4 Sessions

Session 1

slides included

3D Models in Motion Compensation

Hossein Golestani

RWTH Aachen University, Institut für Nachrichtentechnik

slides included

Rate-Distortion Theory for Affine Motion Compensation in Video Coding

Holger Meuel

Leibniz Universität Hannover, Institut für Informationsverarbeitung

Motion Adapted Three-Dimensional Frequency Selective Extrapolation

Andreas Spruck

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

Scalable Lossless Video Coding

Kristian Fischer

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

Session 2

Robust Super-Resolution for Mixed-Resolution Image Data Using Displacement-Compensated Projection

Johannes Bauer

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

Optimising Data for Exemplar-based Inpainting

Pinak Bheed

Saarland University, Mathematical Image Analysis Group

Restricted Boltzmann Machine Image Compression

Markus Küchhold

Technische Universität Berlin, Communication Systems Group

Session 3

Head Pose Estimation using Convolutional Neural Networks
Felix Kuhnke
Leibniz Universität Hannover, Institut für Informationsverarbeitung

Optimization Strategies for Non-Regular Sampling Masks
Simon Grosche
Friedrich-Alexander-Universität Erlangen-Nürnberg, Chair of Multimedia Communications and Signal Processing

slides included
MPEG-G: The Standard for Genomic Information Representation
Jan Voges
Leibniz Universität Hannover, Institut für Informationsverarbeitung

Session 4

slides included
A Comparison of JEM and AV1 with HEVC
Thorsten Laude
Leibniz Universität Hannover, Institut für Informationsverarbeitung

High-resolution and Immersive Video Quality
Alexander Raake
TU Ilmenau

Session 5

Geometry-Corrected Deblocking Filter for 360° Video Coding using Cube Representation
Johannes Sauer
RWTH Aachen University, Institut für Nachrichtentechnik

slides included
Multiple Feature-based Classifications Adaptive Loop Filter
Johannes Erfurt, W. Lim, H. Schwarz, D. Marpe, T. Wiegand
Fraunhofer Heinrich Hertz Institute

Sparse Coding based Frequency Adaptive Loop Filtering for Video Coding
Jens Schneider
RWTH Aachen University, Institut für Nachrichtentechnik

Special session: research at TNT

Robust Image Coding for Erasure Channels
Karsten Vogt
Leibniz Universität Hannover, Institut für Informationsverarbeitung

Fusion of Head and Full-Body Detectors for Multi-Object Tracking
Roberto Henschel
Leibniz Universität Hannover, Institut für Informationsverarbeitung

Coding of Electrical Stimulation Patterns for Binaural Signal Processing in Cochlear Implants
Reemt Hinrichs
Leibniz Universität Hannover, Institut für Informationsverarbeitung

5 Posters

(listed alphabetically by last name of first author)

poster included
Demonstration of Rapid Frequency Selective Reconstruction for Image Resolution Enhancement
Nils Genser, Jürgen Seiler, André Kaup
Friedrich-Alexander-Universität Erlangen-Nürnberg, Chair of Multimedia Communications and Signal Processing

Picture Boundary Handling for Versatile Video Coding
Gao Han
Huawei Technologies, Technische Universität München

CNNs for Video Intra Prediction
Maria Meyer
RWTH Aachen University, Institut für Nachrichtentechnik

poster included
Computer-assisted Optical Verification of Precast and Reinforced Concrete Structures
Marco Munderloh, Armel Dedjouong, Suraja K.P., Sascha Bahlau, Katharina Klemt-Albert, Jörn Ostermann
Leibniz Universität Hannover, Institut für Informationsverarbeitung

poster included
Video Transmission, an Overview of Video Compression and Communication Systems
Yasser Samayoa, Jörn Ostermann
Leibniz Universität Hannover, Institut für Informationsverarbeitung

Restricted Boltzmann Machine Image Compression
Maik Simon
Technische Universität Berlin, Communication Systems Group

6 Collection of presentation slides and posters

(listed alphabetically by last name of first author)

MULTIPLE FEATURE-BASED CLASSIFICATIONS ADAPTIVE LOOP FILTER (MCALF)

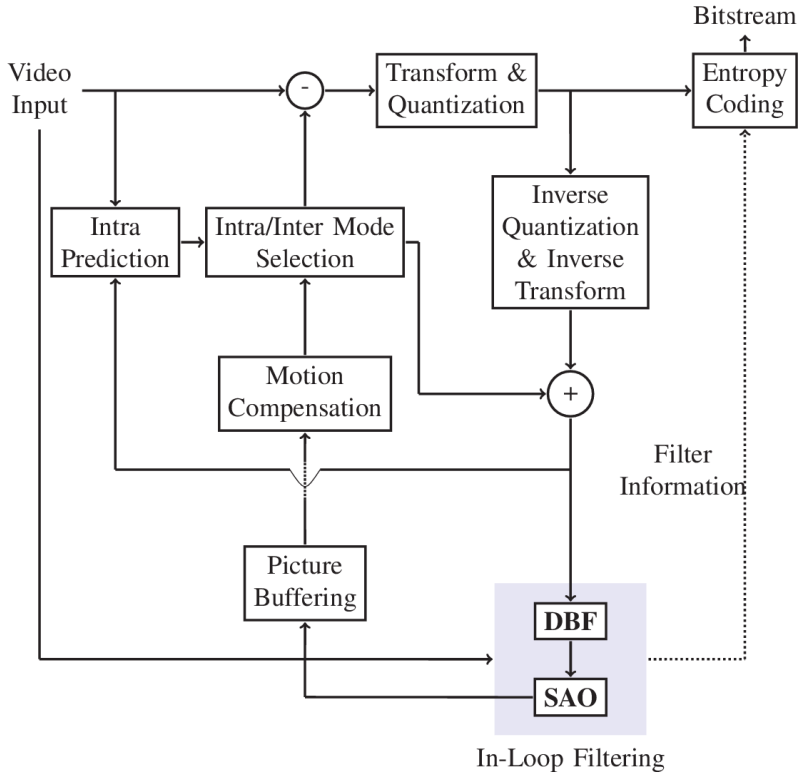
Johannes Erfurt, W. Lim, H. Schwarz, D. Marpe, T. Wiegand

CONTENT

- Introduction to In-Loop Filtering
- Adaptive Loop Filtering
- ALF and GALF
- Multiple Feature-based Classifications
  - Feature Descriptors
  - Concept of Confidence Level
- Simulation Results
- Conclusion

Introduction to In-Loop Filtering

- In-loop filtering is applied after reconstruction of coding blocks
- Filtered picture is stored in decoded picture buffer and may be used for prediction



HEVC Encoder Block Diagram

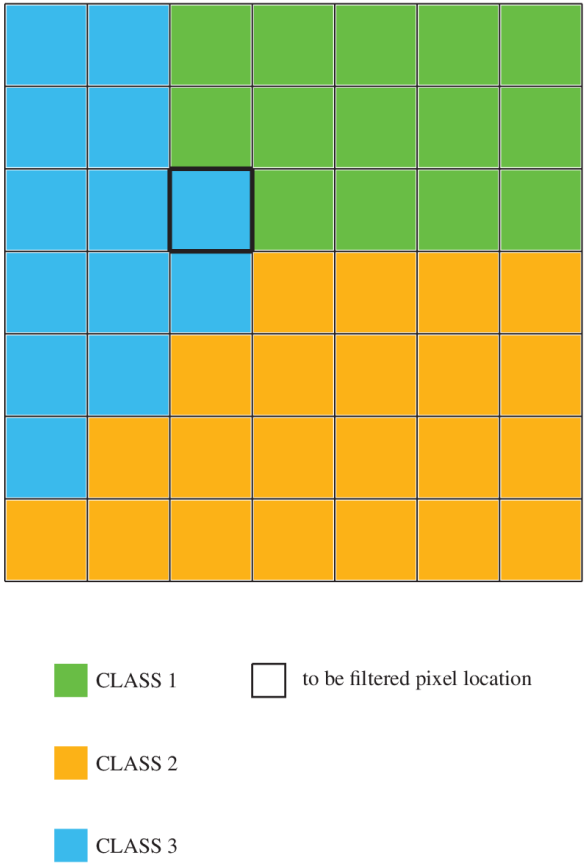
DBF = Deblockling Filter

SAO = Sample Adaptive Offset

Adaptive Loop Filtering

X = original samples, **Y** = reconstruced samples

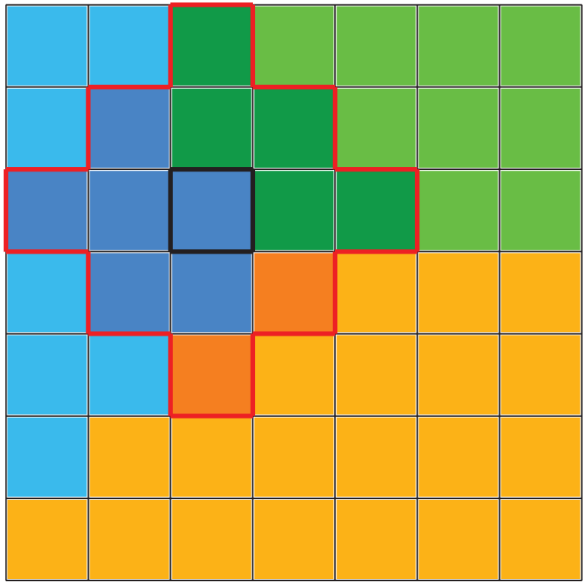
- Each pixel location is classified into one of L classes $\mathcal{C}_1, ..., \mathcal{C}_L$ based on local features
- Estimate multiple Wiener filters F_l for each $\mathcal{C}_l$, $l = 1...L$
- F_l minimizes mean square error (MSE) between X and $\tilde{X}$
- $\tilde{X} = \sum_{\ell=1}^{L} \chi_{\mathcal{C}_\ell} \cdot (Y * F_l)$ with $\chi_{\mathcal{C}_\ell}(i,j) = \begin{cases} 1, & \text{if } (i,j) \in \mathcal{C}_\ell \\ 0, & \text{otherwise} \end{cases}$



Adaptive Loop Filtering

X = original samples, **Y** = reconstruced samples

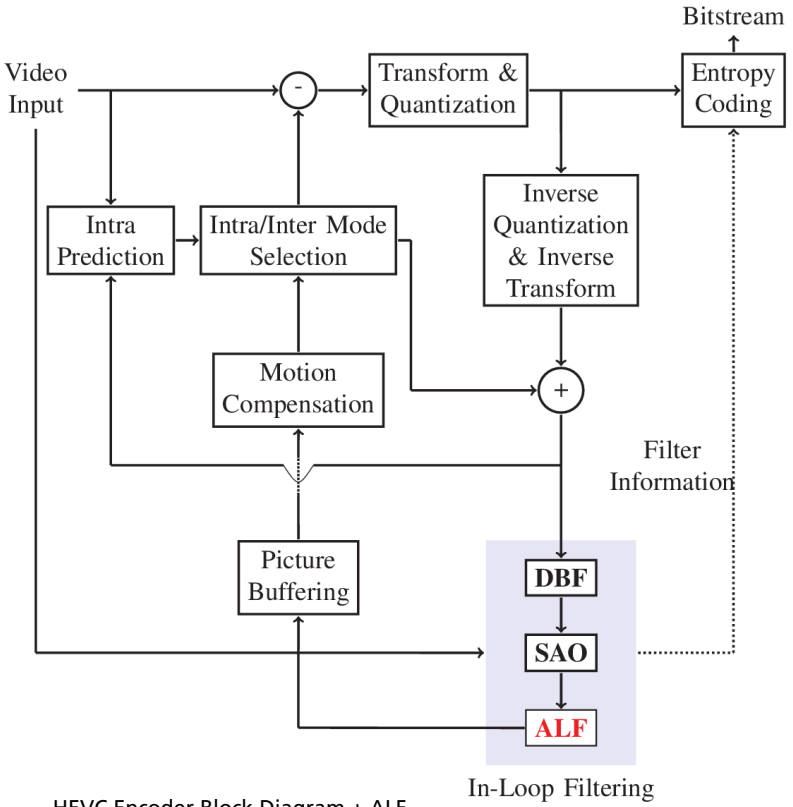
- Each pixel location is classified into one of L classes $\mathcal{C}_1, ..., \mathcal{C}_L$ based on local features
- Estimate multiple Wiener filters F_l for each $\mathcal{C}_l$, $l = 1...L$
- F_l minimizes mean square error (MSE) between X and $\tilde{X}$
- $\tilde{X} = \sum_{\ell=1}^{L} \chi_{\mathcal{C}_\ell} \cdot (Y * F_l)$ with $\chi_{\mathcal{C}_\ell}(i,j) = \begin{cases} 1, & \text{if } (i,j) \in \mathcal{C}_\ell \\ 0, & \text{otherwise} \end{cases}$



How do we perform classification into $\mathcal{C}_1, \ldots, \mathcal{C}_L$?

ALF and GALF

- ALF was proposed for HEVC standard and further developed resulting in GALF
- Certain coding tools make ALF/GALF very efficient:
  - Classification including directional gradients
  - Adaptively chosen filter support for each frame (5x5, 7x7, 9x9)
  - Block-wise on/off-flag
  - Temporal prediction: Use previously coded filter coefficients
  - Class merging



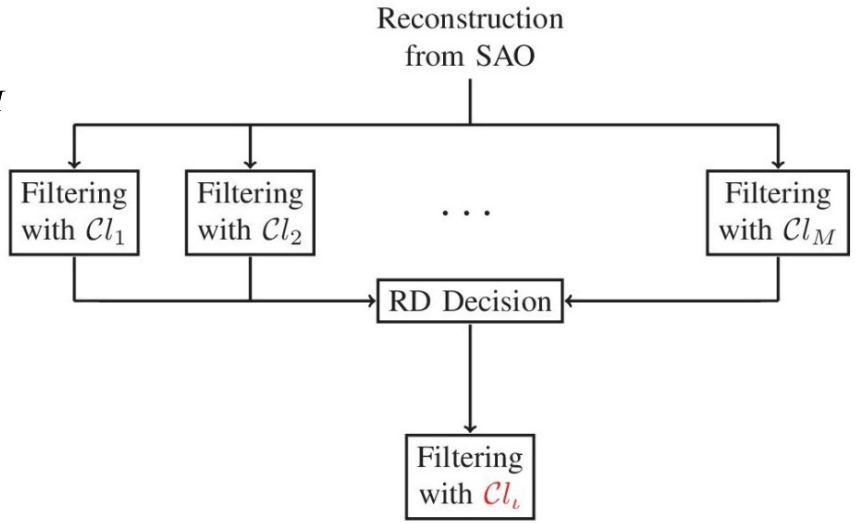
HEVC Encoder Block Diagram + ALF

Multiple Feature-based Classifications

- MCALF: Multiple Feature-based Classifications ALF
- Test at encoder side M classifiers $\mathcal{C}l_1, ..., \mathcal{C}l_M$
- Each classifier has a certain feature descriptor D to group each pixel location into classes

$$\mathcal{C}_\ell = \{(i,j) \in I : D(i,j) = \ell\} \quad \text{for } \ell = 1, \ldots, K$$

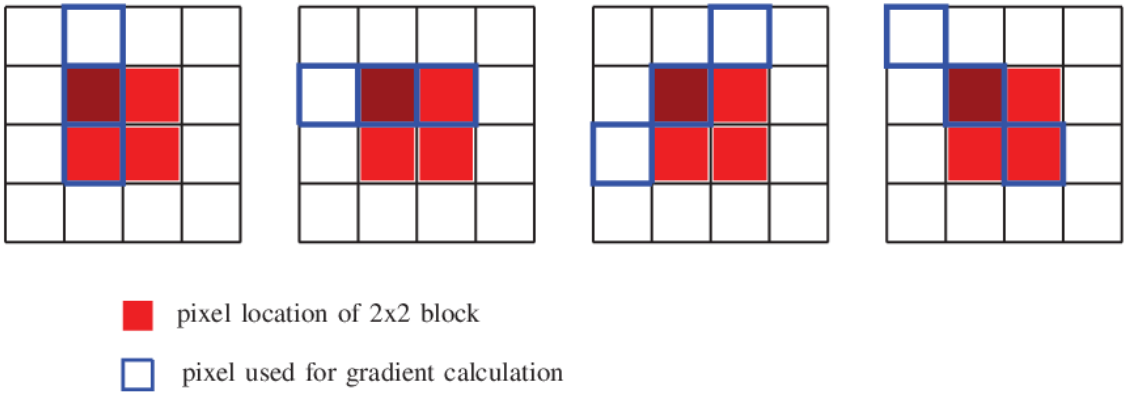
- Classifier with best RD performance chosen
- Filter index and possible filter information are signaled



Feature Descriptors

- Each classifier is descripted through different feature descriptors D
- Laplacian feature descriptor D_L : includes computation of directional gradients for squared blocks, e.g. in vertical direction:

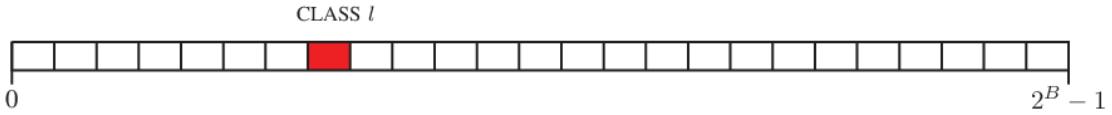
$$g(i,j) = |2 \cdot Y(i,j) - Y(i-1,j) - Y(i+1,j)|$$



Feature Descriptors

- Pixel-based feature descriptor D_P:

$$D_P(i,j) = \left\lfloor \frac{(K-1)}{2^B} Y(i,j) \right\rfloor \quad \text{for bit depth } B \text{ and } K \text{ classes}$$

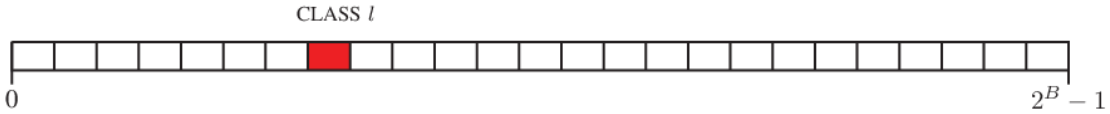


$$Y(i,j) \in [0, 2^B - 1]$$

Feature Descriptors

- Pixel-based feature descriptor D_P:

$$D_P(i,j) = \left\lfloor \frac{(K-1)}{2^B} Y(i,j) \right\rfloor \quad \text{for bit depth } B \text{ and } K \text{ classes}$$



$$Y(i,j) \in [0, 2^B - 1]$$

- Ranking-based feature descriptor D_R:

$$D_R(i,j) = \#\{(k_1,k_2) : Y(i,j) \geq Y(k_1,k_2) \text{for } |k_1 - i| \leq l, |k_2 - j| \leq h\} + 1$$

Feature Descriptors

- Product of two feature descriptors $D_1 : I \rightarrow \{1, \ldots, K_1\}$ and $D_2 : I \rightarrow \{1, \ldots, K_2\}$

$$D(i,j) = (D_1(i,j), D_2(i,j)) \in \{1, ..., K_1\} \times \{1, ..., K_2\}$$

Concept of Confidence Level

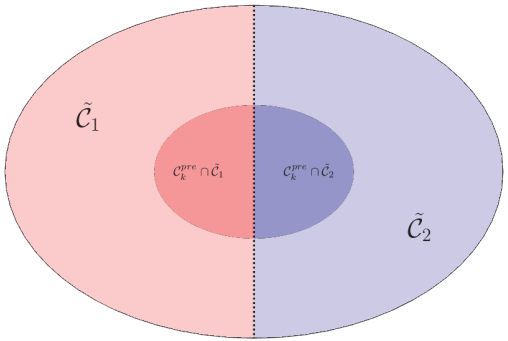
- Optimal classes for $L = 2$:

$$\tilde{\mathcal{C}}_1 = \{(i,j) \in I : Y(i,j) \leq \mathbf{X}(i,j)\},$$
$$\tilde{\mathcal{C}}_2 = \{(i,j) \in I : Y(i,j) > \mathbf{X}(i,j)\}$$

Concept of Confidence Level

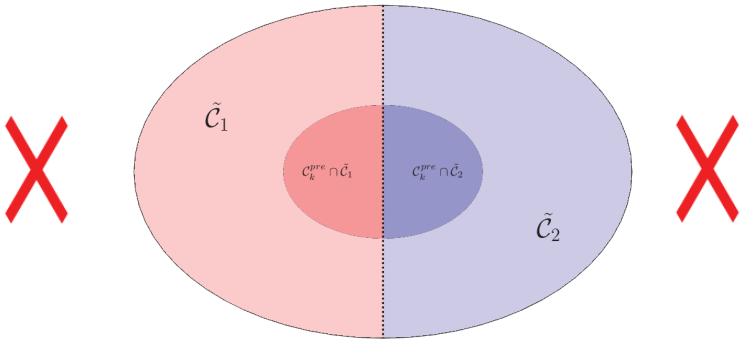
- Optimal classes for $L = 2$: $\tilde{\mathcal{C}}_1 = \{(i,j) \in I : Y(i,j) \leq \mathbf{X}(i,j)\},$

  $\tilde{\mathcal{C}}_2 = \{(i,j) \in I : Y(i,j) > \mathbf{X}(i,j)\}$
- X not available at decoder: Approximate $\tilde{\mathcal{C}}_1$ and $\tilde{\mathcal{C}}_2$
- Use certain feature descriptor D to receive K pre-classes $\mathcal{C}_1^{pre}, \ldots, \mathcal{C}_K^{pre}$



Concept of Confidence Level

- Optimal classes for $L = 2$: $\tilde{\mathcal{C}}_1 = \{(i,j) \in I : Y(i,j) \leq \mathbf{X}(i,j)\}$,
  $\tilde{\mathcal{C}}_2 = \{(i,j) \in I : Y(i,j) > \mathbf{X}(i,j)\}$
- X not available at decoder: Approximate $\tilde{\mathcal{C}}_1$ and $\tilde{\mathcal{C}}_2$
- Use certain feature descriptor D to receive K pre-classes $\mathcal{C}_1^{pre}, \ldots, \mathcal{C}_K^{pre}$

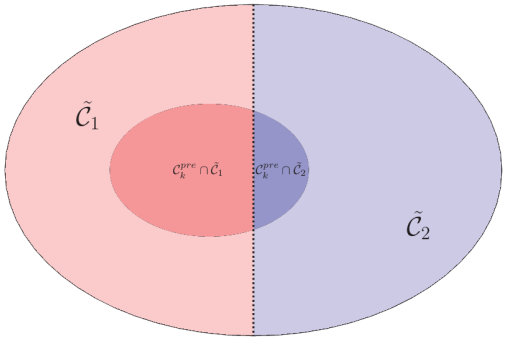


Inner ellipse should be mostly contained in left or right half

Concept of Confidence Level

- Optimal classes for $L = 2$: $\tilde{\mathcal{C}}_1 = \{(i,j) \in I : Y(i,j) \leq \mathbf{X}(i,j)\},$

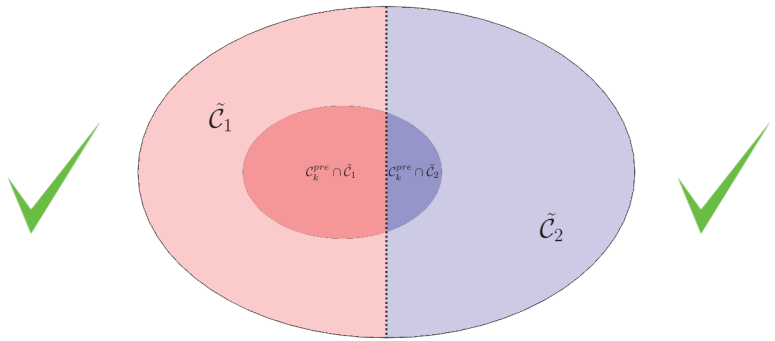
  $\tilde{\mathcal{C}}_2 = \{(i,j) \in I : Y(i,j) > \mathbf{X}(i,j)\}$
- X not available at decoder: Approximate $\tilde{\mathcal{C}}_1$ and $\tilde{\mathcal{C}}_2$
- Use certain feature descriptor D to receive K pre-classes $\mathcal{C}_1^{pre}, \ldots, \mathcal{C}_K^{pre}$



Concept of Confidence Level

- Optimal classes for $L = 2$: $\tilde{\mathcal{C}}_1 = \{(i,j) \in I : Y(i,j) \leq \mathbf{X}(i,j)\},$

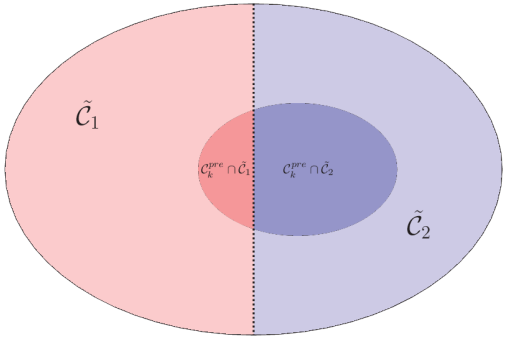
  $\tilde{\mathcal{C}}_2 = \{(i,j) \in I : Y(i,j) > \mathbf{X}(i,j)\}$
- X not available at decoder: Approximate $\tilde{\mathcal{C}}_1$ and $\tilde{\mathcal{C}}_2$
- Use certain feature descriptor D to receive K pre-classes $\mathcal{C}_1^{pre}, \ldots, \mathcal{C}_K^{pre}$



Concept of Confidence Level

- Optimal classes for $L = 2$: $\tilde{\mathcal{C}}_1 = \{(i,j) \in I : Y(i,j) \leq \mathbf{X}(i,j)\},$

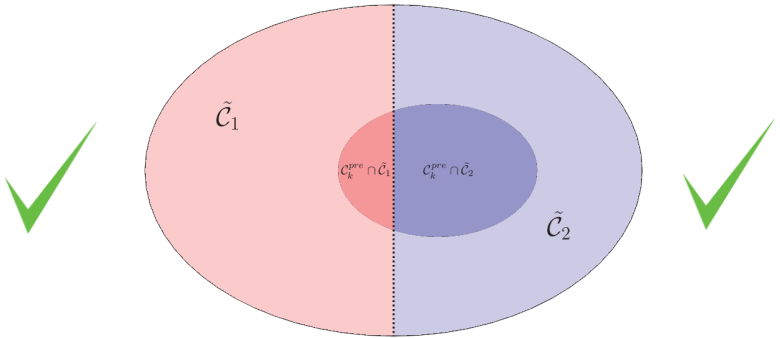
  $\tilde{\mathcal{C}}_2 = \{(i,j) \in I : Y(i,j) > \mathbf{X}(i,j)\}$
- X not available at decoder: Approximate $\tilde{\mathcal{C}}_1$ and $\tilde{\mathcal{C}}_2$
- Use certain feature descriptor D to receive K pre-classes $\mathcal{C}_1^{pre}, \ldots, \mathcal{C}_K^{pre}$



Concept of Confidence Level

- Optimal classes for $L = 2$: $\tilde{\mathcal{C}}_1 = \{(i,j) \in I : Y(i,j) \leq \mathbf{X}(i,j)\},$

  $\tilde{\mathcal{C}}_2 = \{(i,j) \in I : Y(i,j) > \mathbf{X}(i,j)\}$
- X not available at decoder: Approximate $\tilde{\mathcal{C}}_1$ and $\tilde{\mathcal{C}}_2$
- Use certain feature descriptor D to receive K pre-classes $\mathcal{C}_1^{pre}, \ldots, \mathcal{C}_K^{pre}$



Concept of Confidence Level

- Calculate for each class $\mathcal{C}_k^{pre}$ confidence level $p_{k,1}$ and $p_{k,2}$ of D

$$p_{k,1} = \frac{\#(\mathcal{C}_k^{pre} \cap \tilde{\mathcal{C}}_1)}{\#(\mathcal{C}_k^{pre})}, \quad p_{k,2} = \frac{\#(\mathcal{C}_k^{pre} \cap \tilde{\mathcal{C}}_2)}{\#(\mathcal{C}_k^{pre})}$$

Concept of Confidence Level

- Calculate for each class $\mathcal{C}_k^{pre}$ confidence level $p_{k,1}$ and $p_{k,2}$ of D

$$p_{k,1} = \frac{\#(\mathcal{C}_k^{pre} \cap \tilde{\mathcal{C}}_1)}{\#(\mathcal{C}_k^{pre})}, \quad p_{k,2} = \frac{\#(\mathcal{C}_k^{pre} \cap \tilde{\mathcal{C}}_2)}{\#(\mathcal{C}_k^{pre})}$$

- For $p \in (1/2, 1)$ define map $P_D : \{1, \ldots, K\} \longrightarrow \{0, 1, 2\}$ with $P_D(k) = \begin{cases} 1 & p_{k,1} > p \\ 2 & p_{k,2} > p \\ 0 & \text{otherwise.} \end{cases}$

Concept of Confidence Level

- Calculate for each class $\mathcal{C}_k^{pre}$ confidence level $p_{k,1}$ and $p_{k,2}$ of D

$$p_{k,1} = \frac{\#(\mathcal{C}_k^{pre} \cap \tilde{\mathcal{C}}_1)}{\#(\mathcal{C}_k^{pre})}, \quad p_{k,2} = \frac{\#(\mathcal{C}_k^{pre} \cap \tilde{\mathcal{C}}_2)}{\#(\mathcal{C}_k^{pre})}$$

- For $p \in (1/2, 1)$ define map $P_D : \{1, \ldots, K\} \longrightarrow \{0, 1, 2\}$ with $P_D(k) = \begin{cases} 1 & p_{k,1} > p \\ 2 & p_{k,2} > p \\ 0 & \text{otherwise.} \end{cases}$

Concept of Confidence Level

- Calculate for each class $\mathcal{C}_k^{pre}$ confidence level $p_{k,1}$ and $p_{k,2}$ of D

$$p_{k,1} = \frac{\#(\mathcal{C}_k^{pre} \cap \tilde{\mathcal{C}}_1)}{\#(\mathcal{C}_k^{pre})}, \quad p_{k,2} = \frac{\#(\mathcal{C}_k^{pre} \cap \tilde{\mathcal{C}}_2)}{\#(\mathcal{C}_k^{pre})}$$

- For $p \in (1/2, 1)$ define map $P_D : \{1, \ldots, K\} \longrightarrow \{0, 1, 2\}$ with $P_D(k) = \begin{cases} 1 & p_{k,1} > p \\ 2 & p_{k,2} > p \\ 0 & \text{otherwise.} \end{cases}$

- $\mathcal{C}_\ell^e = \{(i,j) \in I : P_D(D(i,j)) = \ell\}$ for $\ell = 1, 2$ high confidence classes

Concept of Confidence Level

- Calculate for each class $\mathcal{C}_k^{pre}$ confidence level $p_{k,1}$ and $p_{k,2}$ of D

$$p_{k,1} = \frac{\#(\mathcal{C}_k^{pre} \cap \tilde{\mathcal{C}}_1)}{\#(\mathcal{C}_k^{pre})}, \quad p_{k,2} = \frac{\#(\mathcal{C}_k^{pre} \cap \tilde{\mathcal{C}}_2)}{\#(\mathcal{C}_k^{pre})}$$

- For $p \in (1/2, 1)$ define map $P_D : \{1, \ldots, K\} \longrightarrow \{0, 1, 2\}$ with $P_D(k) = \begin{cases} 1 & p_{k,1} > p \\ 2 & p_{k,2} > p \\ 0 & \text{otherwise.} \end{cases}$

- $\mathcal{C}_\ell^e = \{(i,j) \in I : P_D(D(i,j)) = \ell\}$ for $\ell = 1, 2$ high confidence classes

Concept of Confidence Level

- Calculate for each class $\mathcal{C}_k^{pre}$ confidence level $p_{k,1}$ and $p_{k,2}$ of D

$$p_{k,1} = \frac{\#(\mathcal{C}_k^{pre} \cap \tilde{\mathcal{C}}_1)}{\#(\mathcal{C}_k^{pre})}, \quad p_{k,2} = \frac{\#(\mathcal{C}_k^{pre} \cap \tilde{\mathcal{C}}_2)}{\#(\mathcal{C}_k^{pre})}$$

- For $p \in (1/2, 1)$ define map $P_D : \{1, \ldots, K\} \longrightarrow \{0, 1, 2\}$ with $P_D(k) = \begin{cases} 1 & p_{k,1} > p \\ 2 & p_{k,2} > p \\ 0 & \text{otherwise.} \end{cases}$

- $\mathcal{C}_\ell^e = \{(i,j) \in I : P_D(D(i,j)) = \ell\}$ for $\ell = 1, 2$ high confidence classes

- $\mathcal{C}_\ell = \{(i,j) \notin \mathcal{C}_1^e \cup \mathcal{C}_2^e : \tilde{D}(i,j) = \ell\}$ for $\ell = 1, \ldots, \tilde{K}$ classes for remaining pixel locations

Concept of Confidence Level

- Calculate for each class $\mathcal{C}_k^{pre}$ confidence level $p_{k,1}$ and $p_{k,2}$ of D

$$p_{k,1} = \frac{\#(\mathcal{C}_k^{pre} \cap \tilde{\mathcal{C}}_1)}{\#(\mathcal{C}_k^{pre})}, \quad p_{k,2} = \frac{\#(\mathcal{C}_k^{pre} \cap \tilde{\mathcal{C}}_2)}{\#(\mathcal{C}_k^{pre})}$$

- For $p \in (1/2, 1)$ define map $P_D : \{1, \ldots, K\} \longrightarrow \{0, 1, 2\}$ with $P_D(k) = \begin{cases} 1 & p_{k,1} > p \\ 2 & p_{k,2} > p \\ 0 & \text{otherwise.} \end{cases}$

- $\mathcal{C}_\ell^e = \{(i,j) \in I : P_D(D(i,j)) = \ell\}$ for $\ell = 1, 2$ high confidence classes

- $\mathcal{C}_\ell = \{(i,j) \notin \mathcal{C}_1^e \cup \mathcal{C}_2^e : \tilde{D}(i,j) = \ell\}$ for $\ell = 1, \ldots, \tilde{K}$ classes for remaining pixel locations

- This gives $\tilde{K} + 2$ classes $\mathcal{C}_1^e, \mathcal{C}_2^e, \mathcal{C}_1, \ldots, \mathcal{C}_{\tilde{K}}$

Simulation Results

- Test conditions:
  - JEM-7.0 with QP points: 27, 32, 37, 42
  - Random Access Main 10 (RA)
- 5 Classifiers for MCALF:
  - $D_P(K = 27)$
  - Product of D_R and $D_P(K = 27)$
  - $D_L\ (K = 25)$
  - D_P for $\mathcal{C}_1^e, \mathcal{C}_2^e\ (p = 0.63, K = 20)$ and $D_L\ (\tilde{K} = 25)$
  - D_R for $\mathcal{C}_1^e, \mathcal{C}_2^e\ (p = 0.63, K = 9)$ and $D_L\ (\tilde{K} = 25)$

Test Sequence	BD Rate (Y) RA
BQTerrace 1920 x 1080	**-1.34%**
MarketPlace 1920 x 1080	**-0.90%**
Rollercoaster 3840 x 2160	**-2.33%**
Encoder run-time	**107%**
Decoder run-time	**100%**

Coding gains of MCALF (5 classifiers) with reference GALF (1 classifier D_L)

Conclusion

- Performance of adaptive loop filter highly depends on classification
- Multiple classifications can better adapt to local features in video sequence
- Classification is performed through feature descriptors such as D_L, D_P or D_R and classification with confidence level
- We can get more than **2%** coding gain on top of GALF with **no increase of decoder runtime** and only small increase of encoder runtime

Thank you!

More Results

Test Sequence	BD Rate (Y) RA
BQTerrace 1920 x 1080	**-5.85%**
MarketPlace 1920 x 1080	**-3.35%**
Rollercoaster 3840 x 2160	**-6.31%**

Coding gains of MCALF (5 classifiers) with reference JEM-7.0 - GALF

Demonstration of Rapid Frequency Selective Reconstruction for Image Resolution Enhancement

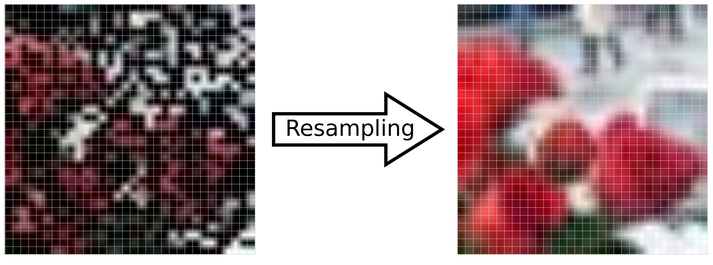
Nils Genser, Jürgen Seiler, and André Kaup

Multimedia Communications and Signal Processing
Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Cauerstr. 7, 91058 Erlangen, Germany

1. Motivation

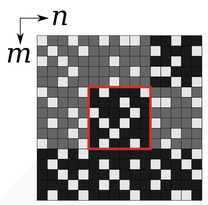
- Many scenarios, where amplitude information of an image is not available on a regular, rectangular grid
  - Optical Cluster Eye
  - Micro-Optical Artificial Compound Eyes
  - Reducing visible influence of aliasing
  - Super-Resolution techniques [1]

- For further processing or displaying, a resampling to the full regular grid is required



2. Frequency Selective Reconstruction (FSR) [2]

- Iterative sparse model generation by superimposing weighted Fourier basis functions: $g[m,n] = \sum_{k \in \mathcal{K}} \hat{c}_k \varphi_k[m,n]$



- Spatial weighting function

$$w[m,n] = \begin{cases} \hat{\rho}\sqrt{\left(m-\frac{M-1}{2}\right)^2+\left(n-\frac{N-1}{2}\right)^2}, & (m,n) \in \mathcal{A} \\ \delta\hat{\rho}\sqrt{\left(m-\frac{M-1}{2}\right)^2+\left(n-\frac{N-1}{2}\right)^2}, & (m,n) \in \mathcal{R} \\ 0, & (m,n) \in \mathcal{B}_i \cup \mathcal{B}_o \end{cases}$$

support area
reconstructed area
loss area

- Fixed frequency prior to favor low frequencies

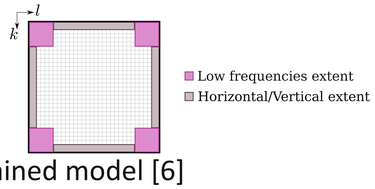
$$w_f[k,l] = \left(1 - \sqrt{2}\sqrt{\frac{\tilde{k}^2}{M^2} + \frac{\tilde{l}^2}{N^2}}\right)^2$$

→ Approximation of the Optical Transfer Function
→ Improved reconstruction quality

3. Demonstration

Algorithmic Enhancements

- YUV color space dependent reconstruction [3]
  - Linear interpolation for chroma channels
  - FSR for luminance channel, only
- Restrict number and set of basis functions [4], [5]
- Real-valued, spectrally constrained model [6]
- Texture dependent block partitioning and parameter estimation [7]



Software Optimizations

- C++ implementation using FFTW3 library
- Parallelization using OpenMP pragmas to make use of all processor cores
- Vectorization
  - Manual preparation of loops
    → Avoid control flows
  - GCC's autovectorization features and OpenMP's SIMD pragmas

Test Systems

Notebook	
CPU	i7-6700HQ
Speed	2.60 GHz
Cores	4
RAM	8 GB
Xeon	
CPU	2 × Xeon E5-2630v4
Speed	2.20 GHz
Cores	2 × 10
RAM	32 GB

4. Conclusion

Reconstruction quality in dB PSNR for different subsampling densities (TECNICK)

	5%	10%	20%	30%	40%	50%	60%	70%	80%	90%	95%
FSR	**27.03**	**29.69**	**32.83**	**34.94**	**36.66**	**38.23**	**39.79**	**41.50**	**43.59**	**46.75**	**49.55**
SKR	16.69	28.11	31.81	33.23	34.29	35.30	36.41	37.75	39.55	42.55	45.42
CLS	24.82	26.94	29.85	32.06	34.03	35.92	37.87	40.01	42.58	46.28	49.37
SI	24.30	27.17	30.87	33.31	35.33	37.16	38.98	40.93	43.23	46.60	49.46

SKR: Steering Kernel Regression, CLS: Constrained Split Augmented Lagrangian Shrinkage Algorithm, SI: Sparse Wavelet Inpainting

- FSR well suited for image reconstruction problems
- Visually noticeable gains of several dB PSNR compared to other state-of-the-art algorithms
- Acceleration of the reconstruction process by
  → Reducing the number of basis functions
  → Lowering the required set of basis functions
  → Texture dependent block partitioning
  → Partitioning based parameter estimation
  → Setting up a real-valued, spectrally constrained model
  → Software optimizations

Accelerated reconstruction process:
→ 15 fps using test system Notebook
→ 25 fps for test system Xeon

Reference implemenation:

References

[1] M. Schöberl, J. Seiler, S. Foessel, and A. Kaup, "Increasing imaging resolution by covering your sensor," in *Proc. IEEE International Conference on Image Processing*, Brussels, Belgium, Sep. 2011, pp. 1937–1940.

[2] J. Seiler, M. Jonscher, M. Schöberl, and A. Kaup, "Resampling images to a regular grid from a non-regular subset of pixel positions using frequency selective reconstruction," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4540–4555, Nov. 2015.

[3] M. Bätz, A. Eichenseer, M. Jonscher, J. Seiler, and A. Kaup, "Accelerated hybrid image reconstruction for non-regular sampling color sensors," in *Proc. IEEE International Conference on Visual Communications and Image Processing*, Dec. 2014, pp. 217–220.

[4] N. Genser, J. Seiler, and A. Kaup, "Local statistics estimation for rapid frequency selective extrapolation," in *Proc. 2017 International Conference on Systems, Signals and Image Processing*, May 2017, pp. 1–5.

[5] N. Genser, J. Seiler, M. Jonscher, and A. Kaup, "Demonstration of rapid frequency selective reconstruction for image resolution enhancement," in *Proc. International Conference on Image Processing*, Sep. 2017. [Online]. Available: http://sigport.org/2108

[6] N. Genser, J. Seiler, F. Schilling, and A. Kaup, "Signal and loss geometry aware frequency selective extrapolation for error concealment," in *Proc. Picture Coding Symposium*, 2018.

[7] N. Genser, J. Seiler, and A. Kaup, "Spectral constrained frequency selective extrapolation for rapid image error concealment," in *Proc. 25th International Conference on Systems, Signals and Image Processing*, 2018.

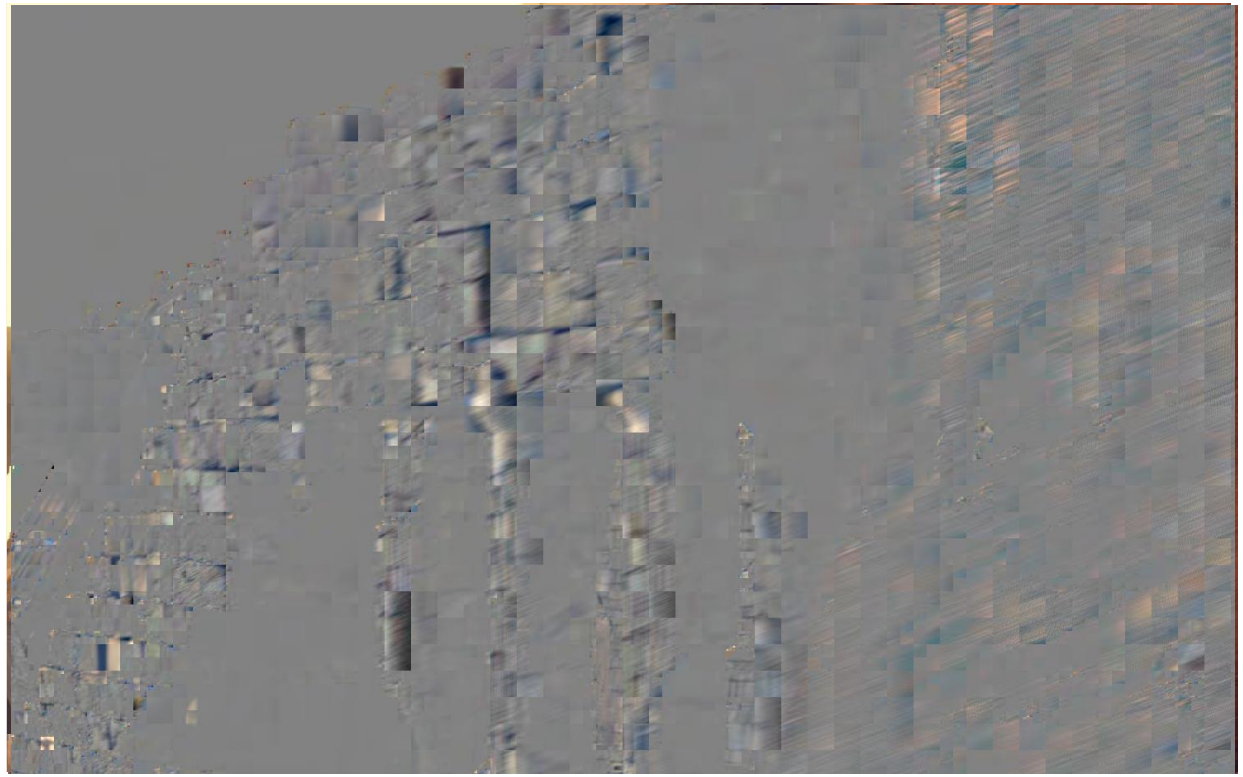
3D Models in Motion Compensation

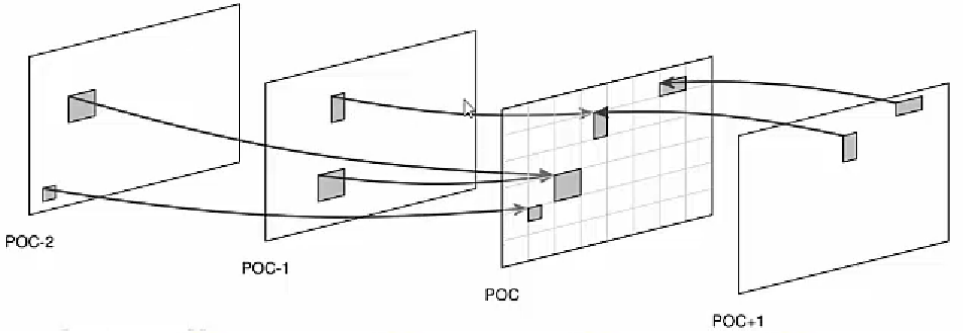
Hossein Golestani
Institut für Nachrichtentechnik
RWTH Aachen University

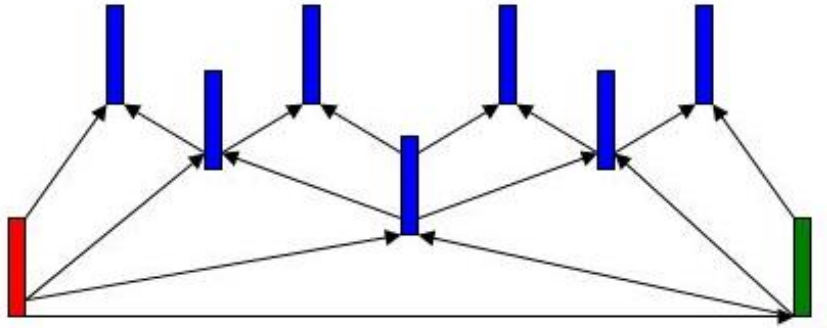
Outline

- Introduction to the main idea
- Structure from Motion (SfM)
- Multi-View Reconstruction (MVR)
- Virtual view synthesis for 2D frame prediction
- Coding results
- Conclusion

Introduction to the main idea







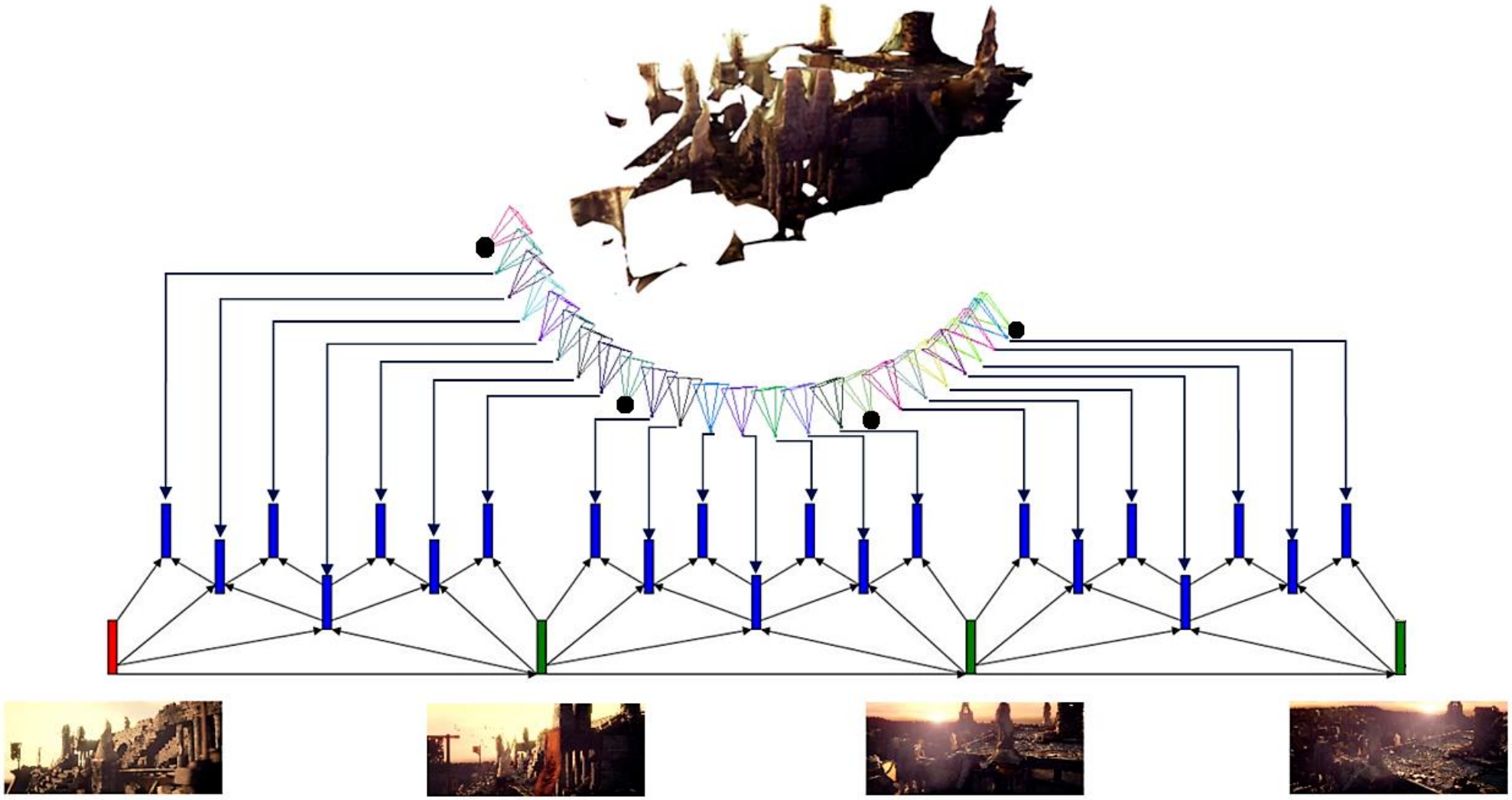
Introduction to the main idea

- Target Sequences (moving camera)

Introduction to the main idea

- 3D model based frame prediction
  - **B-Frame Prediction**
  - Key-Frame Prediction

Introduction to the main idea



Introduction to the main idea

- 3D model based frame prediction
  - B-Frame Prediction
  - **Key-Frame Prediction**

Introduction to the main idea

Outline

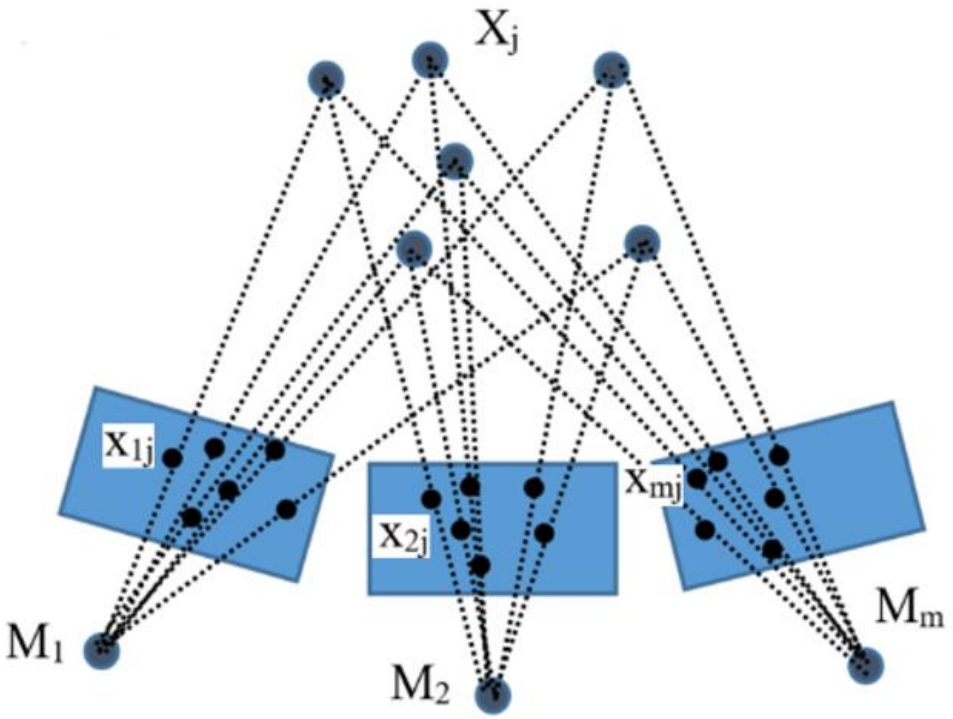
- Introduction to the main idea
- **Structure from Motion (SfM)**
- Multi-View Reconstruction (MVR)
- Virtual view synthesis for 2D frame prediction
- Coding results
- Conclusion

Structure from Motion (SfM)

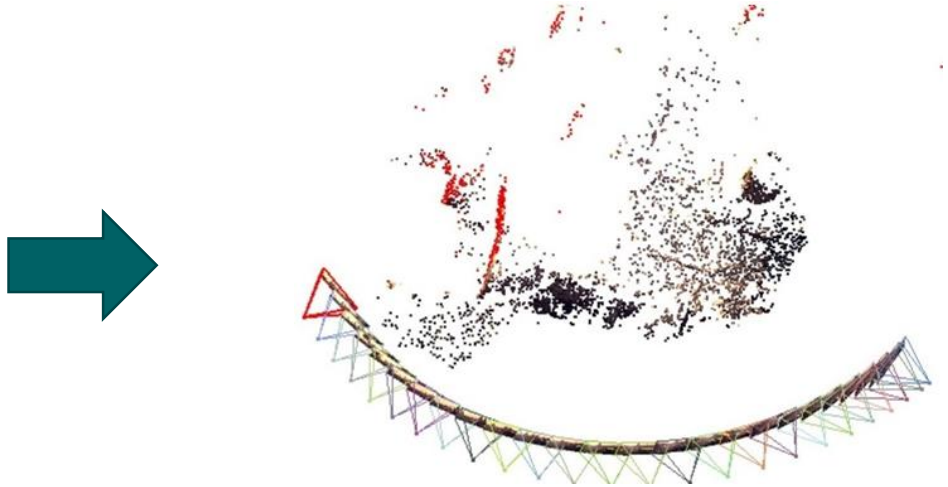
- Structure from Motion (SfM) [1]

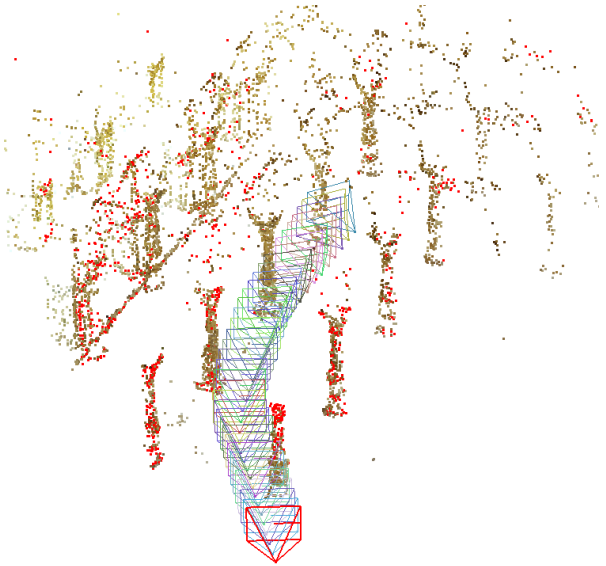
$$x_{ij} = M_i X_j, \qquad i = 1, \dots, m \;\; j = 1, \dots, n$$

- How to solve SfM?
  - Feature extraction & Finding corresponding points
  - Finding Fundamental matrix F (${x'}^T F x = 0$)
  - Estimating projective cameras
  - Triangulating
  - Bundle Adjustment

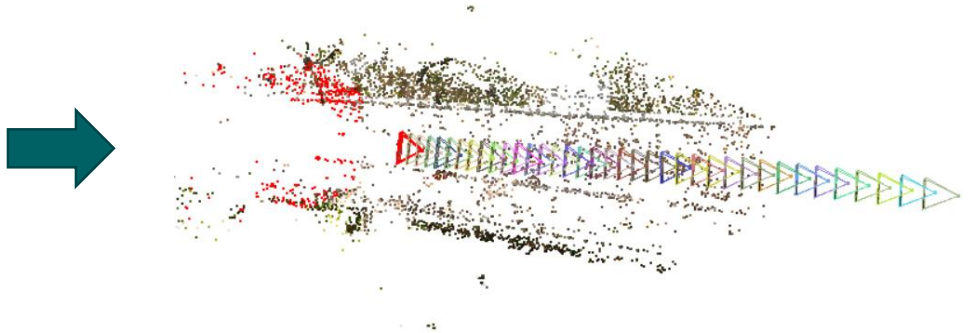


[1] R. Hartley and A. Zisserman, “Multiple view geometry in computer vision“, second edition, Cambridge press, 2013.

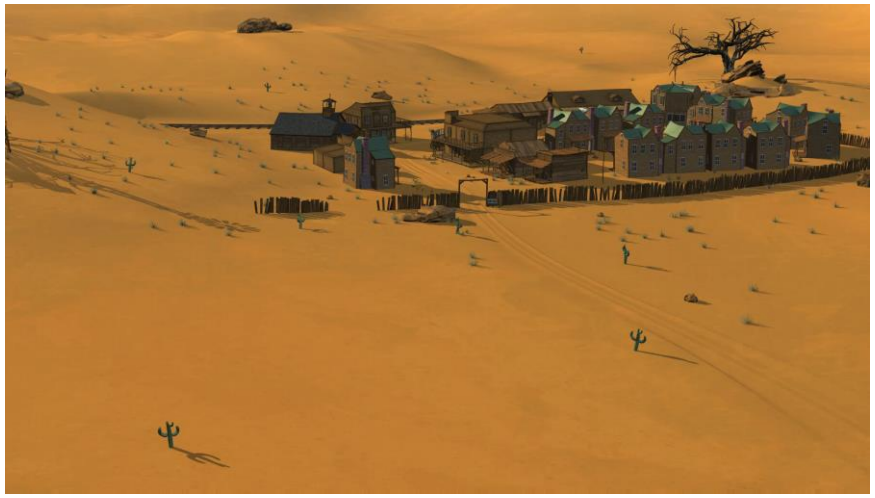


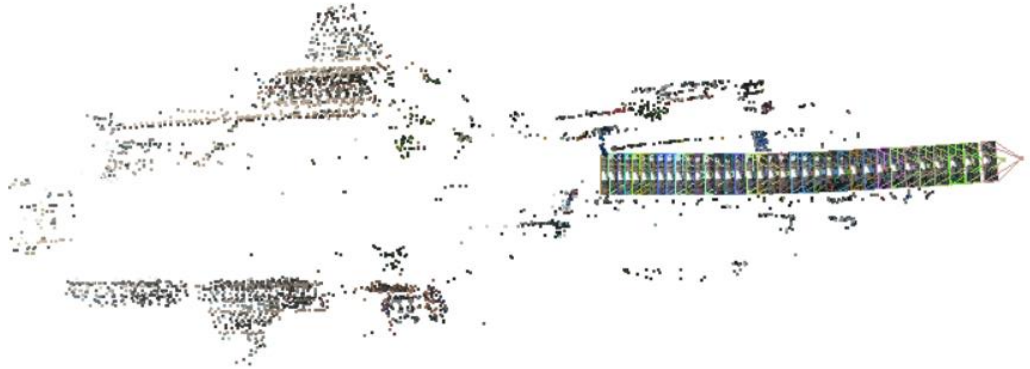


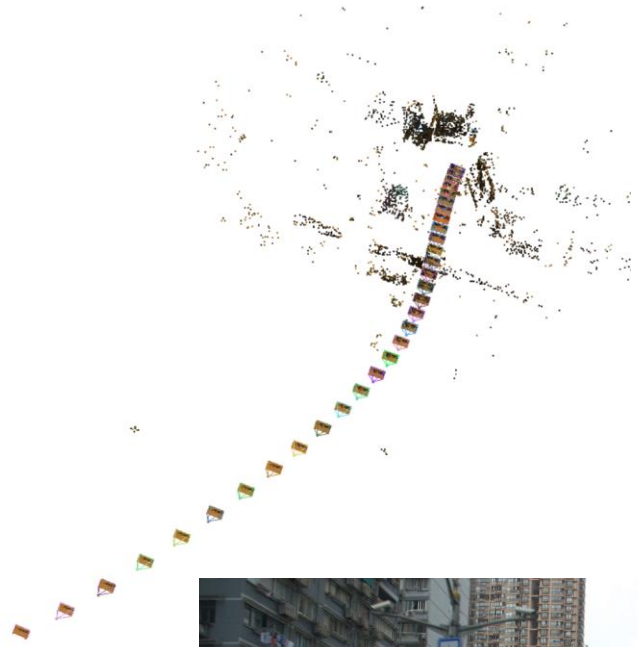
SfM Results

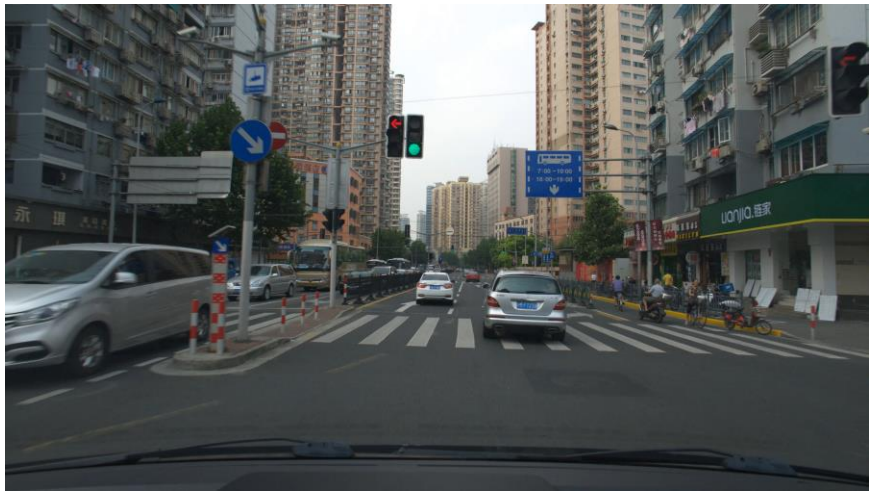


SfM Results









Outline

- Introduction to the main idea
- Structure from Motion (SfM)
- **Multi-View Reconstruction (MVR)**
- Virtual view synthesis for 2D frame prediction
- Coding results
- Conclusion

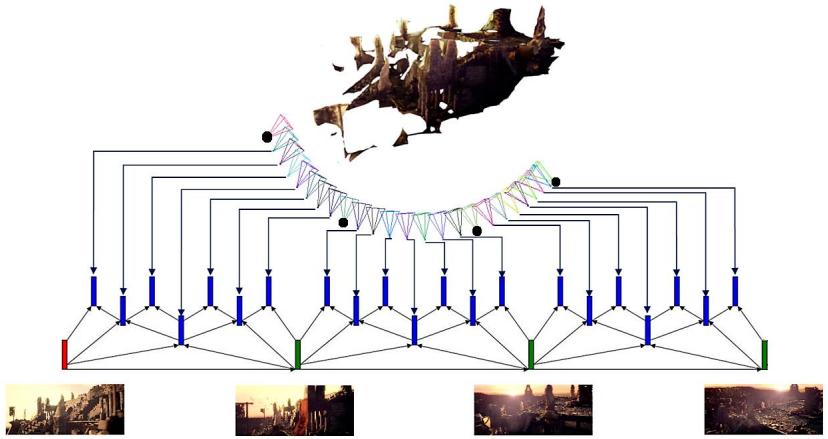
Multi-View Reconstruction (MVR)

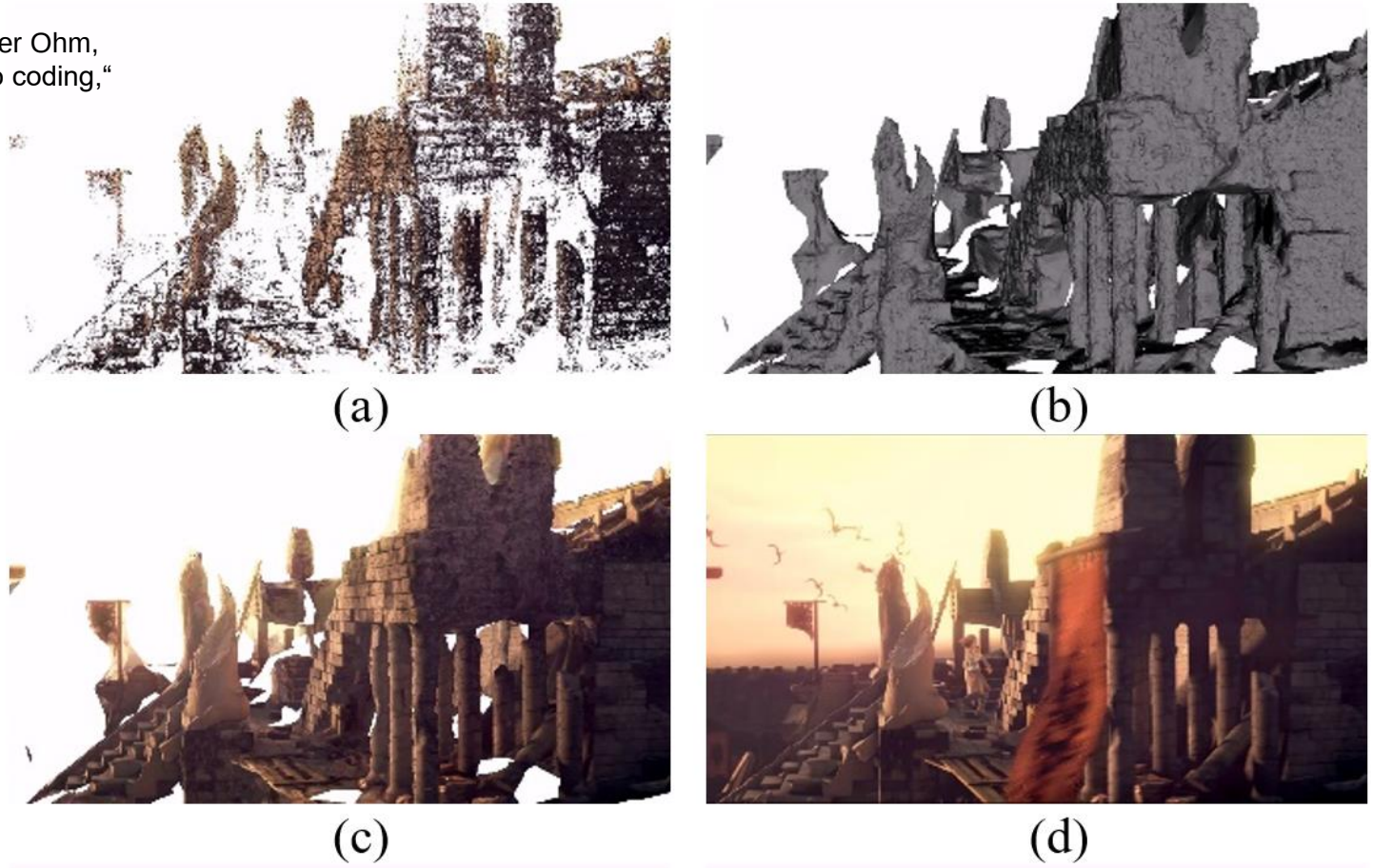
- Generating a quasi-dense point cloud
- Visibility based surface reconstruction
  - Delaunay tetrahedralization
  - Surface optimization with minimum s-t cut [2])
- The variational refinement of mesh to optimize its photo-consistency (Minimizing reprojection error)

[2] H. Vu, P. Labatut, J. Pons and R. keriven, “High accuracy and visibility-consistent dense multiview stereo”, IEEE Transaction on pattern analysis and machine intelligence, vol. 34, no. 5, 2012.

Application in Video Coding?

[3] Hossein B. Golestani, Jens Schneider, Mathias Wien and JensRainer Ohm, „Point cloud estimation for 3D structure based frame prediction in video coding,“ ICME 2017, Hong Kong.





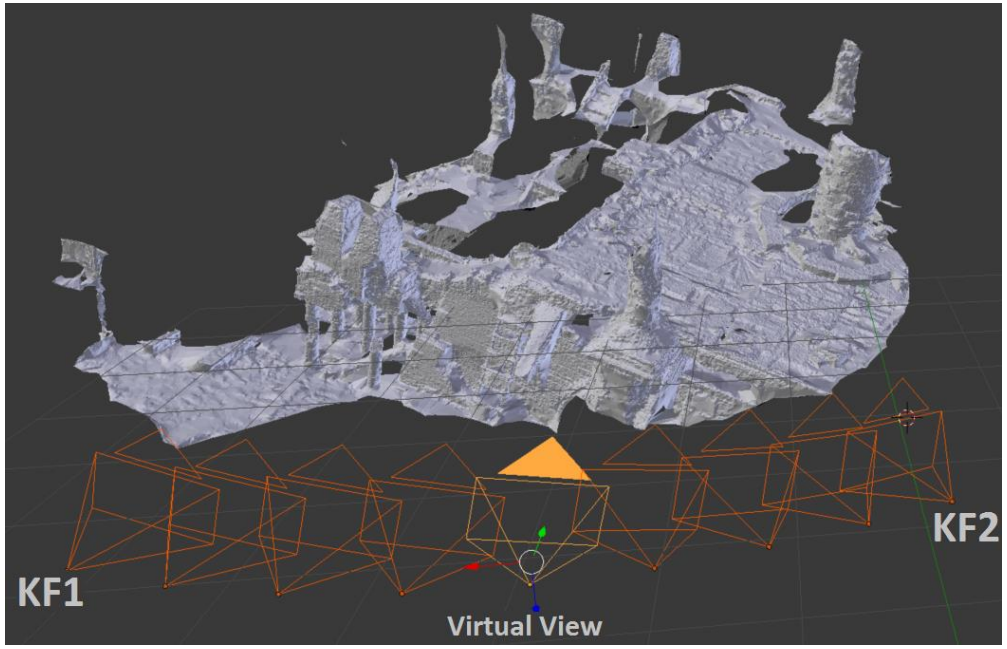
(a) quasi-dense point cloud, (b) shaded surface, (c) textured surface and (d) ground thruth

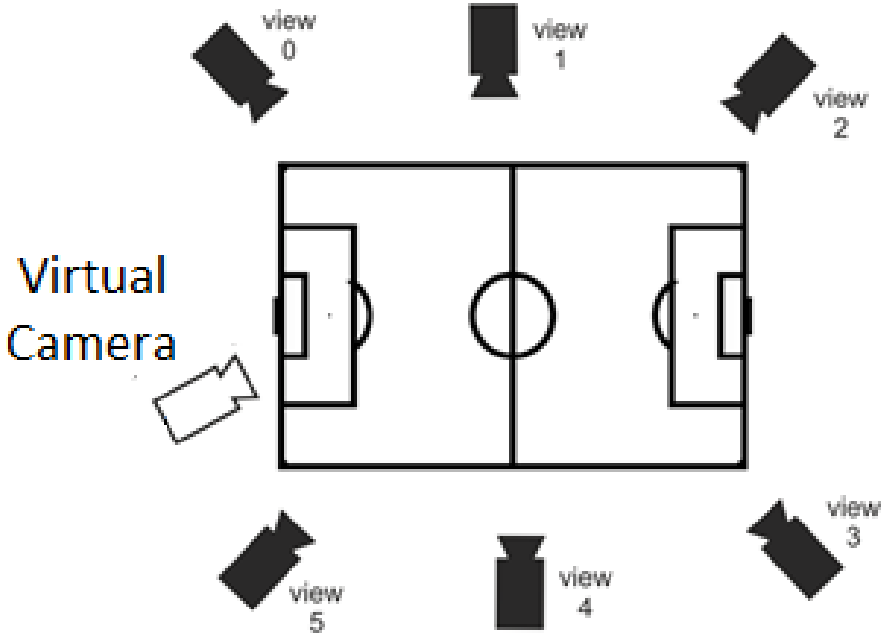
Outline

- Introduction to the main idea
- Structure from Motion (SfM)
- Multi-View Reconstruction (MVR)
- **Virtual view synthesis for 2D frame prediction**
- Coding results
- Conclusion

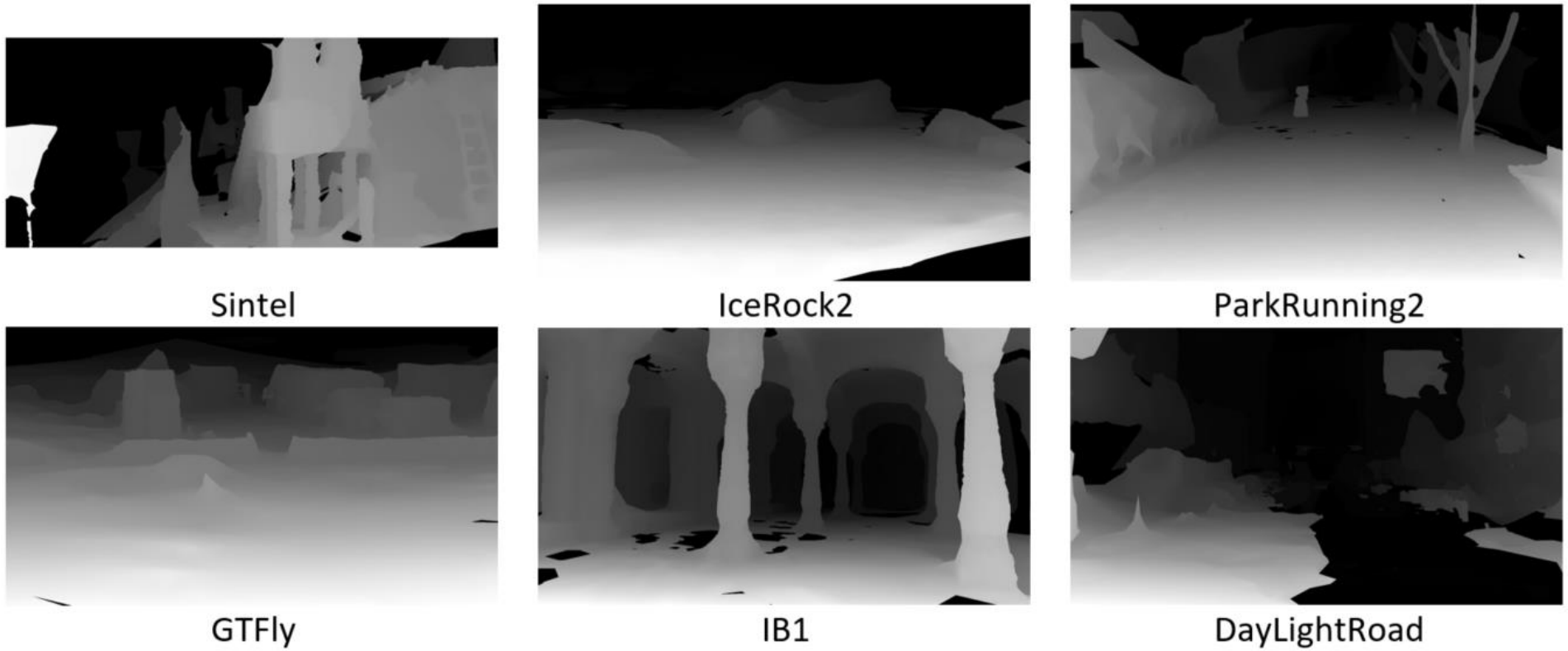
New method for frame prediction: Virtual View Synthesis

- How to improve the quality of predicted frame? Virtual View synthesis
- What does it need?
  - Left and right views
  - Camera parameters for left, right and virtual cameras
  - Depth information of left and right views (depthmaps, Z-min & Z-max)





Estimated Depthmaps



Virtual View Synthesis

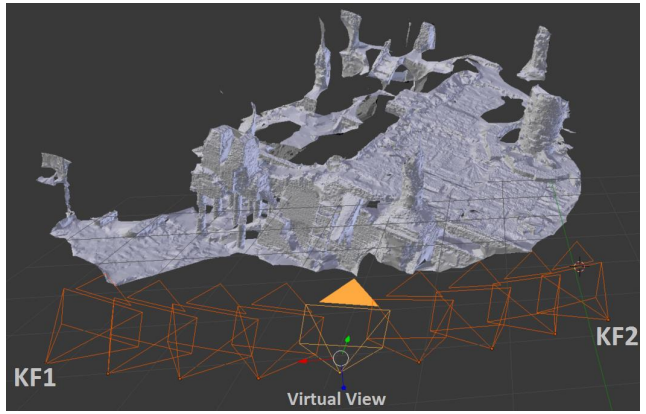
- Steps:
  - Estimating the depthmaps of the virtual view from KFs‘ depthmaps
    - Finding a homography transformation between real and virtual views for each depth level

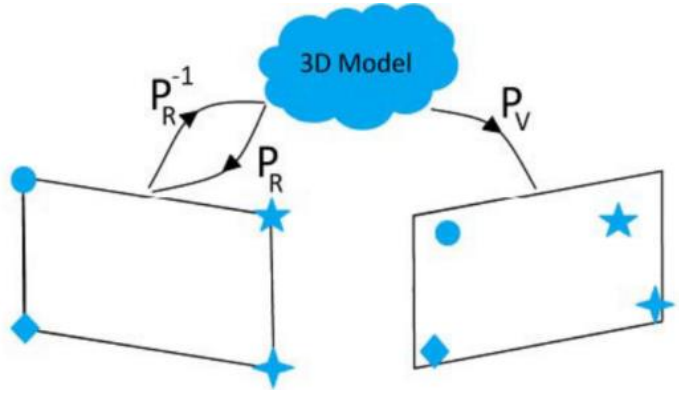
$$X_V[i] = P_V P_R^{-1} X_R[i], \quad i = depth\,levels$$

  - Synthesizing the virtual view by warping textures from left and right key-frames

$$[h, w, 1]^T = H^{-1}[i] \times [u, v, 1]^T, \quad [u, v]: virtual\,view, \quad [h, w]: real\,view$$

  - Blending the synthesized view from left and right

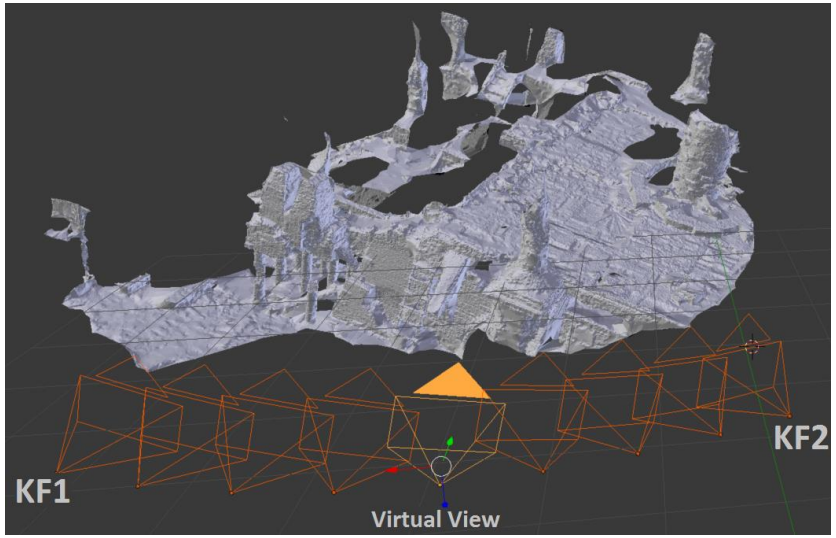




Synthesized frames

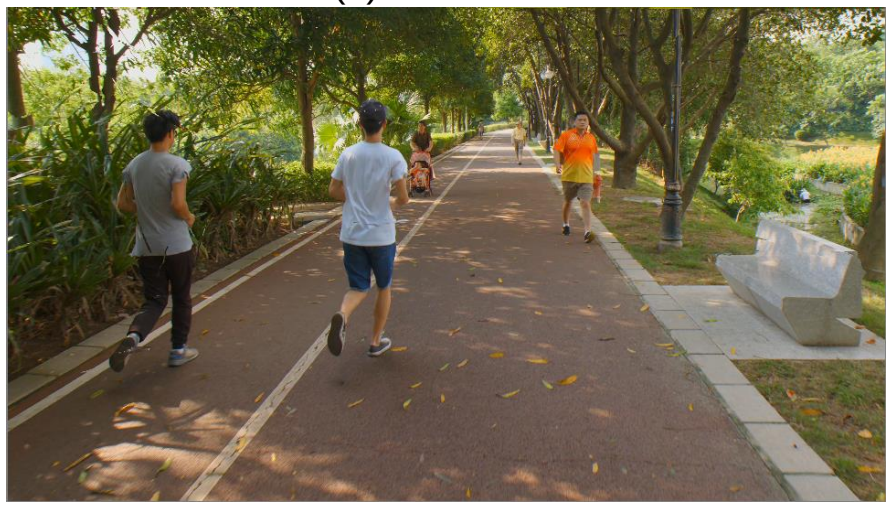
(a) Ground Truth

(b) Synthesized (virtual view)

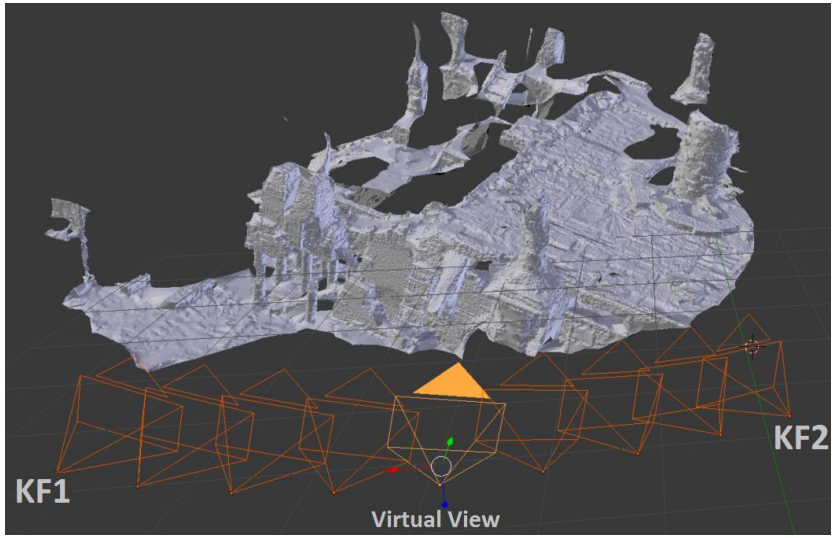


Synthesized frames

(a) Ground Truth



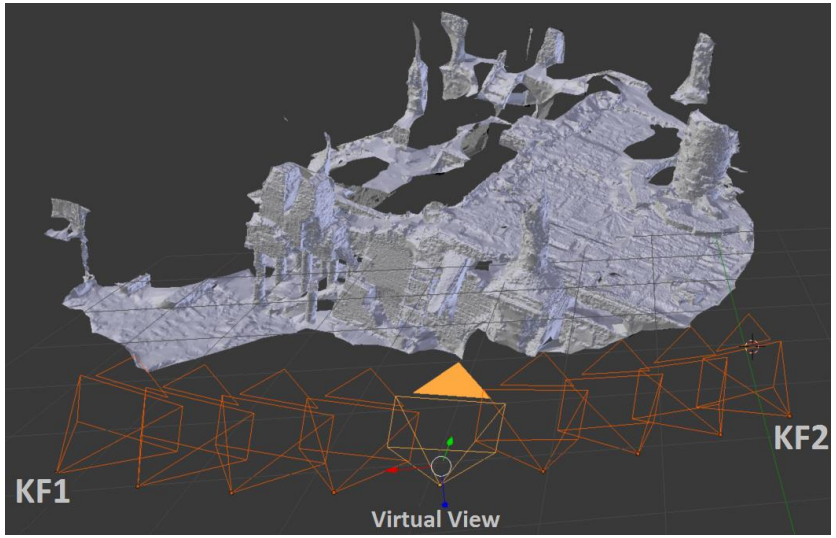
(b) Synthesized (virtual view)



Synthesized frames

(a) Ground Truth

(b) Synthesized (virtual view)

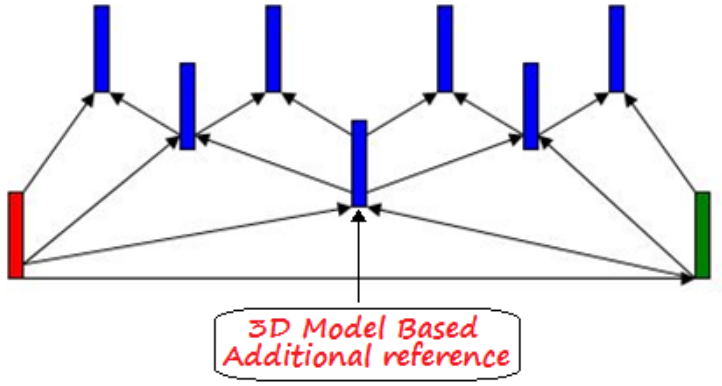


Outline

- Introduction to the main idea
- Structure from Motion (SfM)
- Multi-View Reconstruction (MVR)
- Virtual view synthesis for 2D frame prediction
- **Coding results**
- Conclusion

SfM based frame prediction in Video Coding

- HEVC test Model (HM 16.7)
  - SfM-based prediction is added to L0 and L1
  - The encoder can choose between
    - its built-in reference images
    - The offered SfM-based prediction
  - QP: 25, 29, 33 and 37, Random Acess main profile

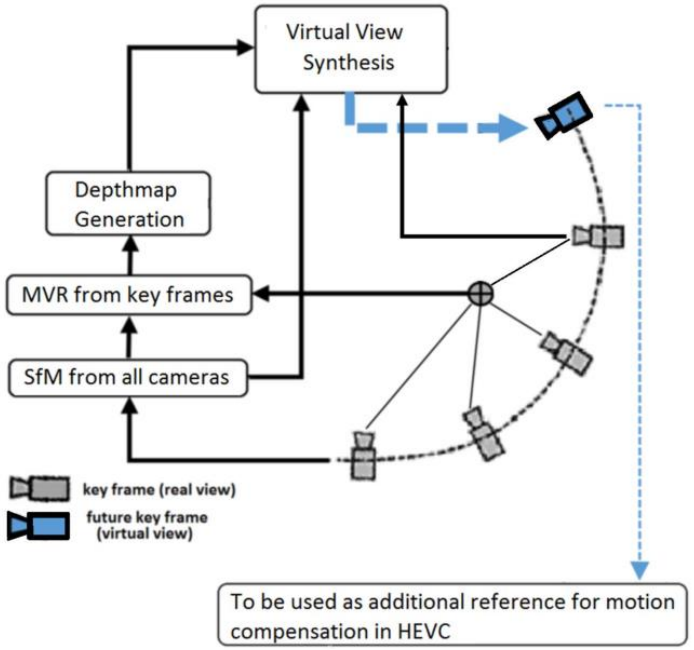


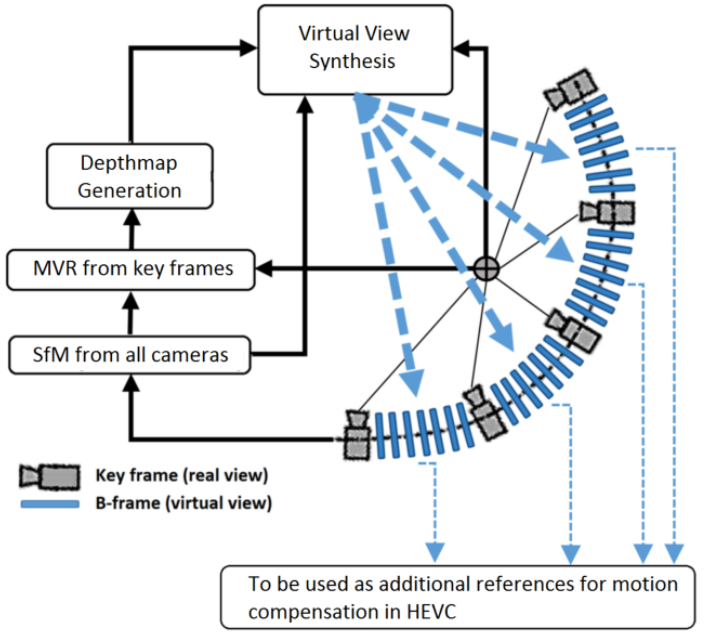

```
POC    24 TId: 0 ( B-SLICE, nQP 38 QP 38 )      48048 bits [Y 39.0652 dB    U 46.8478 dB    V 47.1814 dB] [ET   158 ] [L0 16 8 24(AddRef) ] [L1 16 8 24(AddRef) ]
POC    20 TId: 1 ( B-SLICE, nQP 39 QP 39 )       6552 bits [Y 39.3872 dB    U 46.9528 dB    V 47.2622 dB] [ET   149 ] [L0 16 12 20(AddRef) ] [L1 24 16 20(AddRef) ]
POC    18 TId: 2 ( B-SLICE, nQP 40 QP 40 )       3576 bits [Y 38.7807 dB    U 46.8727 dB    V 47.1718 dB] [ET   161 ] [L0 16 12 18(AddRef) ] [L1 20 24 18(AddRef) ]
POC    17 TId: 3 ( B-SLICE, nQP 41 QP 41 )        992 bits [Y 38.8425 dB    U 46.8946 dB    V 47.2129 dB] [ET   144 ] [L0 16 18 17(AddRef) ] [L1 18 20 17(AddRef) ]
POC    19 TId: 3 ( B-SLICE, nQP 41 QP 41 )       1464 bits [Y 38.5592 dB    U 46.8896 dB    V 47.1888 dB] [ET   158 ] [L0 18 16 19(AddRef) ] [L1 20 24 19(AddRef) ]
POC    22 TId: 2 ( B-SLICE, nQP 40 QP 40 )       3232 bits [Y 39.1846 dB    U 46.9233 dB    V 47.1872 dB] [ET   145 ] [L0 20 16 22(AddRef) ] [L1 24 20 22(AddRef) ]
POC    21 TId: 3 ( B-SLICE, nQP 41 QP 41 )       1576 bits [Y 38.8811 dB    U 46.9153 dB    V 47.1979 dB] [ET   158 ] [L0 20 16 21(AddRef) ] [L1 22 24 21(AddRef) ]
POC    23 TId: 3 ( B-SLICE, nQP 41 QP 41 )       1656 bits [Y 38.7014 dB    U 46.8517 dB    V 47.1399 dB] [ET   144 ] [L0 22 20 23(AddRef) ] [L1 24 22 23(AddRef) ]
POC    32 TId: 0 ( P-SLICE, nQP 37 QP 37 )     383656 bits [Y 38.9582 dB    U 46.6173 dB    V 46.8710 dB] [ET   226 ] [L0 32(AddRef) 32(AddRef) 32(AddRef) ] [L1 ]
```

Seq.	Sintel	IceRock2	GTFly	ParkRunning2	DayLightRoad	IB1
BD-Rate (%)	-2.21	-10.76	-6.64	-2.55	-2.80	-5.91
BD_PSNR (dB)	+0.08	+0.32	+0.19	+0.08	+0.04	+0.07

Conclusion

- An SfM-based frame prediction scheme has been presented.
- SfM for camera calibration
- MVS for surface reconstruction.
- Virtual view synthesis for synthesizing novel views.
- SfM-based prediction is added to reference lists (L0 and L1)
- So far, 2% to 10% bitrate reduction is achieved.
- Estimating better depthsmaps will improve results.





Vielen Dank
für Ihre Aufmerksamkeit

Leibniz Universität Hannover

A Comparison of JEM and AV1 with HEVC

Thorsten Laude

Leibniz Universität Hannover
Institut für Informationsverarbeitung

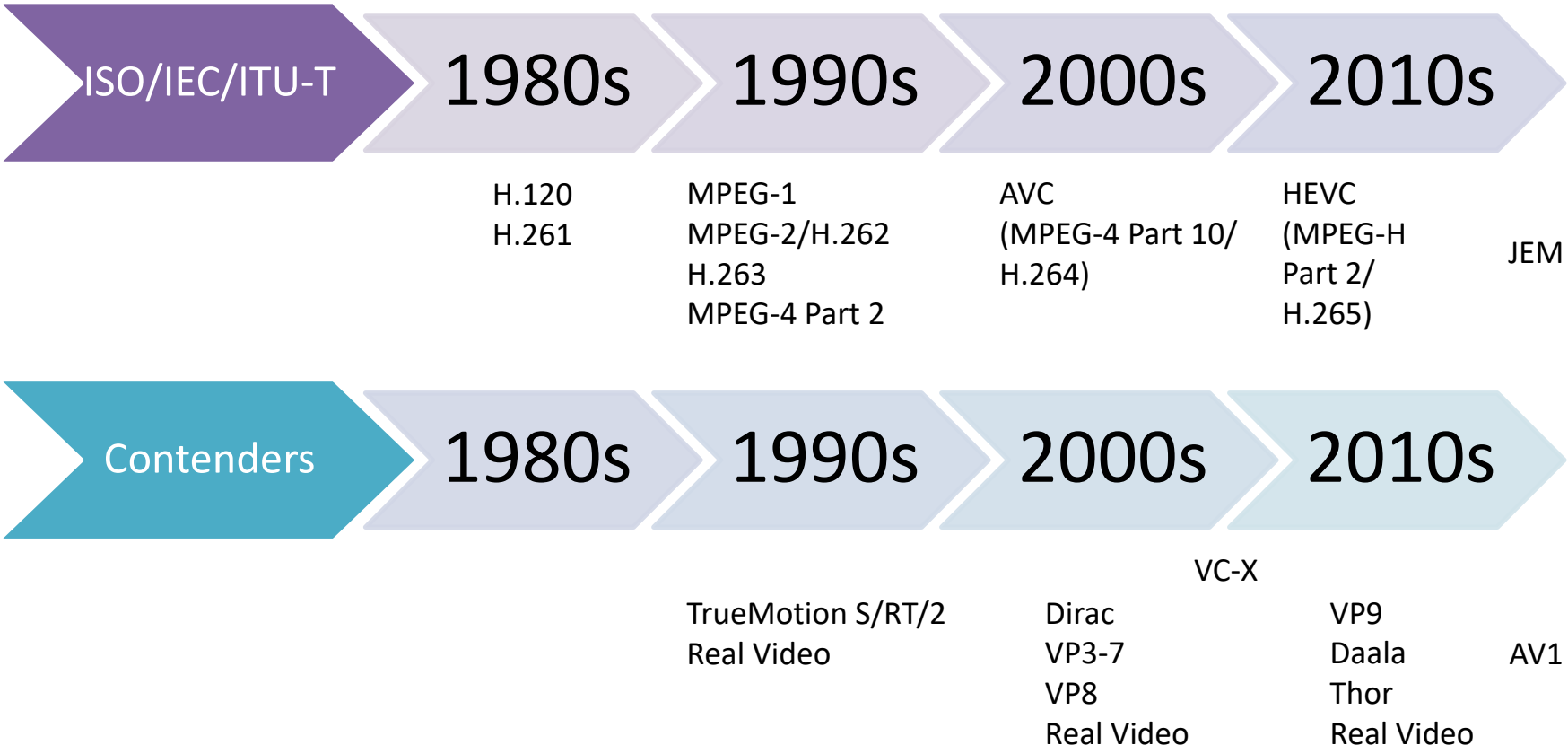
Overview

Coding Tools

Coding Efficiency

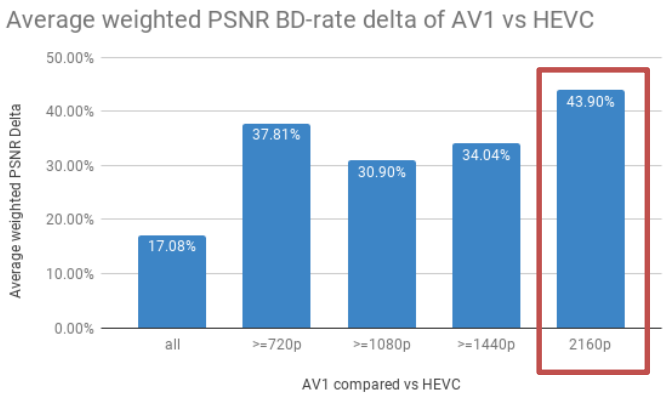
Complexity

History of Video Codecs



Comparison of the latest video codecs (JEM/AV1) with HEVC

On the Difficulty of Comparing Video Codecs



AV1 is up to 43% better than HEVC

Source: Feldmann, "Multi-Codec DASH Dataset: An Evaluation of AV1, AVC, HEVC and VP9", Bitmovin Blog, 2018

"In terms of PSNR, the average BD-rate savings of AV1 relative to […] x264 high […] are […] 45.8% […]On the other hand, the encoding computational complexity […] was increased by factors of […] 5869.9x"

Source: Liu, "AV1 beats x264 and libvpx-vp9 in practical use cases", Facebook Blog, 2018

	anchor			
test candidate	AV1	JEM	VP9	HM
AV1		111.8%	-17.1%	47.7%
JEM	-51.4%		-62.0%	-29.8%
VP9	21.0%	173.7%		92.5%
HM	-30.6%	43.4%	-46.6%	

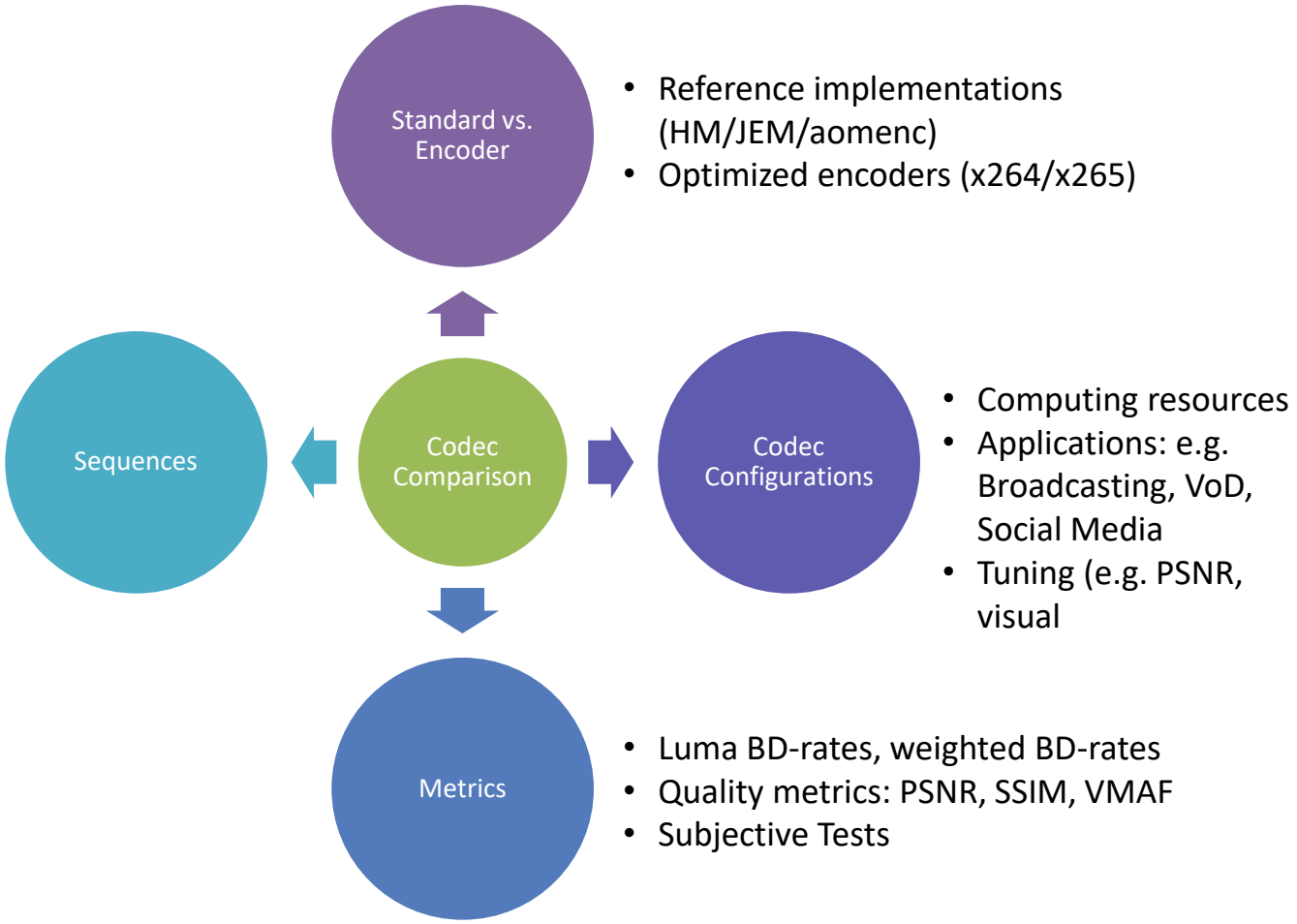
HEVC is 30% better than AV1

Source: Grois et al., "Performance Comparison of AV1, JEM, VP9 and HEVC Encoders", Proceedings of SPIE, 2017

Content	AV1 vs HEVC/H.265	
	BD-PSNR	BD-MOS
Campfire Party	-23.2%	-19.0%
Runners	-2.6%	1.5%
Traffic Flow	5.7%	4.9%
Tree Shade	-5.8%	-9.3%
Sintel2	35.6%	38.0%
Average	**1.9%**	**3.2%**

Source: Akyazi and Ebrahimi, "Comparison of compression efficiency between HEVC/H.265 and VP9 based on subjective assessments", QoMEX, 2018

On the Difficulty of Comparing Video Codecs



- Reference implementations (HM/JEM/aomenc)
- Optimized encoders (x264/x265)

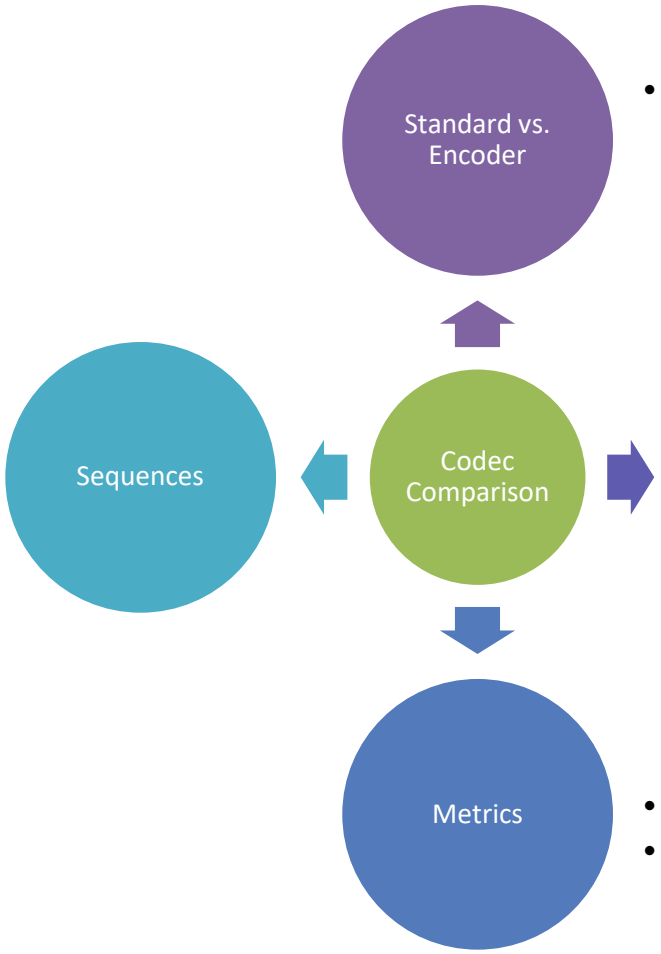
Codecs perform differently good for different content

- Computing resources
- Applications: e.g. Broadcasting, VoD, Social Media
- Tuning (e.g. PSNR, visual

- Luma BD-rates, weighted BD-rates
- Quality metrics: PSNR, SSIM, VMAF
- Subjective Tests

Test Conditions for this Comparison

Class	Sequence
A1 (4K)	Tango2
	Drums100
	Campfire
	ToddlerFountain2
A2 (4K)	CatRobot
	TrafficFlow
	DaylightRoad2
	Rollercoaster2
B (1080p)	Kimono
	ParkScene
	Cactus
	BasketballDrive
	BQTerrace
C (WVGA)	BasketballDrill
	BQMall
	PartyScene
	RaceHorses
D (WQVGA)	BasketballPass
	BQSquare
	BlowingBubbles
	RaceHorses
E (720p)	FourPeople
	Johnny
	KristenAndSara
F (Screen/ Mixed Content)	BasketballDrillText
	ChinaSpeed
	SlideEditing
	SlideShow



- Reference implementations (HM/JEM/aomenc)

- HM/JEM: Common Test Conditions (CTC)
- AV1

```
--auto-alt-ref=1 --psnr
--tune=psnr --i420 -p 1
-t 1 --fps=<?> --bit-
depth=<?> --input-bit-
depth=<?> --cq-
level=<?> --kf-min-
dist=<?> --kf-max-
dist=<?> -w <?> -h <?>
```

- Luma BD-rates
- Quality metrics: PSNR

Coding Tools

JEM

Partitioning
- Quaternary and binary splits
- Bigger block size

Inter coding
- Overlapped block motion compensation
- Higher order motion model
- Sub-CU MV prediction

Intra coding
- Additional directions
- Cross-component linear model

Transform coding
- Adaptive multiple transforms
- Non-separable secondary transform
- Signal-dependent transform

AV1

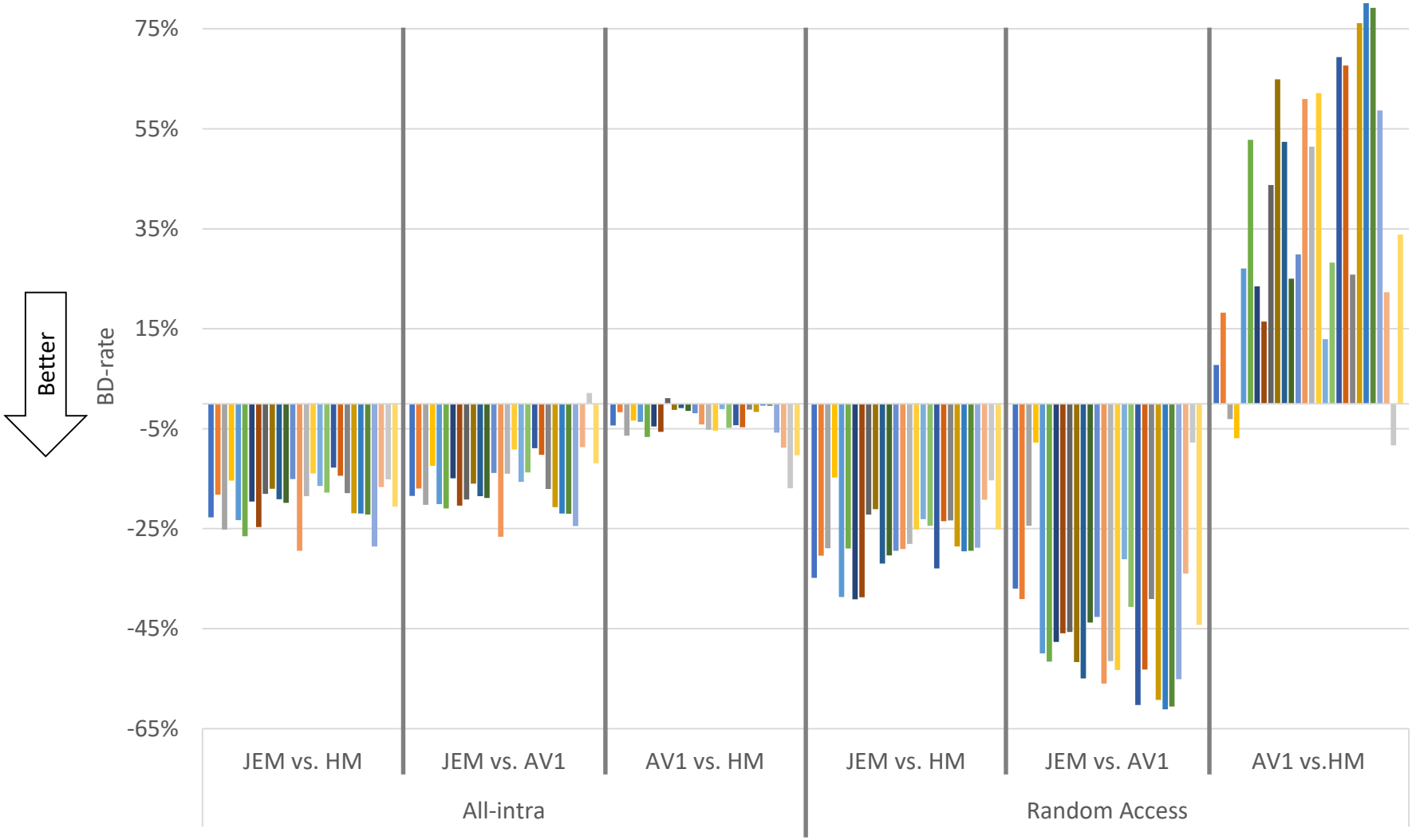
Partitioning
- Quaternary and binary splits
- Bigger block size

Inter coding
- Overlapped block motion compensation
- Higher order motion models
- Wedge mode partitioning
- Compound intra-inter prediction

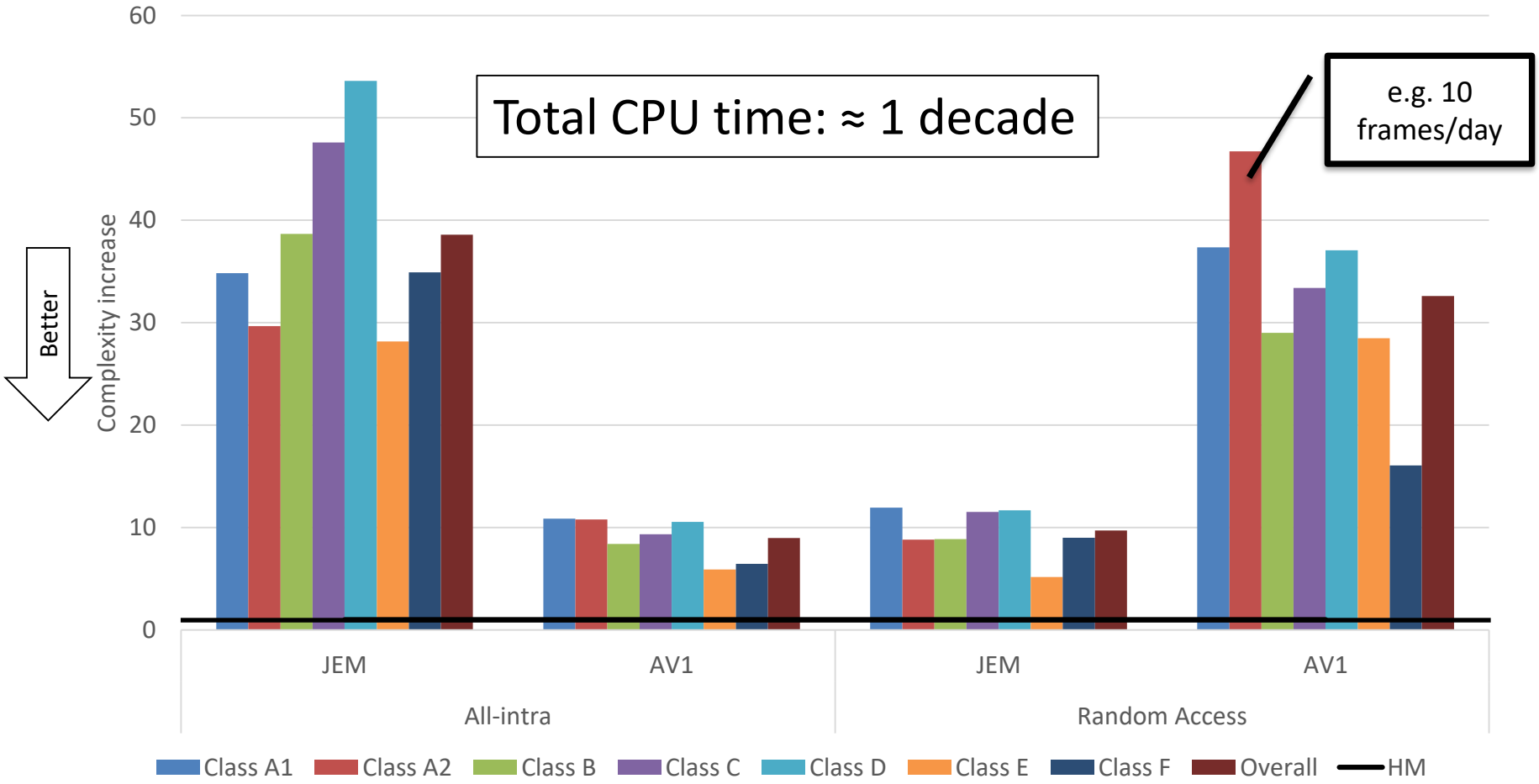
Intra Coding
- Directional, Paeth, Smooth prediction
- Intra block copy
- Palette mode

Transform coding
- DCT, DST, Identity
- Independent horizontal/vertical transforms

Coding Efficiency

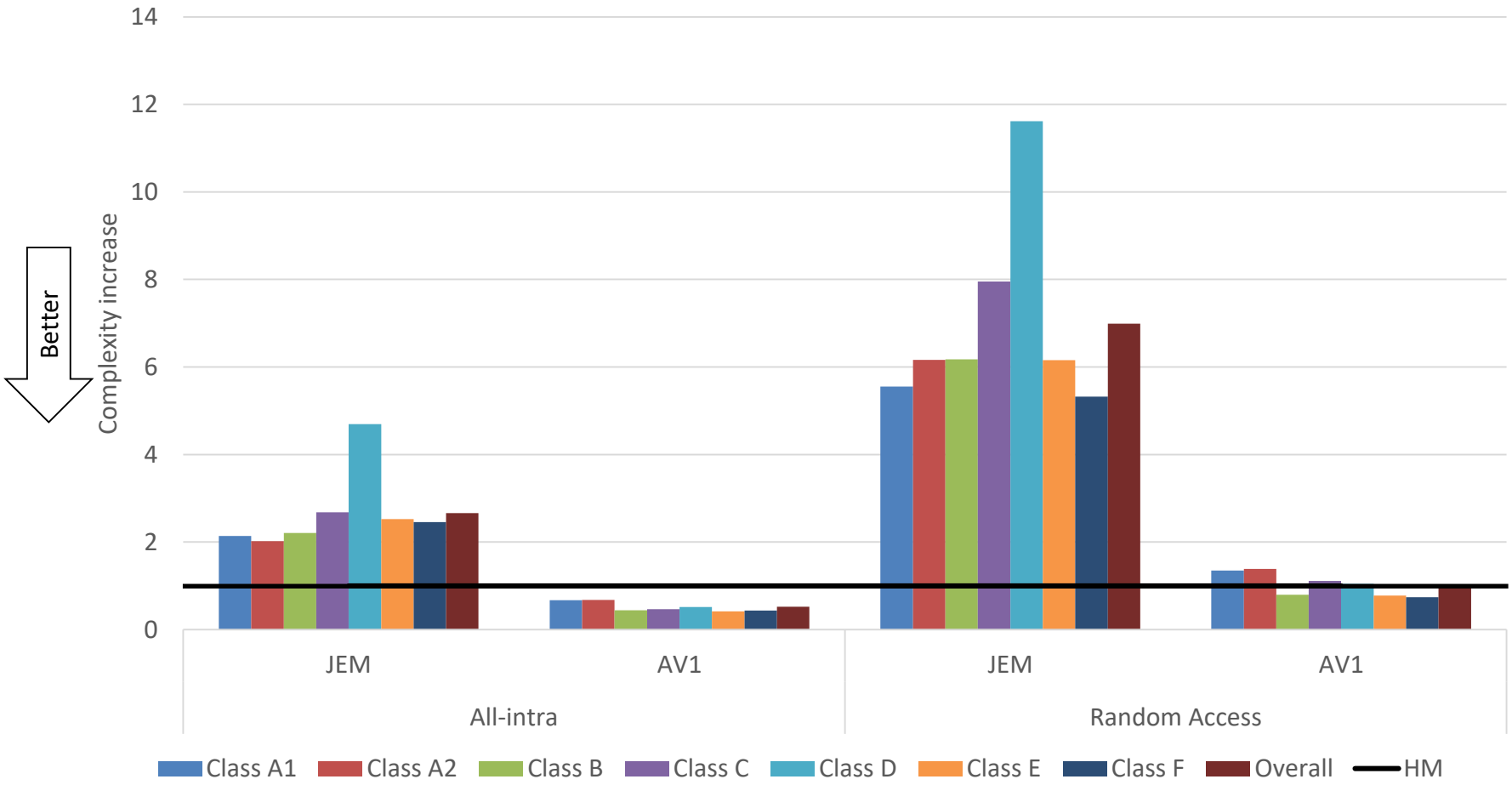


Encoder Runtimes



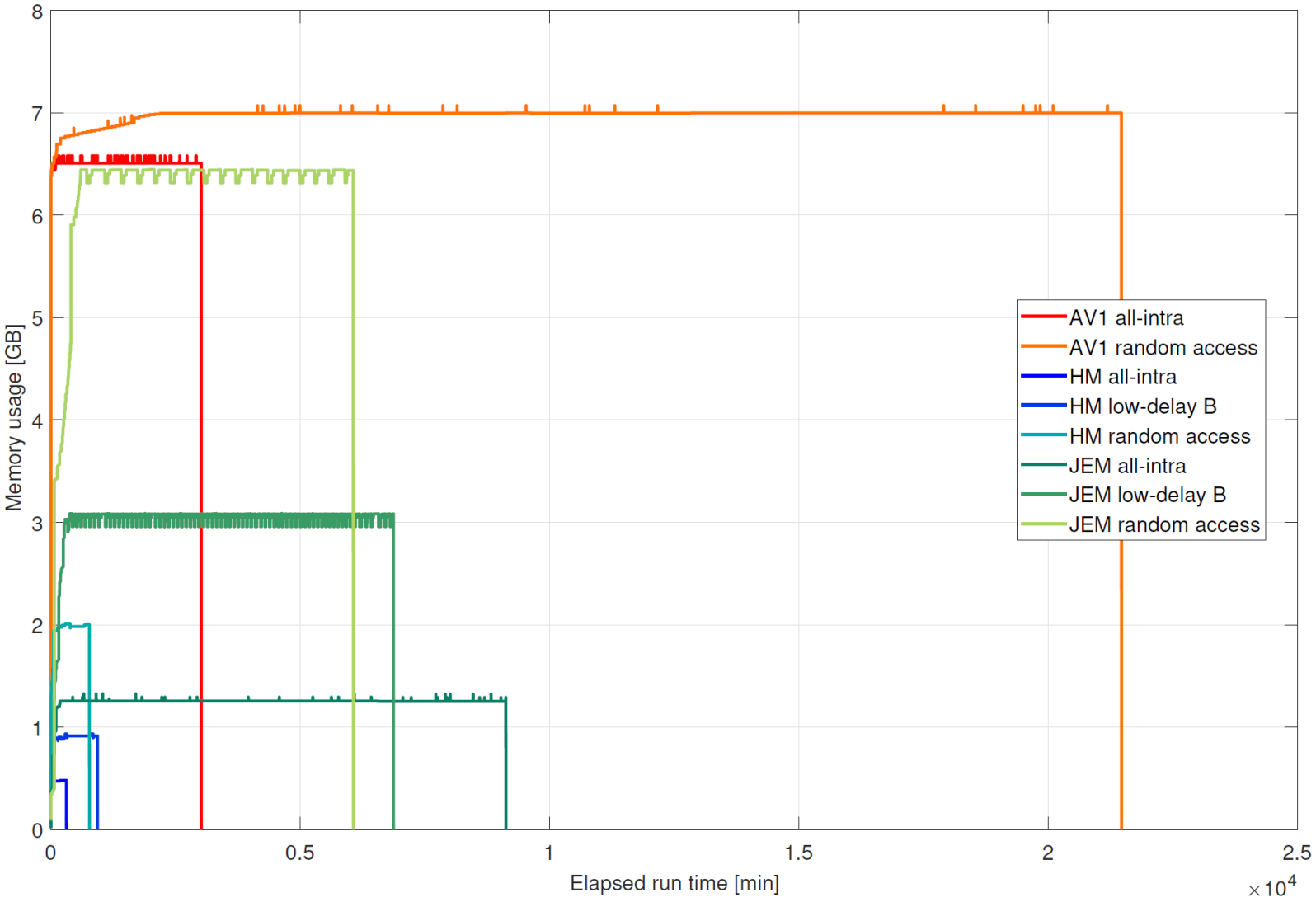
Relative factors to HM, i.e. HM=1

Decoder Runtimes

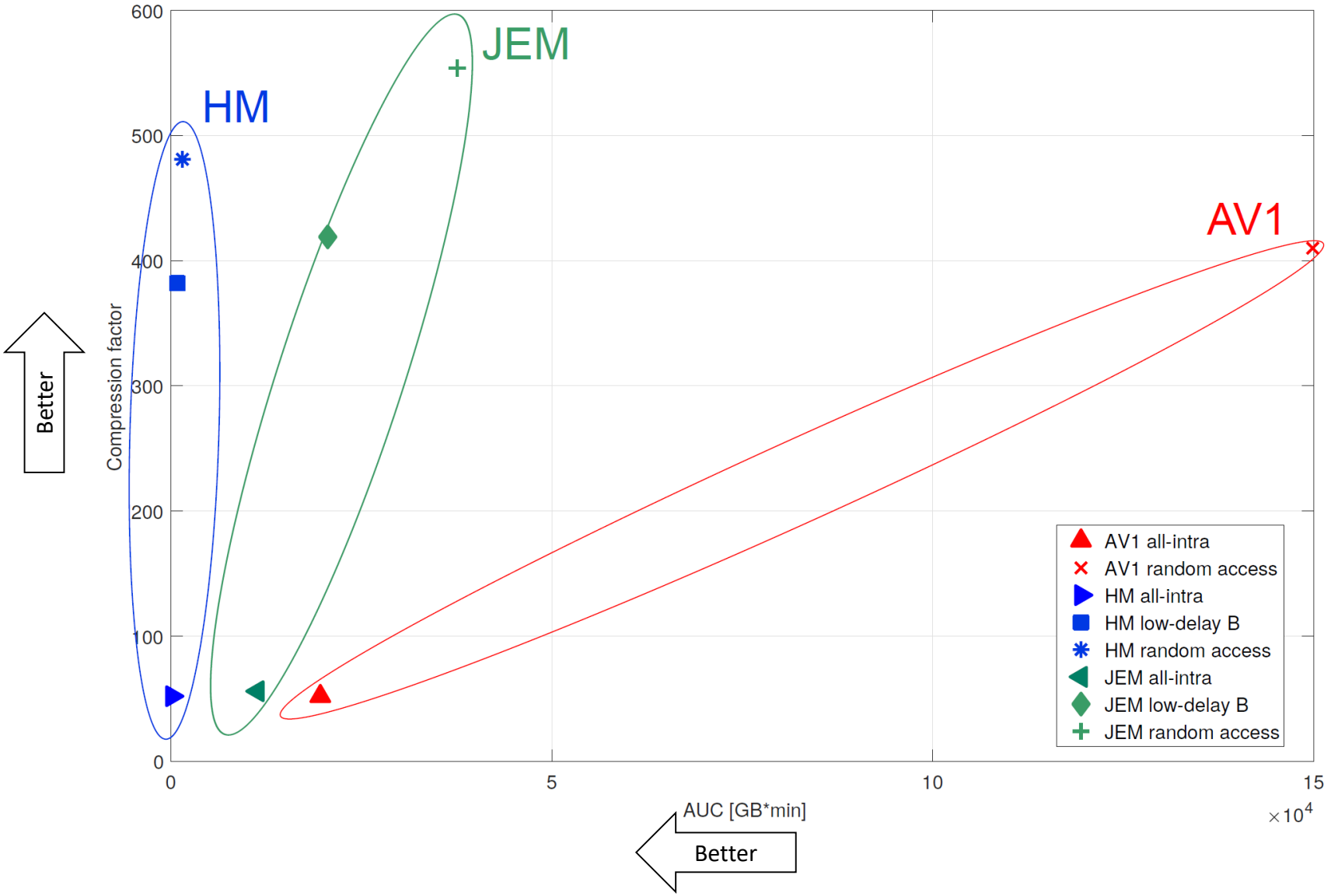


Relative factors to HM, i.e. HM=1

Runtime-memory Complexity



Trade-off Coding Efficiency vs. Complexity



Summary

Coding Efficiency

Comparison vs. HM

All intra (AI)
JEM: 20% gain
AV1: 4% gain

Random Access (RA)
JEM: 28% gain
AV1: 38% loss

Runtimes

Comparison vs. HM

Encoder
JEM: 39 × (AI)/10 × (RA) slower
AV1: 9 × (AI)/32 × (RA) slower

Decoder
JEM: 3 × (AI)/7 × (RA) slower
AV1: 2 × faster (AI)/same (RA)

Closing remarks

- Results are a snapshot of summer 2017 → AV1 finalization in March 2018 and JVET CfP evaluation in April 2018
  - Since last summer, AV1 has gained additional 5% (based on 80 preliminary data points)
- Complexity: Reference implementations vs. product implementations

Details: Laude, T., Adhisantoso, Y. G., Voges, J., Munderloh, M., & Ostermann, J. (2018). A Comparison of JEM and AV1 with HEVC: Coding Tools , Coding Efficiency and Complexity. In *Picture Coding Symposium (PCS)*.

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

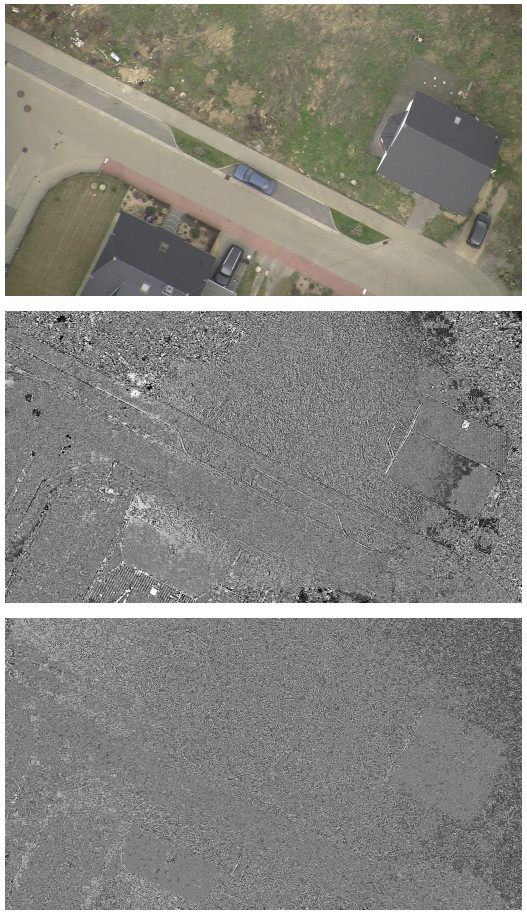
  $\Rightarrow$ High prediction error

  $\Rightarrow$ High entropy

  $\Rightarrow$ High bit rate

Goal:

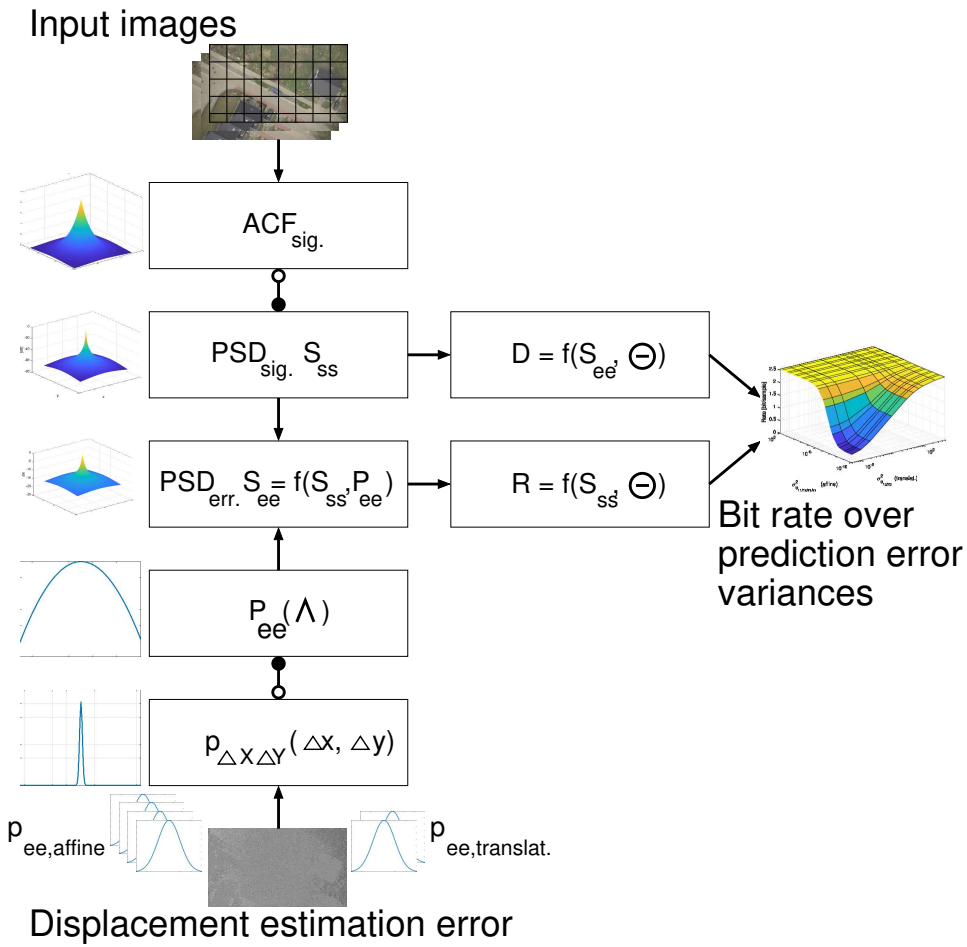
Model the prediction error bit rate as a function of the displacement estimation error for an **affine motion model**



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 $\Rightarrow$ 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

Efficiency Analysis of Affine Motion Compensated Prediction
- Model of the Probability Density Function (p. d. f.)
- Model of Power Spectral Densities (P. S. D.s)
- Rate-Distortion Analysis

Simulations

Conclusion

Outline

Efficiency Analysis of Affine Motion Compensated Prediction
- Model of the Probability Density Function (p. d. f.)
- Model of Power Spectral Densities (P. S. D.s)
- Rate-Distortion Analysis

Simulations

Conclusion

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

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

$$\Delta x = e_{11} \cdot x' + e_{12} \cdot y' + e_{13} \tag{3}$$
$$\Delta y = e_{21} \cdot x' + e_{22} \cdot y' + e_{23} \tag{4}$$

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}:

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

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:

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

Probability Density Function of the Displacement Estimation Error

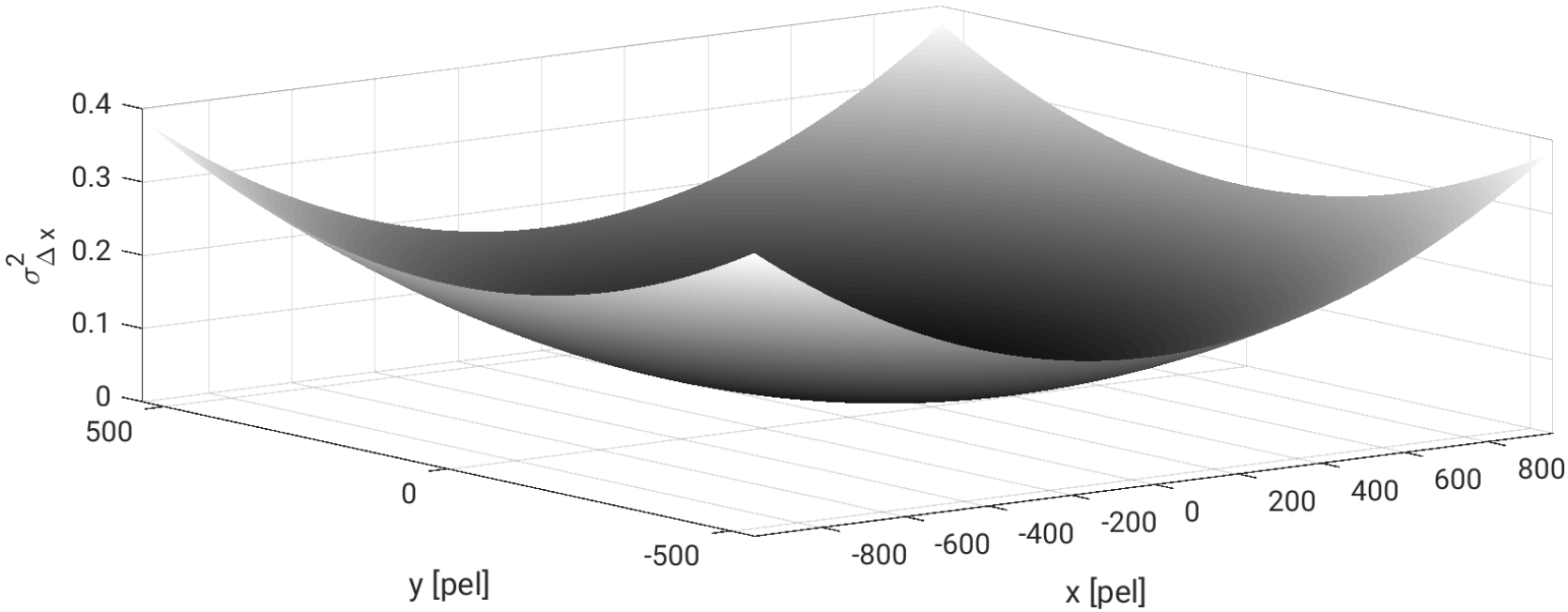
$$p_{\Delta X,\Delta Y}(\Delta x, \Delta y) = \frac{1}{2\pi\sigma_{\Delta x}\sigma_{\Delta y}} \cdot \exp\left(-\frac{\Delta x^2}{2\sigma_{\Delta x}^2}\right) \cdot \exp\left(-\frac{\Delta y^2}{2\sigma_{\Delta y}^2}\right) \quad (8)$$

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

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

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

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

Efficiency Analysis of Affine Motion Compensated Prediction
- Model of the Probability Density Function (p. d. f.)
- Model of Power Spectral Densities (P. S. D.s)
- Rate-Distortion Analysis

Simulations

Conclusion

Signal and Error Power Spectral Density Functions

- Assumption of isotropic autocorrelation function[4]:

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

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

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

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

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

---

[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

Efficiency Analysis of Affine Motion Compensated Prediction
- Model of the Probability Density Function (p. d. f.)
- Model of Power Spectral Densities (P. S. D.s)
- Rate-Distortion Analysis

Simulations

Conclusion

Rate-Distortion Theory[5]

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

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

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

Efficiency Analysis of Affine Motion Compensated Prediction
- Model of the Probability Density Function (p. d. f.)
- Model of Power Spectral Densities (P. S. D.s)
- Rate-Distortion Analysis

Simulations

Conclusion

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).

Variances of Affine Transformation Parameters

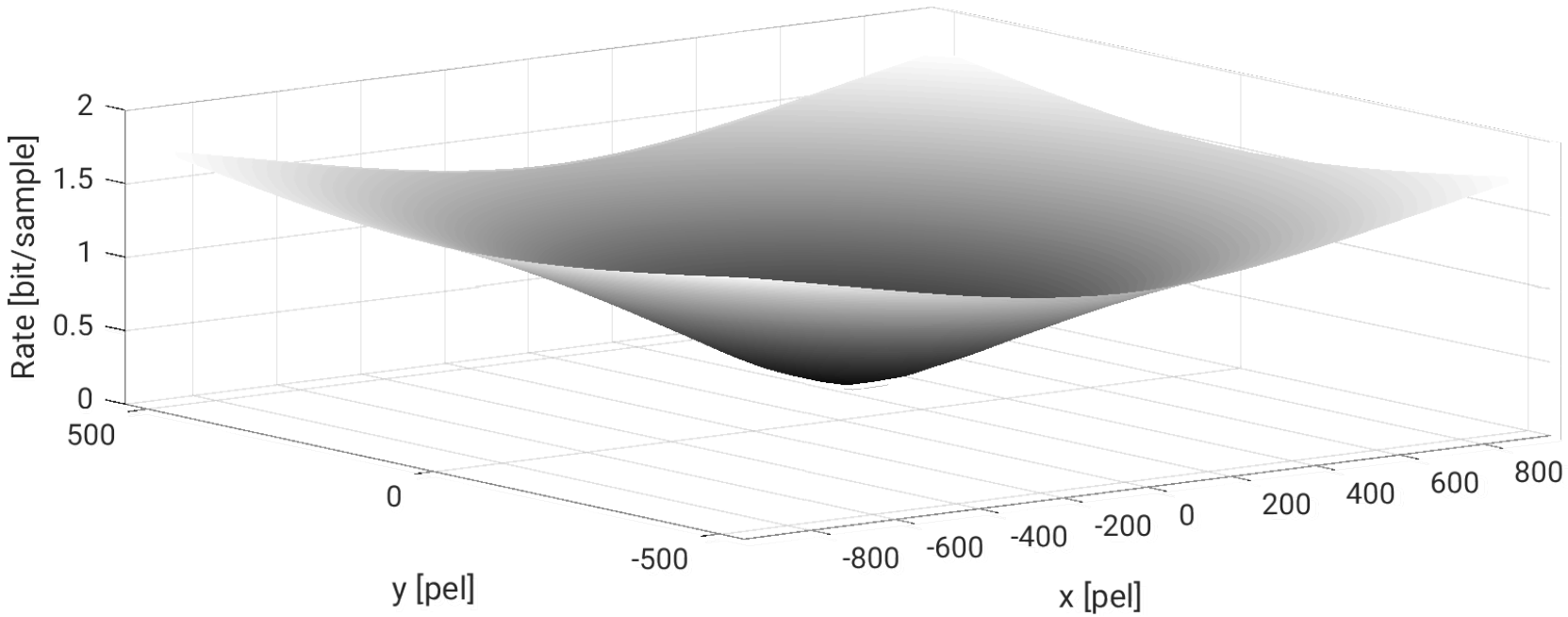
	$\sigma^2_{e_{11}}$	$\sigma^2_{e_{12}}$	$\sigma^2_{e_{21}}$	$\sigma^2_{e_{22}}$	mean $(\sigma^2_{e_{11}}, \sigma^2_{e_{22}})$	mean $(\sigma^2_{e_{12}}, \sigma^2_{e_{21}})$
350m seq.	2.03e−7	6.03e−7	6.59e−7	2.24e−7	**2.13e−7**	**6.31e−7**
500m seq.	1.94e−7	5.09e−7	3.63e−7	1.94e−7	**1.94e−7**	**4.35e−7**
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/`

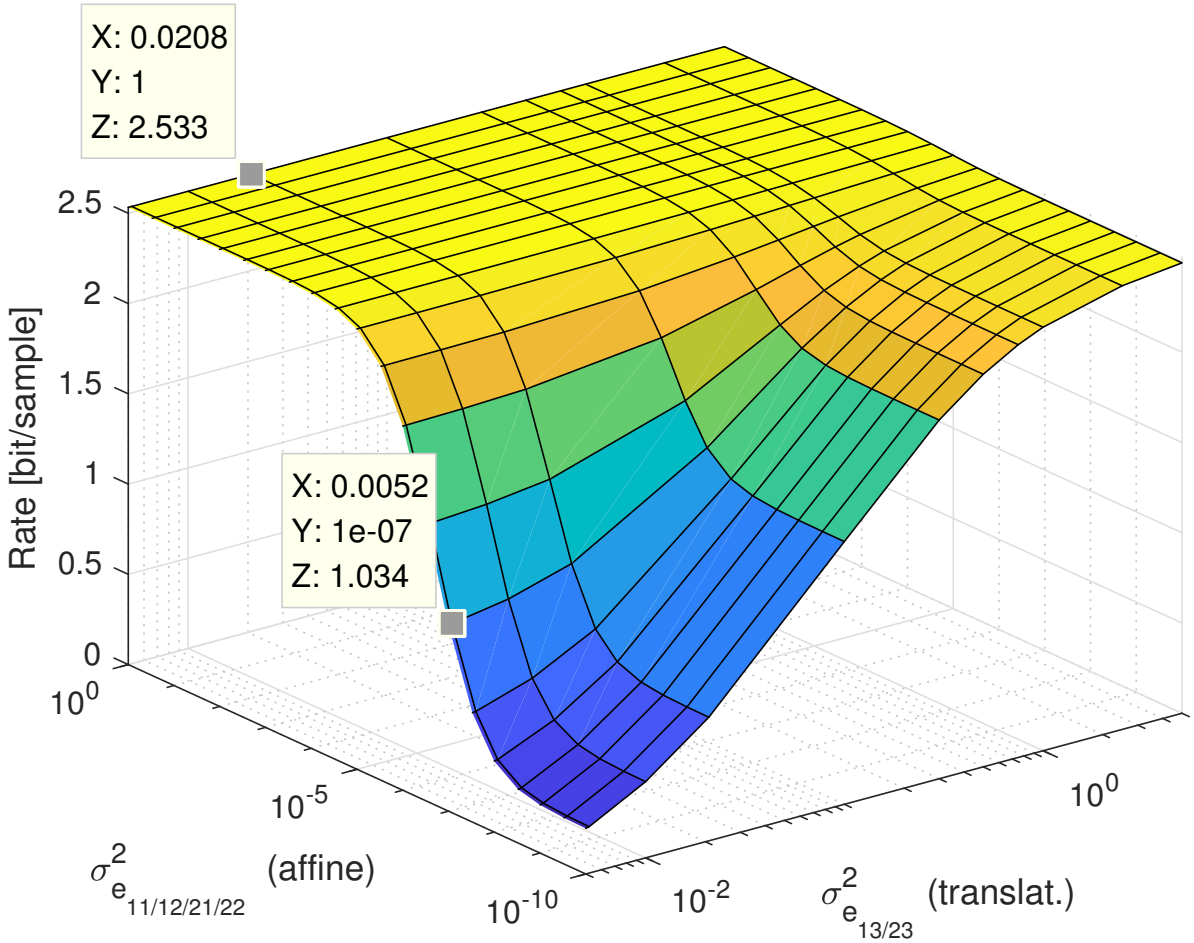
Simulated Location Dependent Bit Rate



Bit rate

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

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

Efficiency Analysis of Affine Motion Compensated Prediction
- Model of the Probability Density Function (p. d. f.)
- Model of Power Spectral Densities (P. S. D.s)
- Rate-Distortion Analysis

Simulations

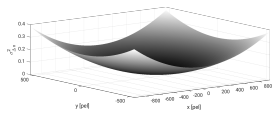
Conclusion

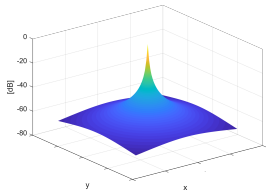
Summary: RD Theory for Affine MCP in Video Coding

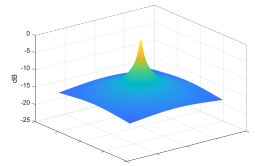
Model for affine motion compensation in video coding:

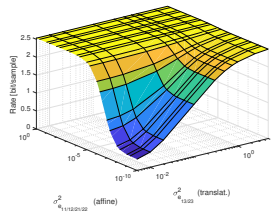
- Model the displacement estimation error as a function of the motion estimation accuracy:
- Model the p.d.f. of displacement estimation error $p_{\Delta x,\Delta Y}(\Delta x, \Delta y)$
- Model the P.S.D. of video signal S_{ss}
- Derivation of power spectral density of displacement estimation error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

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









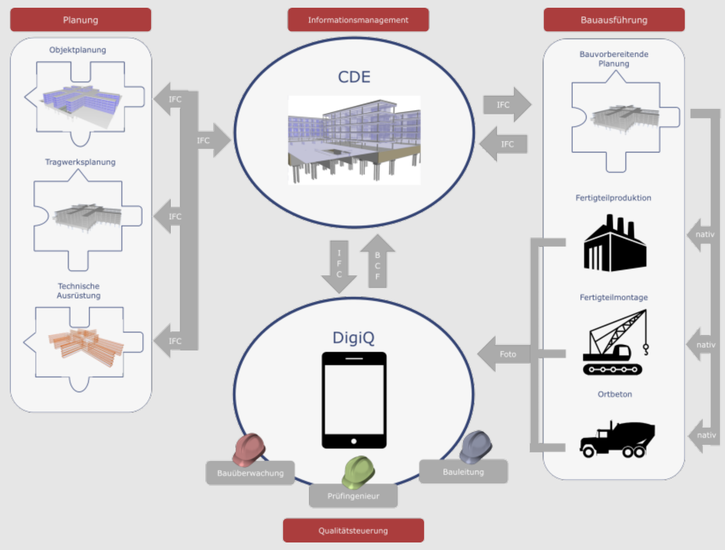
Optical Validation of Precast and Reinforced Concrete

Marco Munderloh · Armel Dedjouong · Suraja K.P. · Sascha Bahlau · Katharina Klemt-Albert · Jörn Ostermann

Approach

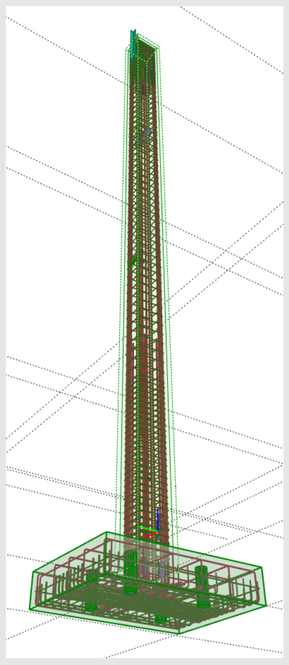
Building Information Modeling (BIM):
- Building design, project planning and construction on one common dataset
- Detailed 3D construction plans down to reinforcement bars
- Usage is spreading rapidly

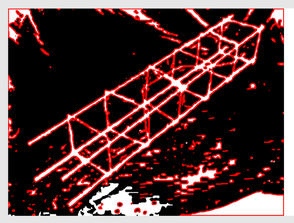
Idea:
- Automatic validation of construction to BIM
- Enrichment of plans by tolerance values and building regulations
- Usage of optical sensors and computer vision to validate construction



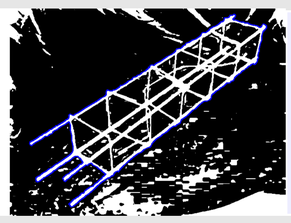
Automatic Optical Validation of Reinforcement Bars

- Image segmentation using canny edge detection
- Segmentation of object by contour detection (largest)
- Iterative optimization of external camera parameters
  - Initialization by bounding box
  - Analyses by syntheses approach
  - Rendering of 3D model using estimated camera
  - Minimize distance of every object pixel to image pixel using 2D Hausdorff metric
- Visualization of final result as augmented reality

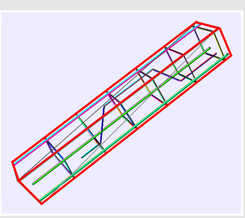




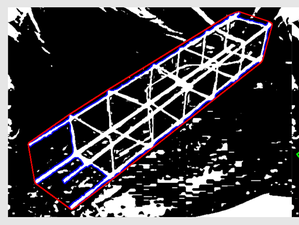
Edge detection



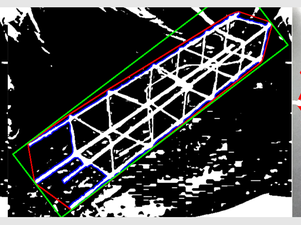
Contour extraction



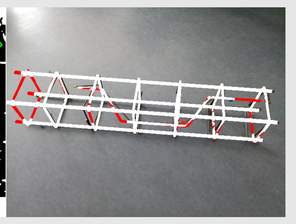
3D bounding box



Convex hull



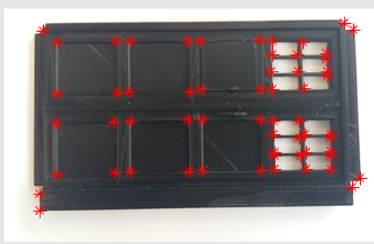
2D bounding box



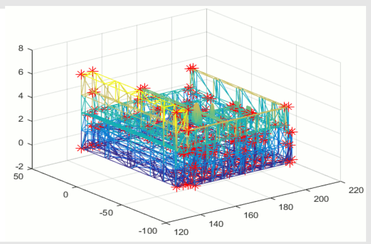
Augmented result

Automatic Optical Validation of Precast Concrete Structures

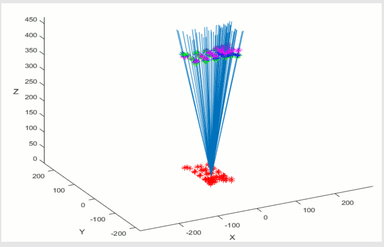
- Extraction of 2D corners from image
- Extraction of 3D corners from 3D model (3 faces w/ ~90°∡)
- Projection of 2D point cloud to 3D camera space
- Initial point cloud matching using procrustes / ICP
- Minimization of distances between 3D model points and rays from estimated camera center through 2D points
- Minimize distance of every object pixel to image pixel using 2D Hausdorff metric



2D corners



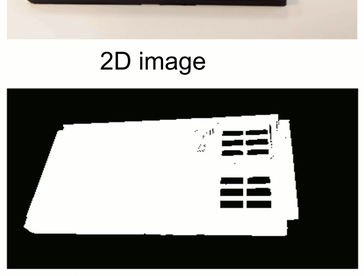
3D corners



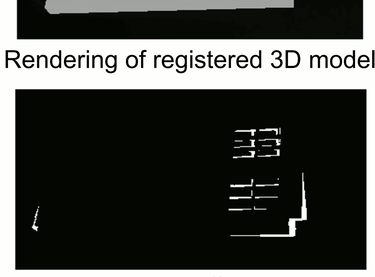
Rays through 2D points

2D image

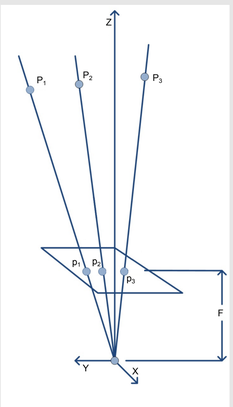
Rendering of registered 3D model



Segmented 2D image



Highlighted differences



Results

- 2D-3D registration of reinforcement bars via bounding box
- 2D-3D registration of precast concrete elements via 2D/3D corners
- Validation and difference detection using 2D Hausdorff distances
- Visualization via augmented reality and highlighting of errors

Future Work

- Test algorithms on construction site in rough environment
- Integrate algorithms into mobile demonstrator (toughbook)
- Integrate Comon Data Environment (CDE) interface for data and tolerance storage

Institut für Informationsverarbeitung
Dr.-Ing. Marco Munderloh
munderl@tnt.uni-hannover.de
https://www.tnt.uni-hannover.de/~munderl

Leibniz Universität Hannover

Video Transmission

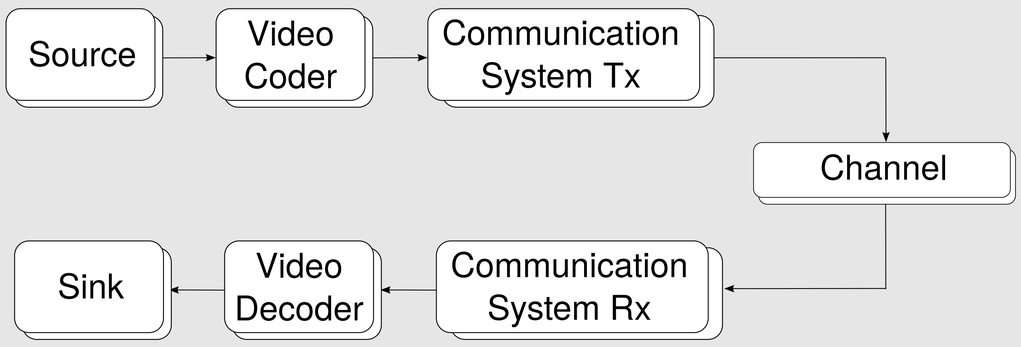
an overview of Video Compression and Communication Systems

Yasser Samayoa · Jörn Ostermann

1. Motivation

Video compression and communciation

- Real time data transmission
- Realistic channel
- Adaptive system
- Feedback channel for controlling and management



Classical separation principle

- Video (source) coding: operate closely to the rate-distortion bound
- Communication system: operate closely to the channel capacity

Assumptions
(i) long block lengths for source and channel codes
(ii) high computational resources and associated delays

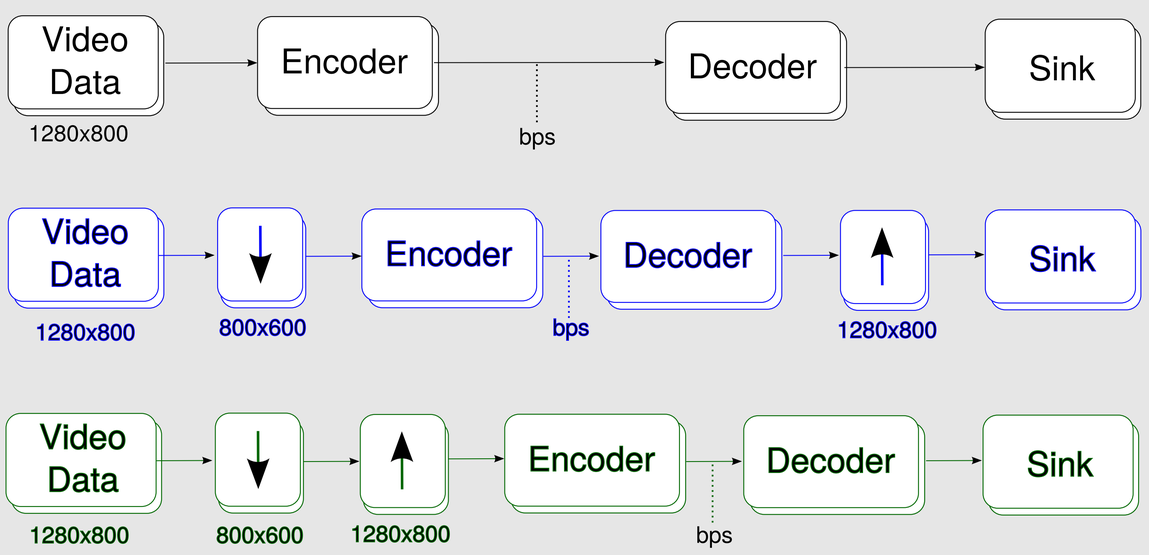
Assumptions do not hold in practice

Goal

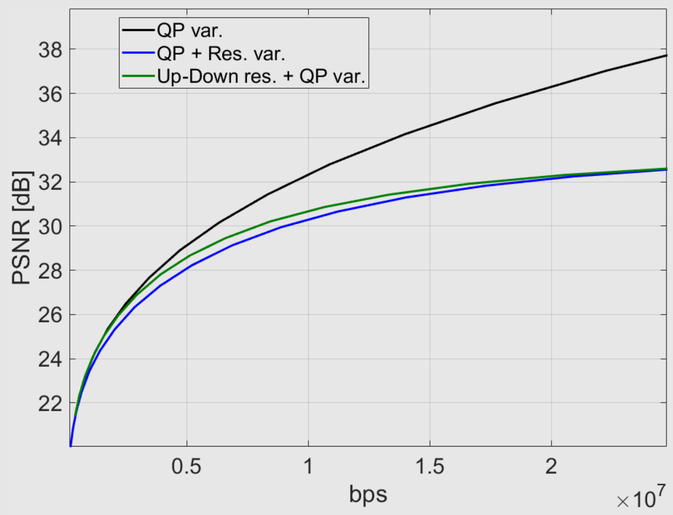
Minimize the end-to-end distortion of the delivered copy of the source under some constraints: bandwidth, transmission power or energy, delay and complexity.

2. Videocoding system: HEVC

- Adaptive parameters, e.g., space resolution and QP



- PSNR comparison

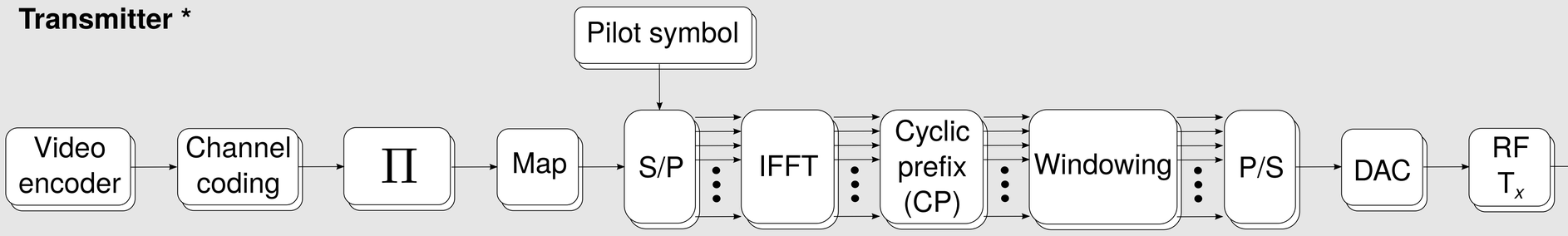


Sequence: PeopleOnStreet,
resolution: 1280x800 pix.
Fps: 50

- Other parameters, e.g., time resolution, GOP structur, etc.

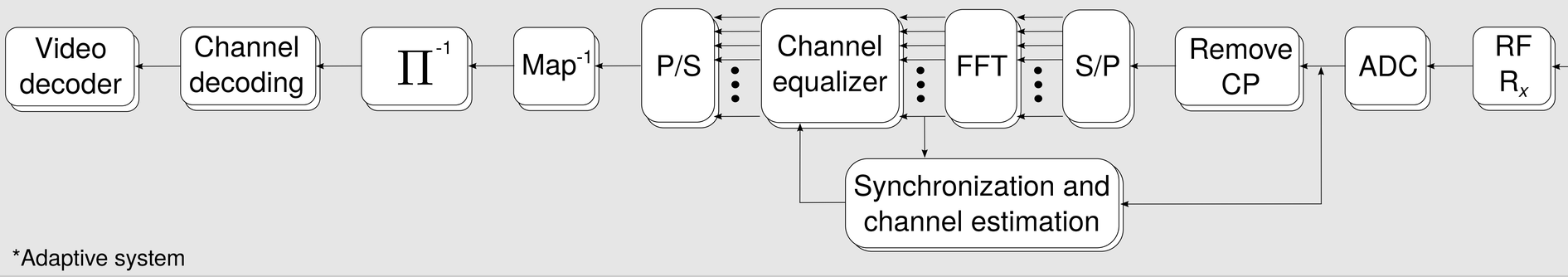
3. Communication system: OFDM block diagram

Transmitter *



Channel $h(\tau, t)$

Receiver *

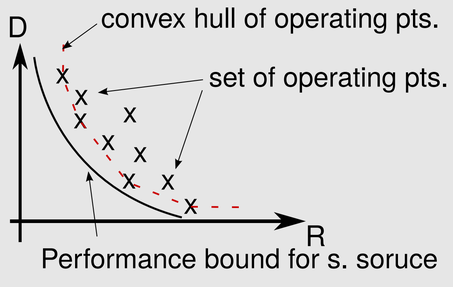


*Adaptive system

Realistic channel, i.e., time/frequency-selective

4. Optimization

- Lagrangian Optimization
- Dynamic Programming



5. Conclusions and future work

- min **Expectation[Distortion]** *subject to* {bandwidth, Tx power, delay}
  { joint paramters }
- Search for optimization procedures
- Bit-sensitivity study

Institut für Informationsverarbeitung
Yasser Samayoa
samayoa@tnt.uni-hannover.de
http://www.tnt.uni-hannover.de/∼samayoa

Leibniz Universität Hannover

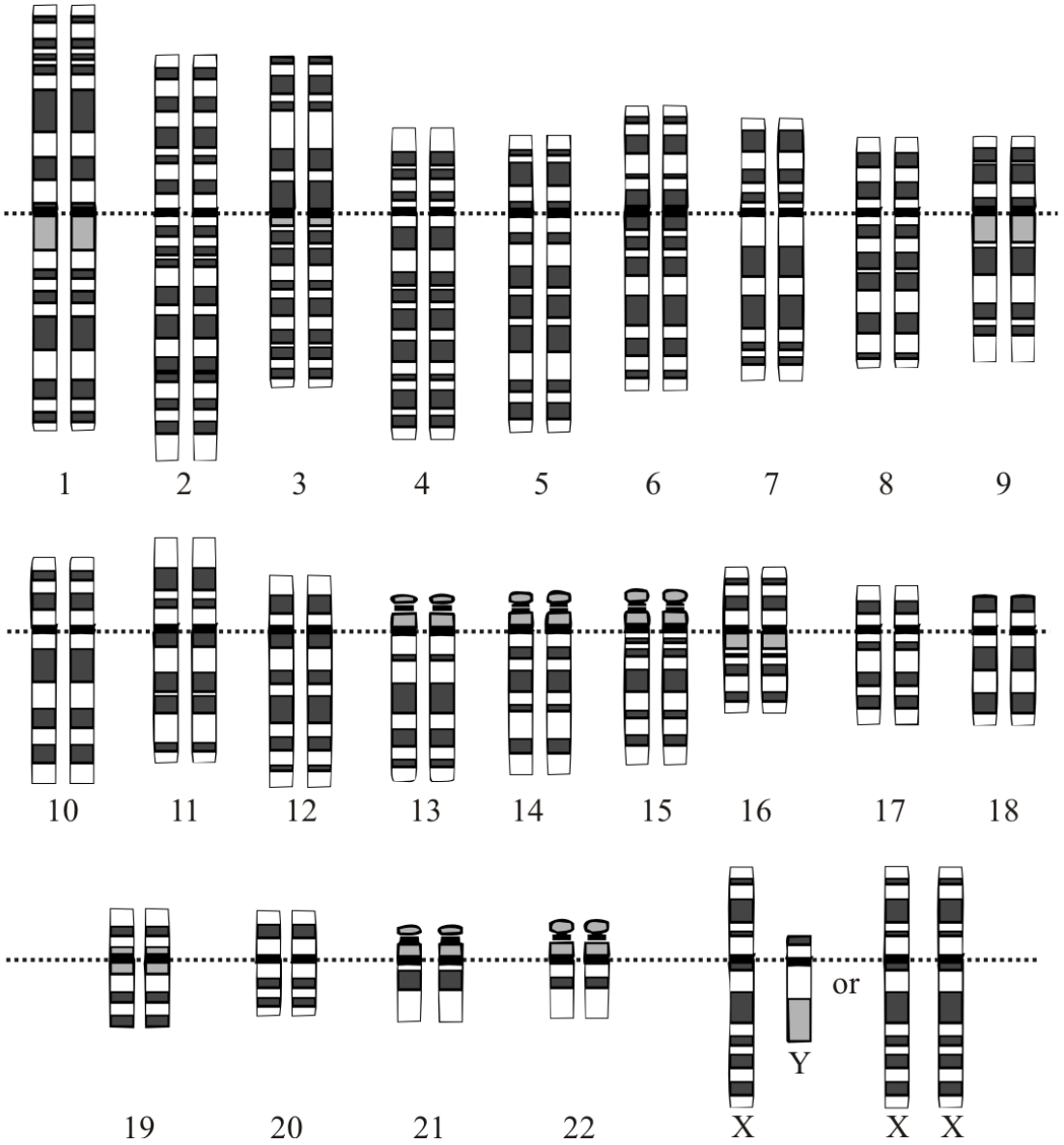
Leibniz Universität Hannover

MPEG-G: The Standard for Genomic Information Representation

Jan Voges

Institut für Informationsverarbeitung

Human karyogram



- Diploid genome
- 46 chromosomes
- 22 pairs of autosomes (1-22)
- 1 allosome pair (XY | XX)
- 4 bases (A, C, G, T)
- ~3 billion base pairs

DNA sequencing

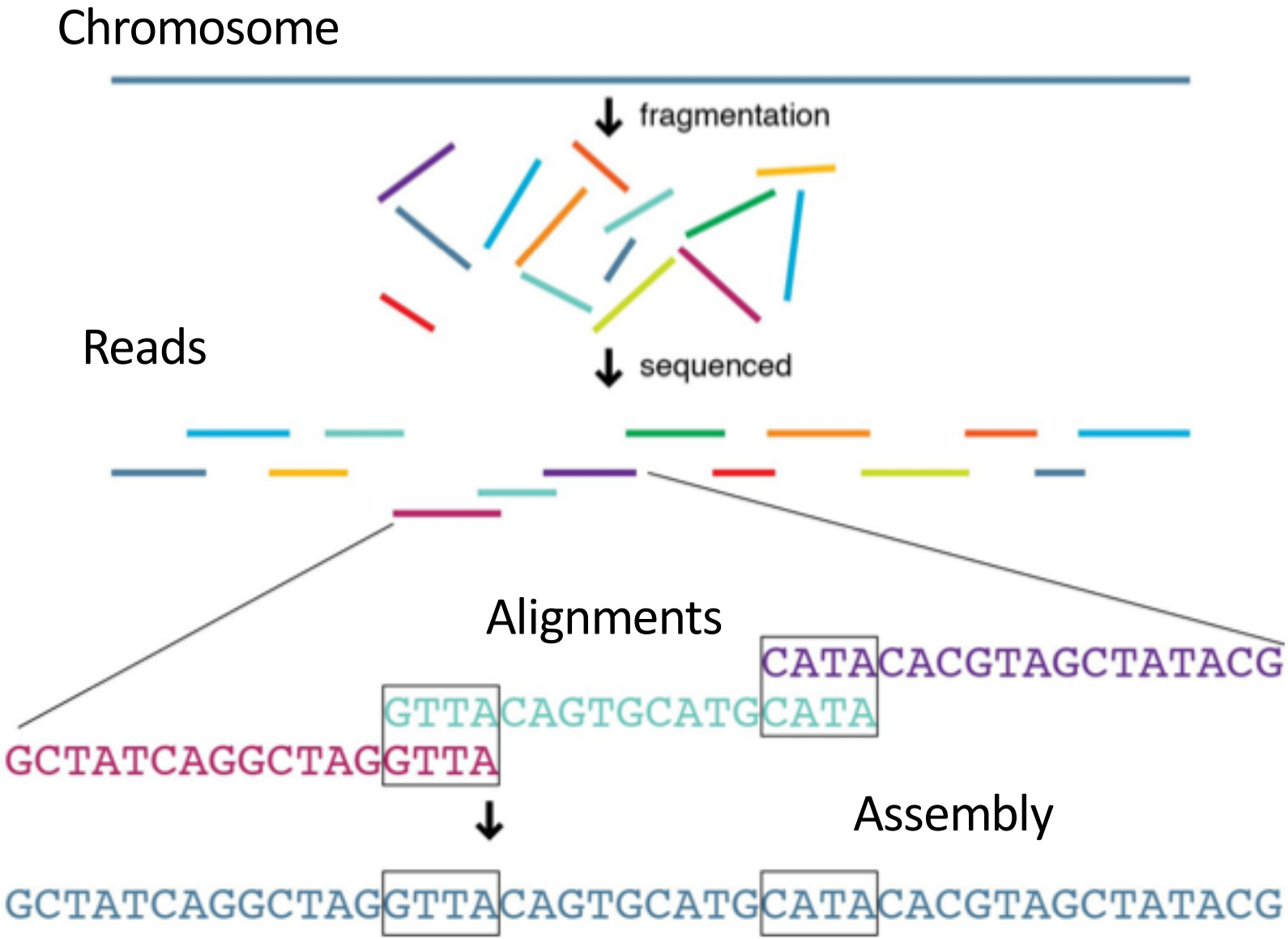
Human genome: ~3 billion base pairs x 2 bits per base = **~750 MB**

+ read-out redundancy
~500 GB

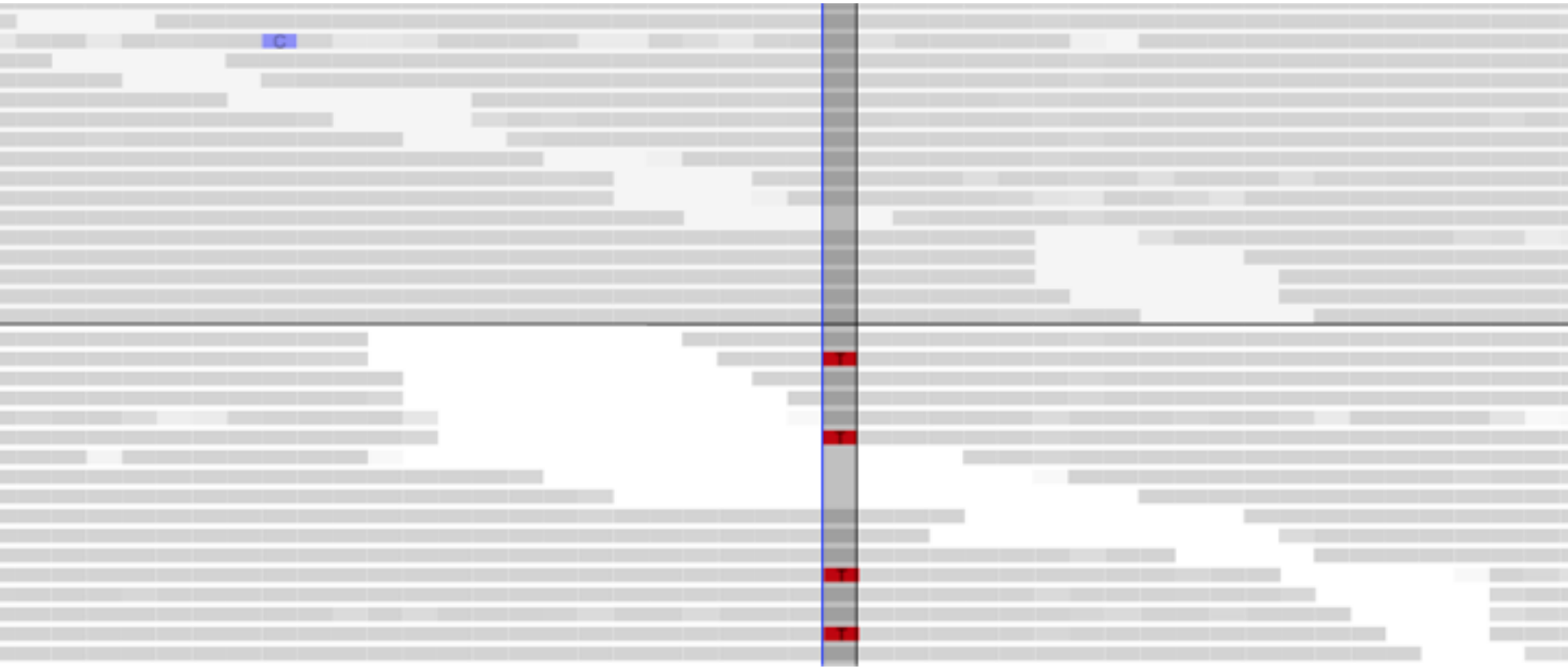
+ meta information
~1 TB

+ alignment information
~1.5 TB

Whole genome sequencing



Alignment



Evolution of genome sequencing

Sequencing technology

	2009	2017/2018
Cost/genome	$100k	$1k
Coverage	~30x	> 200x
Number of reads	~1 billion	> 6 billion
Size of raw sequencing files	~0.25 TB	> 1.5 TB

Storage & transmission infrastructure

	2009	2017/18
Cost/TB	$100	$50
Download speed	10 Mbps	100 Mbps

No technology is keeping with the pace of genome sequencing!

What is MPEG-G?

- MPEG-G = International Standard ISO/IEC 23092
- Largest coordinated and international effort addressing the problems and limitations of current technologies
- Paves the road towards a truly efficient and economical handling of genomic information
- Utilizing the latest technologies
- Planned release: end of 2018

Objectives of MPEG-G

- **Interoperable selective access to data in the compressed domain** by means of standard APIs:
  - Genomic region
  - Class of data
  - …
- On top of compression **higher performance is provided by a specific file format and transport format.**

Support the evolution of a SW/HW ecosystem so that compression technology becomes a commodity for the users.

Benefits provided by MPEG-G

- Selective access to compressed data
- Data streaming
- Genomic studies aggregation
- Enforcement of privacy rules
- Selective encryption of sequencing data and metadata
- Annotation and linkage of genomic segments in the compressed domain
- Interoperability with main existing technologies and legacy formats
- Incremental update of sequencing data and metadata
- Compressed file concatenation

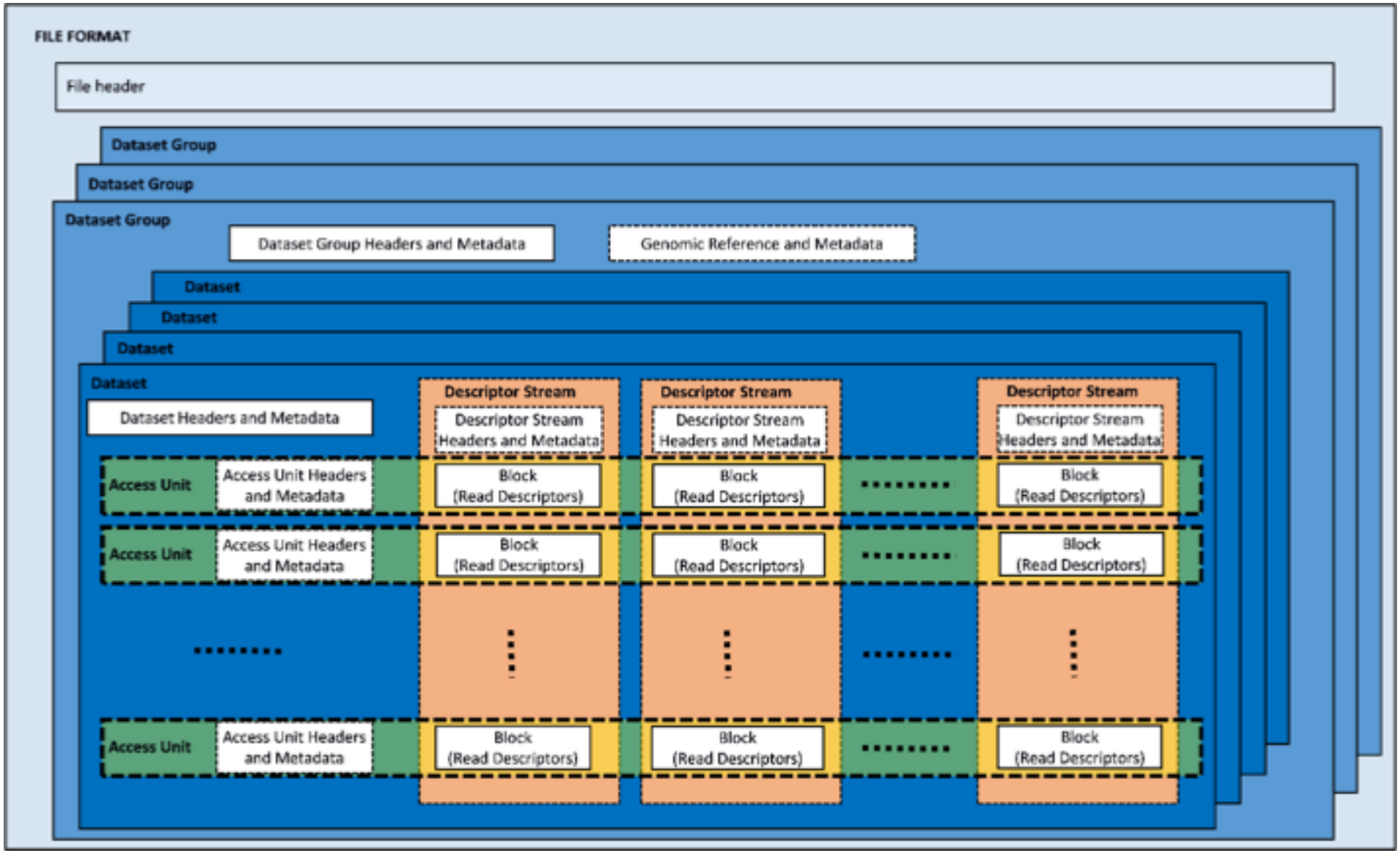
Structure of the MPEG-G standard

- **Part 1: File and Transport Format**
  - The technology to transport and access data
- **Part 2: Genomic Information Representation**
  - The compressed representation
- **Part 3: APIs**
  - The standard interfaces with genomic data applications and legacy formats
- **Part 4: Conformance**
  - The methodology to test compliance with the standard
- **Part 5: Reference Software**
  - The standard support to the implementation of applications

Structure of the MPEG-G standard

- **Part 1: File and Transport Format**
  - The technology to transport and access data
- **Part 2: Genomic Information Representation**
  - The compressed representation
- **Part 3: APIs**
  - The standard interfaces with genomic data applications and legacy formats
- **Part 4: Conformance**
  - The methodology to test compliance with the standard
- **Part 5: Reference Software**
  - The standard support to the implementation of applications

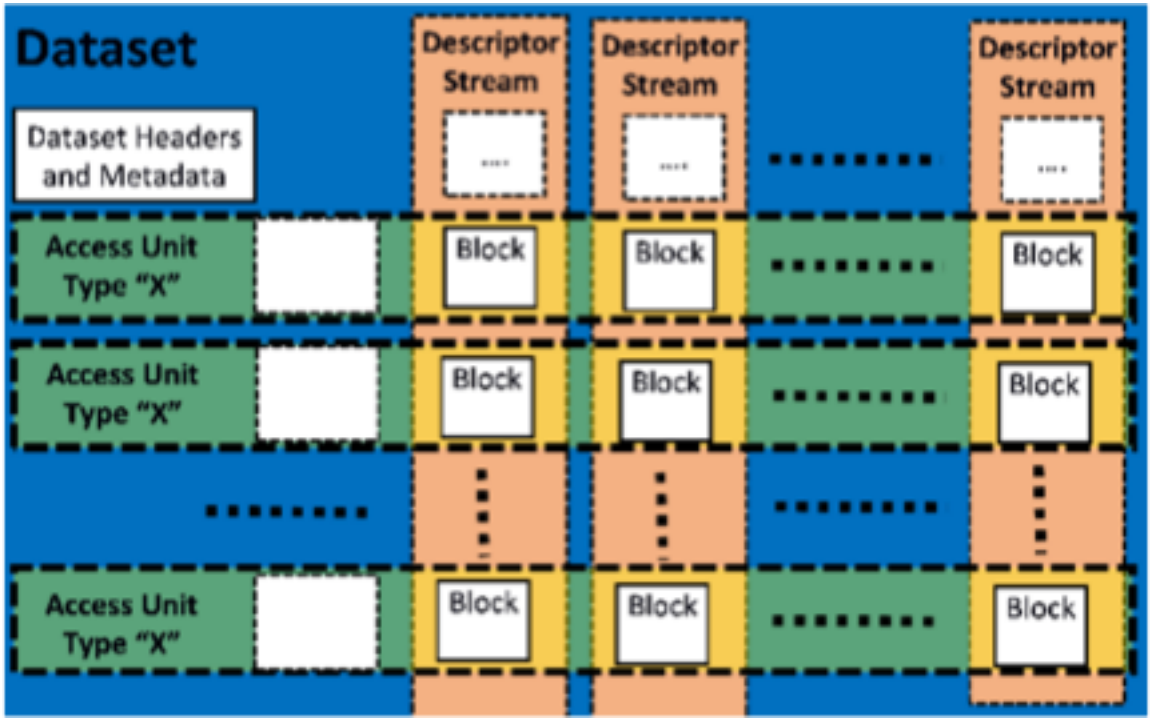
MPEG-G file format



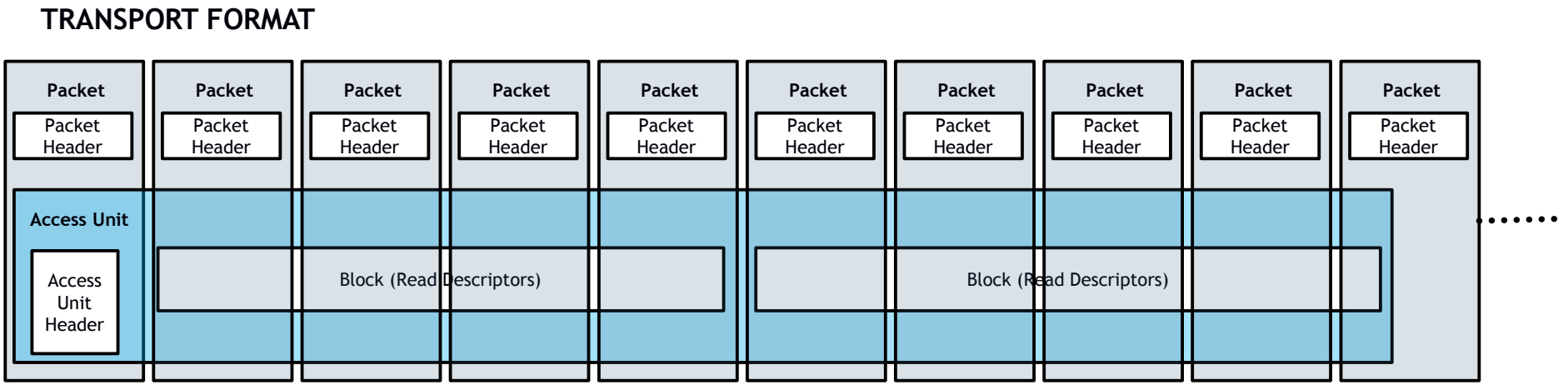
Selective access

The indexing tools embedded in an MPEG-G file enable several types of selective access that can be combined in the same query, e.g.:

- Genomic interval in terms of start to end mapping position on a given reference sequence
- Type of data (i.e., a single data class)



MPEG-G transport format / streaming



MPEG-G streaming features:

- Packet size adaptation to the channel characteristics/state
- Error detection and support of re-transmission of erroneous/incomplete data for error-free delivery
- Support of out-of-order delivery
- Packet-based filtering of genomic data
- Full convertibility of file and transport formats

Structure of the MPEG-G standard

- **Part 1: File and Transport Format**
  - The technology to transport and access data
- **Part 2: Genomic Information Representation**
  - The compressed representation
- **Part 3: APIs**
  - The standard interfaces with genomic data applications and legacy formats
- **Part 4: Conformance**
  - The methodology to test compliance with the standard
- **Part 5: Reference Software**
  - The standard support to the implementation of applications

genie

- **genie** = **GEN**omic **I**nformation **E**ncoding
- Joint collaborative effort to produce a standard-compliant open source encoder

Leibniz Universität Hannover

tnt

Stanford University

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

UPC
UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

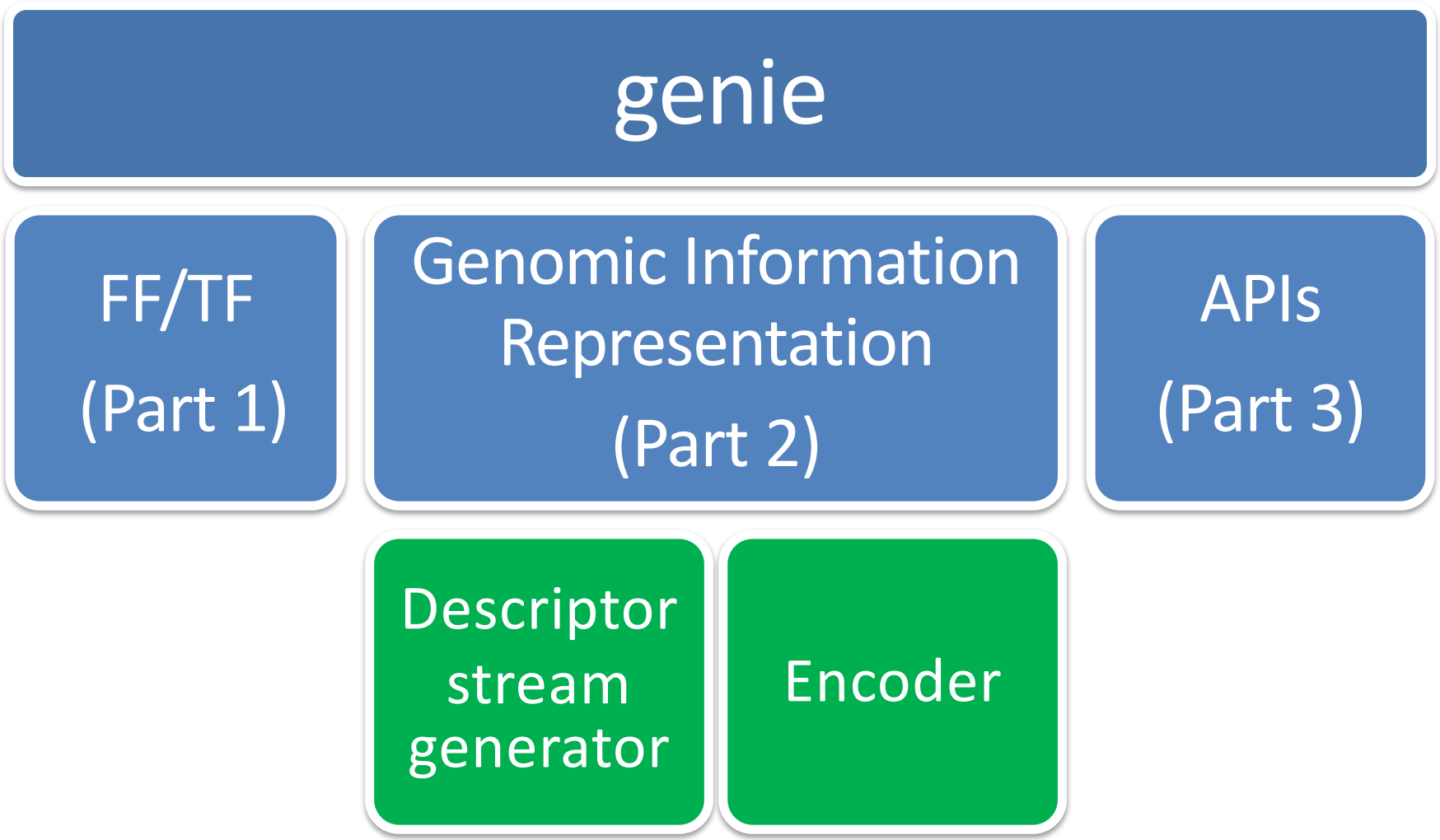
MPEG-G implementations

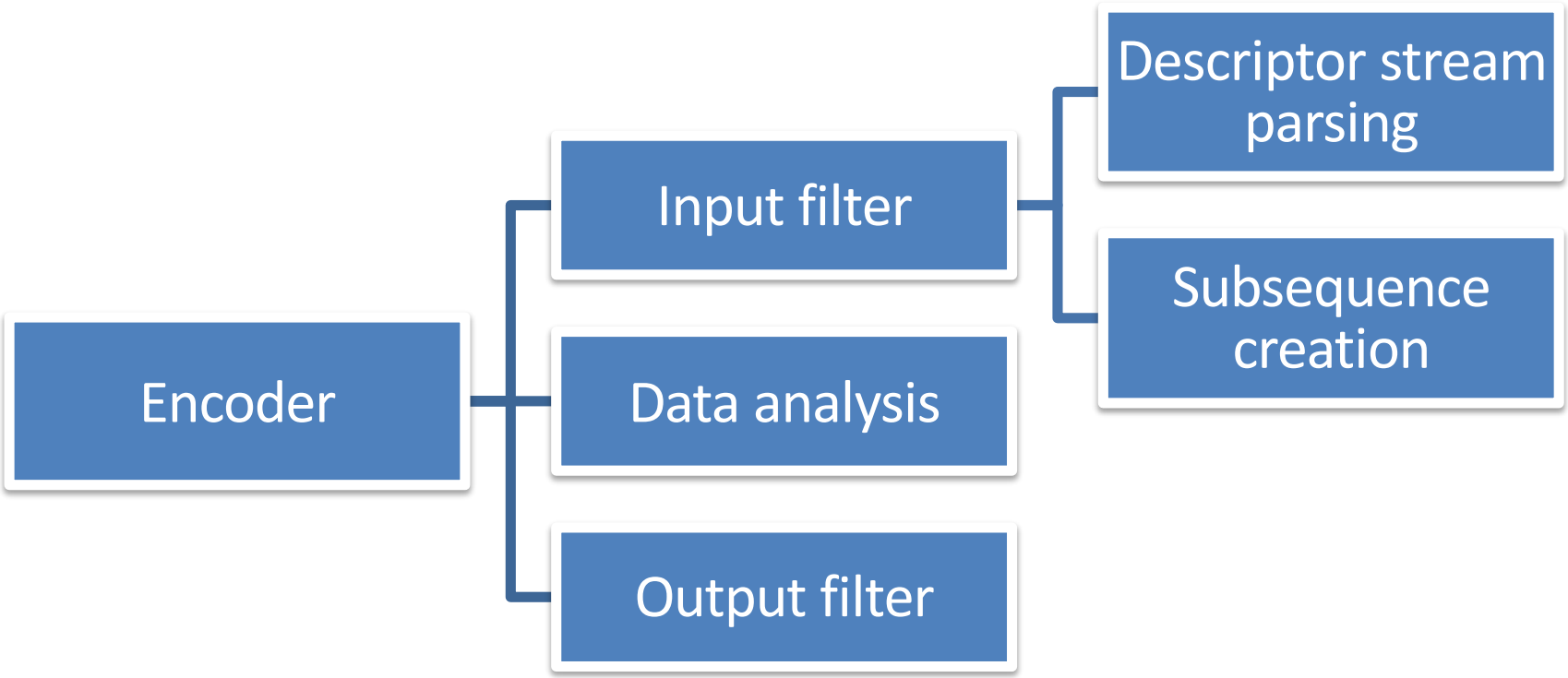
Aginome

MPEG

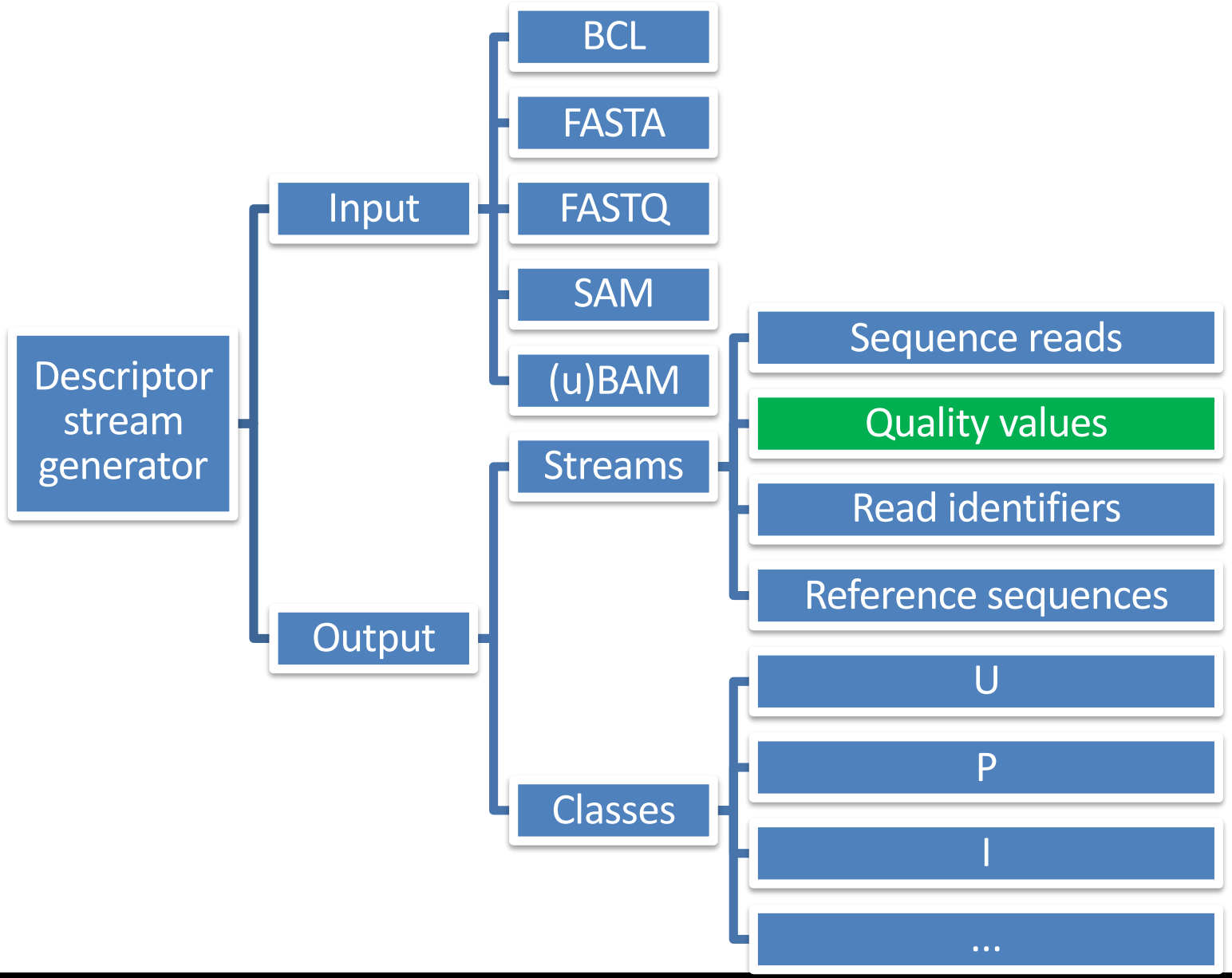
GenomSoft

GenomSys


genie
FF/TF
(Part 1)
Genomic Information
Representation
(Part 2)
APIs
(Part 3)
Descriptor
stream
generator
Encoder


Encoder
Input filter
Data analysis
Output filter
Descriptor stream parsing
Subsequence creation

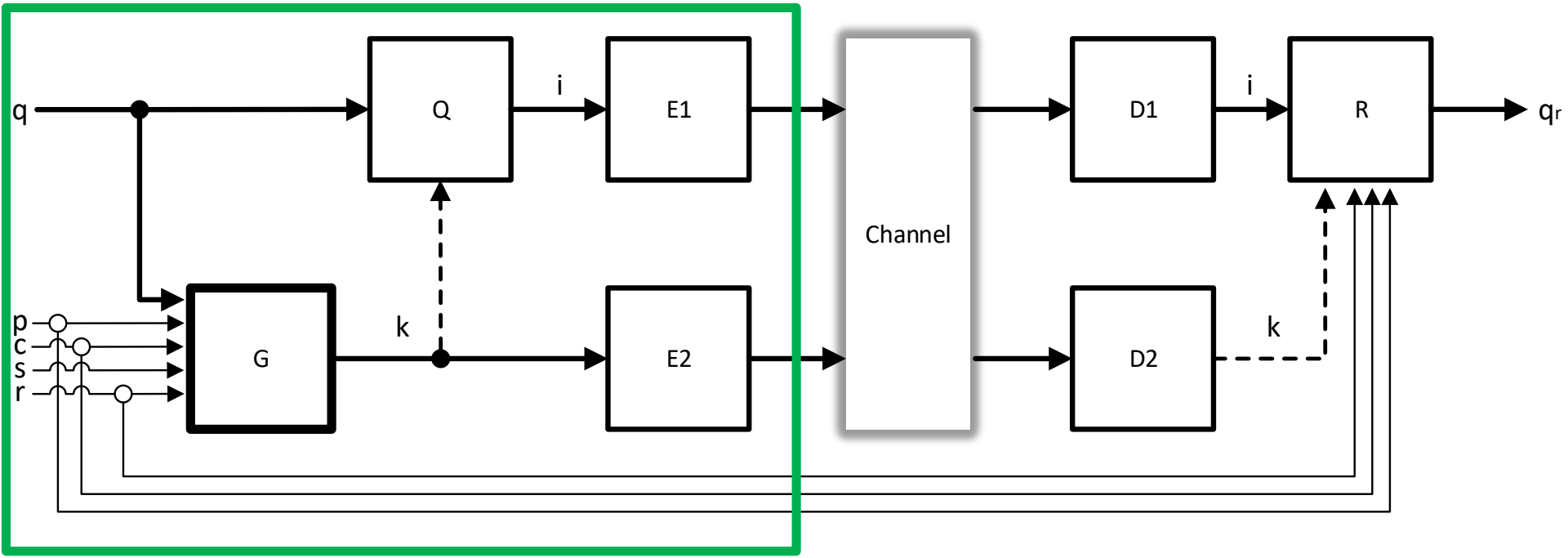
genie


Descriptor stream generator
Input
BCL
FASTA
FASTQ
SAM
(u)BAM
Output
Streams
Sequence reads
Quality values
Read identifiers
Reference sequences
Classes
U
P
I
...

Coding of quality values

Year	Tool
2011	• SlimGene (Kozanitis et al.)
2012	• SCALCE (Hach et al.)
2013	• QualComp (Ochoa et al.) • BEETL (Janin et al.) • Fastqz (Bonfield et al.)
2014	• Illumina's binning • P-/R-Block (Cánovas et a.l.)
2015	• Quartz (Yu et al.) • QVZ (Malysa et al.)
2016	• Crumble (Bonfield)
2017	• **CALQ** (Voges et al.)
2018	• QSComp (Voges et al.)

CALQ



Legend

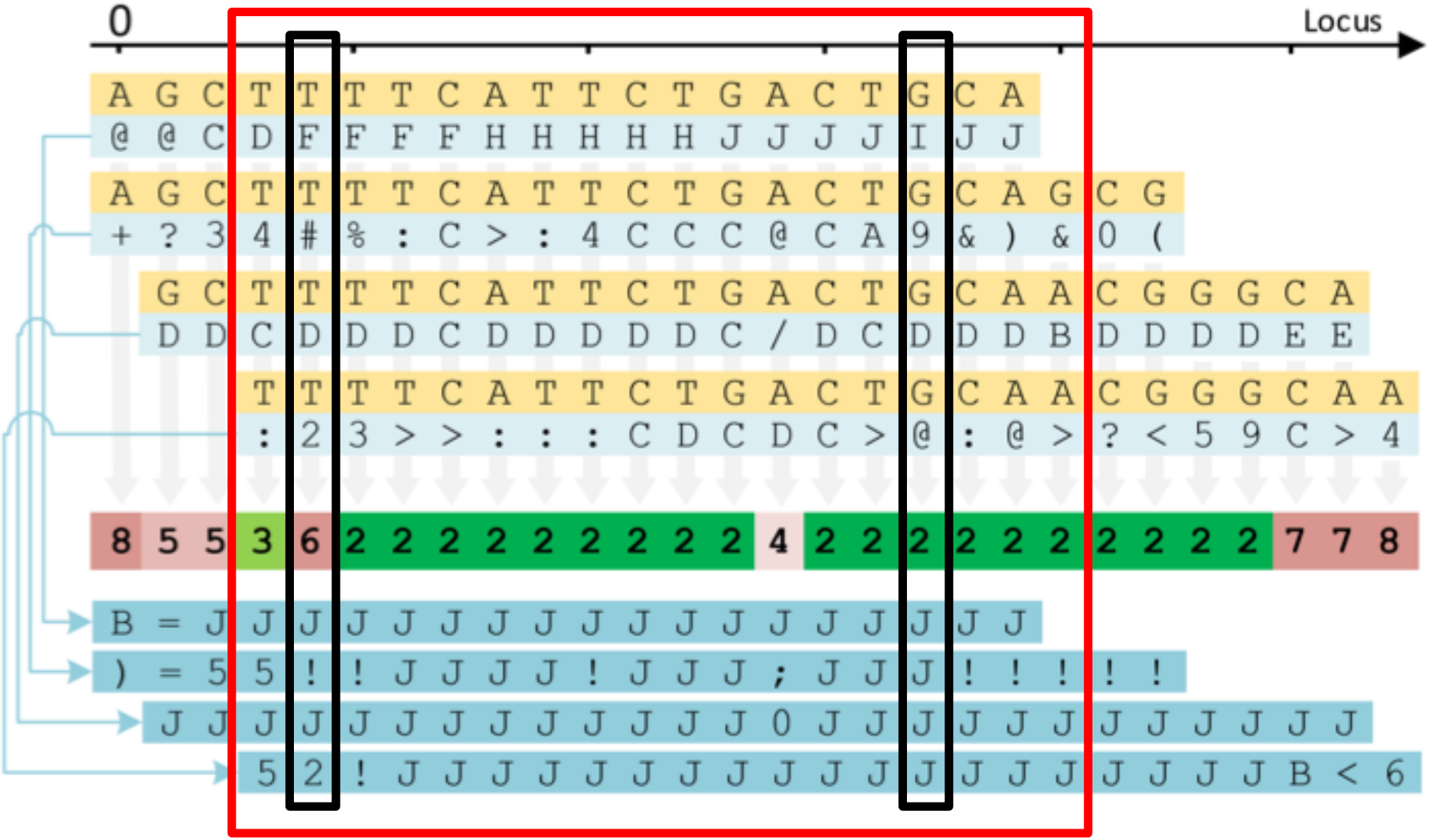
q	quality values
q_r	reconstructed quality values
p, c, s, r	side information (mapping positions, CIGAR strings, donor sequences, reference sequence(s))
k	quantizer indexes
i	quantization indexes

signal

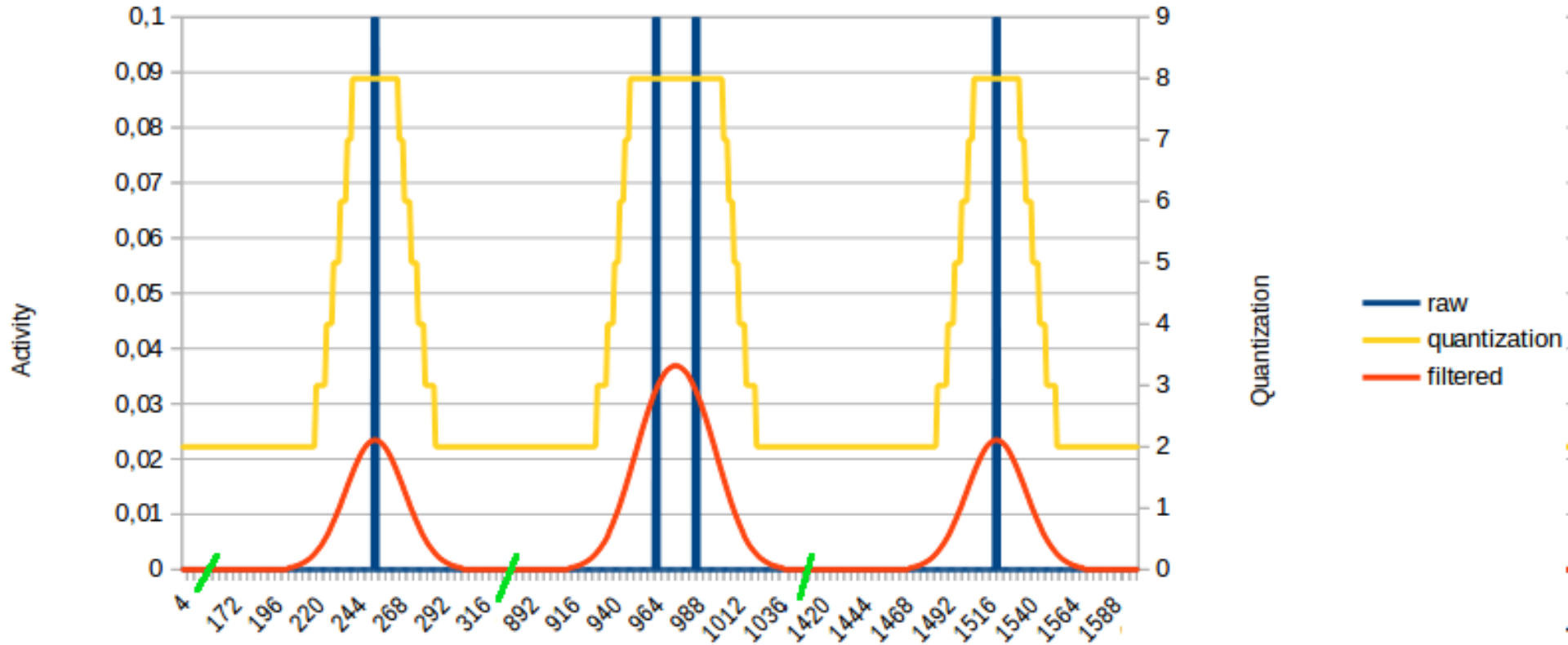
control signal

side signal

CALQ

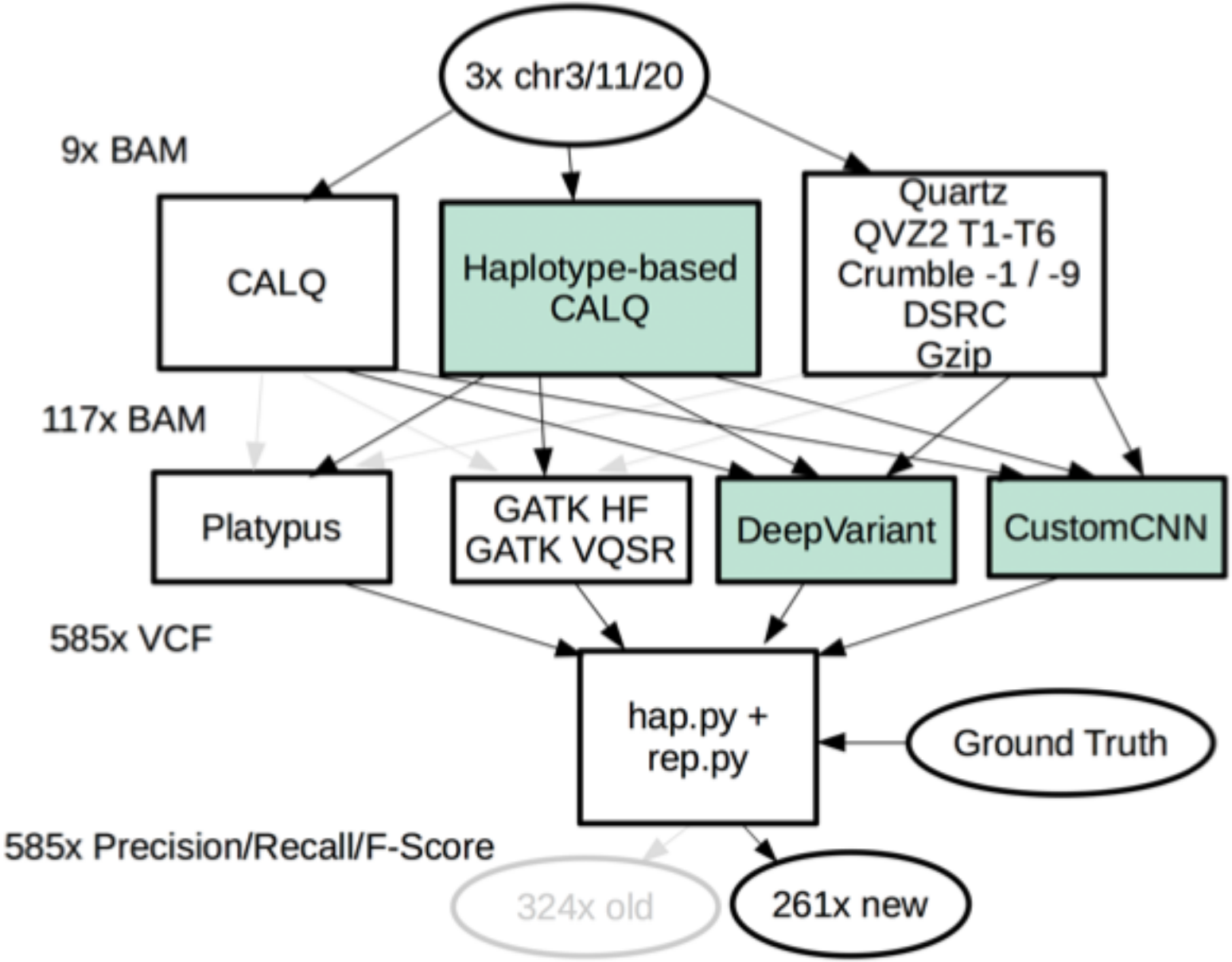

0
Locus
A G C T T T T C A T T C T G A C T G C A
@ @ C D F F F F H H H H H J J J J I J J
A G C T T T T C A T T C T G A C T G C A G C G
+ ? 3 4 # % : C > : 4 C C C @ C A 9 &) & 0 (
G C T T T T C A T T C T G A C T G C A A C G G G C A
D D C D D D C D D D D D C / D C D D D B D D D D E E
T T T T C A T T C T G A C T G C A A C G G G C A A
: 2 3 > > : : : C D C D C > @ : @ > ? < 5 9 C > 4
8 5 5 3 6 2 2 2 2 2 2 2 2 2 4 2 2 2 2 2 2 2 2 2 2 7 7 8
B = J J J J J J J J J J J J J J J J J J J
) = 5 5 ! ! J J J J ! J J J ; J J J ! ! ! ! !
J J J J J J J J J J J J J 0 J J J J J J J J J J J
5 2 ! J J J J J J J J J J J J J J J J J J B < 6

CALQ v2: sequence activity and quantization

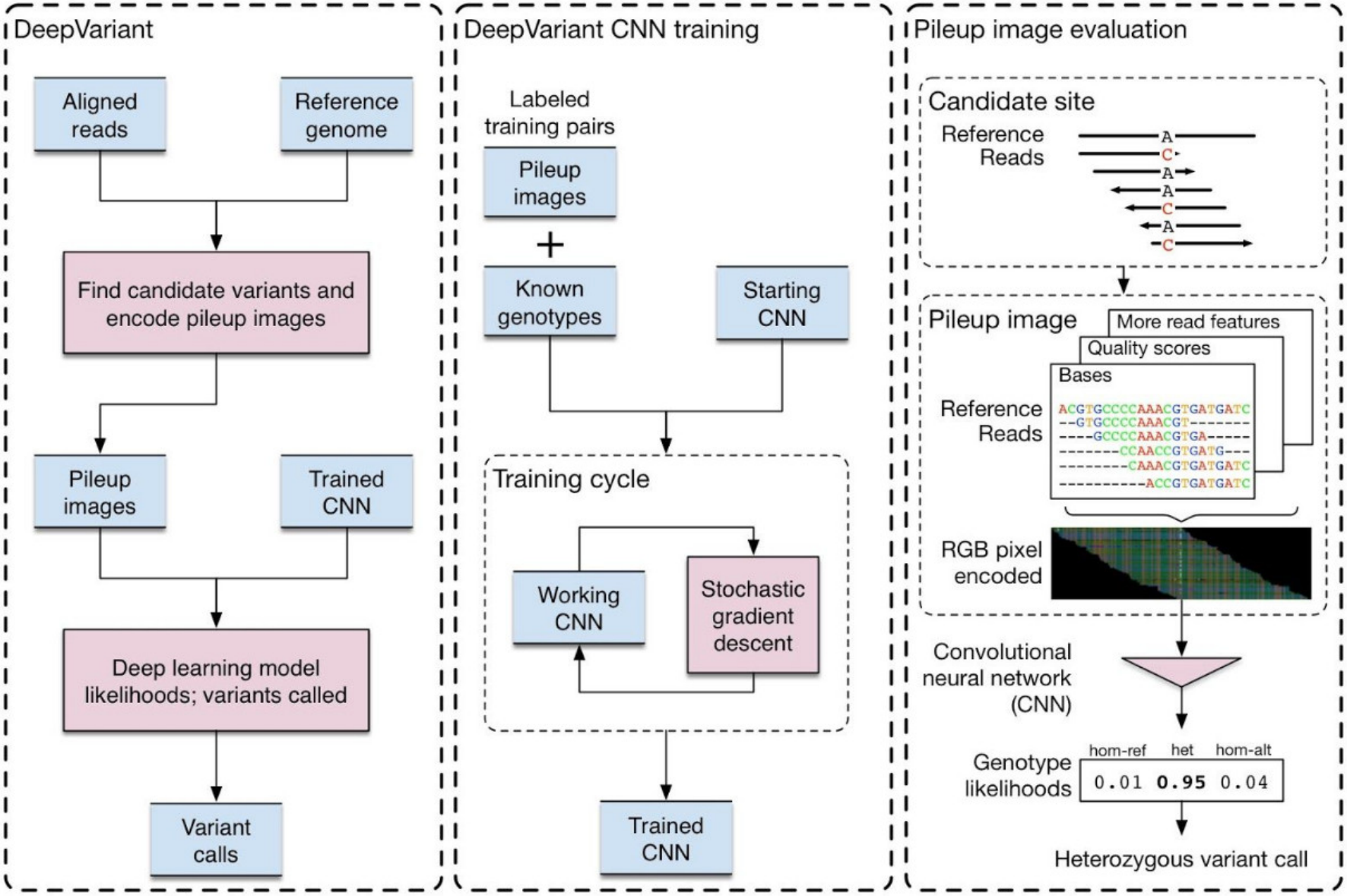


Dipl.-Ing. Jan Voges
voges@tnt.uni-hannover.de

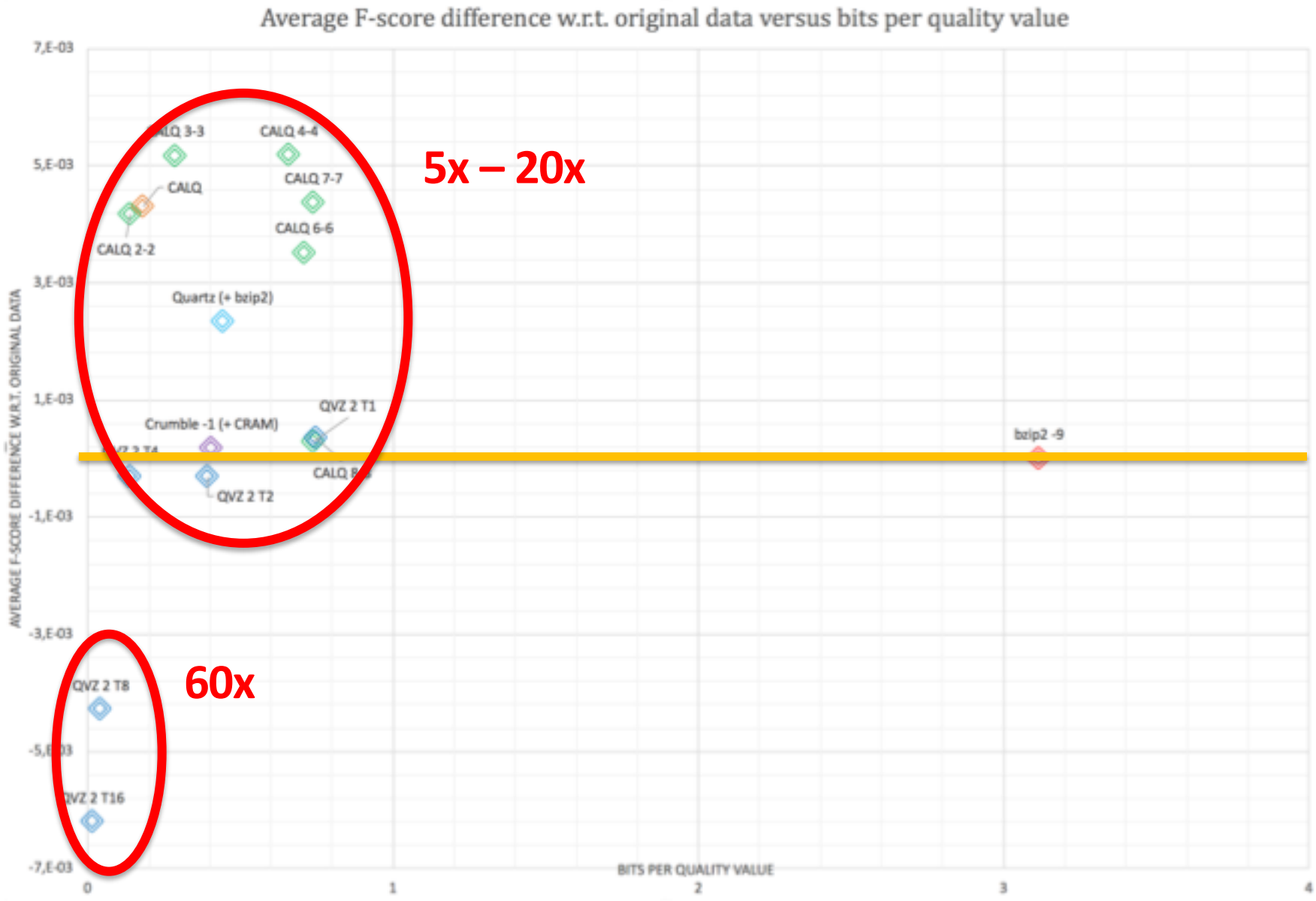
Evaluation of lossy quality value compression



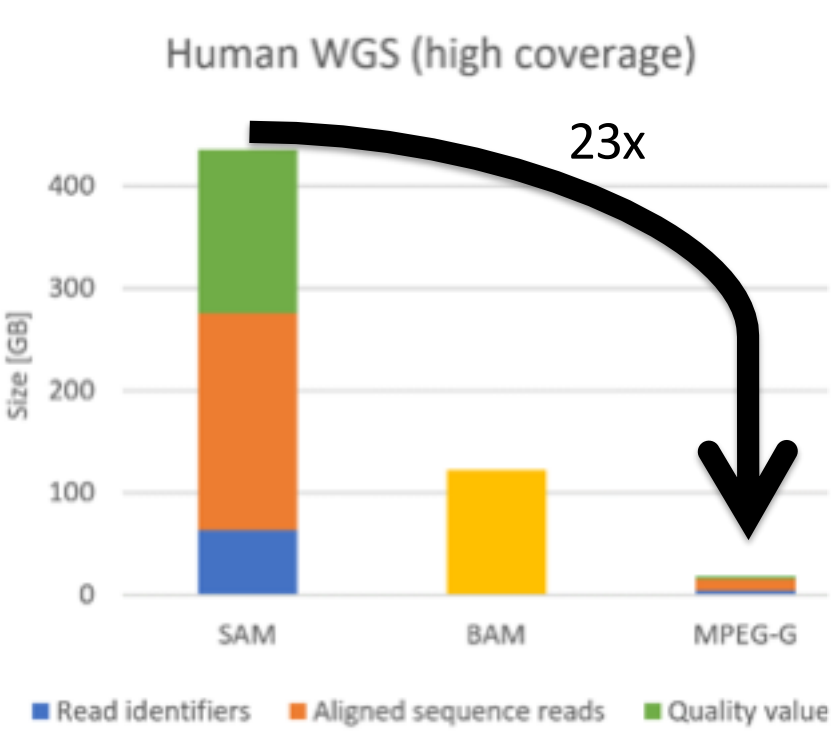
Variant calling using neural networks

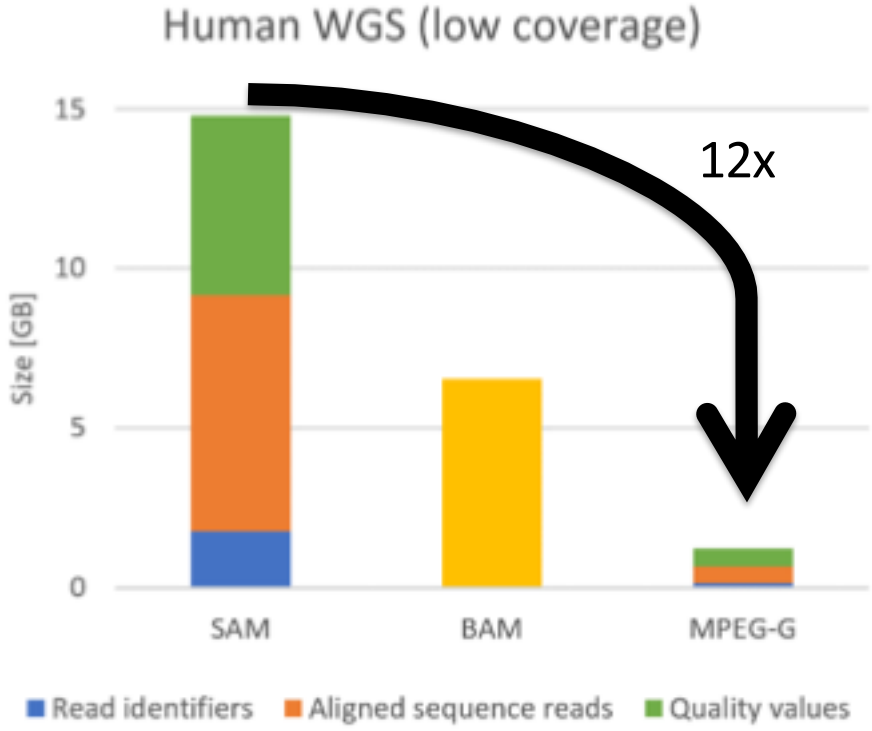


RD diagram for quality values

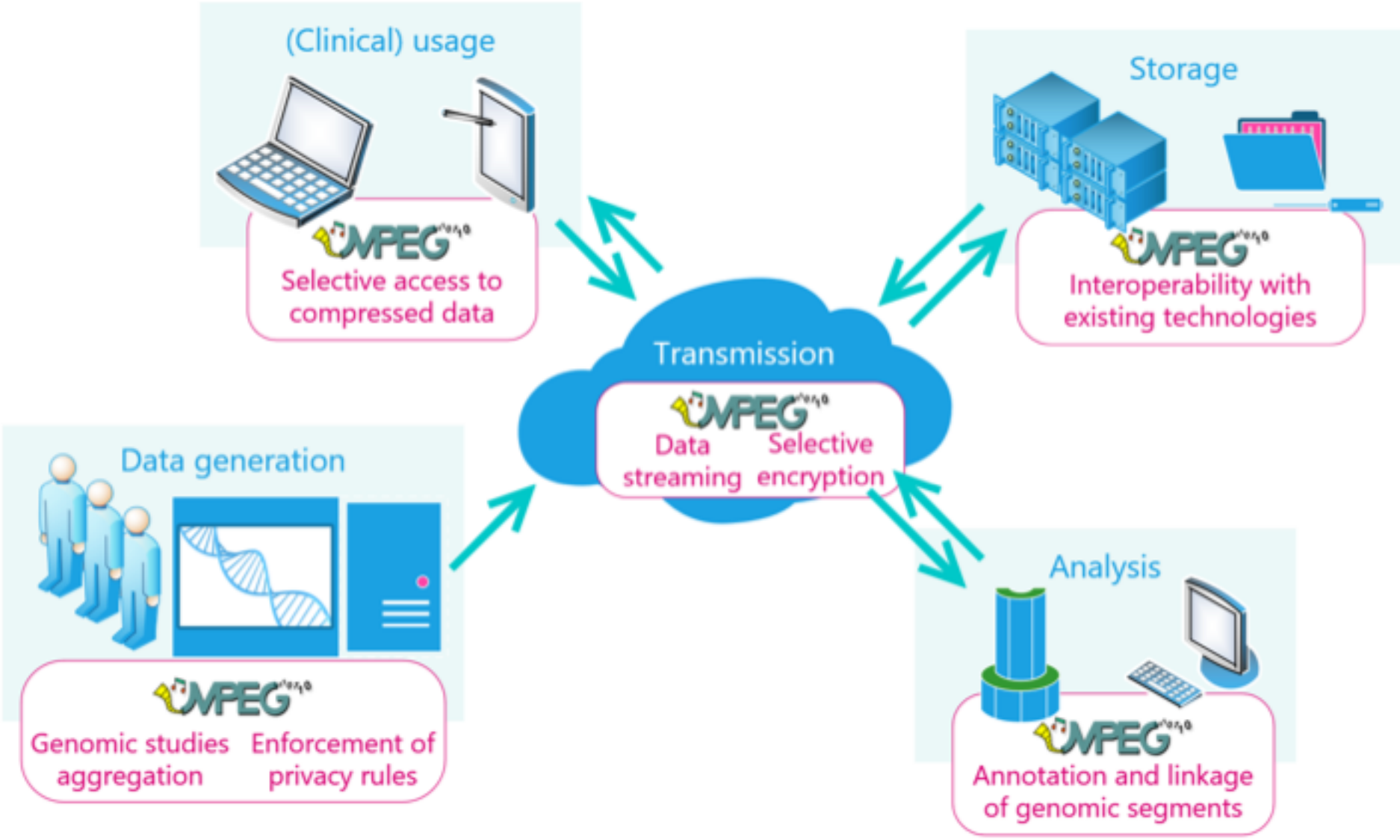


MPEG-G performance





A genomic ecosystem fueled by MPEG-G



The future of genomic data

And if we try to guess the future of genomic data:

- The technology to compress genomic information will **change over time**
- Genomic information compression performance will **improve over time**
- The MPEG-G Systems technologies will evolve and improve, but **the main functionality will stay** and support the evolution of analysis application

A new logo?