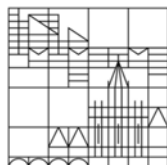


Contributions

5th ITG/VDE Summer School Video Compression and Processing (SVCP) June 17 – 19, 2019 in Konstanz



Universität
Konstanz



University of Konstanz, June 2019

Scientific Chairs

Dietmar Saupe, University of Konstanz, Germany

André Kaup, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

Jens-Rainer Ohm, RWTH Aachen University, Germany

Organizing Committee

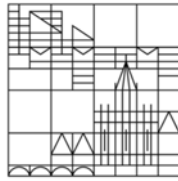
Ingrid Baiker, University of Konstanz, Germany

Hanhe Lin, University of Konstanz, Germany

Oliver Wiedemann, University of Konstanz, Germany

Organizer

Universität
Konstanz



University of Konstanz



VDE Verband der Elektrotechnik, Elektronik und Informationstechnik



SFB-TRR 161 - Quantitative Methods for Visual Computing

Oral Presentations

Artificial Neural Networks for Intra-Frame Prediction

Fabian Brand, Friedrich-Alexander-University Erlangen-Nürnberg

Design Techniques for Incremental Non-Regular Sampling Patterns

Simon Grosche, Friedrich-Alexander-University Erlangen-Nürnberg

Frequency-Selective Mesh-to-Grid Resampling

Viktoria Heimann, Friedrich-Alexander-University Erlangen-Nürnberg

Quantized and Regularized Optimization for Coding Images Using Steered Mixtures-of-Expert

Rolf Jongbloed, Technical University of Berlin

Non-linear Contour-based Multidirectional Intra Coding

Thorsten Laude, Leibniz University Hannover

Visual Quality Assessment for Motion-compensated Frame Interpolation

Hui Men, University of Konstanz

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

Holger Meuel, Leibniz University Hannover

Architectures and Training Methods for Neural Network-based Intra Prediction

Maria Meyer, RWTH Aachen University

Performance of Objective Metrics on 360VR Contents

Marta Orduna, Universidad Politécnica de Madrid

An Affine-Linear Intra Prediction with Memory Constraints

Michael Schäfer, Fraunhofer HHI Berlin

Dictionary Learning based Adaptive Resolution Change in Video Coding

Jens Schneider, RWTH Aachen

High-precision Camera Calibration for Professional Augmented-Reality Applications

Benjamin Spitschan, Leibniz University Hannover

Potential of Deep Learning in the Field of Industrial Quality Assurance

Andreas Spruck, Friedrich-Alexander-University Erlangen-Nürnberg

Optimization Strategy for MPEG-G Compliant Entropy Encoding

Jan Voges, Leibniz University Hannover

Foveated Video Coding for Real Time Streaming Applications

Oliver Wiedemann, University of Konstanz

Poster Session

Scalable Multi-Image 3D Reconstruction using Plane Sweep
Johannes Bauer, Friedrich-Alexander-University Erlangen-Nürnberg

Video Coding with Spatial Downscaling and Super-Resolution
Kristian Fischer, Friedrich-Alexander-University Erlangen-Nürnberg

Model-Based Compression of Genomic Sequences
Michael Gatzen, RWTH Aachen University

Geometrically Compensated Reference Picture Synthesis for Video Sequences with Camera Motion
Hossein Golestani, RWTH Aachen University

JND-based Video Quality Assessment and its Applications
Mohsen Jenadeleh, University of Konstanz

Content Adaptive Wavelet Lifting for Scalable Lossless Video Coding
Daniela Lanz, Friedrich-Alexander-University Erlangen-Nürnberg

MLSP-IQA: Weak Supervision for Deep Distortion-aware IQA Features
Hanhe Lin, University of Konstanz

Bit Allocation on Real Time Video Communication System over Wireless Channel
Yasser Samayoa, Leibniz University Hannover

Padding Usage Information for Geometry Padding of 360° Videos
Johannes Sauer, RWTH Aachen University

Abstracts

Invited Presentation

Christopher Schroers

Disney Research Zürich

Neural Video Processing in Post Production

Abstract: Post production pipelines for feature films are comprised of numerous complex processing steps. Nowadays, these steps are often solved with classical image processing and computer vision algorithms but deep learning based approaches offer great potential in increasing quality and efficiency. In this presentation, I will give an overview of our recent research in the space of neural video processing targeting post production tasks such as rate conversion, upscaling, denoising, and video compression

Contributed Presentations

Johannes Bauer

Friedrich-Alexander-University Erlangen-Nürnberg

Scalable Multi-Image 3D Reconstruction using Plane Sweep

Abstract: During the last decades, a lot of research has been conducted in the field of reconstructing 3D scene information from two or more images. State-of-the-art multi-view-stereo (MVS) algorithms allow for elaborate scene analysis and yield optically impressive results, but usually come at high computational complexity. Therefore, a plane sweeping approach is proposed, offering high scalability in the system size up to large-scale application, use of heterogeneous camera systems, as well as an easy trade-off between spatial resolution and computational cost. In contrast to most MVS methods, it is targeted at machine vision tasks such as object detection / tracking, where complexity and operating speed is more crucial than visual quality.

Fabian Brand

Friedrich-Alexander-University Erlangen-Nürnberg

Artificial Neural Networks for Intra-Frame Prediction

Abstract: Neural Networks are able to learn complex structures and are therefore used in many applications. Recently there has been research of how to use them for intra-frame prediction in image and video coding. This presentation will give an overview over different methods which can be found in literature, including the use of CNNs for prediction refinement, recurrent networks and dense feed forward networks. There are many new problems arising when neural networks are used instead of traditional intra prediction modes, mode-prediction being only one of them. The presentation will conclude with a few examples of my own research into this topic.

Kristian Fischer

Friedrich-Alexander-University Erlangen-Nürnberg

Video Coding with Spatial Downscaling and Super-Resolution

Abstract: Commonly, video encoders compress the video in the same resolution as the video was captured or stored. However, there are scenarios where it is feasible to subsample the video before encoding in order to reduce the number of pixels that have to be processed. Consequently, the decoded video has to be upsampled to the target resolution at the receiver side to reach the starting resolution. For this purpose, super-resolution algorithms based on neural-networks are implemented in such a coding chain with spatial up- and downscaling. By doing so, it is investigated under which circumstances such a coding chain is feasible considering high resolution videos.

Michael Gatzen

RWTH Aachen University

Model-Based Compression of Genomic Sequences

Abstract: The emergence of high-throughput sequencing technologies has greatly reduced the cost of analyzing the human genome. The vast amount of data produced by an increasing number of institutions poses a significant challenge to data storage infrastructure. Considering various efforts to employ efficient compression algorithms for genomic data, a context-adaptive arithmetic coder is used, accounting for statistical features in the underlying signal. This method is examined and later extended by a block-based framework employing prediction-based methods known from digital signal processing. The feasibility to apply these signal processing techniques to genomic data is investigated and compared to other existing compression frameworks.

Franz Götz-Hahn

University of Konstanz

Video Quality Assessment based on Multi-Level Spatially Pooled Frame Features

Abstract: The presentation shows a novel way of predicting video quality by training on features extracted from off-the-shelf (CNNS). Additionally, KonVid-150k is presented, a massive ecologically valid VQA database. Using this database, the performance of this novel deep learning-based VQA method is compared to classical feature-based no-reference VQA methods. The tradeoff between gathering more videos with fewer human judgments of quality is evaluated based on a fixed budget for annotation.

Hossein Golestani

RWTH Aachen University

Geometrically Compensated Reference Picture Synthesis for Video Sequences with Camera Motion

Abstract: In the case of camera motion, the content of a current frame could be very different from its reference pictures and consequently, it may lead to a more difficult motion compensation. The main idea of this topic is to estimate the 3D geometry of the scene captured by a monocular moving camera, and employ it in order to assist a video encoder to improve its rate-distortion performance. This goal is pursued by synthesizing virtual geometrically compensated predictions and adding them to the HEVC reference pictures lists. Our simulation results show more than 11% bitrate reduction compared to HEVC.

Simon Grosche

Friedrich-Alexander-University Erlangen-Nürnberg

Design Techniques for Incremental Non-Regular Sampling Patterns

Abstract: Non-regular sampling can be used instead of regular sampling to reduce aliasing and therefore increase the resolution per pixel. From the measured data, the missing pixels need to be reconstructed on a high-resolution grid subsequent to the acquisition. One possible application is the acquisition of scanning electron microscopy images, where non-regular acquisition allows to reduce dose and/or measurement time.

It turns out, that the actual choice of the sampling pattern has a strong influence on the reconstruction quality. Based on evaluations of less optimal sampling strategies, we will elaborate on approaches leading to optimized sampling patterns. In terms of the reconstruction method, we highlight Frequency Selective Reconstruction being well-suited for such tasks and leading to a high reconstruction quality.

Viktoria Heimann

Friedrich-Alexander-University Erlangen-Nürnberg

Frequency-Selective Mesh-to-Grid Resampling

Abstract: In many applications that are used in image processing, pixel values are mapped from a regular grid of pixel positions onto arbitrary noninteger positions, called mesh. As pixel values lying on the mesh cannot be displayed on a digital screen, the pixel values have to be resampled onto a regular grid of pixels. Therefore, Frequency-Selective Mesh-to-Grid Resampling (FSMR) is used. FSMR generates a model of weighted basis functions iteratively. However, samples on floating mesh positions lead to a severe overfitting problem as nonorthogonal weighted bases are sampled at noninteger positions. FSMR overcomes this problem by incorporating adaptively weighted initial estimates.

Mohsen Jenadeleh

University of Konstanz (until 2018)

JND-based Video Quality Assessment and its Applicants

Abstract: We will discuss the challenges and choices for subjective evaluation of a large-scale authentically distorted video dataset using just-noticeable-difference (JND) methodology. Such a database will enable developing objective methods for accurate JND estimations of video sequences acquired in unconstrained environment. Also, since, one of the main applications of the JND-based quality assessment is video coding, we will discuss the applications of such JND-based video dataset for devising new approaches and technologies for the compression of videos using machine learning approaches and their potential to produce more accurate and visually pleasing video frame reconstructions at a higher compression rate.

Rolf Jongeblod

Technical University of Berlin

Quantized and Regularized Optimization for Coding Images Using Steered Mixtures-of-Expert

Abstract: Compression algorithms that employ Mixtures-of-Experts depart drastically from standard hybrid block-based transform domain approaches as in JPEG and MPEG coders. In previous works we introduced the concept of

Steered Mixtures-of-Experts (SMoEs) to arrive at sparse representations of signals. SMoEs are gating networks trained in a machine learning approach that allow individual experts to explain and harvest directional long-range correlation in the N-dimensional signal space. Previous results showed excellent potential for compression of images and videos but the reconstruction quality was mainly limited to low and medium image quality. In this paper we provide evidence that SMoEs can compete with JPEG2000 at mid- and high-range bit-rates. To this end we introduce a SMoE approach for compression of color images with specialized gates and steering experts. A novel machine learning approach is introduced that optimizes RD-performance of quantized SMoEs towards SSIM using fake quantization. We drastically improve our previous results and outperform JPEG by up to 42%.

Daniela Lanz

Friedrich-Alexander-University Erlangen-Nürnberg

Content Adaptive Wavelet Lifting for Scalable Lossless Video Coding

Abstract: Wavelet-based video coding decomposes an input sequence into a lowpass and a highpass subband by filtering along the temporal axis. So far, the number of total decomposition levels is determined for the entire input sequence in advance. However, if the motion in the video sequence is strong or if abrupt scene changes occur, a further decomposition leads to a low-quality lowpass subband. Therefore, we propose a content adaptive wavelet transform, which locally adapts the depth of the decomposition to the content of the input sequence.

Thorsten Laude

Leibniz University Hannover

Non-linear Contour-based Multidirectional Intra Coding

Abstract: Intra coding is an essential part of all video coding algorithms and applications. Additionally, intra coding algorithms are predestined for an efficient still image coding. To overcome limitations in existing intra coding algorithms (such as linear directional extrapolation, only one direction per block, small reference area), we propose non-linear Contour-based Multidirectional Intra Coding (COMIC). This coding mode is based on four different non-linear contour models, on the connection of intersecting contours, and on a boundary recall-based contour model selection algorithm. The different contour models address robustness against outliers for the detected contours and evasive curvature changes. Additionally, the information for the prediction is derived from already reconstructed pixels in neighboring blocks. The achieved coding efficiency is superior to those of related works from the literature. Compared to the closest related work, BD rate gains of 2.16% are achieved on average.

Hanhe Lin

University of Konstanz

MLSP-IQA: Weak Supervision for Deep Distortion-Aware IQA Features

Abstract: Current artificially distorted image quality assessment (IQA) databases are small in size and limited in content. To address the limitation, we create two datasets, the Konstanz Artificially Distorted Image quality Database (KADID-10k) and the Konstanz Artificially Distorted Image quality Set (KADIS-700k). The former contains 81 pristine images, each degraded by 25 distortions in 5 levels. The latter has 140,000 pristine images, with 5 degraded versions each, where the distortions are chosen randomly. We conduct a subjective IQA crowdsourcing study on KADID-10k to yield 30 degradation category ratings (DCRs) per image. We propose a novel deep learning no-reference IQA method that make use of KADID-10k and KADIS-700k by means of weakly supervised learning.

Hui Men

University of Konstanz

Visual Quality Assessment for Motion-compensated Frame Interpolation

Abstract: Current benchmarks for optical flow algorithms evaluate the estimation quality by comparing their predicted flow field with the ground truth, and additionally may compare interpolated frames, based on these predictions, with the correct frames from the actual image sequences. For the latter comparisons, objective measures such as mean square errors are applied. However, for applications like image interpolation, the expected user's quality of experience cannot be fully deduced from such simple quality measures. Therefore, we conducted a subjective quality assessment study by crowdsourcing for the interpolated images provided in one of the optical flow benchmarks, the Middlebury benchmark. Our result shows the necessity of visual quality assessment as another evaluation metric for optical flow and frame interpolation benchmarks.

Holger Meuel

Leibniz University Hannover

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

Abstract: The minimum bit rate for encoding the prediction error in affine motion compensated video coding is derived. For that, the probability density function of the displacement estimation error is calculated as a function of the affine motion parameter estimation errors. The rate-distortion theory is derived and evaluated to determine the minimum bit rate for encoding the prediction error, taking into account the power spectrum density of real image signals. The theoretic findings

are compared to real-world measurements and conclusions for the accuracy of affine motion compensation in video coding are drawn.

Maria Meyer

RWTH Aachen University

Architectures and Training Methods for Neural Network-based Intra Prediction

Abstract: It has been shown recently, that neural networks can improve video intra prediction significantly. Within the last year we therefore further investigated, which network architecture, training method and data is most suitable for this application. This included analyzing the benefits of including cross-component information for chroma prediction and reducing the computational overhead by pruning the applied networks. Likewise, it was shown to be beneficial to train with a transform domain loss function, a combination of coded and uncoded data and a reduced number of low variance samples.

Marta Orduna

Universidad Politécnica de Madrid

Performance of Objective Metrics on 360VR Contents

Abstract: The presentation shows the performance of different video objective metrics on 360VR contents. Through a complete set of tests, we evaluate the behavior of the selected objective metrics looking for the linearity between the subjective scores and the objective outcomes. As a particular case, we are interested in showing the results of Video Multimethod Assessment Fusion (VMAF) to 360VR contents, a full reference metric developed by Netflix initially designed to work with traditional 2D contents. Therefore, through a complete set of tests, we prove that this metric can be successfully used without any specific training or adjustments to obtain the quality of 360VR sequences actually perceived by users.

Yasser Samayoa

Leibniz University Hannover

Bit Allocation on Real Time Video Communication System over Wireless Channel

Abstract: Good performance at a high data rate has become a constant growing prerequisite for deploying video communication systems. Video communication over rate-limited and error-prone wireless channels requires both a high error resilience and high compression solutions. The development of flexible, near-instantaneously adaptive scheme capable of maintaining an acceptable video quality regardless of the channel quality encountered will be the main goal of the talk.

Johannes Sauer

RWTH Aachen University

Padding Usage Information for Geometry Padding of 360° Videos

Abstract: Geometry padding of 360° videos in cube based projections requires reprojection of pixels from neighboring cube faces. Doing so on-the-fly changes the en/decoder at a block level which is not desirable. Applying the padding at a high level (reference picture) can generate pixels which are not actually required by motion compensation. To avoid this inefficiency we add a high level signaling of geometry padding usage information using an SEI message.

Michael Schäfer

Fraunhofer HHI Berlin

An Affine-Linear Intra Prediction with Memory Constraints

Abstract: The author presents a novel method for a data-driven training of neural networks for intra picture prediction. The resulting predictors are affine-linear and use subband decomposition of the input and output samples. Thereby, the architecture allows to share one set of weights across different block shapes. Furthermore, the number of multiplications does not exceed eight per sample to predict. During the training, a loss function modelling the bitrate of the DCT-transformed residuals is used. The obtained predictors are incorporated into the Versatile Video Coding Test Model 4. All Intra BD-rate savings up to 1.2 % across different resolutions are reported.

Jens Schneider

RWTH Aachen University

Dictionary Learning based Adaptive Resolution Change in Video Coding

Abstract: The concept of dynamic resolution change is well known from MPEG-4. However, in MPEG-4 linear filters are used for the upsampling, which is a crucial to coding video at varying resolution. With the rise of machine learning based super resolution methods in the last decade, powerful algorithms outperforming conventional upsampling were developed. This contribution introduces a dynamic resolution change concept for intra frames using a dictionary learning based upsampling method. Thereby, the encoder decides on the CTU-level whether the original CTU or a downsampled version should be coded. Simulation results show that gains with respect to VTM-3.0 reference software can be achieved.

Benjamin Spitschan

Leibniz University Hannover

High-precision Camera Calibration for Professional Augmented-Reality Applications

Abstract: Camera calibration is crucial to most augmented reality (AR) systems. While powerful self-calibration methods are available, professional AR applications such as laparoscopic surgery or assisted industrial maintenance require highest calibration accuracy. Conventional target-based calibration is commonly chosen here. In cases, however, where the camera system has a shallow depth-of-field, calibration must be carried out with strongly blurred images of the target. A robust marker detection method for calibration patterns is presented that is able to cope with strong optical blur, noise, and other perturbations that occur during the imaging process.

Andreas Spruck

Friedrich-Alexander-University Erlangen-Nürnberg

Potential of Deep Learning in the Field of Industrial Quality Assurance

Abstract: With the recent advances in the field of production engineering the need for automated inspection systems rises, as more complex parts can be manufactured, which approach the failure limit quite close. With the progress in machine learning and within the scope of Industry 4.0 the use of deep learning techniques for the inspection of produced items bears a great potential. This novel approach bears the benefit of a very flexible system while existing infrastructure might be reused. By this the presented approach is also appealing for small companies, as the roll-out costs can be kept low.

Jan Voges

Leibniz University Hannover

Optimization Strategy for MPEG-G Compliant Entropy Encoding

Abstract: The research field of genomics and DNA sequencing in particular has made great progress in recent years. The comprehensive use of high-throughput technologies for DNA sequencing opens up new perspectives in the treatment of diseases and enables personalized medicine on unprecedented scales. Since DNA sequencing technologies produce extremely large amounts of raw data, the costs for storing, transmitting and processing sequencing data are very high. To facilitate the widespread use of DNA sequencing at acceptable costs, international standardization organizations developed the MPEG-G standard. The MPEG-G compression pipeline consists of three stages: classification of the input data into clusters, further splitting of the clusters into independent streams, and entropy encoding. The entropy coding in MPEG-G can be configured by many parameters, which results in at least one billion potential combinations for a given input stream. The choice of parameters is a crucial step as it has a high impact on the resulting bitrate. Trying all possible combinations is unfeasible because this would require an entire encoding of the input stream for each combination. I present a method for designing an MPEG-G compliant entropy encoder which balances encoder complexity and encoder efficiency.

Oliver Wiedemann

University of Konstanz

Foveated Video Coding for Real Time Streaming Applications

Abstract: Video streaming with strict real-time constraints is gaining popularity in academic research and in consumer applications such as cloud gaming. Scenarios where future frames are dependent on e.g. user feedback and thus unavailable to the encoder prohibit the application of modern bidirectional coding schemes. We present a framework that utilizes live eye-tracking data in a foveated region-of-interest coding scheme with the goal of retaining perceived visual quality at smaller bitrates under the imposed limitations and constraints.



Artificial Neural Networks for Intra-Frame Prediction

Fabian Brand

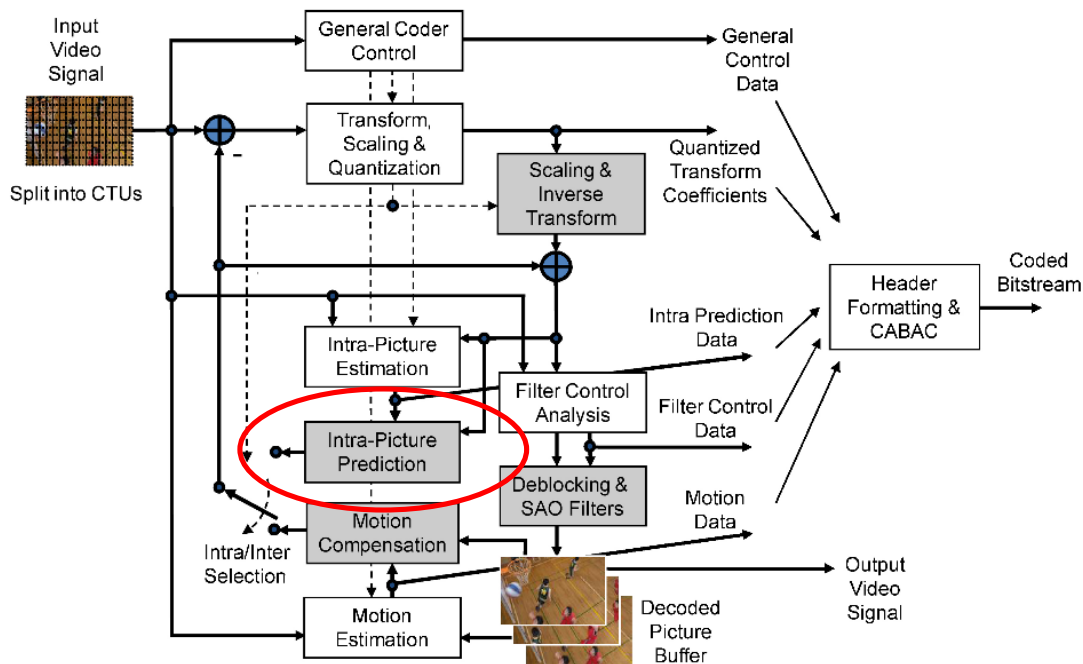
Fabian.Brand@fau.de

Chair of Multimedia Communications
and Signal Processing

Outline

- Introduction
- Intra Frame Prediction
- Applications of Neural Networks in Intra Frame Prediction
 - Additional Modes
 - Full Mode Training
- Training Set Clustering
- Experimental Results
- Conclusion

Introduction - Hybrid Video Coder



G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," Dec 2012.

Intra Prediction

- Reduces Spatial Redundancy
 - Local environment
- Predicting CU from spatial environment (reference area)
- Usually multi-mode prediction
- Transmitting (quantized) residual and mode information
- Spatial mode prediction
 - Large scale correlations



Intra Prediction

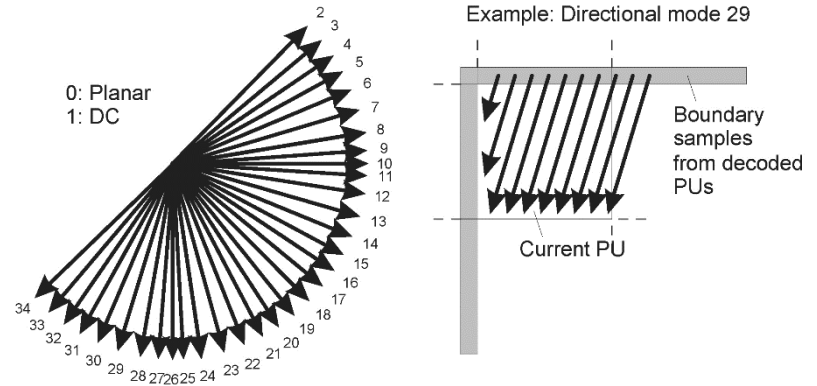


Compute Residual



Intra Prediction HEVC

- 35 modes:
 - DC and planar mode
 - 33 angular modes
- Copying pixels from the reference area in different angles
- Advantages:
 - Low complexity
 - Sharp Edges are preserved
- Disadvantage:
 - Insufficient for complex structures



G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," Dec 2012.

Intra Frame Prediction with Neural Networks

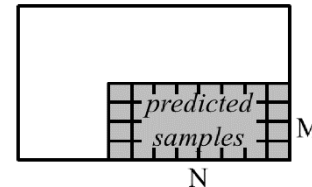
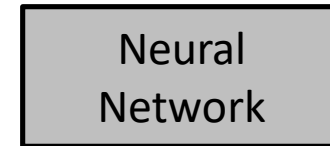
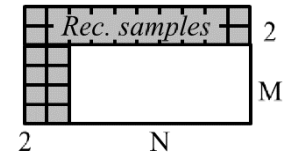
- Interpretation:

- Intra-Prediction as function

$$\hat{y} = f_m(x)$$

- Find mode m which minimizes error of prediction signal

- Neural networks (NNs) can approximate arbitrary functions
- Usually larger reference area (e.g. 2 or 4 pixels wide)



cf. J. Pfaff, P. Helle, S. K. D. Maniry, W. Samek, H. Schwarz, D. Marpe, and T. Wiegand, "Neural network based intra prediction for video coding," 2018

Additional Modes with Neural Networks

- Keeping HEVC modes intact and adding one or two additional modes
- Example: IPFCN by Li *et al.*
 - Shallow 4 layer fully connected (FC) network
- Two proposals: IPFCN-S and IPFCN-D
 - One and two additional modes respectively
 - Comparing with HEVC
- IPFCN-S
 - Training one mode with all available training data
 - Average BD-rate: -2.9%
- IPFCN-D
 - Training one mode from DC/planar blocks
 - Training another mode from angular blocks
 - Average BD-rate: -3.4%
- Gain decreases for second mode

Mode-Based Intra Prediction with Neural Networks

- Replacing all modes with neural network based modes
- Challenges:
 - Computational complexity: Full search becomes more difficult
 - Spatial mode prediction
 - Training set
- Example: Pfaff *et al.*
 - Four-layer networks
 - All modes share the first three layers
 - Improves runtime complexity and memory requirements
 - Another neural network used for mode prediction
 - Comparing with JEM including rectangular blocks
 - Average BD-rate: -3.01%

Spatial Mode Prediction

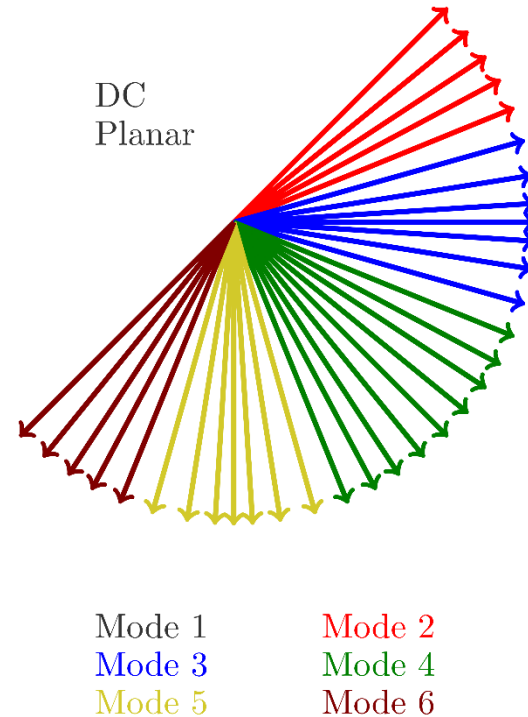
- Angular prediction:
 - Mode semantic clear and equal for all block sizes
 - Modes can be ordered
 - Easy spatial prediction
- Neural Network based prediction:
 - Highly non-linear functions
 - Mode semantic unclear, depending on training sets
 - No trivial ordering possible
 - Independent training for different block sizes leads to different semantics
 - Difficult spatial prediction
 - Pfaff *et al.* use separate network for mode prediction

Training

- Different modes require different training sets
- Requirements:
 - Complete coverage of all contents
 - Clear distinction between the modes
- How can we design suitable training sets?
- Splitting the whole training set

Splitting the Training Set

- How to split the training set to train good predictors
- Not only consider block structure but also the support area
 - Same structure must be predicted differently depending on support area
- Proposed method
 - Cluster blocks according to best HEVC mode



Results

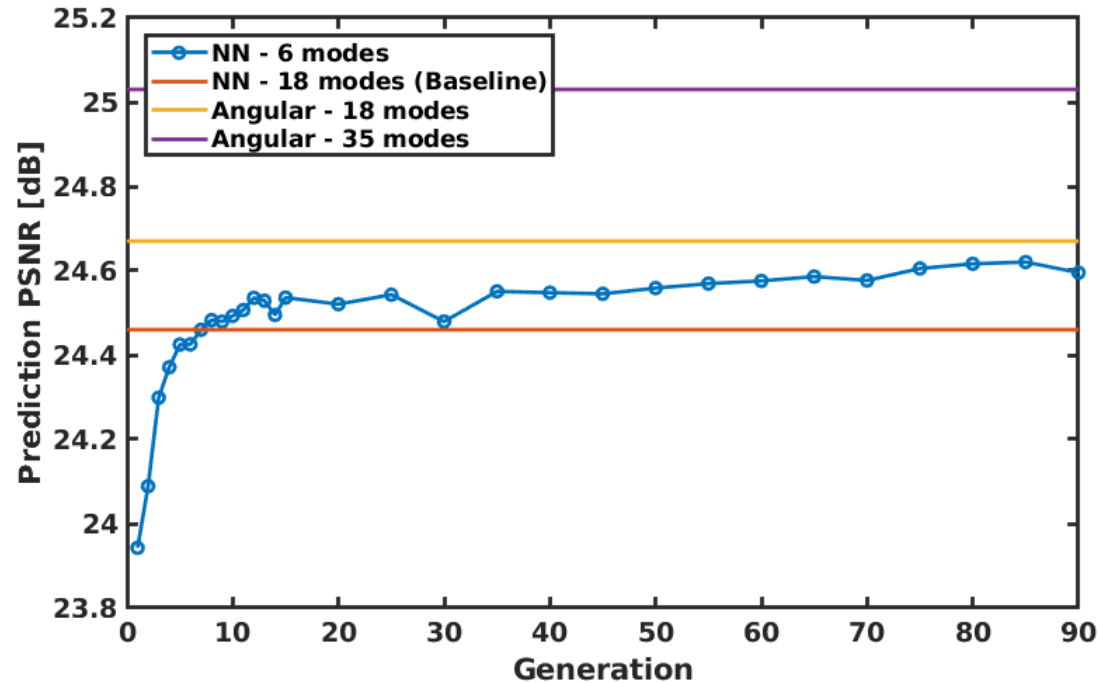
- Evaluating PSNR of prediction
- Training on 90 Images of TECNICK
- Evaluating the remaining images
- Only block size 16x16 tested
- Four-layer fully connected network

Predictor	# Modes	Prediction PSNR [dB]
Angular (HEVC)	35	25.03
NN	35	24.73
Angular	18	24.67
NN	18	24.46
Angular	6	22.71
NN	6	23.94

Iterative Approach

- Clustering according to HEVC modes produces similar modes
- Neural Networks can do more!
- Proposal:
 - Clustering according to previously trained predictor
 - Iterative clustering
 - Many “Generations”
 - Similar to expectation-maximization (EM) algorithm
- High training effort
- Challenges:
 - Keep individual modes from becoming too dominant
 - Ideally: Use different datasets for each generation
 - Practice: Use data augmentation, e.g. by flipping
 - Mode predictability decreases with proceeding generations
 - No solution yet

Results



Summary

- Two concepts for network-based intra prediction
 - Adding additional modes
 - Good results
 - Using new structures
 - Increased side information
 - Replacing all modes
 - High potential
 - High training effort
 - Major Challenges: Training procedure, Mode prediction
- Generally high potential for neural-network-based intra prediction

References

- [1] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, “Overview of the high efficiency video coding (HEVC) standard,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649–1668, Dec 2012.
- [2] J. Li, B. Li, J. Xu, R. Xiong, and W. Gao, “Fully connected network-based intra prediction for image coding,” IEEE Transactions on Image Processing, vol. 27, no. 7, pp. 3236–3247, July 2018.
- [3] J. Pfaff, P. Helle, S. K. D. Maniry, W. Samek, H. Schwarz, D. Marpe, and T. Wiegand, “Neural network based intra prediction for video coding,” in Proc. SPIE, vol. 10752, 2018, pp. 13 – 1–8.



Design Techniques for Incremental Non-Regular Sampling Patterns

Simon Grosche, Jürgen Seiler, and André Kaup

simon.grosche@fau.de

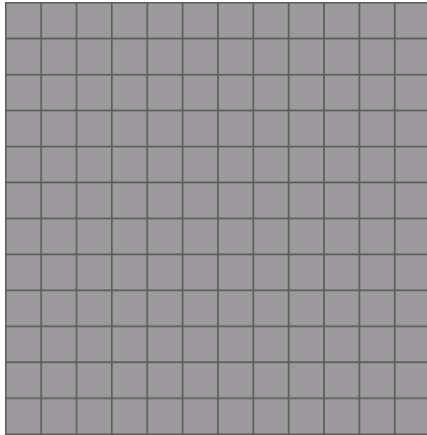
Chair of Multimedia Communications
and Signal Processing

Outline

- (Incremental) Non-Regular Sampling
- Importance of Proper Sampling Patterns
- Design Techniques for Incremental Sampling Patterns
- Simulation & Evaluation
- Conclusion & Outlook

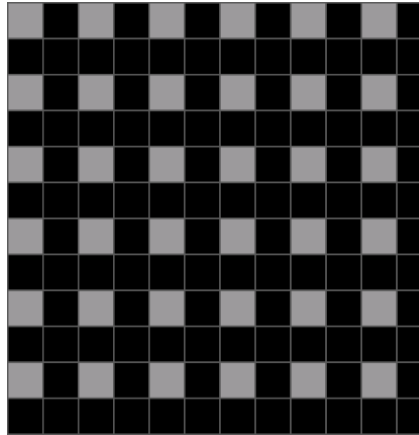
Non-Regular Sampling

High-Resolution Sampling
(N^2 pixels)



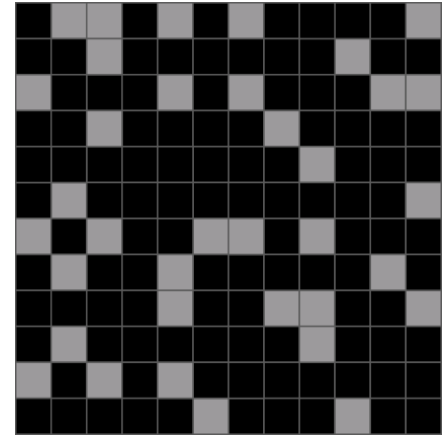
→ Long measurement time,
high data rate

25% Regular Sampling
($N^2/4$ pixels)



→ Interpolate remaining pixels
→ Resolution limited by
aliasing

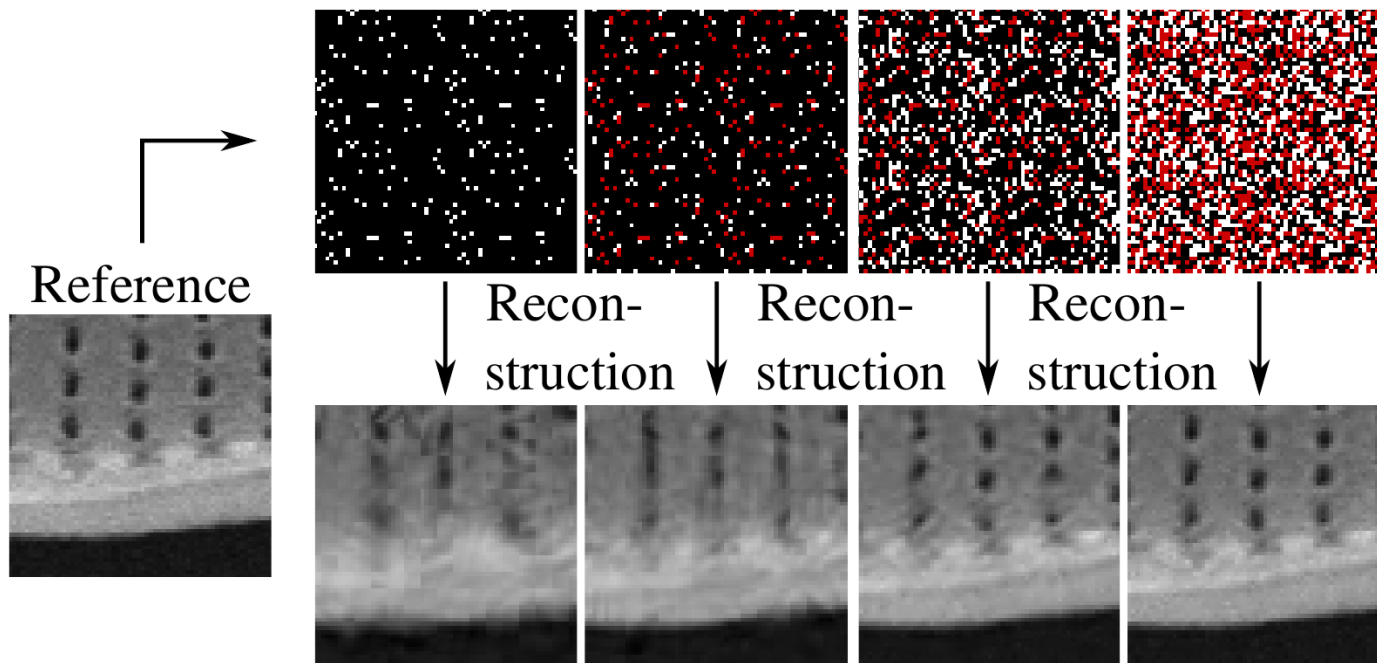
25% Non-Regular Sampling
($N^2/4$ pixels)



→ Higher resolution after
appropriate reconstruction
→ Reduced aliasing

Seiler et al., "Resampling images to a regular grid from a non-regular subset of pixels using frequency selective reconstruction," *IEEE Transactions on Image Processing*, vol. 24., no. 11, pp. 4540-4555, Nov. 2015

Incremental Non-Regular Sampling



Grosche et al., "Design Techniques for Incremental Non-Regular Image Sampling Patterns," in *Proc. IEEE IST*, Krakow, Oct. 2018

Reconstruction Algorithms and Testsets

Reconstruction Algorithms

- Linear Interpolation
 - fast, reasonable quality
- Frequency Selective Reconstruction (FSR)
 - more complex, real-time capable, high quality

Image Testset

- SEM images
 - 8-bit grayscale images
 - 30 images, 1200x1200 pixels
- Tecnick Dataset (2011)
 - Natural 8-bit grayscale images
 - First 30 images, 1200x1200 pixels



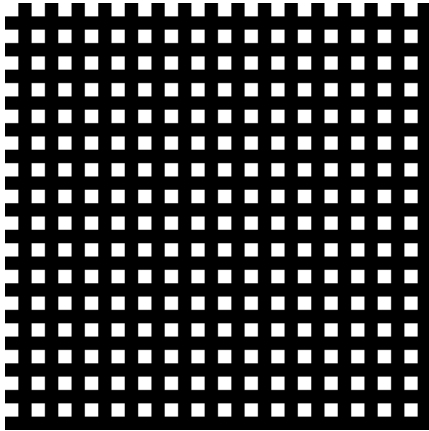
SEM images: Museo del Scienze, online: https://commons.wikimedia.org/wiki/Category:SEM_images_from_MUSE_-_Science_Museum, Apr. 2016; accessed 18-May-2018

Tecnick images: N. Asuni et al., "Testimages: a large-scale archive for testing visual devices and basic image processing algorithms," in *Proc. Smart Tools and Apps for Graphics*, Cagliari, Sep. 2014, pp. 63–70

FSR: Seiler et al., "Resampling images to a regular grid from a non-regular subset of pixel positions using FSR," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4540–4555, Nov. 2015

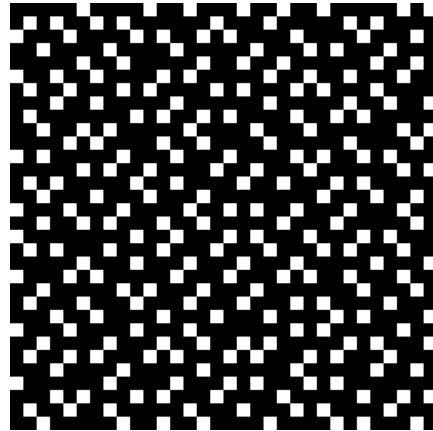
How to choose the Sampling Pattern?

Regular Pattern



→ PSNR 33.2 dB (FSR)

Optimized Quarter Pattern



→ PSNR 34.1 dB (FSR)

Random Sampling Pattern



→ PSNR 32.4 dB (FSR)

→ Observation: Good sampling pattern should be **uniform** and **non-regular**

Optimized Quarter Pattern: Grosche et al., "Iterative Optimization of Quarter Sampling Masks for Non-Regular Sampling Sensors," in *Proc. IEEE ICIP*, Athens, Oct. 2018

How to choose the Sampling Pattern?

Uniformity

- Local density \approx global density
- Details can be anywhere in the image

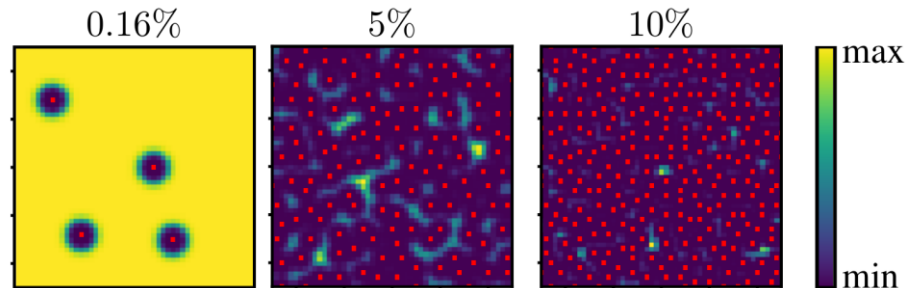
Non-Regularity

- Flat frequency spectrum
- Reduce aliasing

How to combine both properties in a single, incremental sampling pattern?

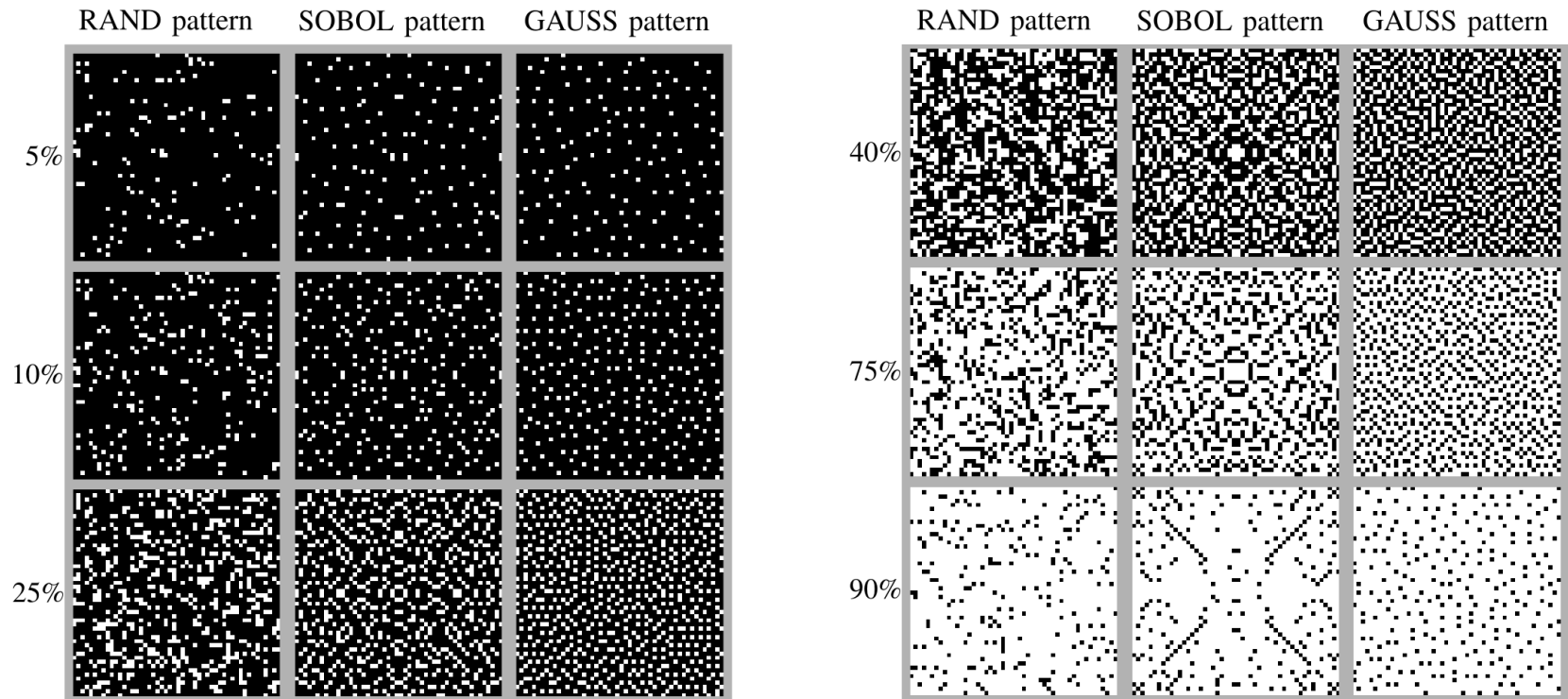
Techniques for Incremental Non-Regular Patterns

- Incremental random sampling patterns (RAND)
 - Draw random sampling positions from uniform probability distribution
- Sobol sampling patterns (SOBOL)
 - Often used in Monte Carlo integration, here discretized
- Proposed incremental Gaussian probability distribution patterns (GAUSS)
 - Draw random sampling positions from Gaussian probability distribution



Grosche et al., "Design Techniques for Incremental Non-Regular Image Sampling Patterns," in *Proc. IEEE IST*, Krakow, Oct. 2018

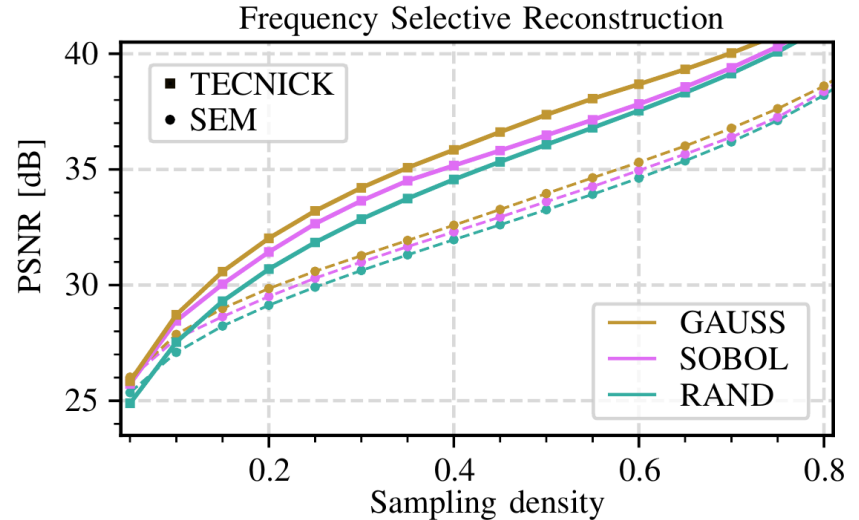
Sampling Patterns (central section)



Grosche et al., "Design Techniques for Incremental Non-Regular Image Sampling Patterns," in *Proc. IEEE IST*, Krakow, Oct. 2018

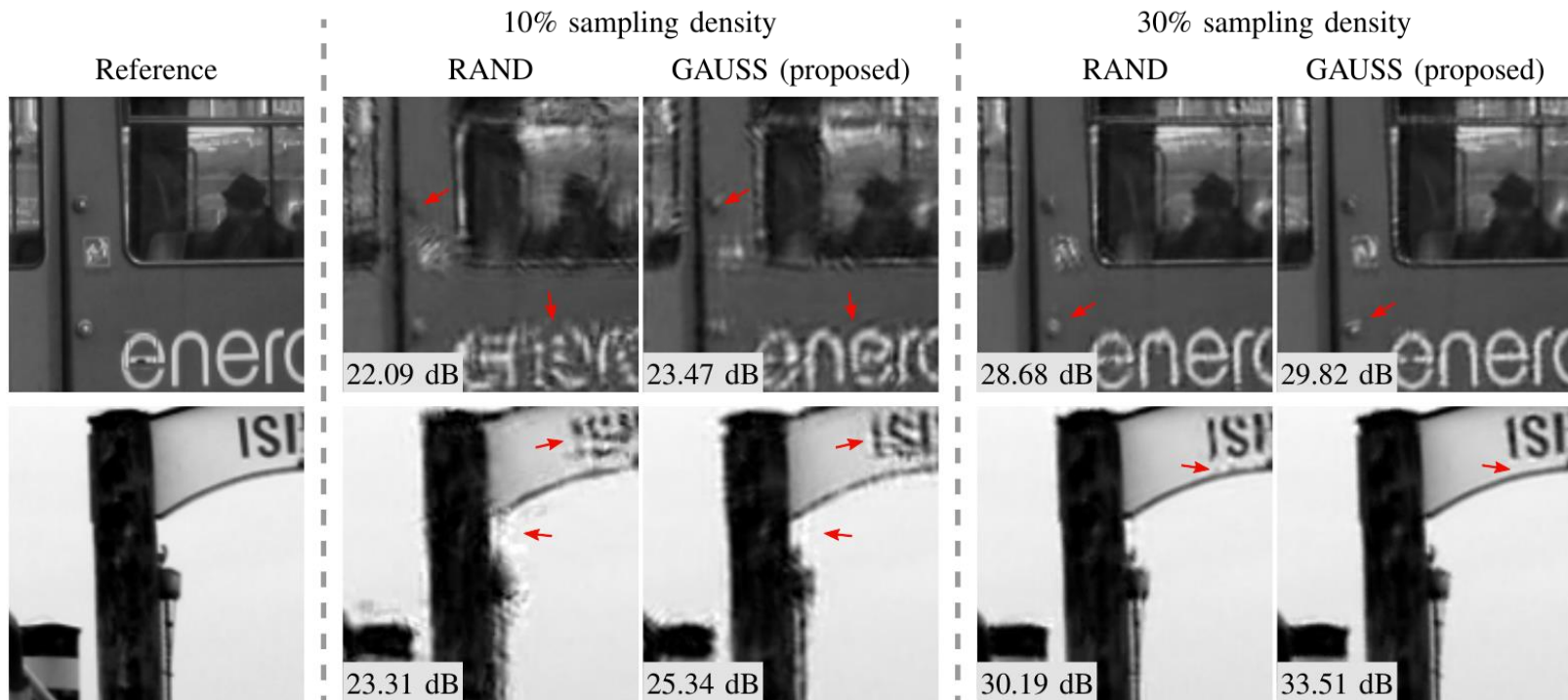
Reconstruction Quality

- Similar for both reconstruction methods
- Similar for both test sets
- GAUSS > SOBOL > RAND
- More than +0.5dB gain



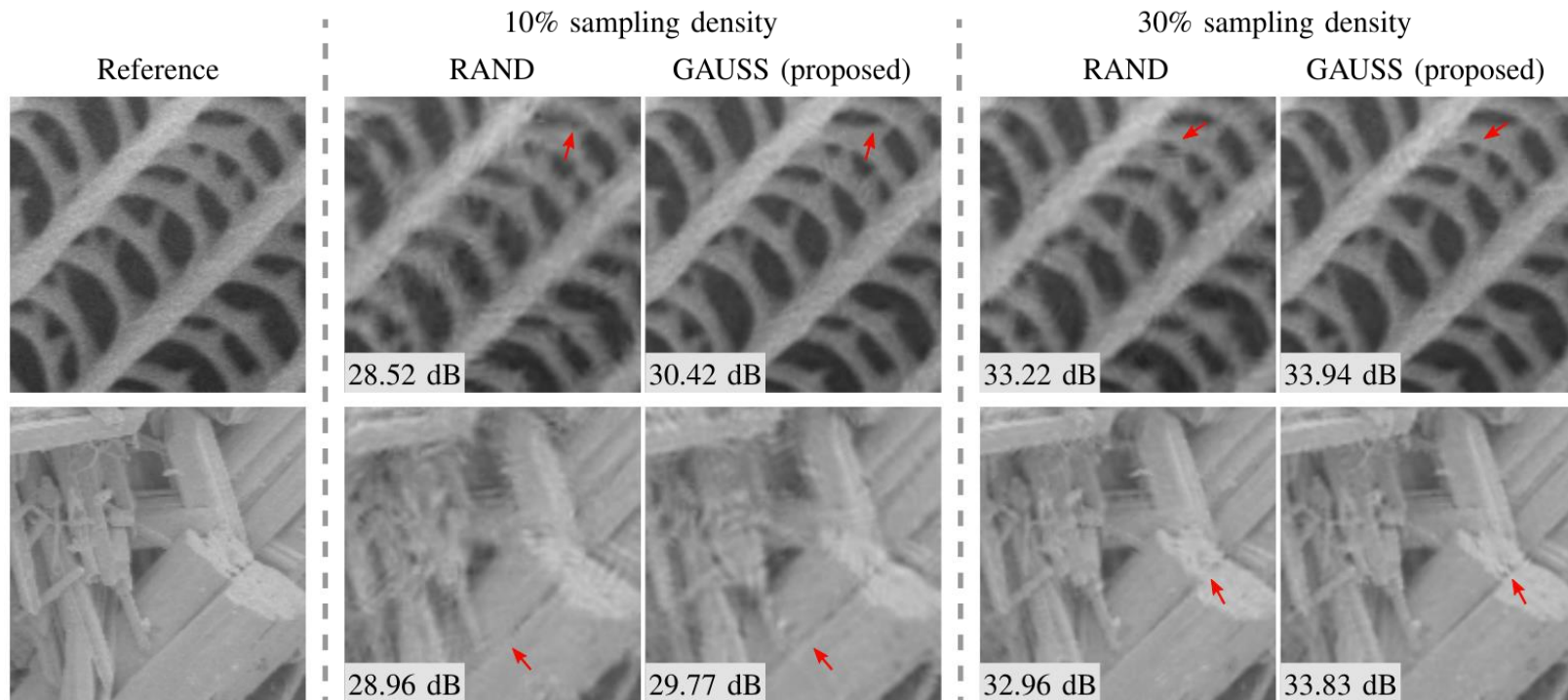
Grosche et al., "Design Techniques for Incremental Non-Regular Image Sampling Patterns," in *Proc. IEEE IST*, Krakow, Oct. 2018

Visual Comparisons



Grosche et al., "Design Techniques for Incremental Non-Regular Image Sampling Patterns," in *Proc. IEEE IST*, Krakow, Oct. 2018

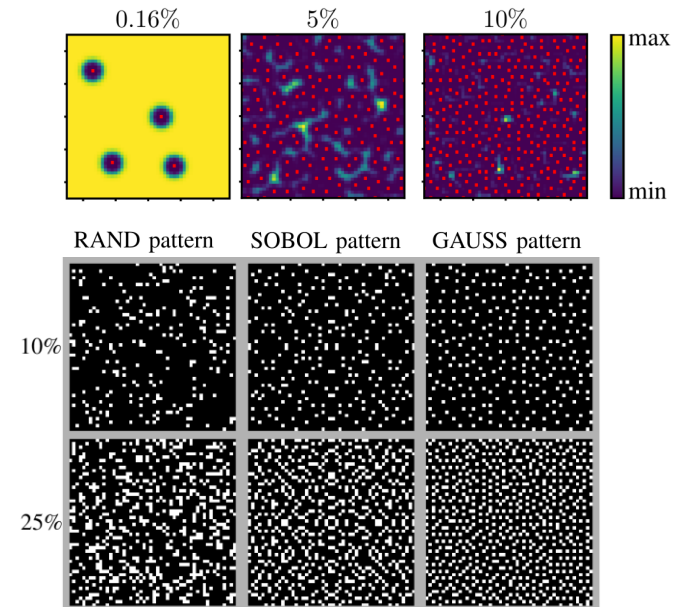
Visual Comparisons



Grosche et al., "Design Techniques for Incremental Non-Regular Image Sampling Patterns," in *Proc. IEEE IST*, Krakow, Oct. 2018

Conclusion

- Non-regular quarter sampling can achieve higher resolution per pixel using an appropriate reconstruction method
 - Observation: Good sampling patterns should be uniform and non-regular
 - Proposal: Incremental sampling patterns
RAND, SOBOL, GAUSS (proposed)
-
- Results similar for both test sets and both reconstruction methods
 - Gain $>+0.5$ dB using GAUSS instead of RAND patterns



Grosche et al., "Design Techniques for Incremental Non-Regular Image Sampling Patterns," in *Proc. IEEE IST*, Krakow, Oct. 2018

Outlook

- Content adaptive patterns
- Extend to 3D-patterns
- Theoretical limitations from compressed sensing

- **FSR Matlab-Reference Implementation available at**
<https://gitlab.lms.tf.fau.de/LMS/Rapid-FSR>
 - Bundles latest research on FSR
 - Dynamic parameter estimation
 - Three quality profiles: fast, compromise, best





Frequency-Selective Mesh-to-Grid Resampling

Viktoria Heimann

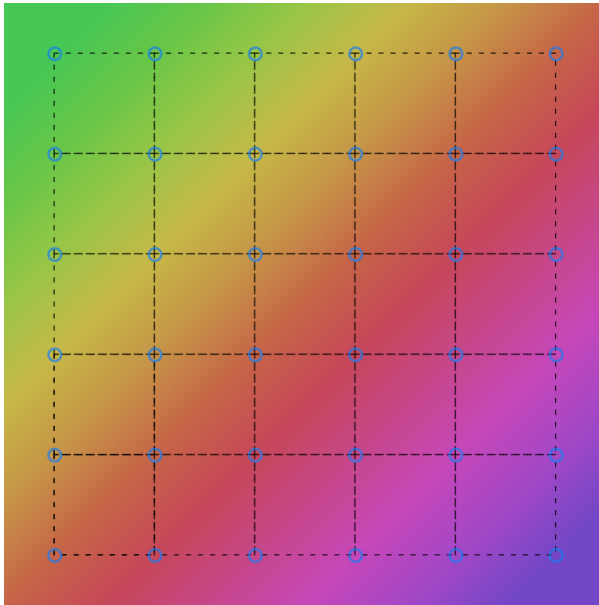
viktoria.heimann@fau.de

Chair of Multimedia Communications
and Signal Processing

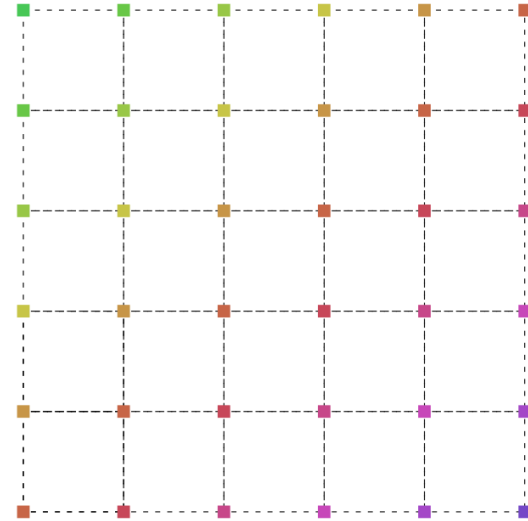
Outline

- Mesh-to-Grid
- Resampling
- Frequency-Selective

Grid

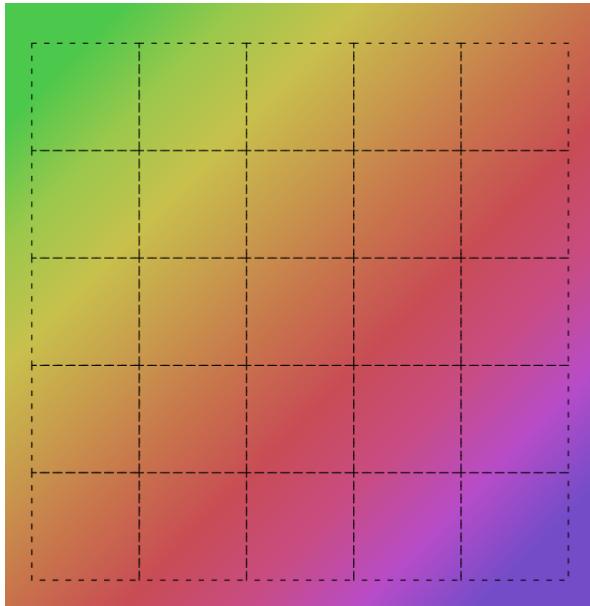


Continuous Image

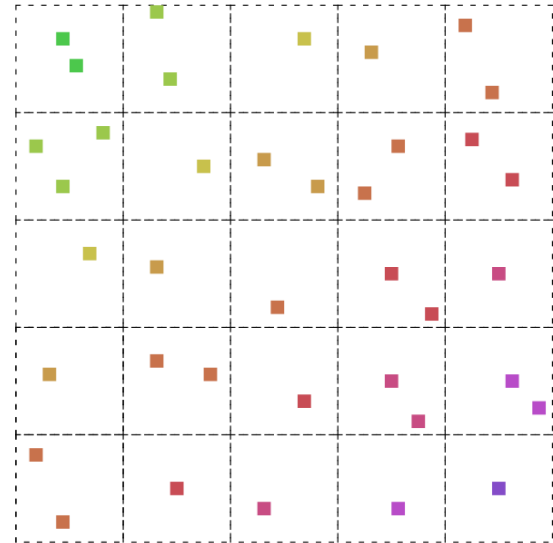


Regularly Sampled Image
= Grid

Mesh

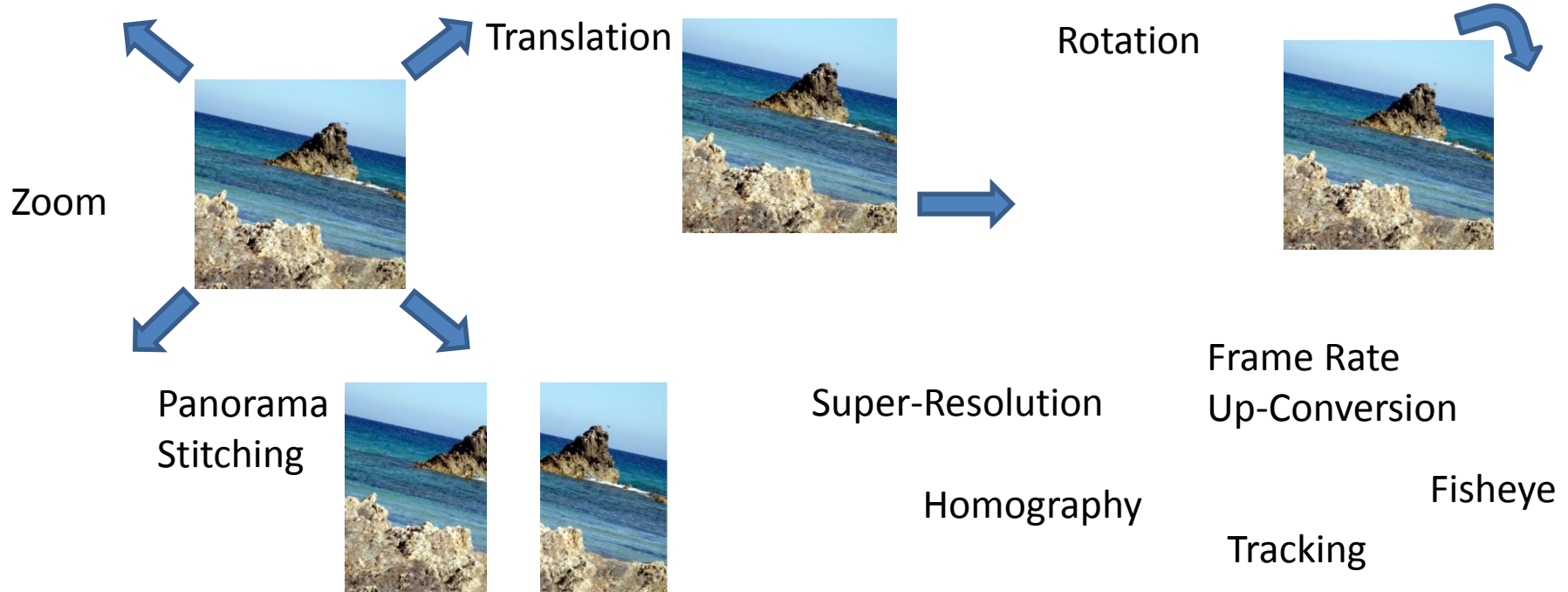


Continuous Image



Irregularly Sampled Image
= Mesh

How to generate a mesh?



Source: <https://testimages.org/>

And many more...

Image Resampling

- Pixels at non-integer positions cannot be displayed nor efficiently stored
→ Resampling is necessary



How can resampling be done?

- Using the classic methods
 - Linear Interpolation
 - Cubic Interpolation
 - Spline Interpolation etc.
- Using frequency-based method
 - Frequency-Selective Mesh-to-Grid Resampling (FSMR) [1]

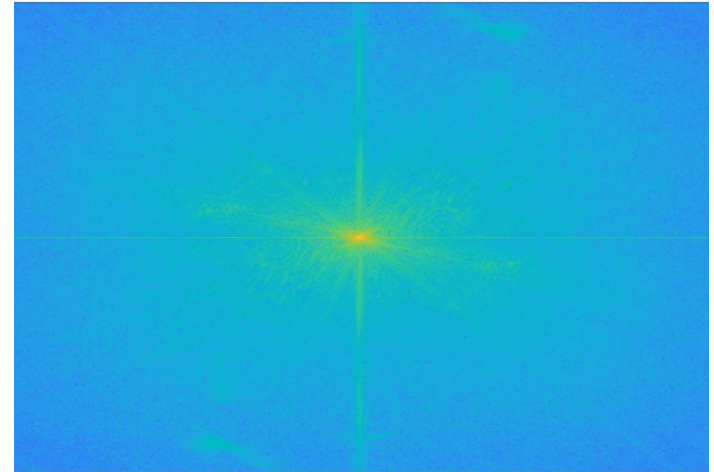
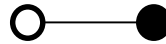
[1] „Frequency-Selective Mesh-to-Grid Resampling for Image Communication“, J.Koloda et al., IEEE Transactions on Multimedia, 2017

Image in Spatial Domain

Main Principle:

$$f[m, n] = \sum_{k \in K} c_k \varphi_k[m, n]$$

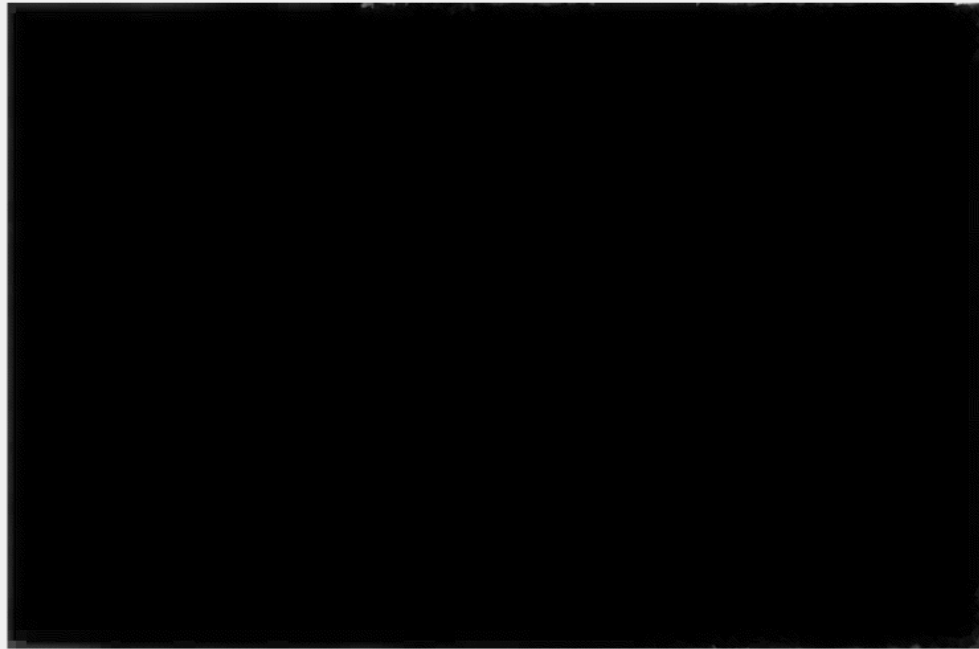
Image Signal = Sum of weighted Basis Functions



Source: <http://r0k.us/graphics/kodak/>

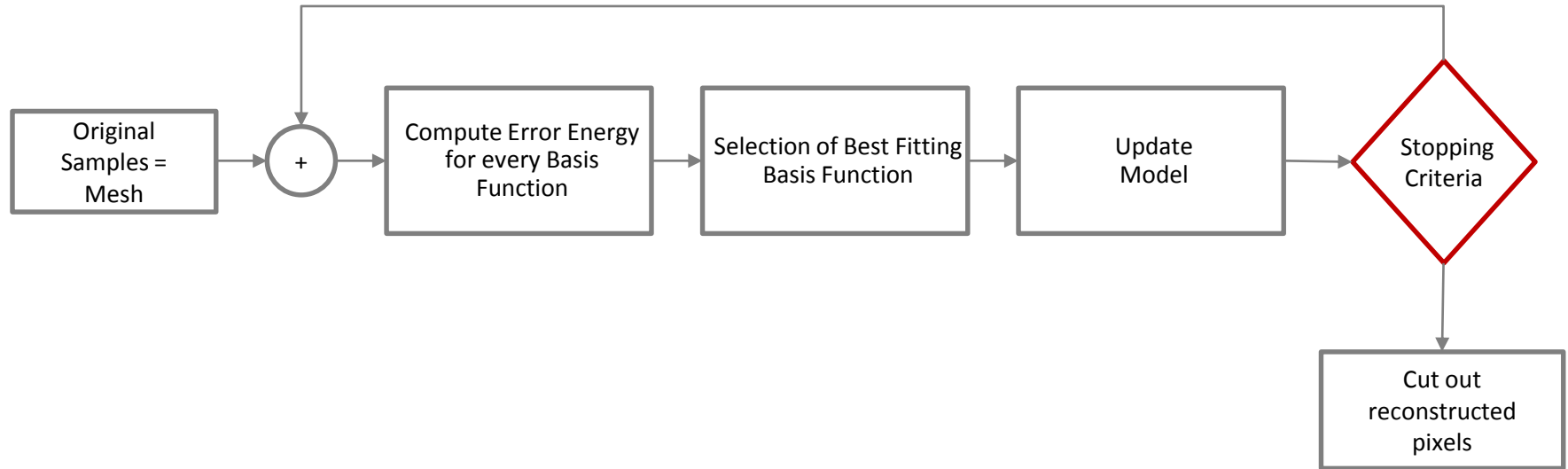
Iterative Model

Iteration 1



$$g^{(v)}[m, n] = g^{(v-1)}[m, n] + \hat{c}_u^{(v)} \varphi_u[m, n]$$

FSMR



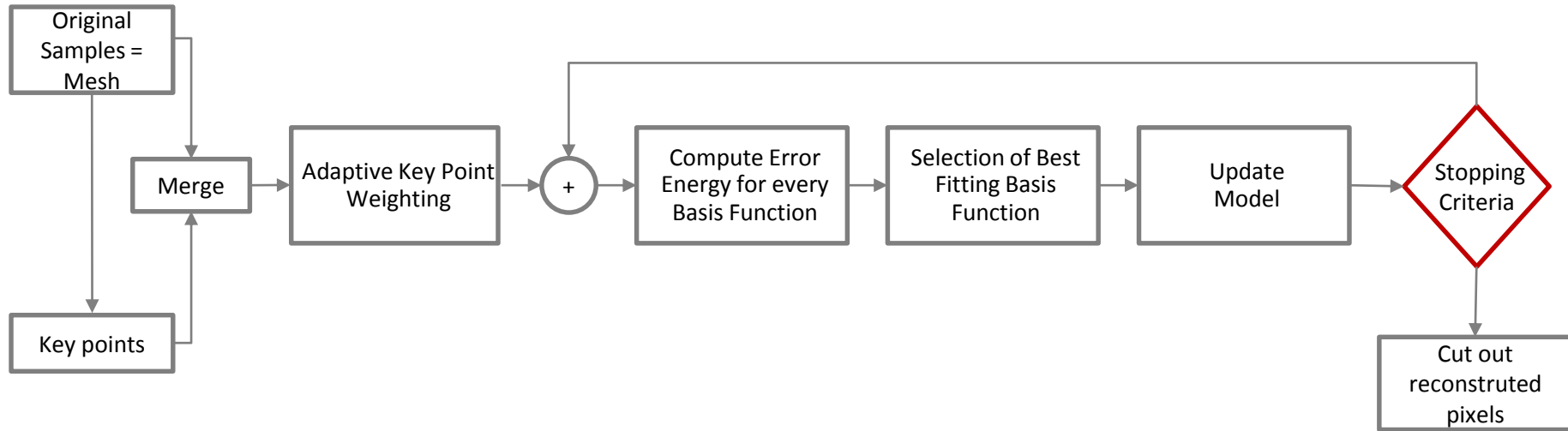
Results



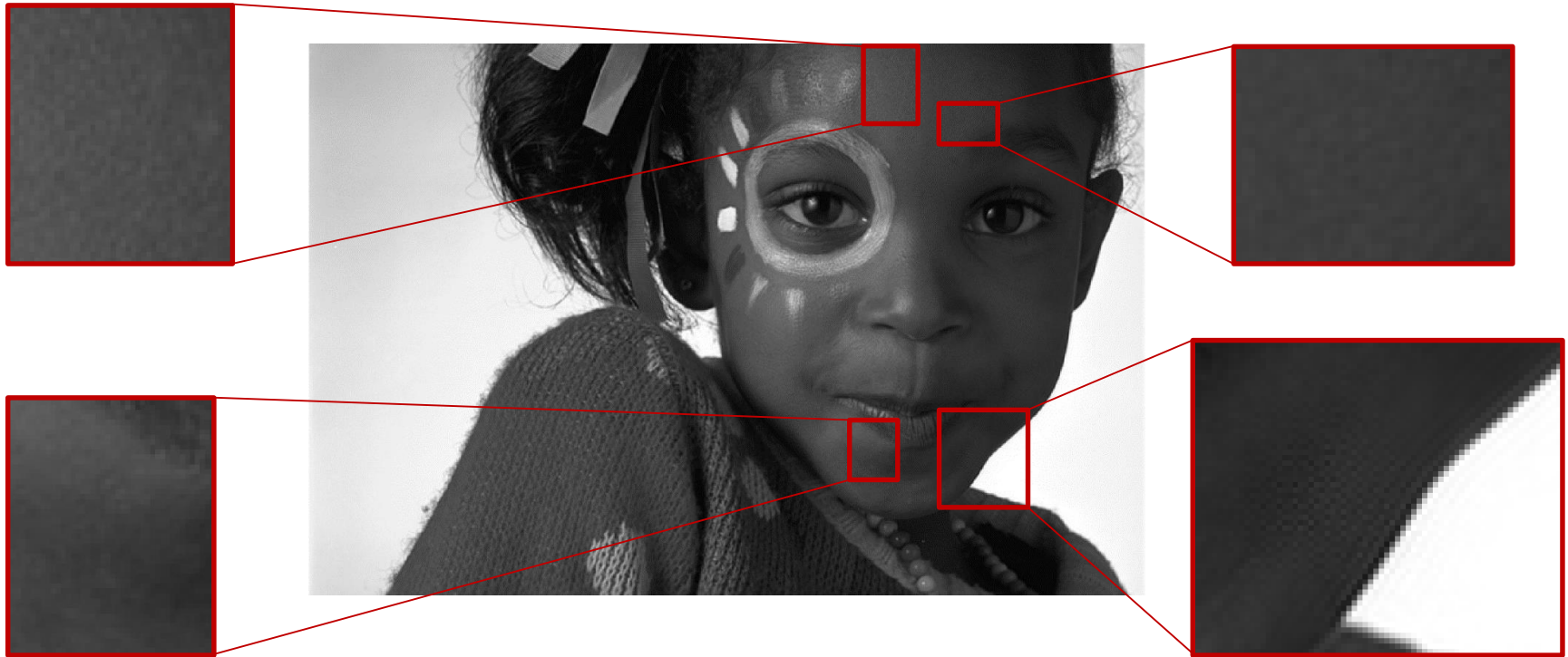
Orthogonality Problem

- Problem: Basis Functions not orthogonal
- Reason: Evaluation of the Basis Function on Floating Mesh and not on Regular Grid
- Idea: Include Grid Points in the Generation of the Model
- How: Easy Interpolation of Grid Points = Key Points
- Attention: Key Points not as reliable as Mesh Points
→ Smaller Weights for Key Points

FSMR using Keypoints



Results with Adaptive Key Point Weighting



Comparison

Average PSNR in dB with respect to linear interpolation for the KODAK data image dataset

	Cubic	Natural Neighbor	Inverse Distance Weighting	Lanczos	FSMR
Rotation	3.30	-0.11	0.56	0.32	9.77
Translation	3.23	-0.13	1.63	3.59	12.06
Zoom	3.65	-0.26	4.20	4.44	7.09

Source: „Frequency-Selective Mesh-to-Grid Resampling for Image Communication“, J.Koloda et al., IEEE Transactions on Multimedia, 2017

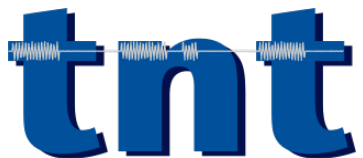
Conclusion

- Pixels at non-integer positions cannot be displayed nor efficiently stored
- Resampling is necessary for many applications
- FSMR takes advantage of frequency information
- FSMR is a powerful method for image resampling

Non-linear Contour-based Multidirectional Intra Coding

5th Summer School on Video Compression and Processing (SVCP) 2019

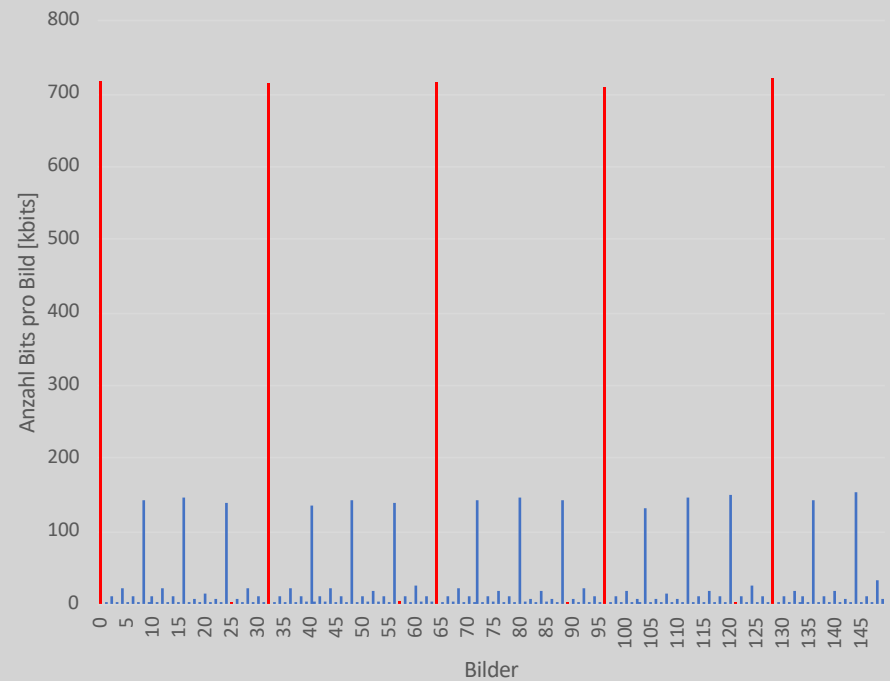
Thorsten Laude



Background

Necessity of intra coding

- Start and random access of/to transmissions
- Error concealment
- Chunk-based bitrate adaptivity
- Coding of newly appearing content
- Predestined for efficient still image coding



Limitations of HEVC Intra Prediction



Relatively small reference area
(1pel width/height)



Directional modes only allow
prediction of linear structures



Only one intra mode
(directional, DC, planar) per
block

CoMIC

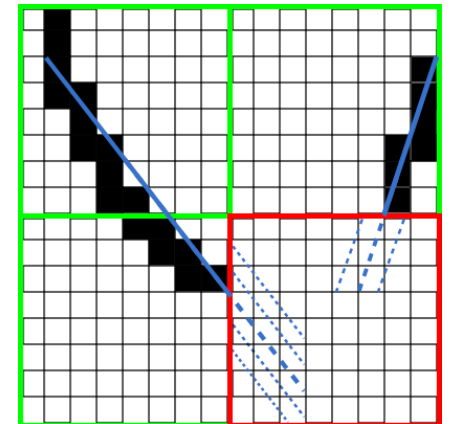
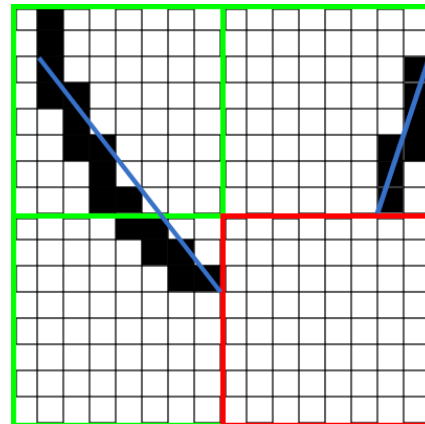
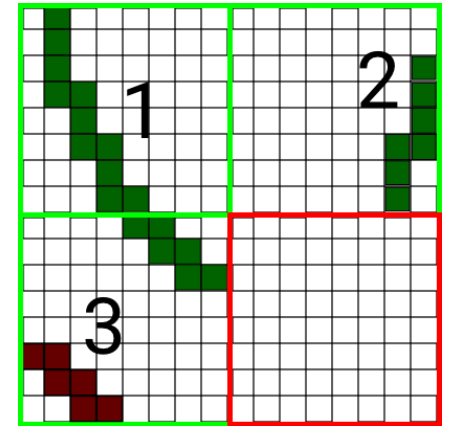
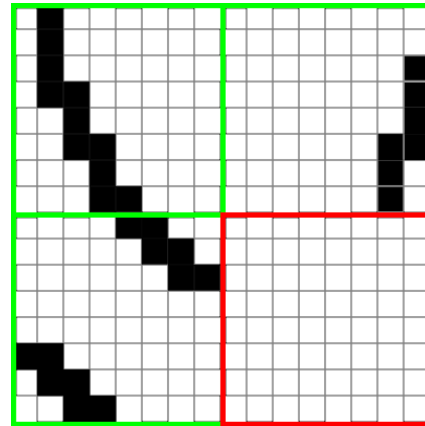
Contour-based
Multidirectional Intra
Coding

CoMIC v1 (PCS 2016)

- Prediction based on information gathered in reference area
- Available at encoder and decoder
- Linear contour modeling
- Reference sample continuation

Difference to HEVC

- ✓ Multiple directions

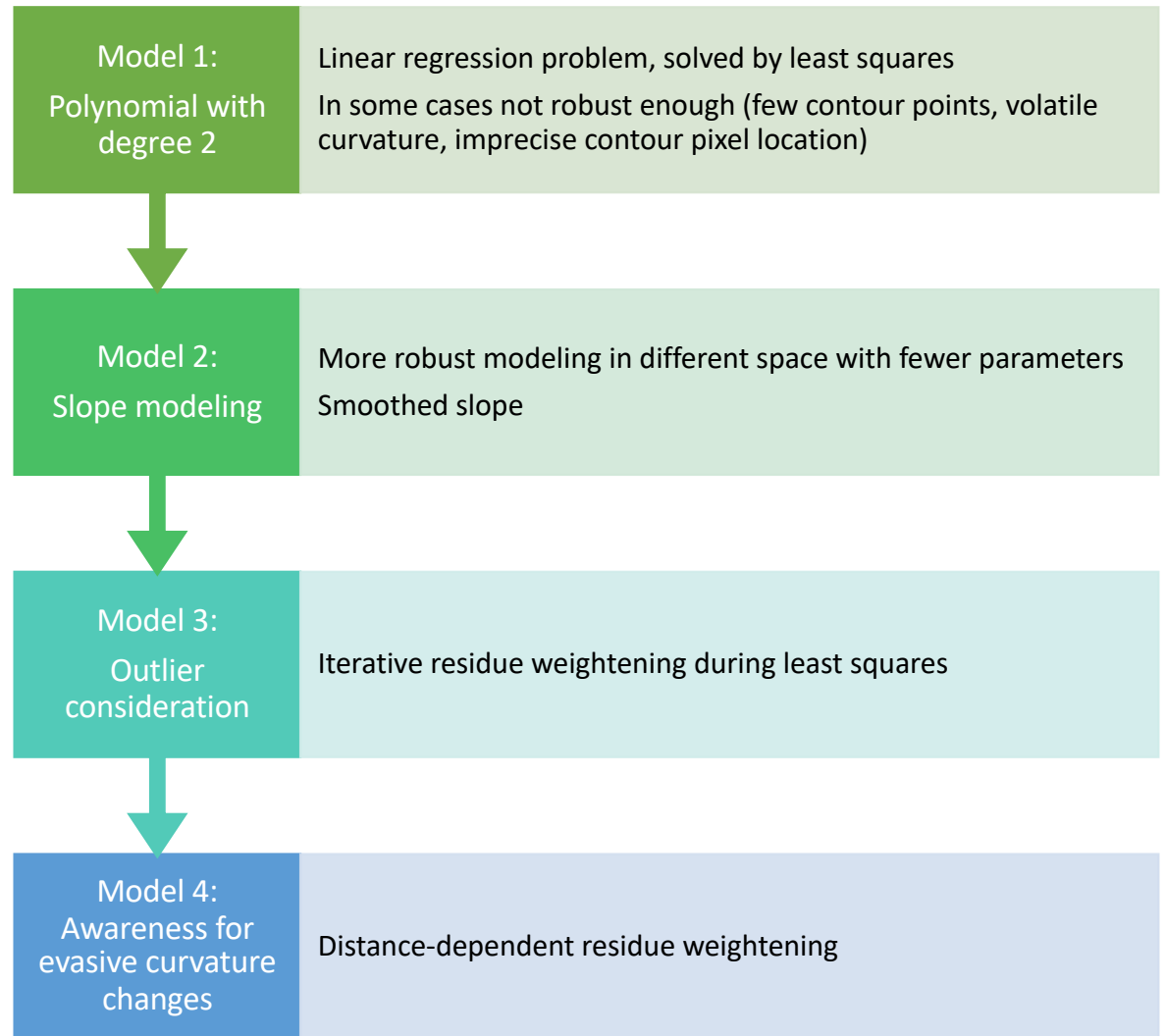


CoMIC v2 (TSIP 2018)

Extension of CoMIC v1 by

- Four non-linear contour models
- Connection of intersecting contours
- Boundary Recall-based contour model selection

Non-linear Contour Models



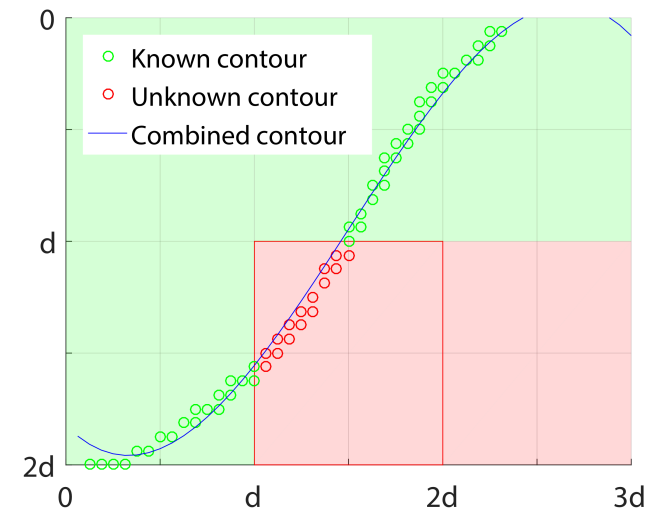
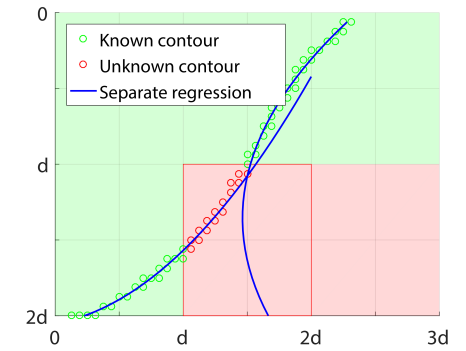
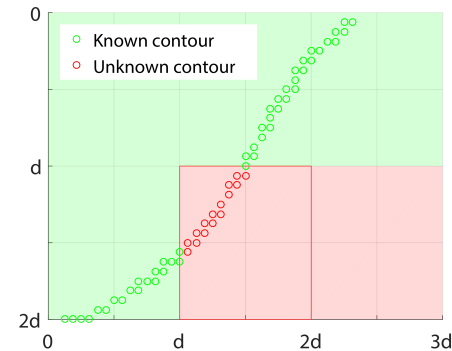
Contour Combination

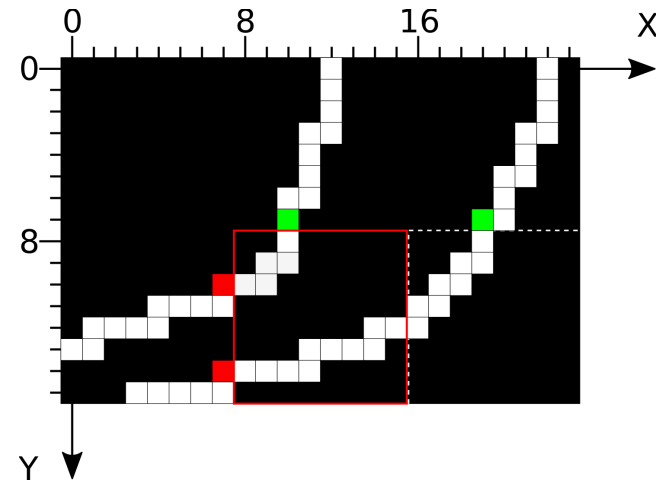
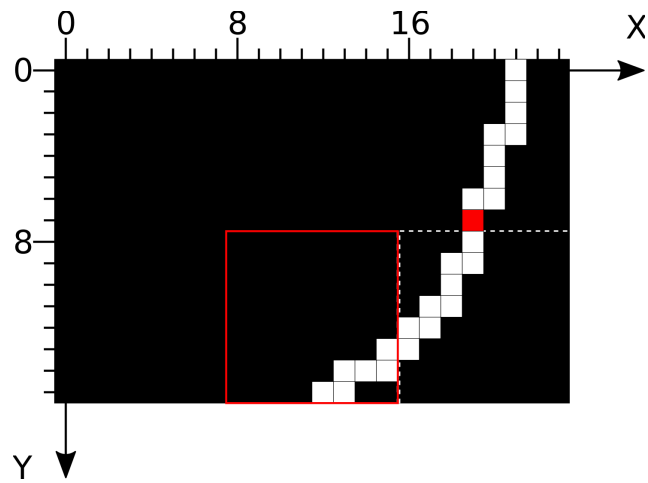
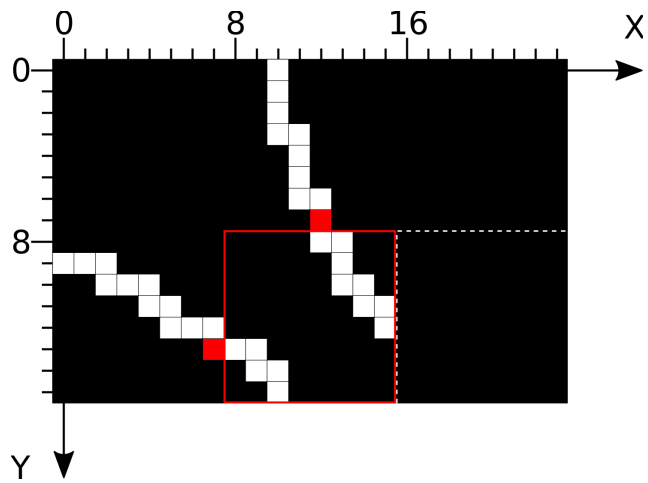
Problem:

Contours can be intersected by current block \rightarrow Two independent contours

Solution:

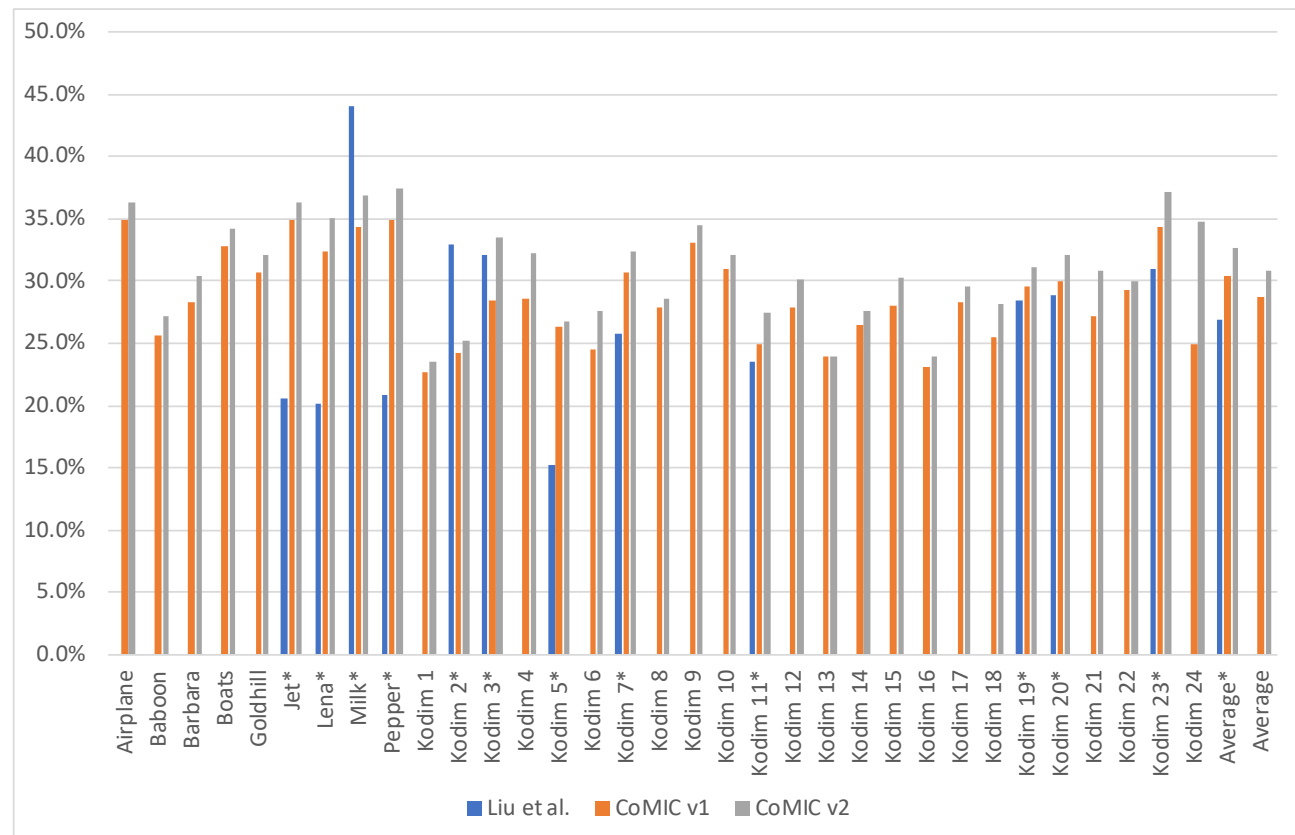
1. Detect intersecting contours after extrapolation
2. Join contours to single contour
3. Interpolation between contours





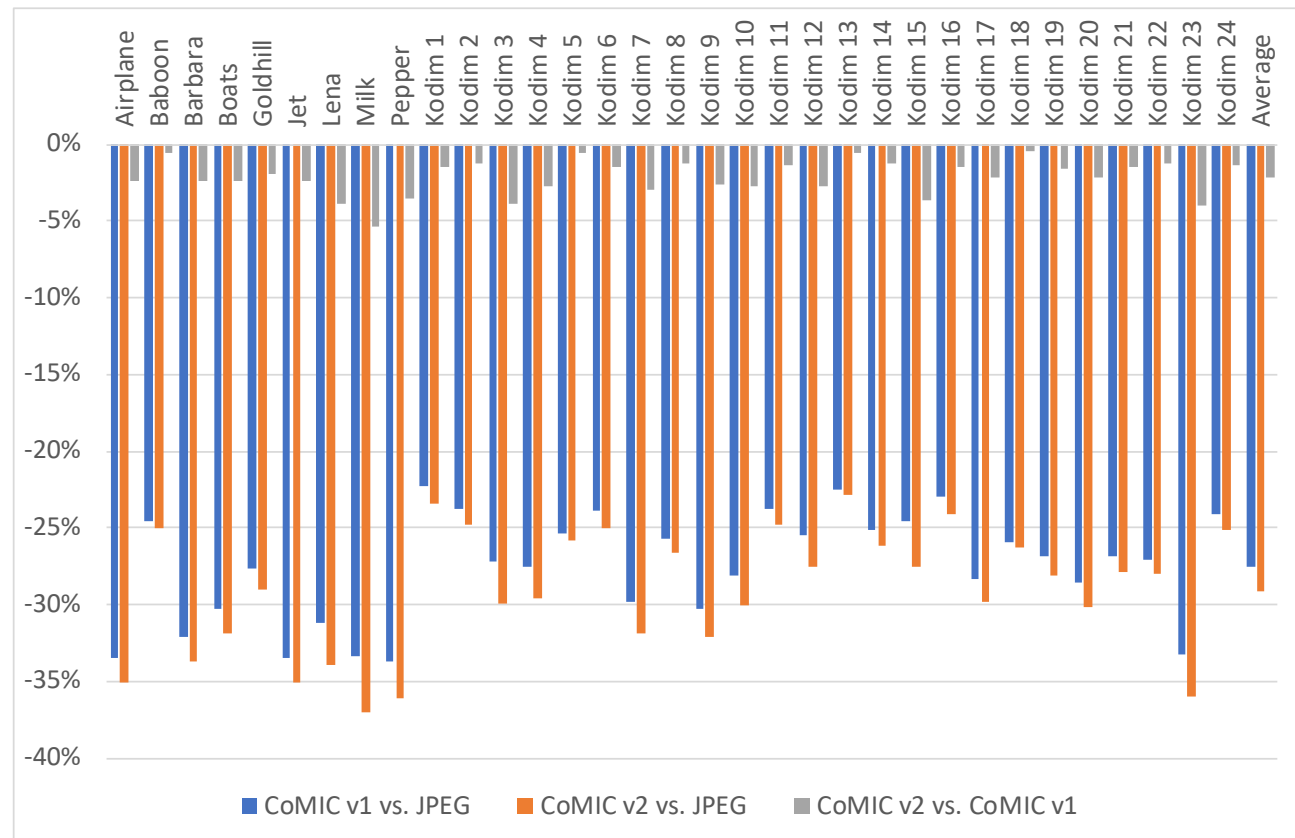
Sample Value Prediction

Coding Efficiency 1



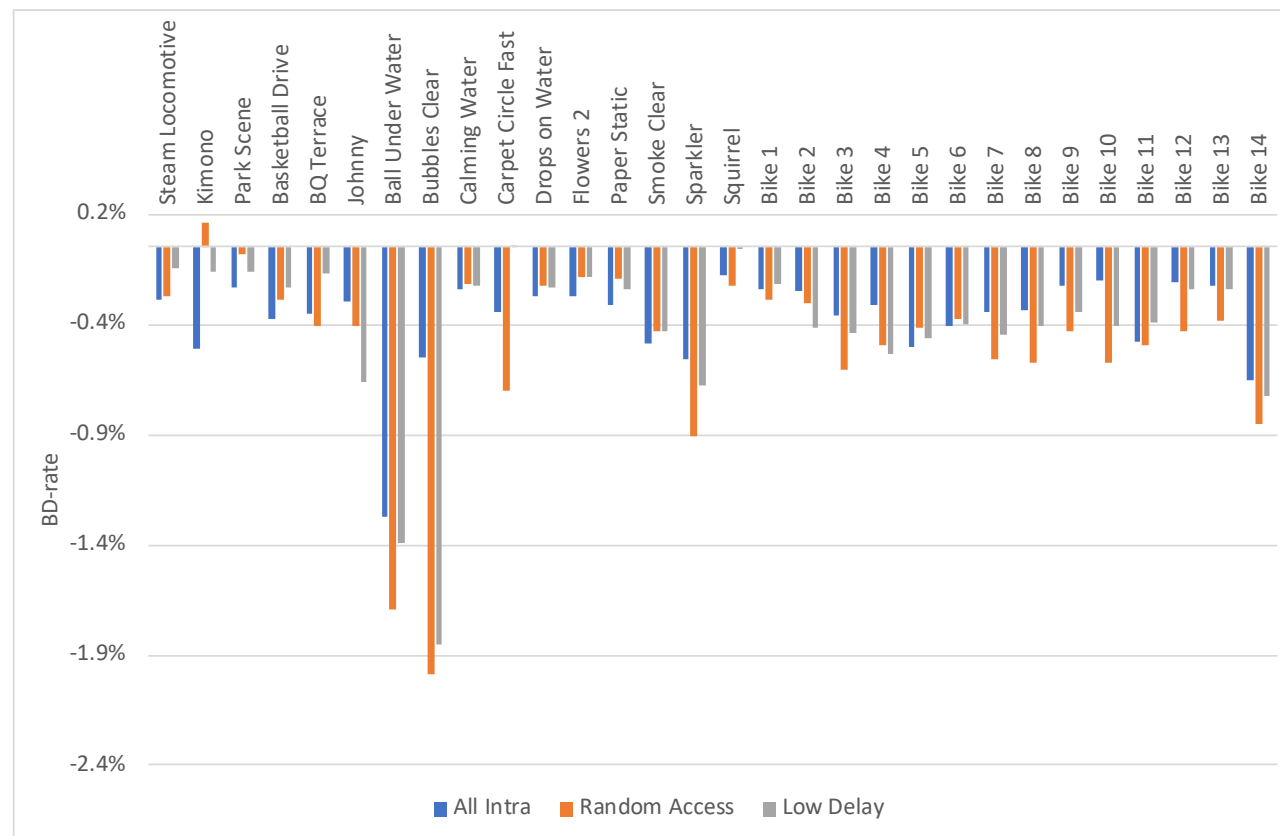
Bit rate savings relative to JPEG at fixed quality

Coding Efficiency 2



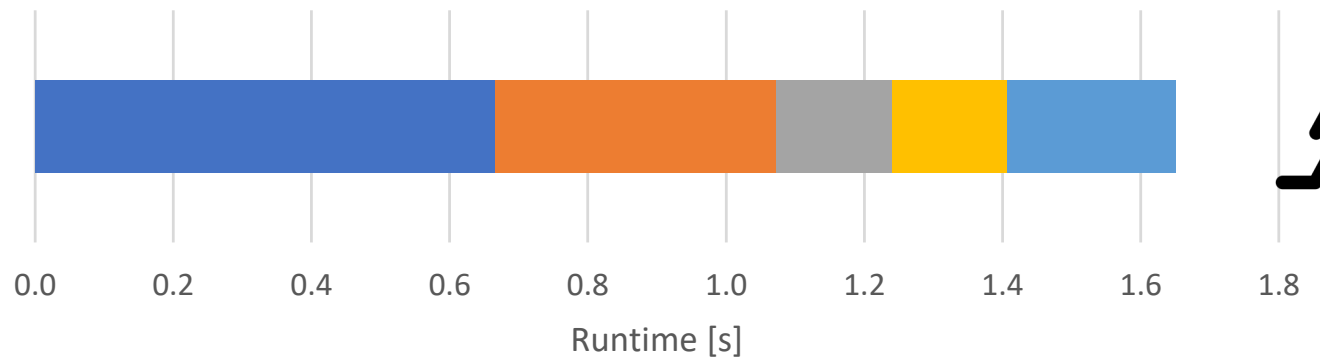
BD rates relative to JPEG

Coding Efficiency 3



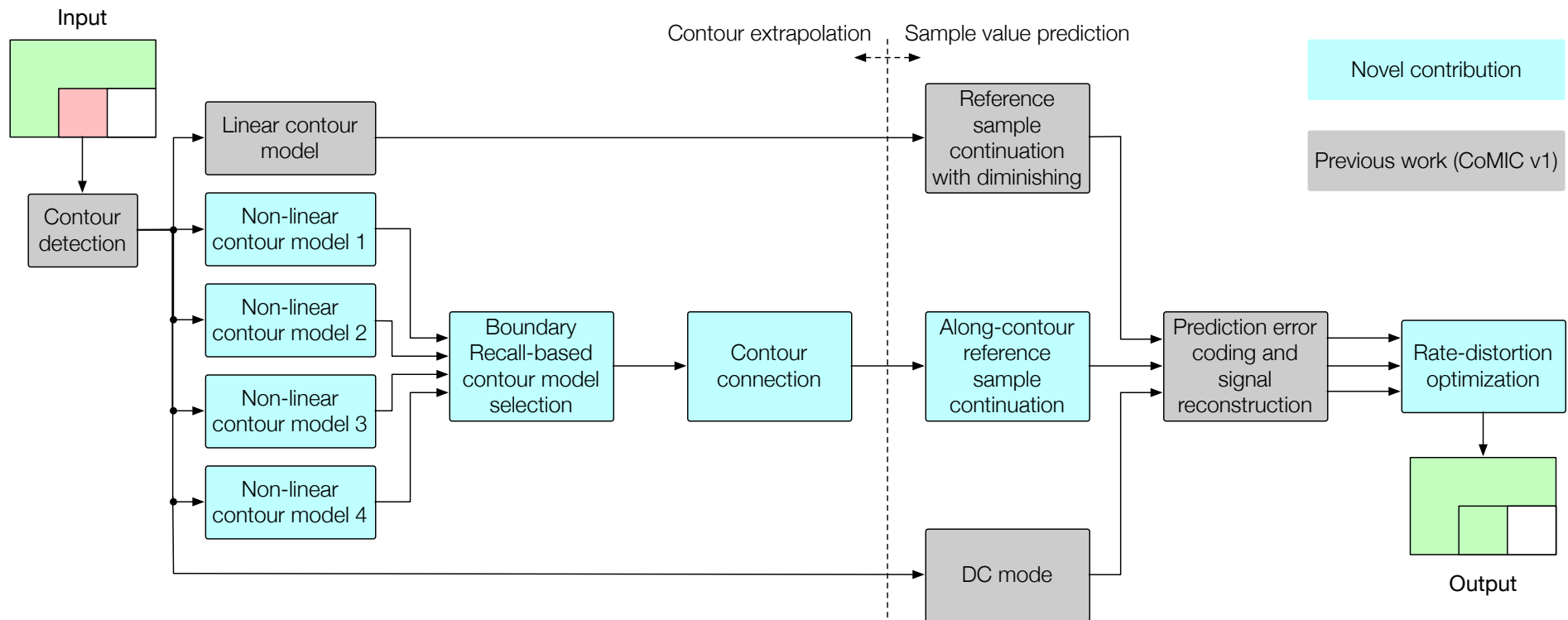
BD rates relative to HM

Complexity



■ Non-linear prediction ■ Linear prediction ■ Transform coding ■ Mode selection ■ Rest

The CoMIC Codec Today

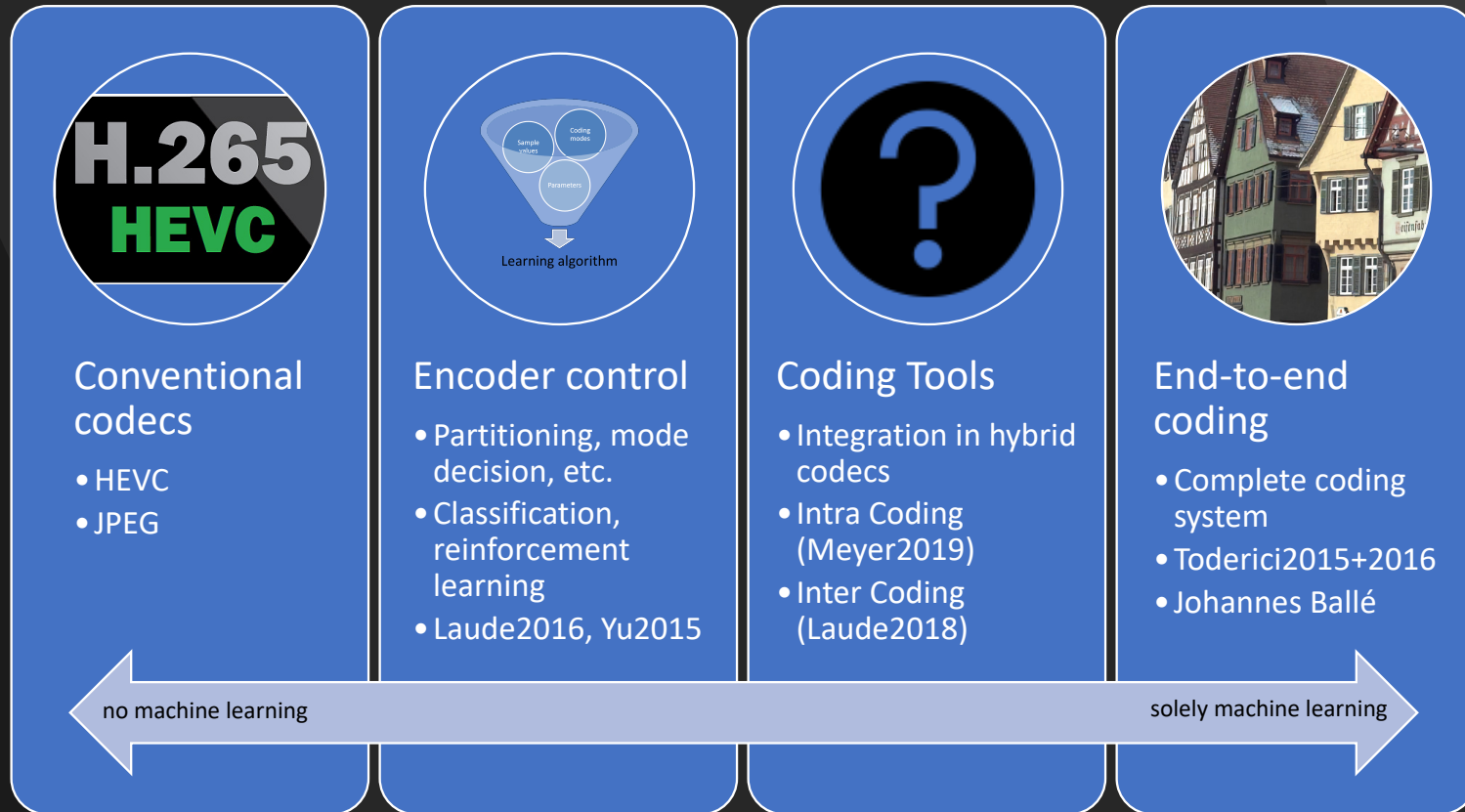


Current Work

Stochastic contour
modeling

Machine learning-based
sample value prediction

How much machine learning is good?



More Information

- Laude, T., Tumbrägel, J., Munderloh, M., & Ostermann, J. (2018). Non-linear contour-based multidirectional intra coding. *APSIPA Transactions on Signal and Information Processing*, 7(11).
<https://doi.org/10.1017/ATSIP.2018.14>
- Laude, T., & Ostermann, J. (2016). Contour-based Multidirectional Intra Coding for HEVC. In *Proceedings of 32nd Picture Coding Symposium (PCS)*. Nuremberg, Germany: IEEE.
<https://doi.org/10.1109/PCS.2016.7906319>

Conclusion



CoMIC



Combination of contour modeling and sample value prediction



Bit rate savings



Good extend of complexity

Visual Quality Assessment for Motion Compensated Frame Interpolation

Hui Men
University of Konstanz

SVCP 2019 | Konstanz | 2019-6-18



Universität Stuttgart

Universität
Konstanz



DFG

Introduction

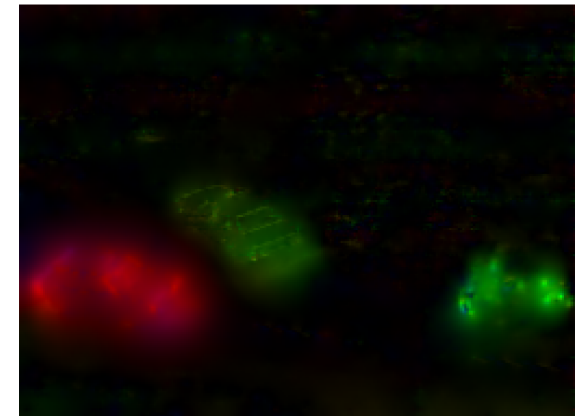
Motion Estimation

- Example: Analysis of Hamburg Taxi Sequence

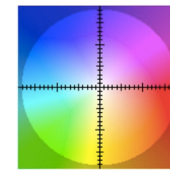


Frame 1 \longrightarrow Frame 2

Movements of the cars ?



Color-coded
Displacement Field
Flow Field

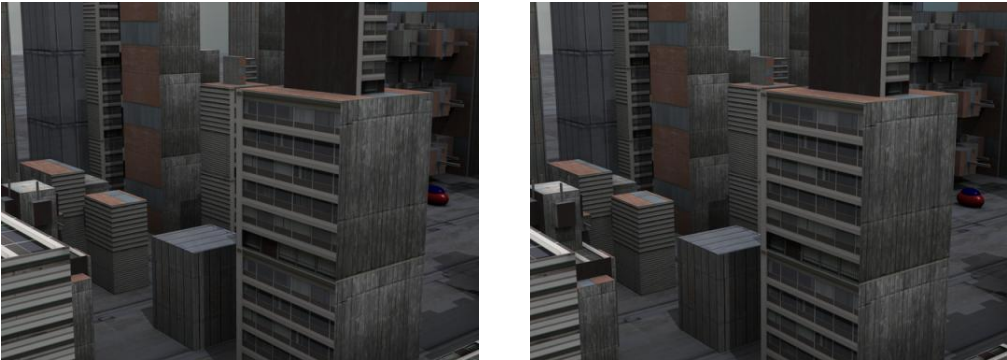


Flow color coding

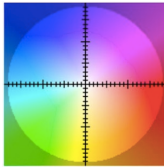
Introduction

Motion Estimation

- Optical Flow



Frame 1 \longrightarrow Frame 2
Displacement ?



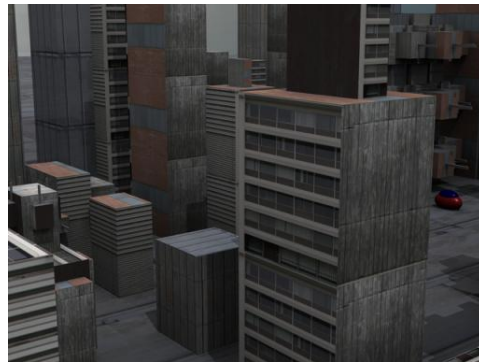
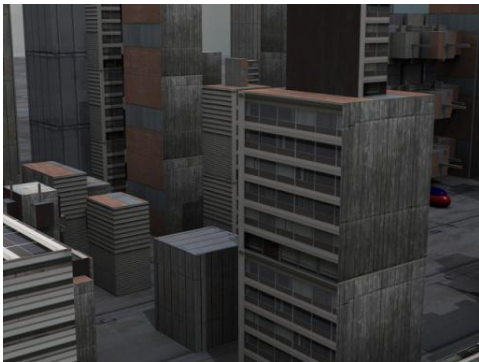
Flow color coding



Optical Flow

Quantitative Evaluation

- Angular error & endpoint error between flow vector & ground-truth flow



Ground-truth flow

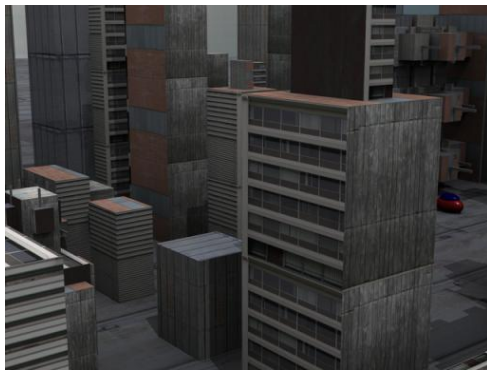
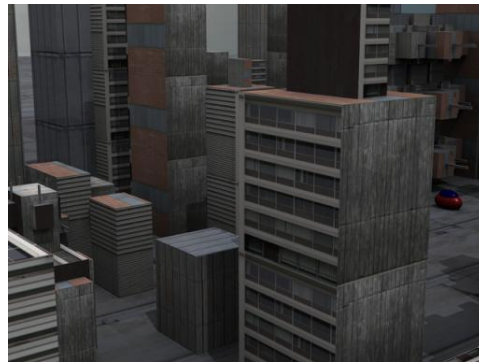
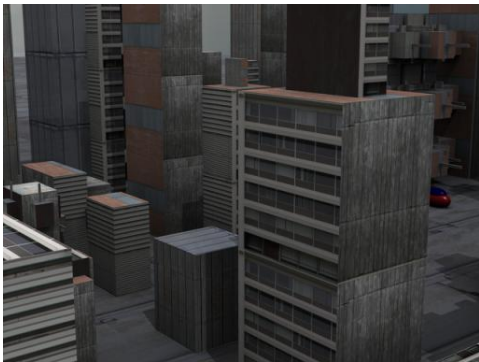


Flow vector

Optical Flow

Quantitative Evaluation

- MSE between interpolated image & ground-truth in-between image



Ground-truth in-between Image

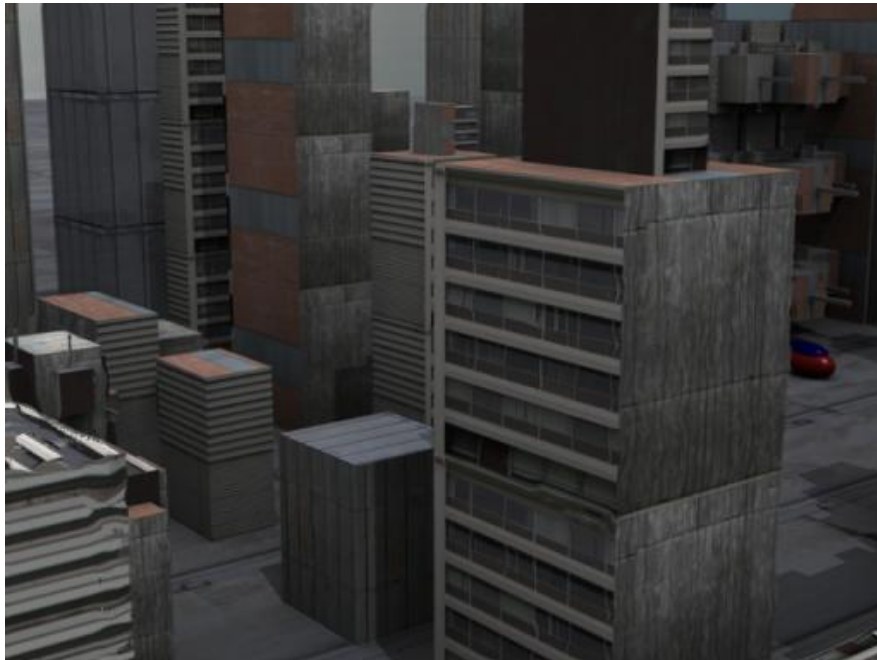


Interpolated Image

Optical Flow

Quantitative Evaluation

- Is MSE enough from the visual quality aspect ?



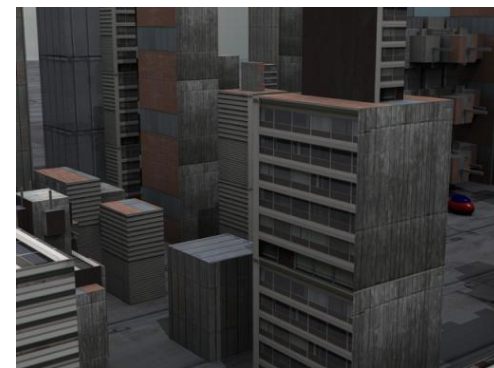
Q: Which image has a better quality?

- Left Right The same

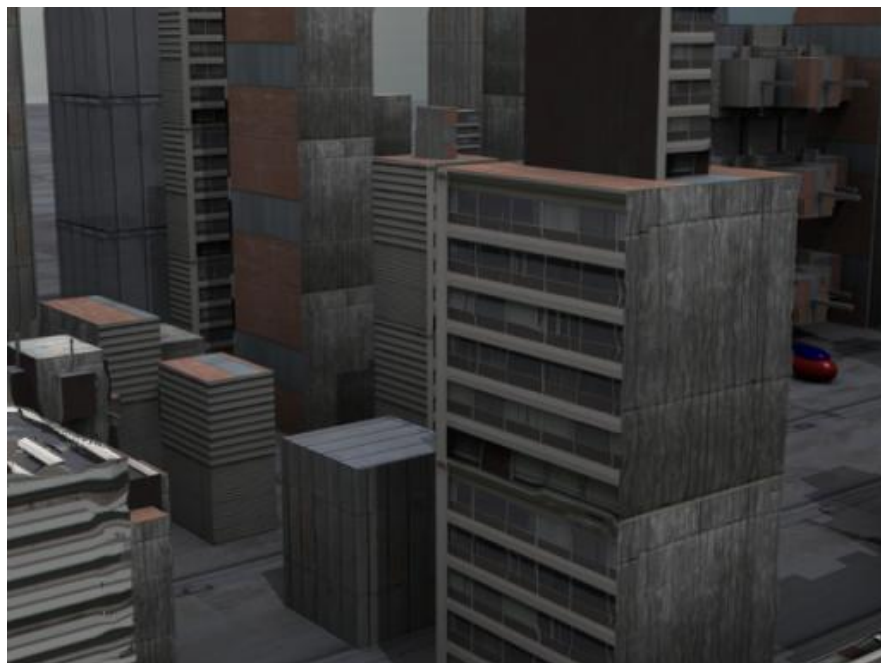
Optical Flow

Quantitative Evaluation

- Is MSE enough from the visual quality aspect ?



Ground-truth in-between image



MSE: 11.9



MSE: 11.9

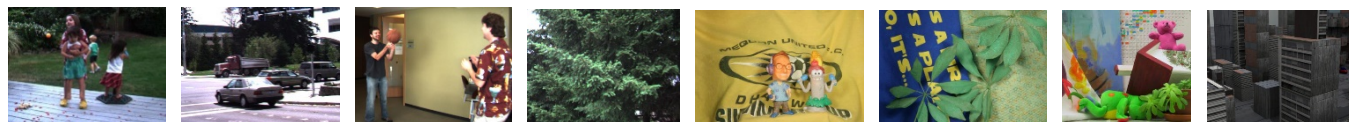
Adopt
Visual Quality Assessment
to
Optical Flow Benchmarks

Middlebury Benchmark

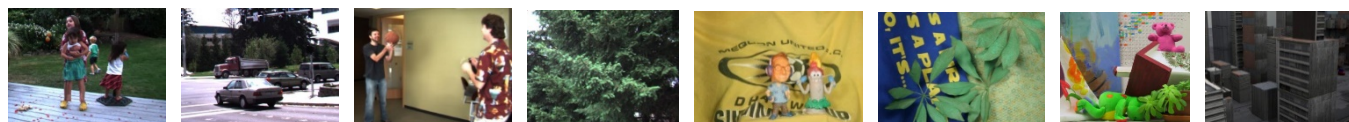
Evaluation Metrics

- Flow Accuracy: Endpoint & Angular Error
- Interpolation Quality: RMSE & Gradient-normalized RMSE

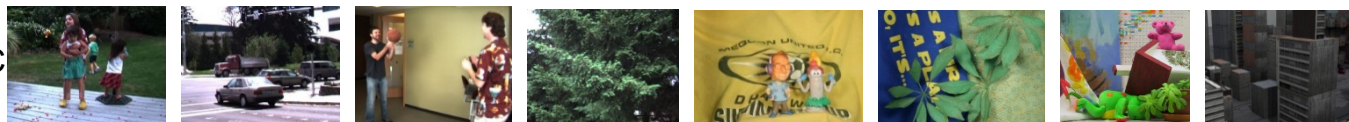
Ground-truth In-between Images



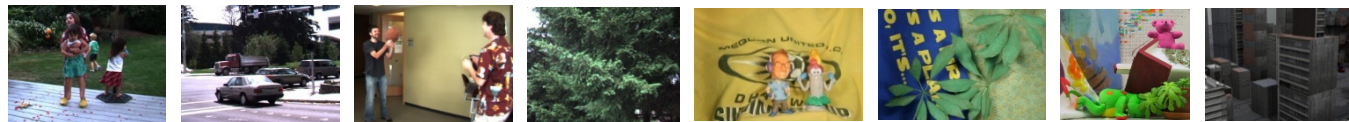
Epicflow



NNF-EAC



Seg-OF



...

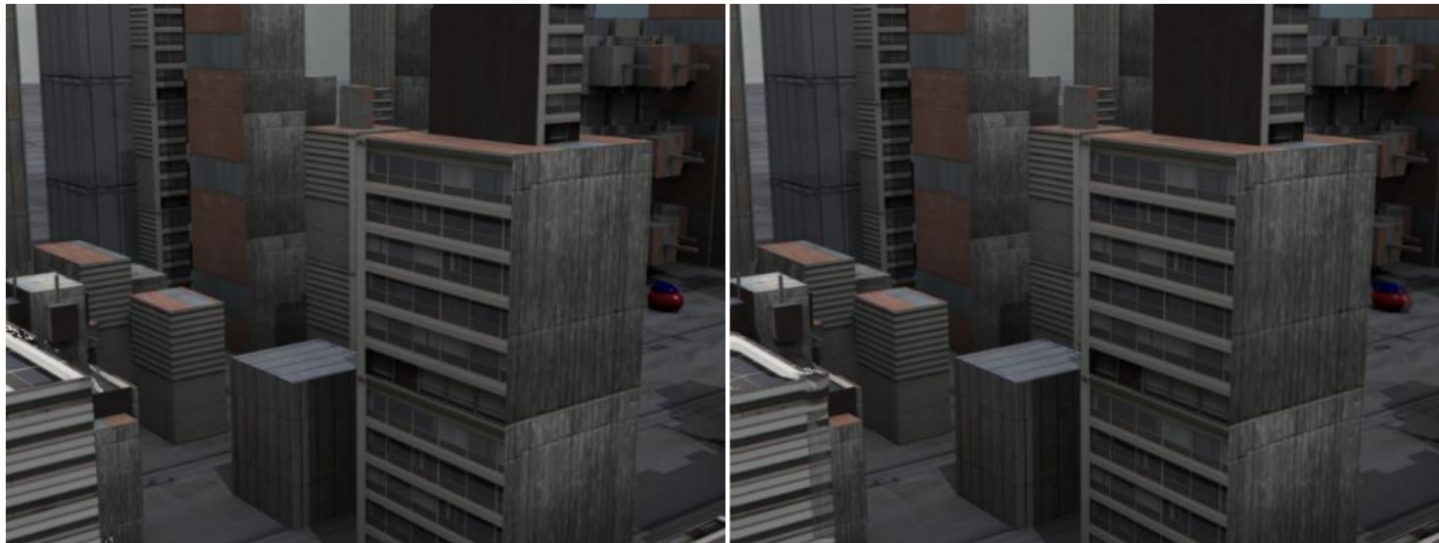
...

Interpolated Images
given by
141 optical flow methods

* In Collab. with SFB-TRR 161 Project B04

Paired Comparisons using Crowdsourcing

Crowdsourcing Interface



Which of the two images has a better quality? (required)

- the left
- the right

Paired Comparisons using Crowdsourcing

Task Instructions

Compare Pairs Of Images

Instructions ▾

Overview

We need your help to compare pairs of images. Try to quickly answer the question. We are interested in your first impression.

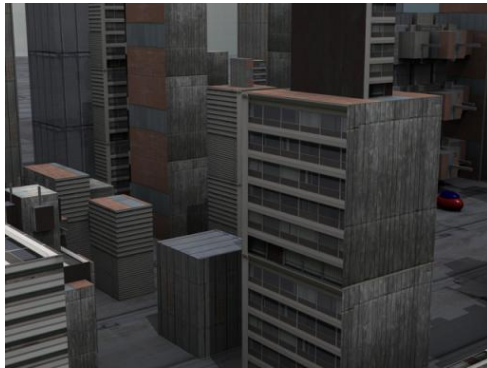
Specifically, you can pay more attention to the parts in the red circles:



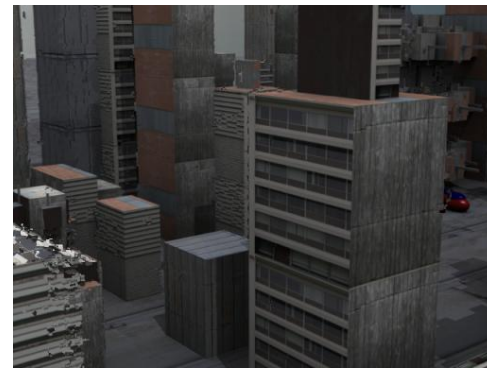
Paired Comparisons using Crowdsourcing

Quality Control

- Test Questions



Ground-truth image



Bad quality image

- Accuracy Requirement: 70%

Paired Comparisons using Crowdsourcing

Experimental Details

- Full pair comparison: 78,960 pairs
 - Time & Money Consuming
- Random connected pair comparison
 - Degree 6: 423 pairs

- # Pairs/page: 20
- # Votes/pair: 30

- Running Time

Running Time	Average	Mequon	Schefflera	Urban	Teddy	Backyard	Basketball	Dumptruck	Evergreen
Hours	29	60	72	10	20	10	3	20	35

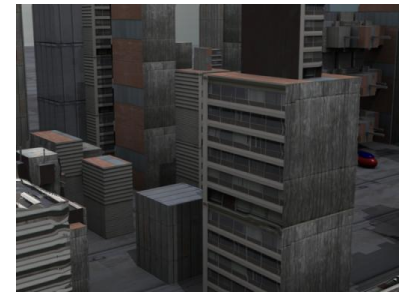
Result Reconstruction

- Adding 2 Anchors



Anchor: worst

423 pairs of images



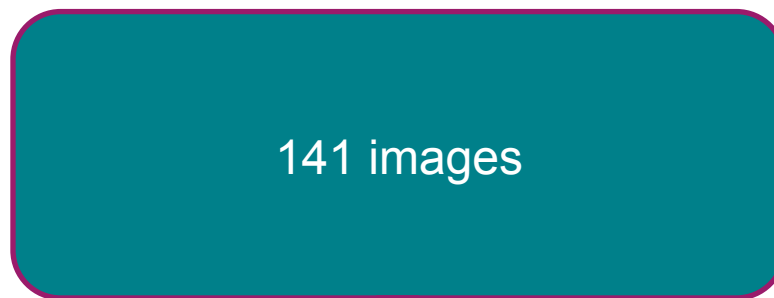
Anchor: best

Result Reconstruction

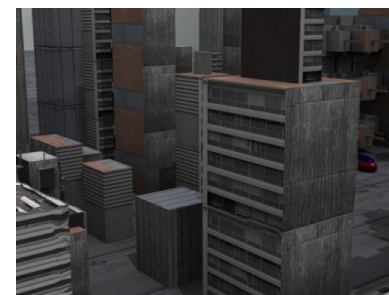
- Adding 2 Anchors



Anchor: worst
0



(0,1)

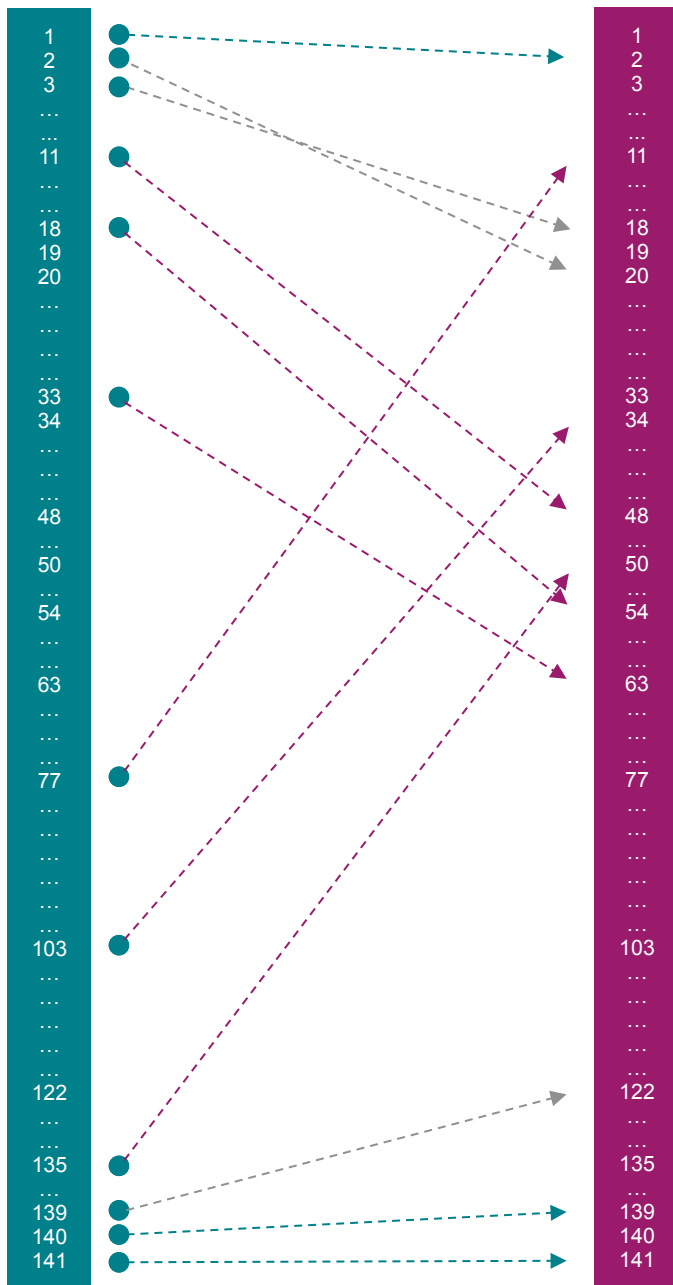


Anchor: best
1

- Reconstruction using Thurstone's Model
- Rescale to $[0,1]$
- Correlations: Average SROCC = 0.598

Re-ranking Result

	Average new / old	Mequon new / old	Schefflera new / old	Urban new / old	Teddy new / old	Backyard new / old	Basketball new / old	Dumptruck new / old	Evergreen new / old
SuperSlomo	1 / 5	3 / 2	71 / 34	1 / 1	2 / 2	6 / 1	8 / 3	1 / 1	32 / 3
CtxSyn	2 / 1	2 / 1	3 / 1	91 / 63	1 / 1	1 / 2	13 / 1	2 / 3	61 / 2
DeepFlow2	3 / 9	26 / 24	39 / 57	9 / 8	23 / 21	26 / 39	3 / 6	18 / 22	9 / 31
SuperFlow	4 / 10	13 / 15	104 / 70	82 / 41	46 / 40	5 / 9	4 / 25	10 / 31	5 / 7
DeepFlow	5 / 7	51 / 22	13 / 56	16 / 7	17 / 26	38 / 40	11 / 10	8 / 4	21 / 30
ALD-Flow	6 / 21	27 / 95	2 / 48	22 / 6	15 / 49	116 / 26	1 / 12	12 / 7	82 / 74
PMMST	7 / 4	7 / 10	31 / 15	17 / 4	77 / 20	19 / 6	5 / 5	32 / 9	68 / 9
Aniso. Huber-L1	8 / 13	18 / 16	32 / 99	35 / 24	12 / 10	47 / 46	24 / 11	11 / 23	10 / 15
SIOF	9 / 28	49 / 38	36 / 98	24 / 52	25 / 28	12 / 4	21 / 21	45 / 16	35 / 48
CBF	10 / 8	17 / 4	40 / 66	12 / 9	14 / 5	27 / 3	40 / 13	29 / 39	30 / 12
Bartels	11 / 77	86 / 115	22 / 72	18 / 22	64 / 77	7 / 12	73 / 102	3 / 17	33 / 58
IROF++	12 / 17	4 / 31	6 / 26	71 / 64	7 / 3	77 / 36	6 / 45	19 / 18	83 / 55
LDOF	13 / 41	25 / 32	98 / 74	89 / 65	4 / 35	2 / 8	17 / 50	80 / 79	41 / 11
RNLOD-Flow	14 / 71	80 / 37	4 / 63	33 / 72	33 / 15	10 / 105	52 / 92	24 / 47	44 / 100
2nd-order prior	15 / 24	14 / 11	72 / 87	38 / 28	60 / 27	23 / 48	70 / 27	5 / 24	16 / 40
SepConv-v1	16 / 6	6 / 3	95 / 23	83 / 50	8 / 30	68 / 7	15 / 4	41 / 2	2 / 1
DF-Auto	17 / 20	95 / 14	7 / 69	41 / 30	55 / 22	3 / 15	28 / 31	78 / 52	31 / 18
MDP-Flow2	18 / 3	19 / 8	89 / 10	20 / 13	13 / 4	28 / 5	49 / 24	13 / 8	95 / 24
CLG-TV	19 / 15	41 / 13	96 / 88	47 / 17	6 / 16	74 / 47	46 / 9	9 / 20	11 / 25
FGIK	20 / 2	15 / 5	113 / 112	112 / 113	16 / 124	15 / 100	12 / 81	7 / 131	59 / 117
IROF-TV	21 / 14	53 / 40	50 / 40	15 / 20	11 / 7	54 / 37	32 / 49	61 / 44	49 / 16
LME	22 / 16	23 / 17	54 / 37	40 / 40	24 / 23	49 / 59	36 / 40	15 / 6	76 / 32
TC/T-Flow	23 / 38	59 / 83	19 / 61	4 / 18	79 / 79	94 / 53	2 / 33	103 / 92	43 / 60
Modified CLG	24 / 35	12 / 7	134 / 106	25 / 49	21 / 46	14 / 25	104 / 48	25 / 14	20 / 35
CombBMOF	25 / 19	87 / 69	9 / 20	61 / 37	18 / 58	65 / 22	16 / 32	50 / 42	54 / 26
CRTflow	26 / 48	40 / 47	34 / 95	79 / 48	68 / 14	82 / 38	18 / 20	20 / 89	25 / 64
...

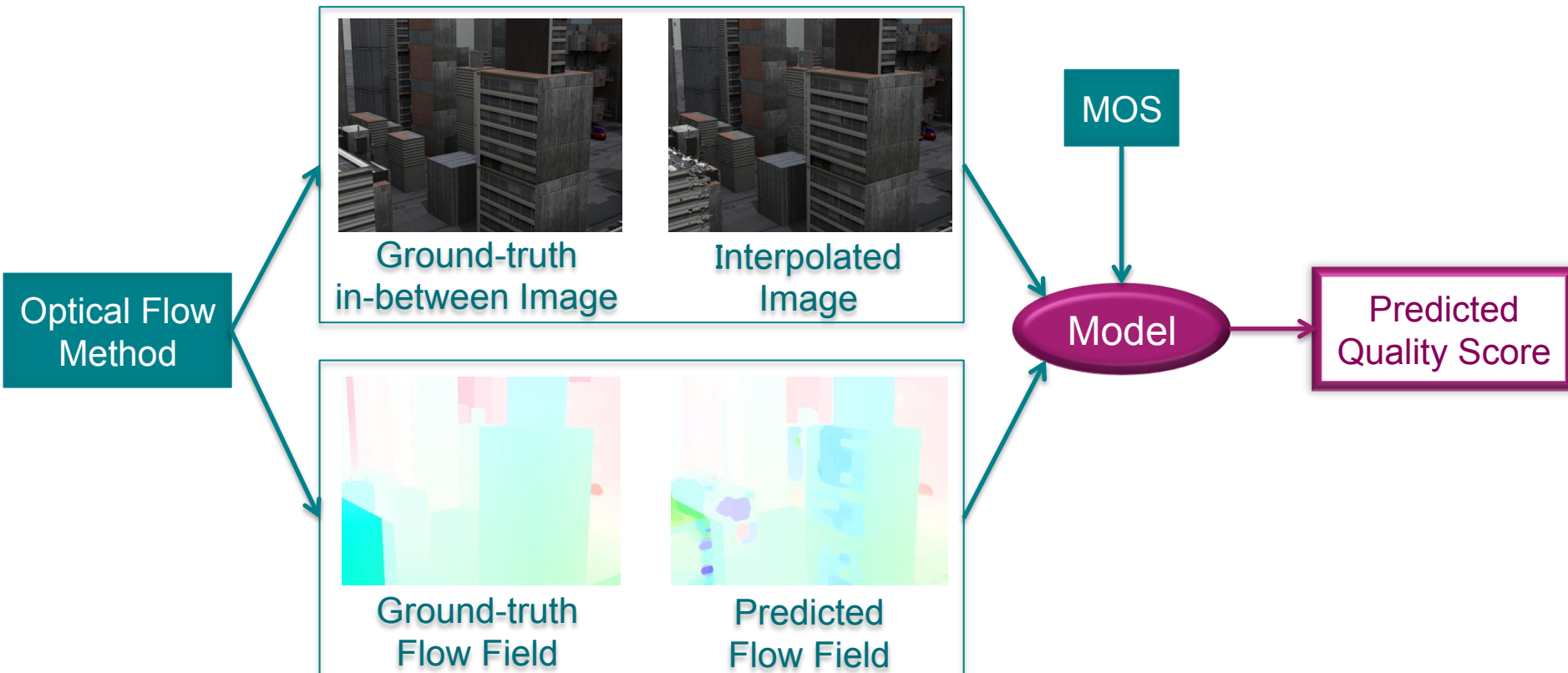


Optical Flow Method	Ranking Old / New	Ranking Differences
CtxSyn	1 / 2	1
FGIK	2 / 20	18
MDP-Flow2	3 / 18	15
...
...
NN-field	11 / 48	37
...
NNF-EAC	18 / 54	36
...
...
F-TV-L1	33 / 63	30
...
...
Bartels	77 / 11	66
...
TI-DOFE	103 / 34	69
...
PGAM+LK	135 / 50	85
...
...
Pyramid LK	139 / 122	17
GroupFlow	140 / 139	1
Periodicity	141 / 141	0

Future Work 1

New FR-IQA Model

- Specifically trained for images interpolated by optical flow
- Can be used as an evaluation metric in the benchmark



Future Work 2

Video Quality Assessment for Optical Flow

- Interpolate videos using optical flow methods
- Evaluate the quality of the interpolated videos



How is the video quality?

Thanks !

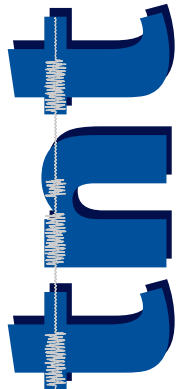
Hui Men, Hanhe Lin, Vlad Hosu, Daniel Maurer, Andrés Bruhn, Dietmar Saupe,
“Visual Quality Assessment for Motion Compensated Frame Interpolation”,
2019 Eleventh International Conference on Quality of Multimedia Experience
(QoMEX)

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

Holger Meuel

Institut für Informationsverarbeitung
Leibniz Universität Hannover, Germany

June 19th, 2019

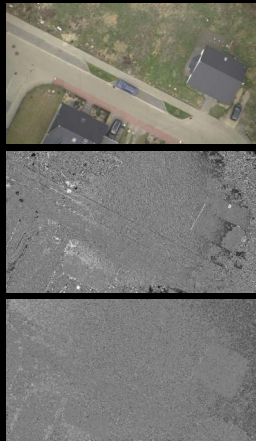


Motivation

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

Goal:

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



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

Outline

Efficiency Analysis of Affine Motion Compensated Prediction

- Overview of the Derivations

- Affine Motion and Error Model

- Model Displacement Estimation Error Probability Density Function (pdf)

- Model Video and Error Signal Power Spectral Densities (PSDs)

- Rate-Distortion Analysis

Simulations

Experiments

Conclusion

Outline

Efficiency Analysis of Affine Motion Compensated Prediction

Overview of the Derivations

Affine Motion and Error Model

Model Displacement Estimation Error Probability Density Function (pdf)

Model Video and Error Signal Power Spectral Densities (PSDs)

Rate-Distortion Analysis

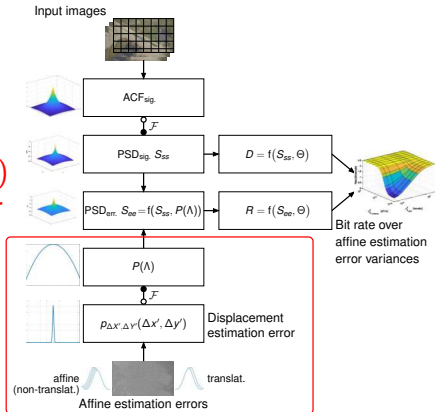
Simulations

Experiments

Conclusion

Overview: Bit Rate Derivation for Affine Estimation Errors

- ▶ Modeling of power spectral density (PSD) of signal
- ▶ Modeling of probability density function (pdf) $p_{\Delta x', \Delta y'}(\Delta x', \Delta y')$ of displacement estimation error
- ▶ Derivation of PSD of displacement estimation error $S_{ee}(\Lambda)$ ¹
- ▶ Application of rate-distortion theory \Rightarrow bit rate²



¹Bernd 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

²Toby 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

Overview of the Derivations

Affine Motion and Error Model

Model Displacement Estimation Error Probability Density Function (pdf)

Model Video and Error Signal Power Spectral Densities (PSDs)

Rate-Distortion Analysis

Simulations

Experiments

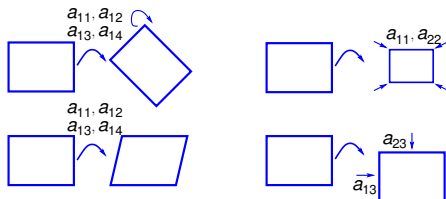
Conclusion

Motion Model

Affine motion model:

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

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



Affine Motion Estimation

Estimated affine motion:

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

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

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

Affine Error Model

Displacement estimation error *in the frame*:

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

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

Outline

Efficiency Analysis of Affine Motion Compensated Prediction

Overview of the Derivations

Affine Motion and Error Model

Model Displacement Estimation Error Probability Density Function (pdf)

Model Video and Error Signal Power Spectral Densities (PSDs)

Rate-Distortion Analysis

Simulations

Experiments

Conclusion

Probability Density Function Derivation

- ▶ Assumption: e_{ij} follow zero-mean Gaussian distributed pdfs
- ⇒ Joint pdf for independent e_{ij} :

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

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

Probability Density Function of the Displacement Estimation Error

With transformation theorem for pdfs:

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

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

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

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

Outline

Efficiency Analysis of Affine Motion Compensated Prediction

Overview of the Derivations

Affine Motion and Error Model

Model Displacement Estimation Error Probability Density Function (pdf)

Model Video and Error Signal Power Spectral Densities (PSDs)

Rate-Distortion Analysis

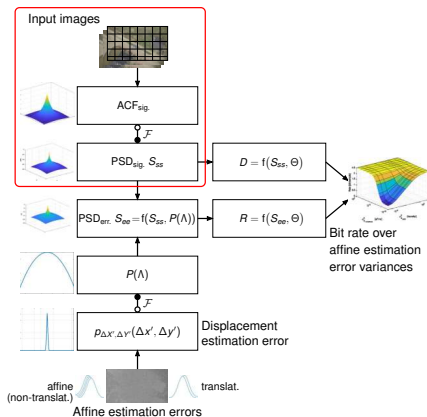
Simulations

Experiments

Conclusion

Signal and Error Power Spectral Density Functions

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



Outline

Efficiency Analysis of Affine Motion Compensated Prediction

Overview of the Derivations

Affine Motion and Error Model

Model Displacement Estimation Error Probability Density Function (pdf)

Model Video and Error Signal Power Spectral Densities (PSDs)

Rate-Distortion Analysis

Simulations

Experiments

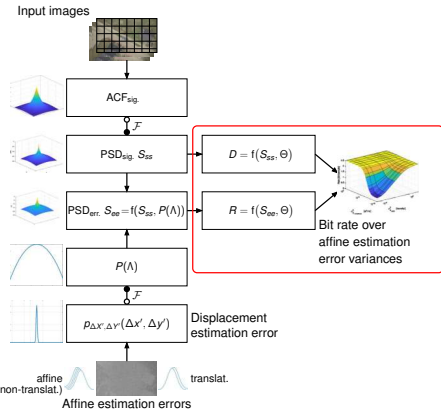
Conclusion

Rate-Distortion Theory³

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

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

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



³based 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

Overview of the Derivations

Affine Motion and Error Model

Model Displacement Estimation Error Probability Density Function (pdf)

Model Video and Error Signal Power Spectral Densities (PSDs)

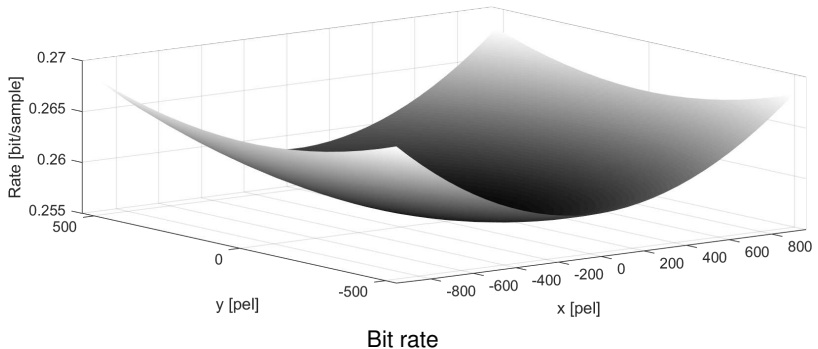
Rate-Distortion Analysis

Simulations

Experiments

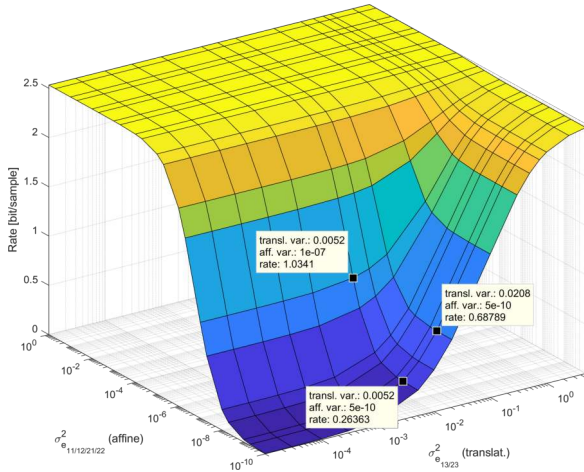
Conclusion

Location Dependent Bit Rate



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

Minimum Required Bit Rate for Prediction Error Coding



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

Outline

Efficiency Analysis of Affine Motion Compensated Prediction

Overview of the Derivations

Affine Motion and Error Model

Model Displacement Estimation Error Probability Density Function (pdf)

Model Video and Error Signal Power Spectral Densities (PSDs)

Rate-Distortion Analysis

Simulations

Experiments

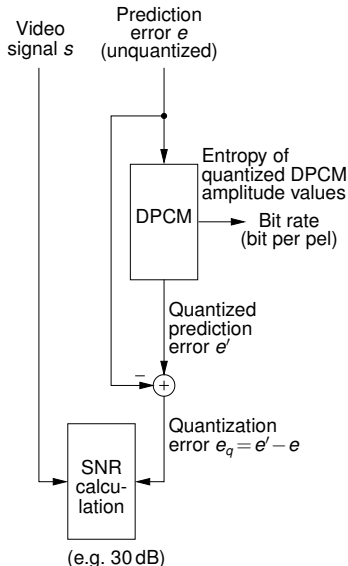
Conclusion

Experimental Setup

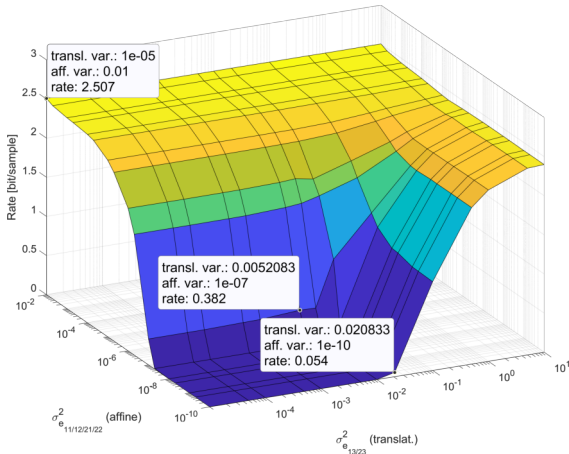
- ▶ Video signal s with artificially introduced motion of specific variances
 - ▶ Most-trivial motion estimation always predicting “no motion”
- ⇒ Introduced motion becomes exactly prediction error e

Experimental accomplishment:

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



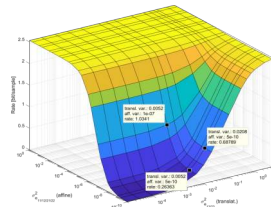
Measured Bit Rates for Encoding the Prediction Error



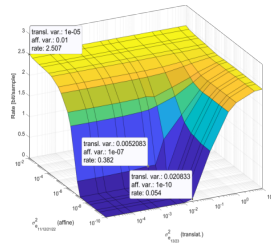
Measured bit rate for encoding the prediction error as a function of the motion estimation error variances, full HD resolution frame

Comparison between Theory and Experimental Data

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



Theory

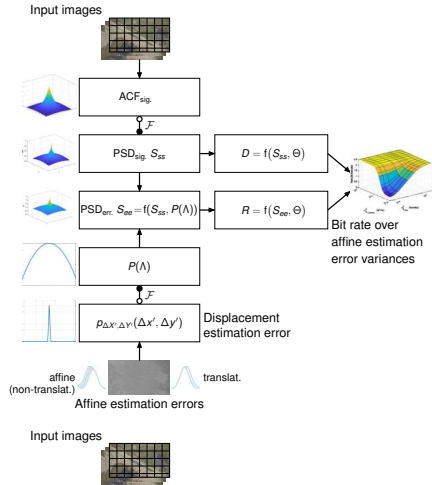


Measurement

Real-World Application of the Model?

Consideration of simplified affine model as used in upcoming VVC

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



Distinct Affine Test Sequences⁴



ShieldsPart, frame 1



ShieldsPart, frame 100



TractorPart, frame 1





TractorPart, frame 100

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

Model vs. Real-World Measurements

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

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

Conclusion:

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

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

Outline

Efficiency Analysis of Affine Motion Compensated Prediction

Overview of the Derivations

Affine Motion and Error Model

Model Displacement Estimation Error Probability Density Function (pdf)

Model Video and Error Signal Power Spectral Densities (PSDs)

Rate-Distortion Analysis

Simulations

Experiments

Conclusion

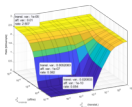
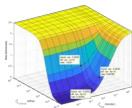
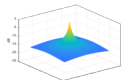
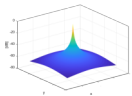
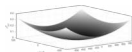
Application of RD Theory for Affine MCP in Video Coding

Model for affine motion compensation in video coding:

- ▶ Modeling of pdf of displacement estimation error
 $p_{\Delta x', \Delta y'}(\Delta x', \Delta y')$
 - ▶ Consideration of power spectral density of video signal
 - ▶ Derivation of power spectral density of displacement estimation error
 - ▶ Application of rate-distortion function
- ⇒ **Minimum bit rate for coding the prediction error**

Experimental verification:

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



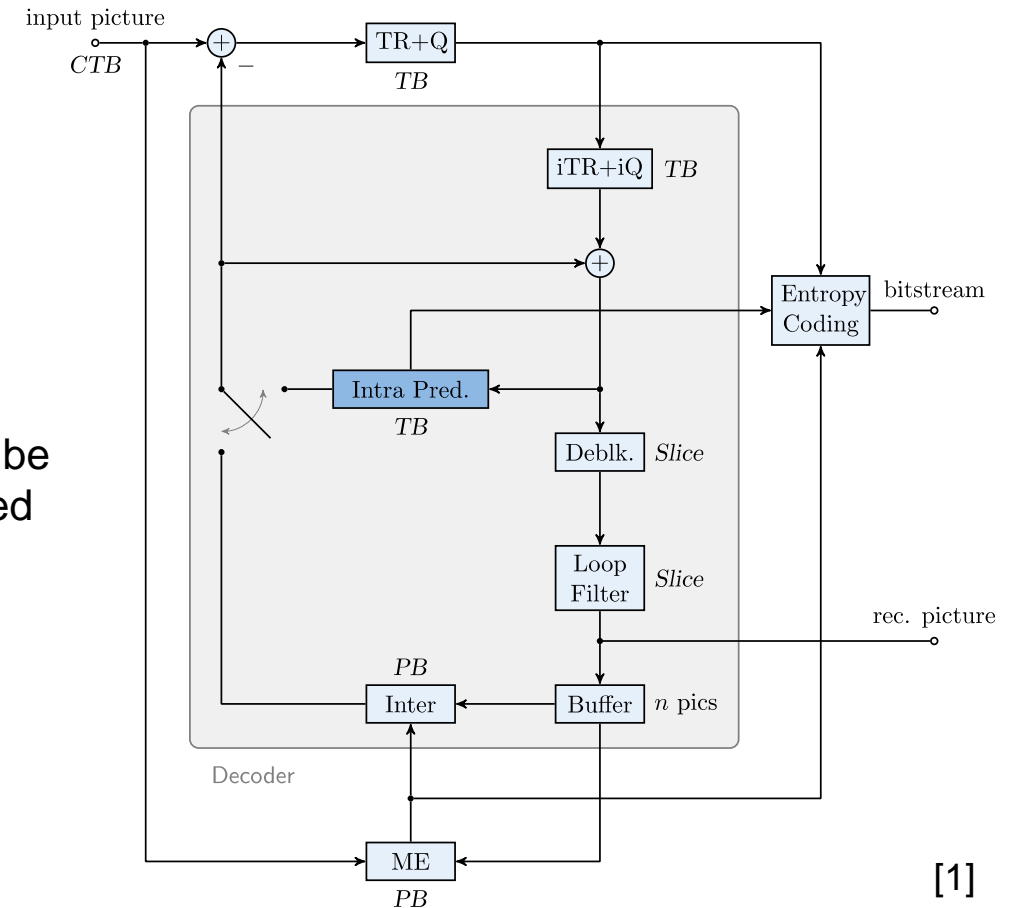
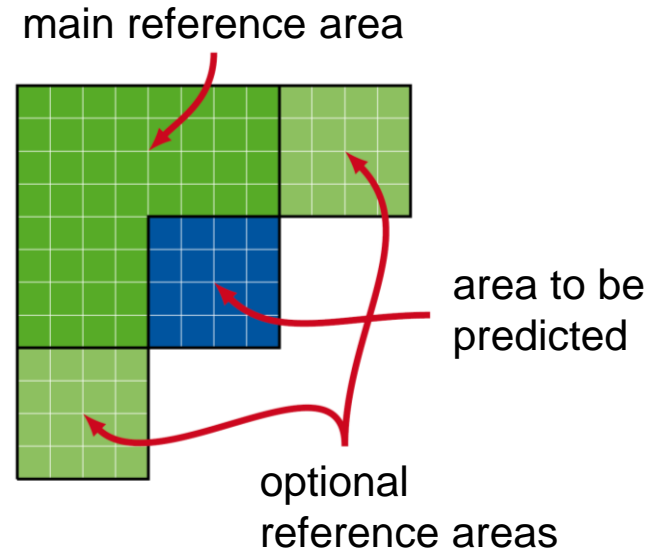
Architectures and Training Methods for Neural Network-based Intra Prediction

Maria Meyer, Jonathan Wiesner, Jens Schneider, Christian Rohlfing

Problem Statement: Neural Networks for Intra Prediction

Additional neural network (NN) - based intra prediction mode for hybrid video codecs:

- Block-based predictions
- Optionally available information
- Channel wise prediction
- Signaling and rate-distortion decisions
- Low Complexity



[1] M. Wien, High Efficiency Video Coding – Coding Tools and Specification. Berlin, Heidelberg: Springer, Sept. 2014

Overview

- Open Problems
- Prediction Network
 - Training Methods
 - Architecture
- Mode Signaling and Codec Integration
- Results and Evaluation
- Conclusion

Open Questions

Architecture:

- Best so far can not be definitely concluded due to different training sets
- Only three types of architectures tried so far

Chroma and Cross-Component Prediction:

- No separate consideration of chroma blocks
- No usage of cross component information

Loss Function:

- So far only sum of absolute transform differences (SATD) and mean square error (MSE) compared

Signaling:

- Flag causes a lot of overhead

[2] J. Li, B. Li, J. Xu, R. Xiong, and W. Gao, "Fully connected network-based intra prediction for image coding," *IEEE Transactions on Image Processing*, vol. 27, no. 7, pp. 3236–3247, July 2018.

[3] Y. Hu, W. Yang, M. Li, and J. Liu, "Progressive spatial recurrent neural network for intra prediction," *Computing Research Repository (CoRR)*, 2018

[4] T. Dumas, A. Roumy, and C. Guillemot, "Context-adaptive neural network based prediction for image compression," *Computing Research Repository (CoRR)*, 2018.

[5] J. Pfaff, P. Helle, D. Maniry, S. Kaltenstadler, W. Samek, H. Schwarz, D. Marpe, and T. Wiegand: *Neural Network based Intra Prediction for Video Coding*, *Proceedings of the SPIE 10752, Applications of Digital Image Processing XLI*, San Diego, USA, vol. 1075213, September 2018

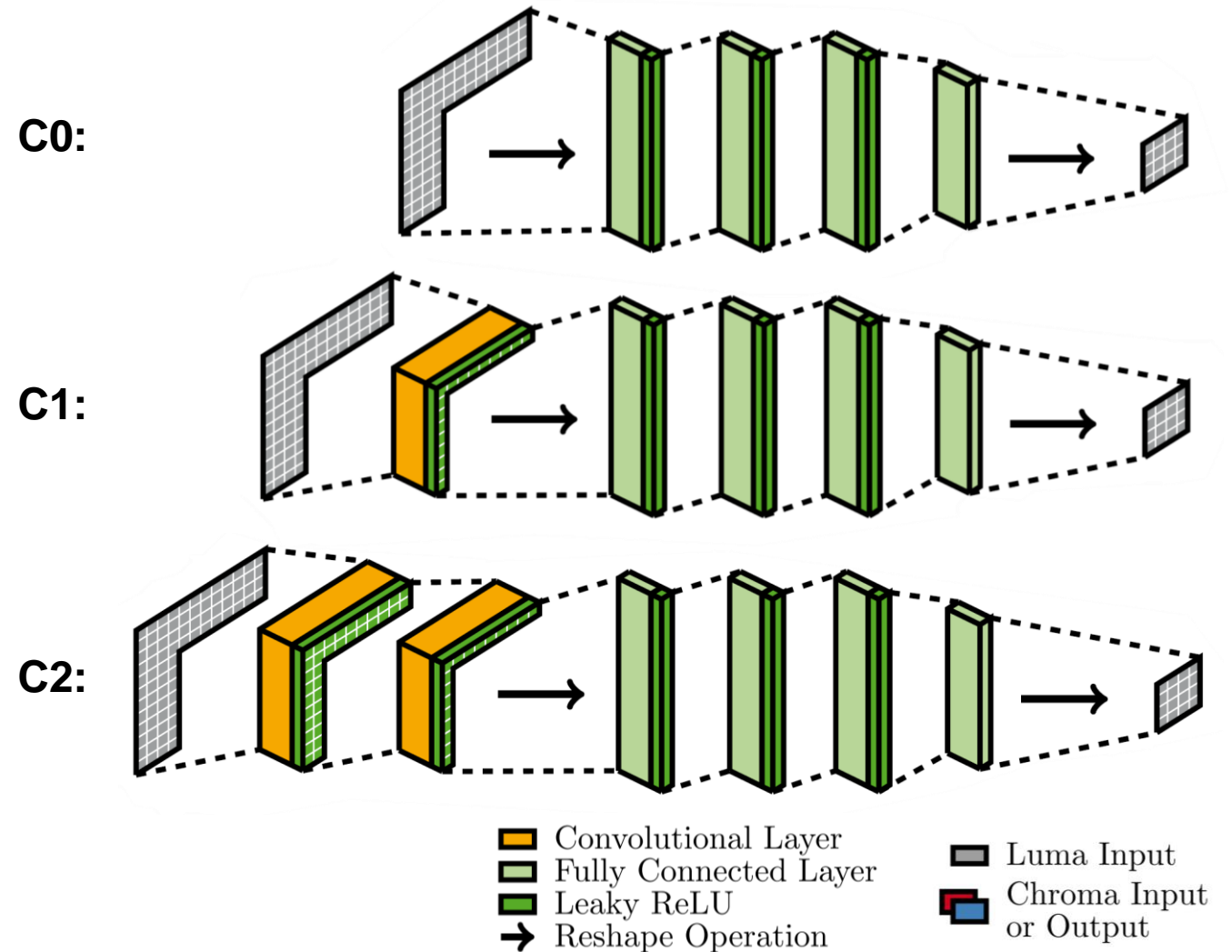
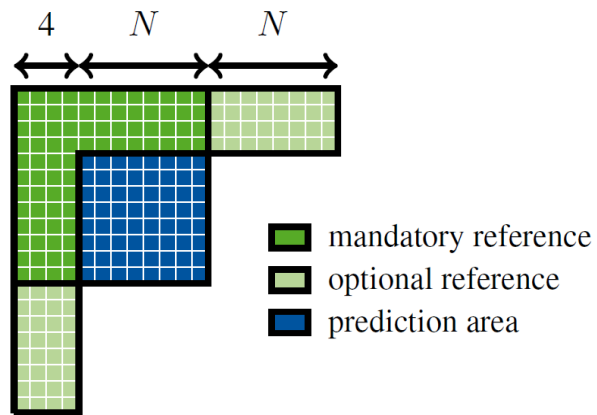
Prediction Network – Luma Architecture

General Settings:

- Four reference lines input
- Separate Networks for each block size

Compared Variants:

- Purely fully-connected architecture (C0)
- Convolutional layers followed by fully-connected ones (C1, C2)



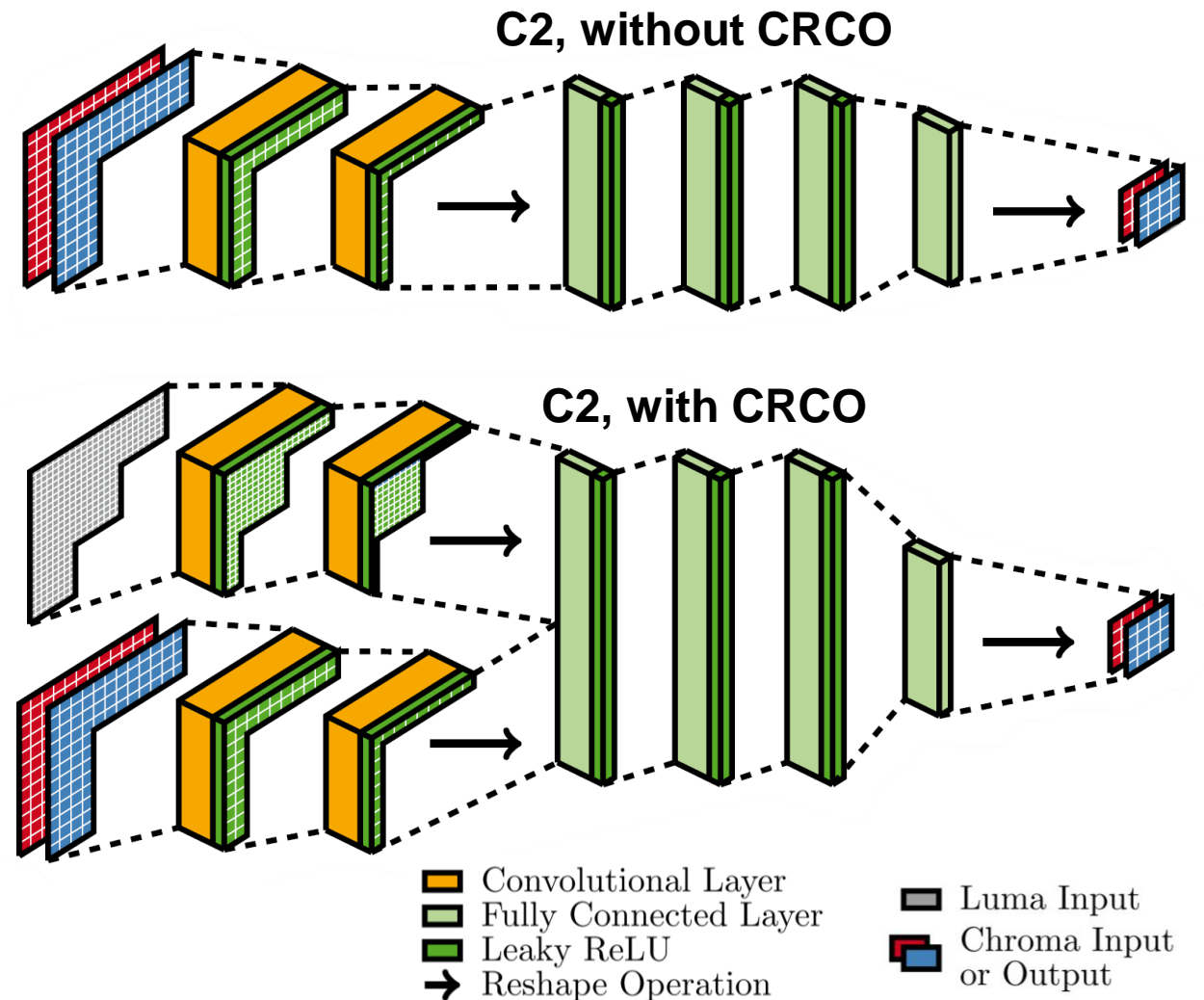
Prediction Network – Chroma Architecture

Joint Chroma Channel Prediction:

- Two input and two output channels
- Otherwise same as luma prediction

Cross-Component Adaptation (CRCO):

- Problems:
 - Different input shape
 - Different resolution
- Architectural Solution:
 - Additional convolutional branch processing luma information
 - Concatenation before first fully connected layer



Prediction Network – Training Methods

Datasets:

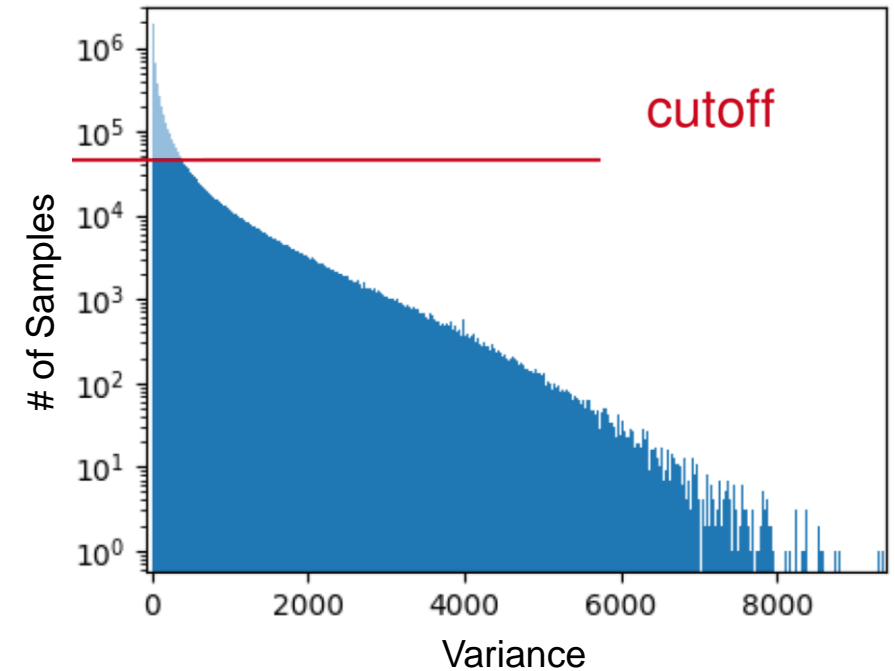
- Extracted samples from 115 raw videos
- Optional input areas masked
- Excluding a portion of the low variance samples possible without loss of bd-rate gains

Training Methods:

- Adam optimizer
- SATD or L1 loss with regularization term

Problems:

- Overfitting for larger chroma blocks



Sample / Parameters Relation
for C2 architecture with CRCO

Blocksize	Luma	Chroma
4	46.29	2.80
8	21.17	0.68
16	7.65	0.12
32	1.10	–

Prediction Examples and Evaluation

Here:

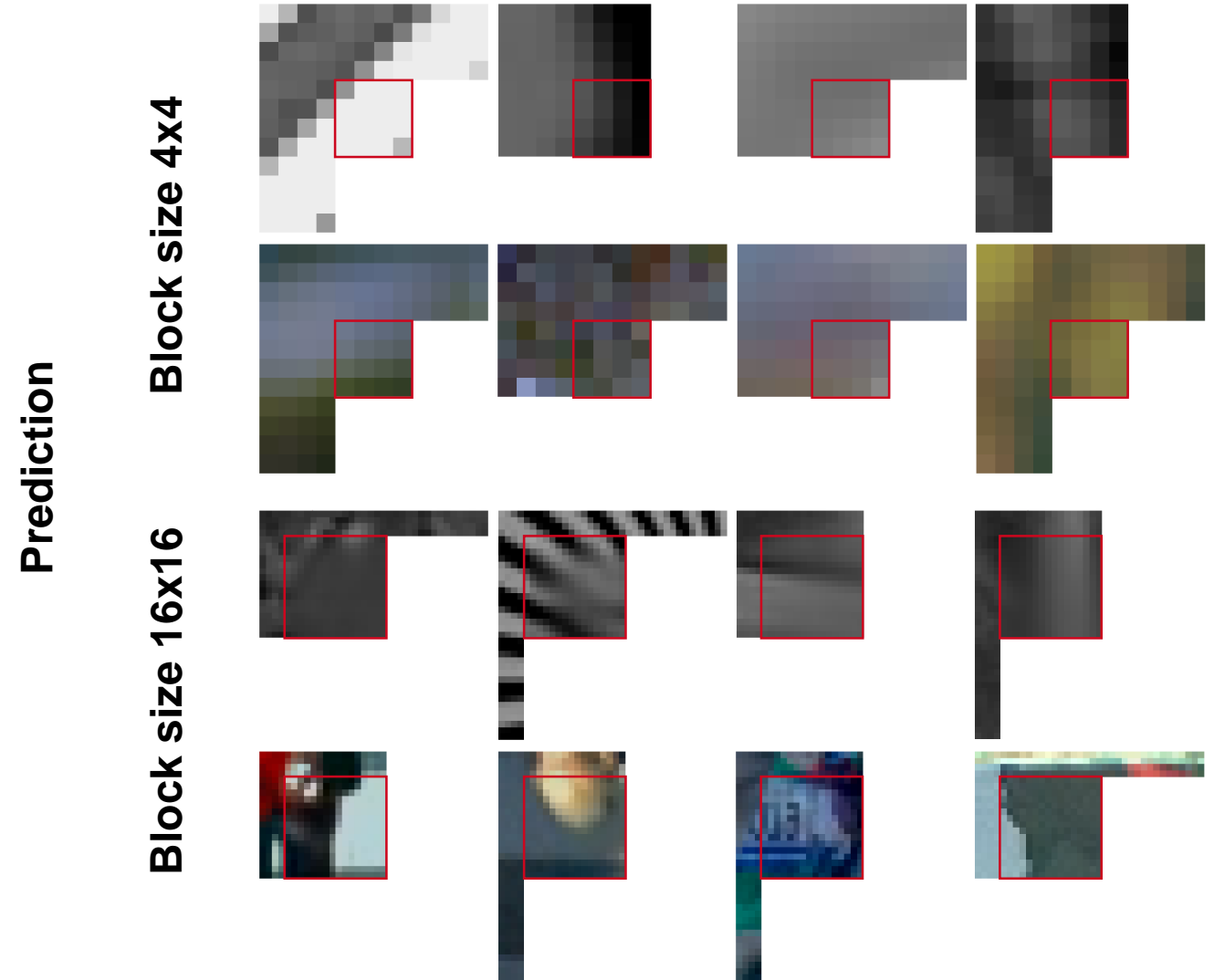
- C2 architecture with CRCO

Luma Samples:

- Enables continuing more than one direction, circles etc.
- Tending towards mean value when continuation unclear/ in bottom right corner

Chroma:

- Enables use of additional luma information



Prediction Examples and Evaluation

Here:

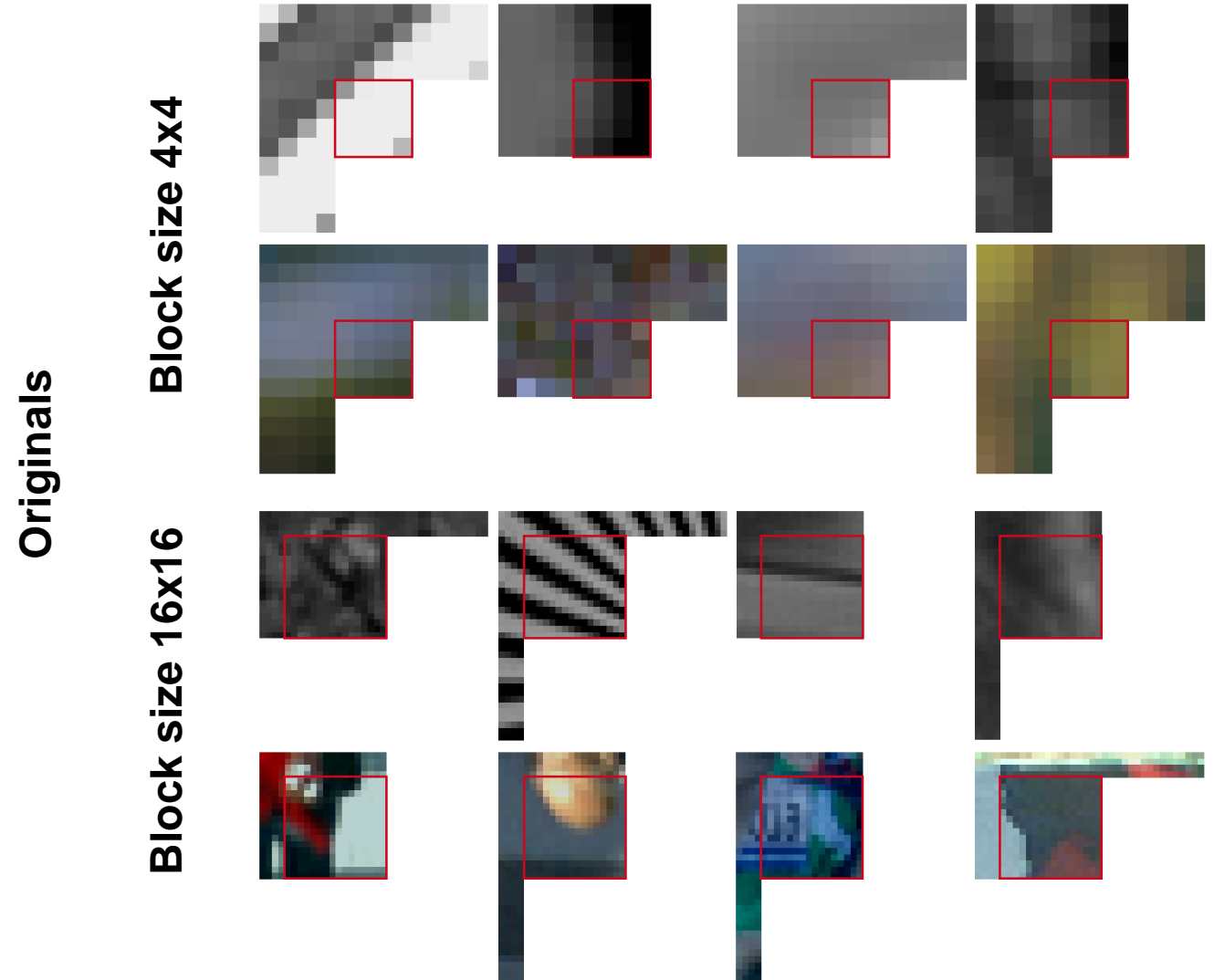
- C2 architecture with CRCO

Luma Samples:

- Enables continuing more than one direction, circles etc.
- Tending towards mean value when continuation unclear/ in bottom right corner

Chroma:

- Enables use of additional luma information



Mode Integration and Signaling

Integration:

- Implemented in HM16.9 as 36th intra mode
- RD-decision as for any other intra mode

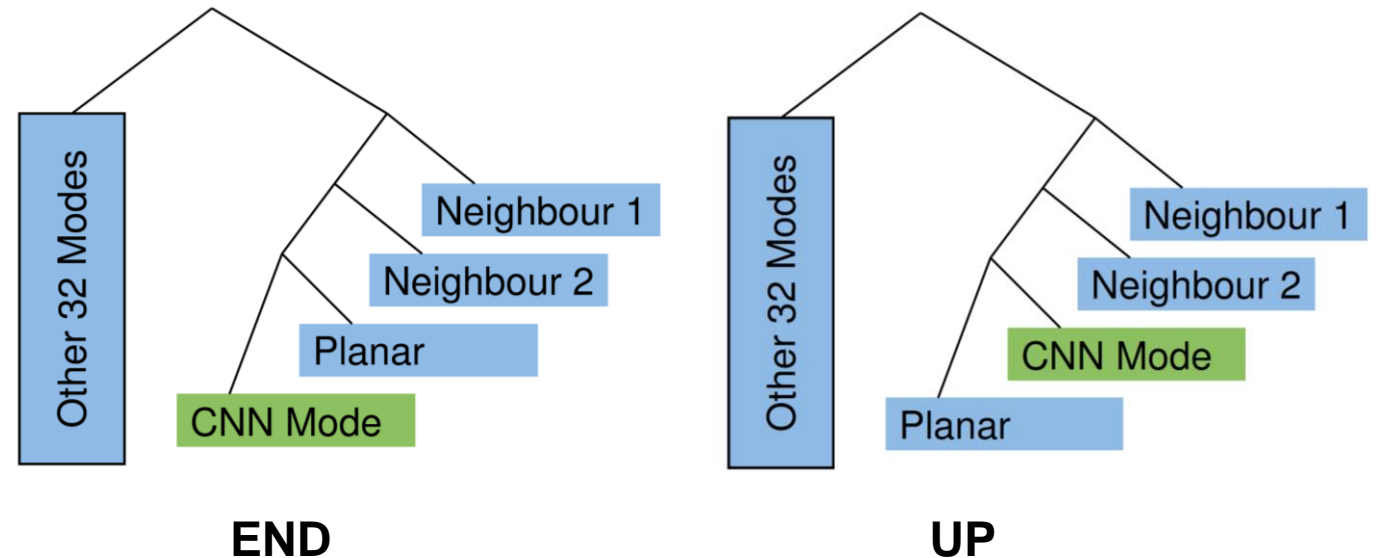
Luma Signaling:

- Most probable mode list extended to four items
- New mode always on MPM-list
- Two variants for MPM-list placement
 - UP: directly behind neighbors
 - END: at the last list position

Chroma Signaling:

- No dedicated signaling for chroma
 - Only useable, when used for luma

Decision Tree Examples



Results – Architecture and Loss

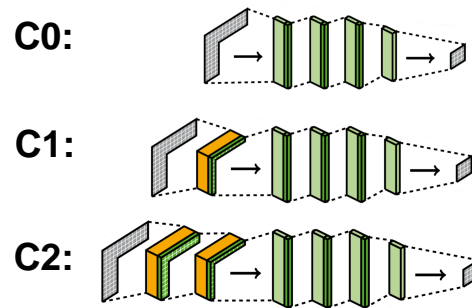
From BD-rates:

- SATD outperforms L1
- C2 outperforms C1 and C0 on average
- C0 better for noisy, high resolution content

Further Analysis:

- C2 always better validation loss
- Difference increasing with block size
- C2 more used for 4x4 blocks, C0 for 32x32 blocks in all class B sequences

Loss Function	L1	SATD
BQTerrace	-1.51 %	-1.61 %
BasketballDrive	-1.97 %	-2.30 %
Cactus	-2.08 %	-2.30 %
Kimono	-2.61 %	-3.17 %
ParkScene	-2.75 %	-2.85 %
AVG Class B	-2.18 %	-2.45 %



Architecture	C2	C1	C0
BQTerrace	-1.79 %	-1.74 %	-1.61 %
BasketballDrive	-2.33 %	-2.28 %	-2.30 %
Cactus	-2.46 %	-2.43 %	-2.30 %
Kimono	-2.66 %	-3.02 %	-3.17 %
ParkScene	-2.55 %	-2.66 %	-2.85 %
AVG Class B	-2.36 %	-2.43 %	-2.45 %
BQMall	-2.00 %	-1.85 %	-1.85 %
BasketballDrill	-1.99 %	-1.96 %	-1.81 %
PartyScene	-1.46 %	-1.39 %	-1.34 %
RaceHorses	-1.89 %	-1.84 %	-1.75 %
AVG Class C	-1.84 %	-1.76 %	-1.69 %
BQSquare	-0.98 %	-0.88 %	-0.79 %
BasketballPass	-1.85 %	-1.51 %	-1.49 %
BlowingBubbles	-1.70 %	-1.74 %	-1.63 %
RaceHorses	-2.43 %	-2.30 %	-2.00 %
AVG Class D	-1.74 %	-1.61 %	-1.48 %
AVG All Classes	-2.01 %	-1.97 %	-1.91 %

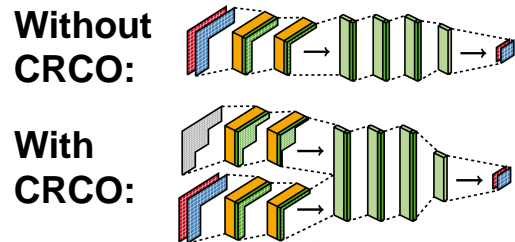
Results – Dedicated Chroma Prediction

Luma Comparison:

- Small improvement (-0.2%) without CRCO
- 3 times more gain (-0.6%) with CRCO

Chroma Comparison:

- Again small improvement (-0.37%) without CRCO
- Nearly -1% with CRCO



Version Channel	with CRCO			without CRCO			no chroma IntraNN		
	Y	U	V	Y	U	V	Y	U	V
BQTerrace	-1.79%	-0.84%	-0.36%	-1.66%	-0.75%	-0.79%	-1.57%	-0.26%	-0.03%
Basket.Drive	-2.33%	-1.64%	-1.97%	-1.83%	-0.15%	-0.88%	-1.34%	-0.05%	-0.56%
Cactus	-2.46%	-1.89%	-2.05%	-1.99%	-1.24%	-1.12%	-1.60%	-0.89%	-0.50%
Kimono	-2.66%	-2.46%	-1.84%	-1.71%	-1.60%	-1.41%	-1.62%	-1.46%	-1.26%
ParkScene	-2.55%	-1.75%	-1.91%	-1.87%	-1.27%	-1.82%	-1.88%	-0.79%	-1.15%
AVG Class B	-2.36%	-1.72%	-1.63%	-1.81%	-1.00%	-1.20%	-1.59%	-0.69%	-0.62%
BQMall	-2.00%	-1.62%	-1.57%	-1.71%	-1.22%	-1.05%	-1.58%	-1.37%	-0.36%
BasketballDrill	-1.99%	-2.42%	-2.26%	-1.21%	-0.38%	-0.74%	-0.63%	-0.17%	-0.21%
PartyScene	-1.46%	-0.86%	-0.96%	-1.31%	-0.68%	-0.70%	-1.21%	-0.68%	-0.65%
RaceHorses	-1.89%	-1.28%	-1.44%	-1.55%	-0.90%	-0.60%	-1.22%	-0.69%	-0.49%
AVG Class C	-1.84%	-1.55%	-1.56%	-1.45%	-0.80%	-0.77%	-1.16%	-0.73%	-0.43%
BQSquare	-0.98%	-0.65%	-0.28%	-1.04%	-0.55%	0.00%	-1.00%	-0.28%	0.20%
BasketballPass	-1.85%	-1.67%	-1.36%	-1.39%	-1.37%	-0.82%	-1.21%	-0.26%	-0.58%
BlowingBubbles	-1.70%	-1.54%	-0.97%	-1.40%	-0.60%	-0.43%	-1.32%	-0.58%	-0.37%
RaceHorses	-2.43%	-1.40%	-2.13%	-1.91%	-1.34%	-1.34%	-1.58%	-0.79%	-0.78%
AVG Class D	-1.74%	-1.32%	-1.19%	-1.44%	-0.97%	-0.65%	-1.28%	-0.48%	-0.38%
AVG All Classes	-2.01%	-1.54%	-1.47%	-1.58%	-0.93%	-0.90%	-1.37%	-0.57%	-0.52%

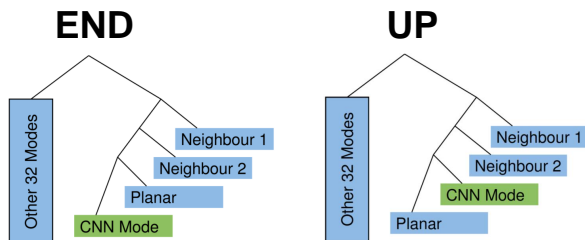
Results – Signaling and Final Evaluation

Signaling:

- UP outperforms end version
 - Mode must be used frequently
- Difference not huge

General Evaluation:

- Hard to compare to other approaches due to training sets
- Beating other approaches in terms of U and V BD-rate gains



Version	END, with CRCO			UP, with CRCO		
	Y	U	V	Y	U	V
BQTerrace	-1.75%	-0.69%	-0.76%	-1.79%	-0.84%	-0.36%
Basket.Drive	-2.24%	-1.64%	-2.08%	-2.33%	-1.64%	-1.97%
Cactus	-2.35%	-1.95%	-2.06%	-2.46%	-1.89%	-2.05%
Kimono	-2.42%	-2.33%	-1.75%	-2.66%	-2.46%	-1.84%
ParkScene	-2.44%	-1.46%	-1.99%	-2.55%	-1.75%	-1.91%
AVG Class B	-2.24%	-1.61%	-1.73%	-2.36%	-1.72%	-1.63%
BQMall	-1.97%	-1.63%	-1.56%	-2.00%	-1.62%	-1.57%
BasketballDrill	-2.00%	-2.17%	-2.03%	-1.99%	-2.42%	-2.26%
PartyScene	-1.46%	-0.83%	-0.95%	-1.46%	-0.86%	-0.96%
RaceHorses	-1.84%	-1.20%	-1.66%	-1.89%	-1.28%	-1.44%
AVG Class C	-1.82%	-1.46%	-1.55%	-1.84%	-1.55%	-1.56%
BQSquare	-1.00%	-0.56%	-0.01%	-0.98%	-0.65%	-0.28%
BasketballPass	-1.78%	-1.76%	-1.19%	-1.85%	-1.67%	-1.36%
BlowingBubbles	-1.69%	-1.74%	-0.71%	-1.70%	-1.54%	-0.97%
RaceHorses	-2.35%	-1.92%	-2.14%	-2.43%	-1.40%	-2.13%
AVG Class D	-1.71%	-1.50%	-1.01%	-1.74%	-1.32%	-1.19%
AVG All Classes	-1.94%	-1.53%	-1.45%	-2.01%	-1.54%	-1.47%

Conclusion and Outlook

Conclusion:

- Useful to train separate networks for chroma channel prediction and integrate cross component information
- Best Architecture depends on content and complexity restrictions
- SATD loss better approximation than L1
- Proposed new signaling with less overhead

Outlook:

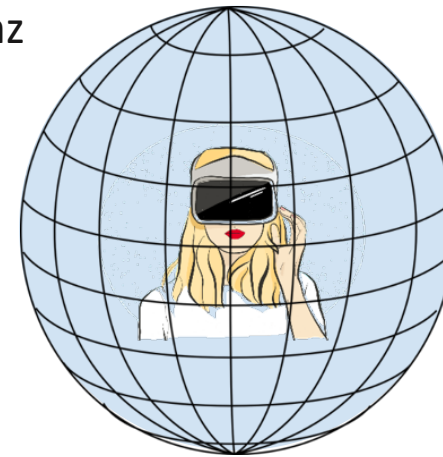
- More architectures, loss functions
- Multiple predictions
- Complexity reduction

Performance of objective metrics on 360VR contents. Special case: Video Multimethod Assessment Fusion (VMAF)

SVCP 2019 – Universität Konstanz

Marta Orduna

moc@gti.ssr.upm.es



Grupo de Tratamiento de Imágenes (GTI)
Universidad Politécnica de Madrid (UPM)



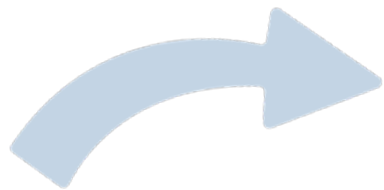
- Motivation
- Review of quality metrics on 360VR contents
- Video Multimethod Assessment Fusion (VMAF)
- Work approach
- 360VR Quality Assessment and Analysis
 - Test material
 - Objective analysis
 - Subjective analysis
- Results
- Conclusions



- Main challenge:
 - to provide omnidirectional content guaranteeing an immersive experience and saving bit rate



- Main solutions:
 - Definition of different perceptible levels of quality
 - Efficient delivery schemes
 - Users' behavior → attention maps
 - Exploitation of peculiarities of the type of projection



All these solutions require **a quality metric**

Review of quality metrics on 360VR contents



- Traditional metrics
 - PSNR (PSNR)
 - Structural Similarity Index (SSIM)
 - Multi-Scale SSIM (MS-SSIM)
 - **VMAF**
- Adaptations of quality metrics to 360VR contents
 - Weighted to Spherically - PSNR (WS-PSNR)
 - Craster Parabolic Projection - PSNR (CPP-PSNR)

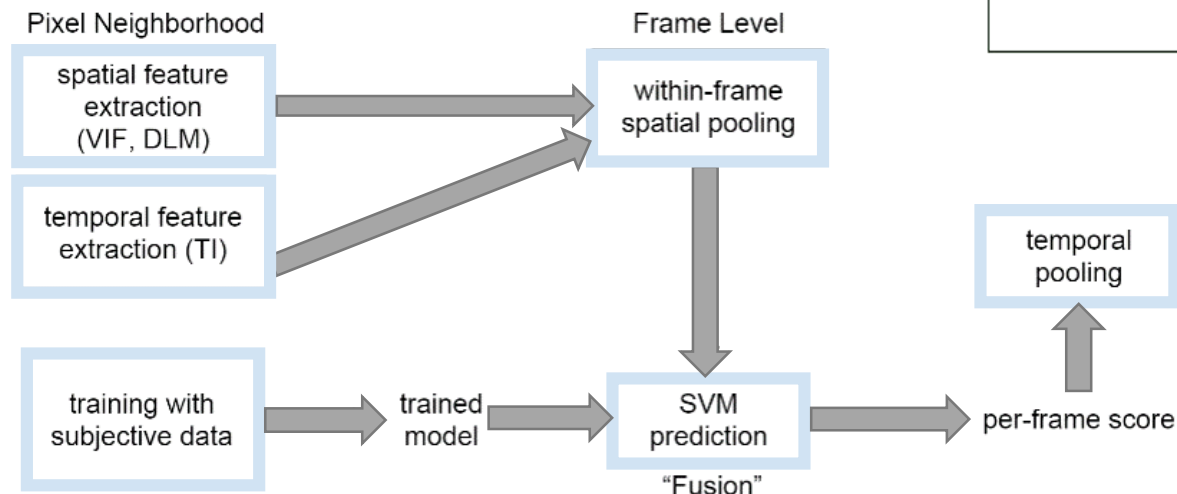
Video Multimethod Assessment Fusion (VMAF)



VMAF is an objective metric able to exploit the benefits of different known elementary metrics, combining them using a machine-learning algorithm, trained with subjective data, and finally, providing the VMAF final score

VMAF has provided significantly good results on different types of non-immersive contents and viewing condition

DATASET CHARACTERISTICS	
Number of reference videos	34
Duration	6 seconds
Encoding	H.264/AVC
Resolution	384x288 – 1920x1080
Bitrate	375 kbps – 20 Mbps
Total of distorted videos: 300	



Developed by:



- *Research question:* can VMAF be applied to omnidirectional content without making any specific adjustment?
- *Underlying hypothesis:* There is a monotonic relationship between 2D-VMAF and 360VR-VMAF (non-existing)
- If so, we can avoid:
 - generating a large specific 360VR video dataset
 - carrying out numerous subjective quality assessments
 - performing the training and testing stages



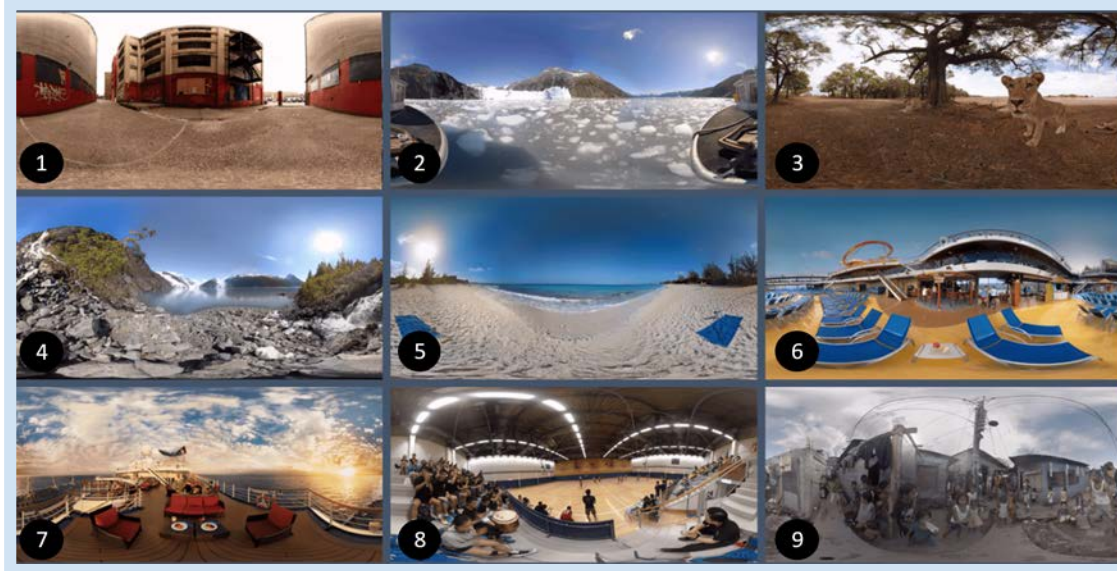
The validation of VMAF on 360VR contents is carried out in two steps:

Objective Analysis

VMAF application to omnidirectional sequences encoded with constant Quantization Parameter (QP) in the whole range of possible values

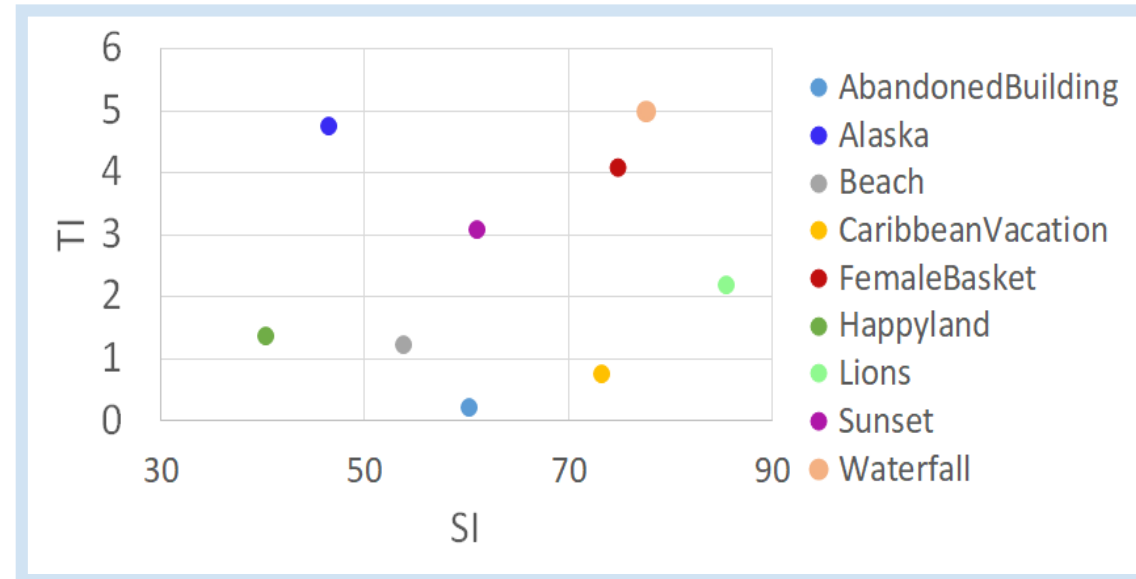
Subjective Assessment

VMAF scores validation through a subjective assessment



A wide range of contents selected with different features in terms of color, texture, camera motion, composition, and content in the scenes

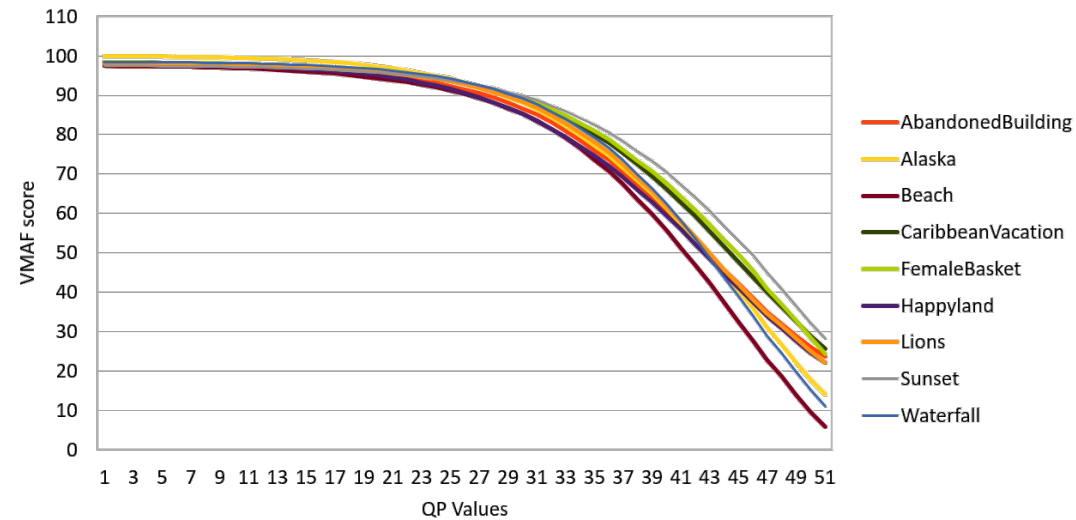
Spatial Information (SI)
and Temporal (TI)
Information indicators

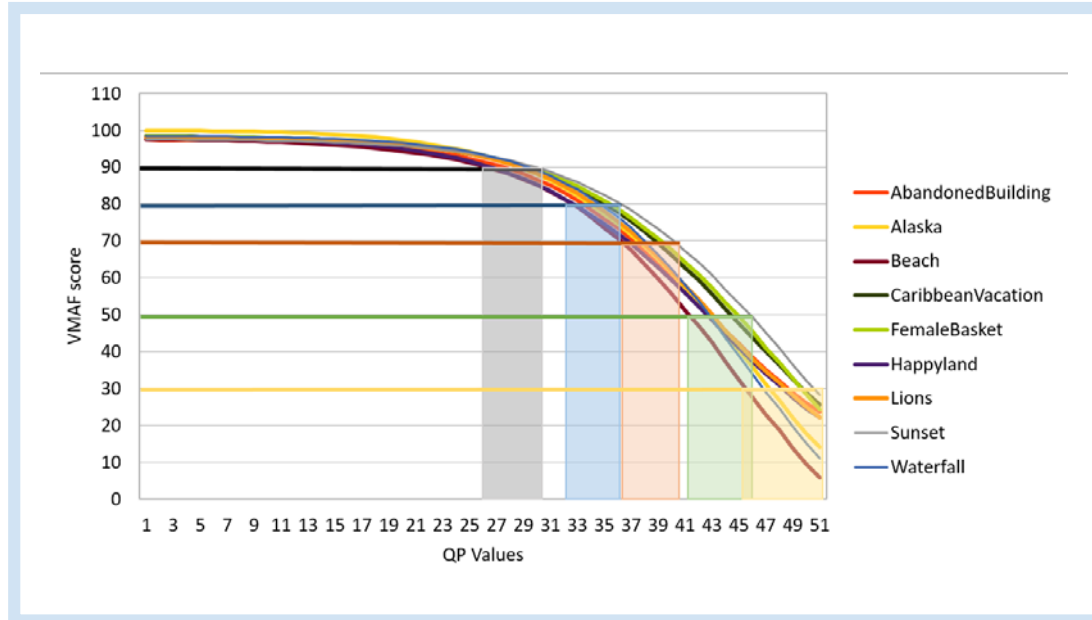


- <http://vhil.stanford.edu/360-video-database/>
- C. Wu, Z. Tan, Z. Wang, and S. Yang, "A Dataset for Exploring User Behaviors in VR Spherical Video Streaming," in Proceedings of the 8th ACM on Multimedia Systems Conference, 2017, pp. 193–198

Number of reference videos	9
Duration	10 seconds
Encoding	H.265/HEVC
Resolution	4K (3840x1920)
Hypothetical Reference Circuits (HRCs)	QP range (1-51)
Framerate	25 fps
Total number of videos: 459	

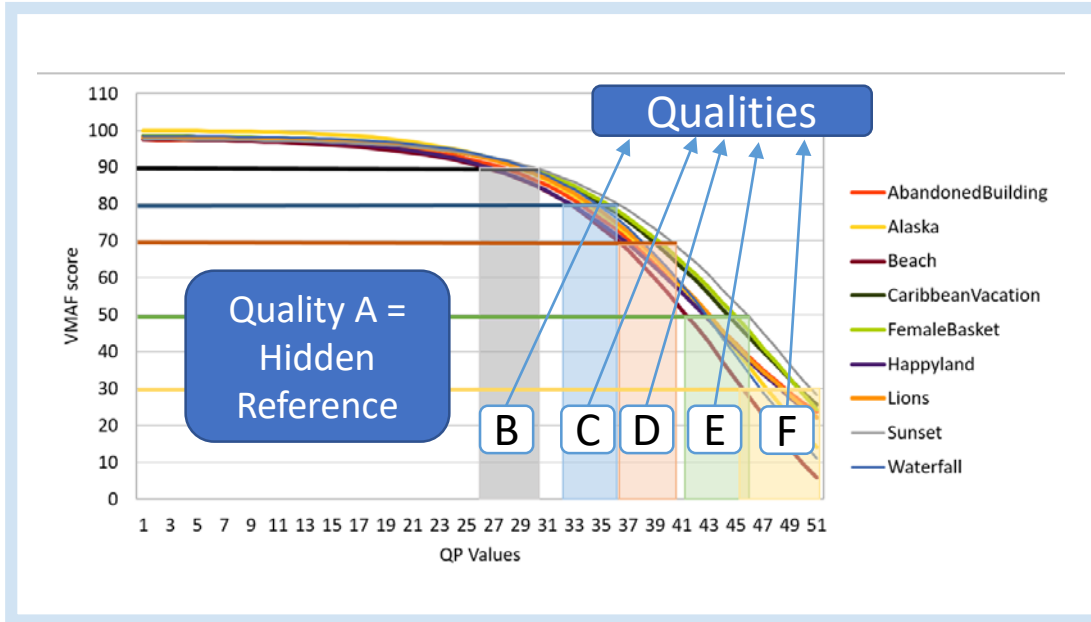
- No temporal pooling challenge
- 4K throughout the process





DATASET CHARACTERISTICS	
Number of source videos	9
Duration	10 seconds
Encoding	H.265/HEVC
Resolution	4K (3840x1920)
Number of QP values	6
Total of videos: 54	

VMAF ~ 90, 80, 70, 50, 30 + Reference



DATASET CHARACTERISTICS	
Number of source videos	9
Duration	10 seconds
Encoding	H.265/HEVC
Resolution	4K (3840x1920)
Number of QP values	6
Total of videos: 54	

VMAF ~ 90, 80, 70, 50, 30 + Reference

Methodology

ACR-HR

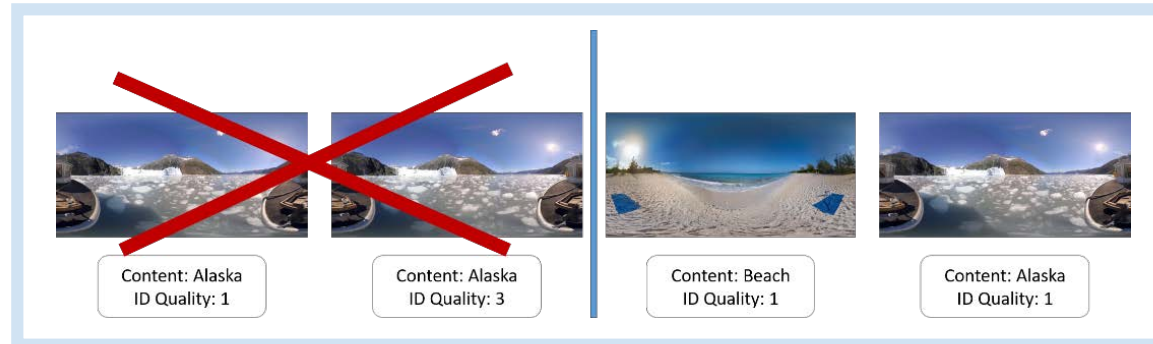
Five Grade Scale - Quality	
5	Excellent
4	Good
3	Fair
2	Poor
1	Bad

START	VIDEO 1	VOTE 1	...	VIDEO 17	VOTE 17	THE END
-------	---------	--------	-----	----------	---------	---------

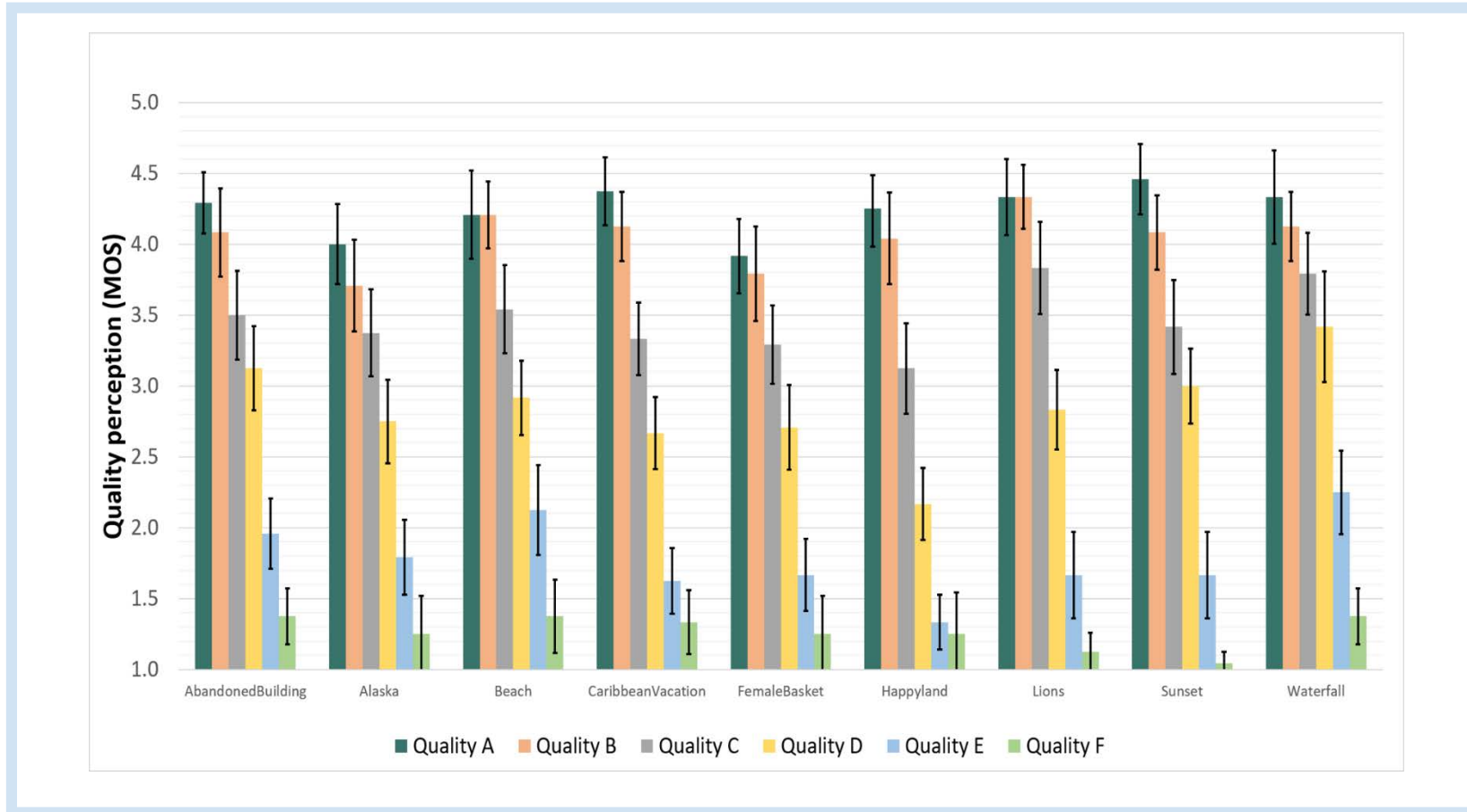
- No training session
- All videos are viewed by each subject
- Duration ~ 15 minutes (assuming 5 seconds for evaluation)
- 24 observers (average age of 26)
- 1 subject was removed because of being considered an outlier

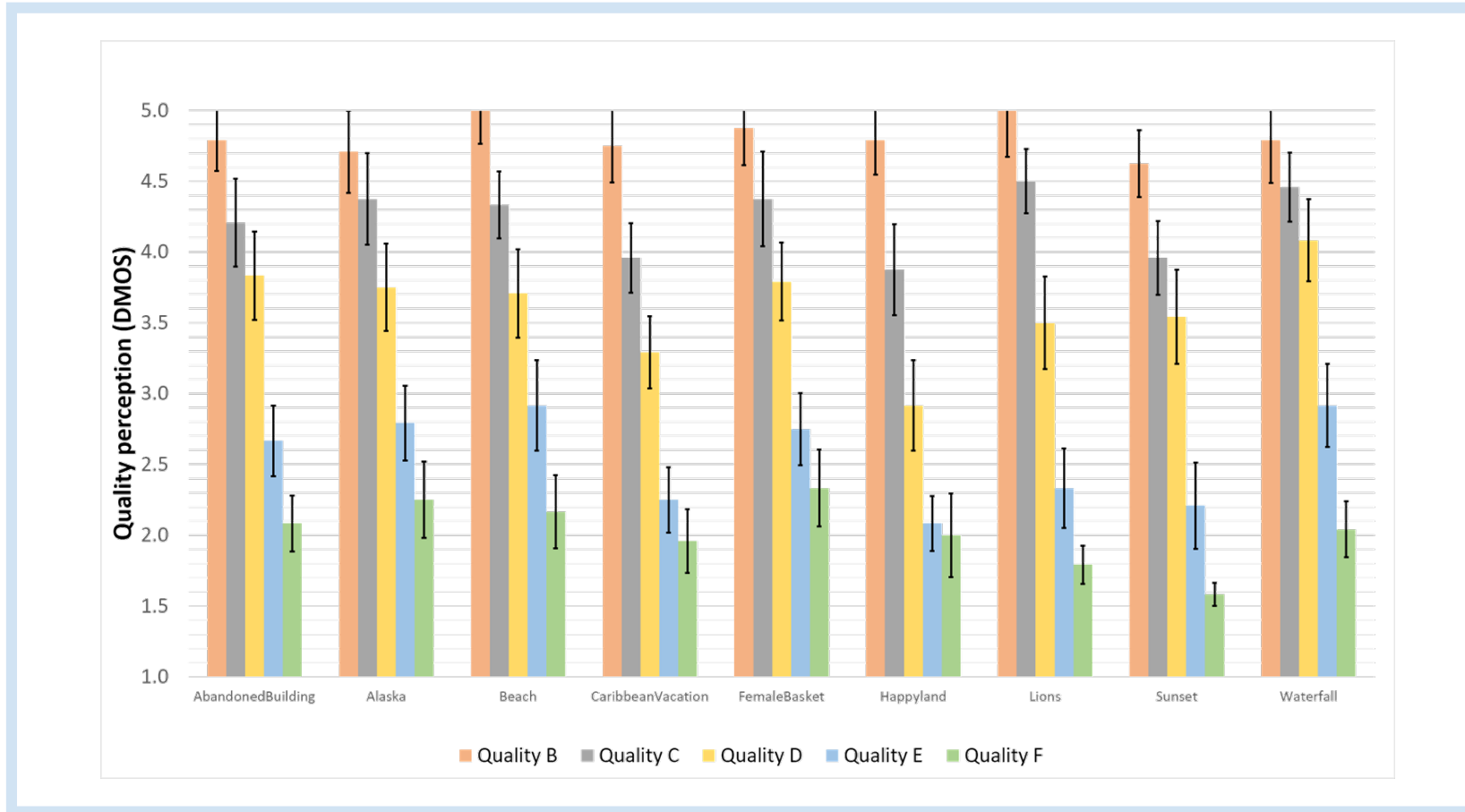


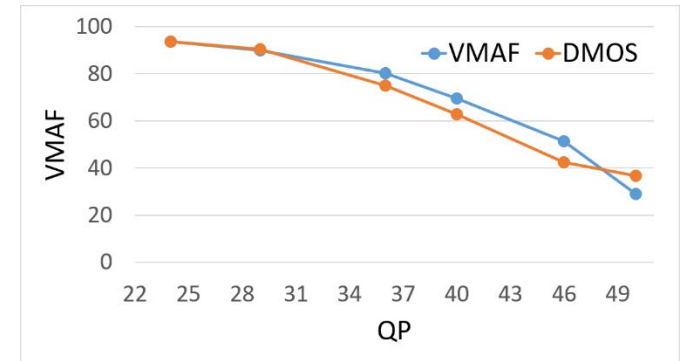
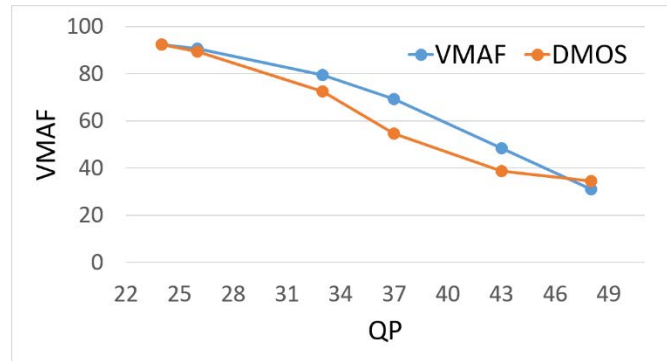
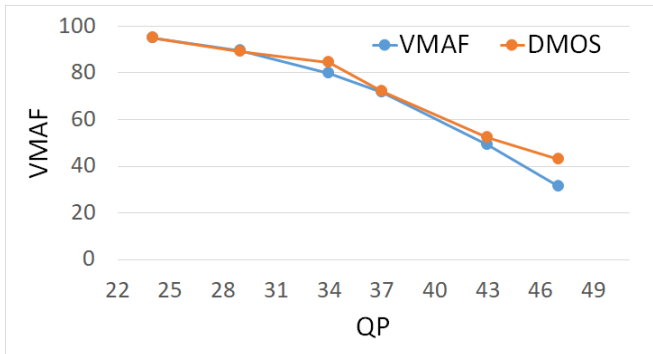
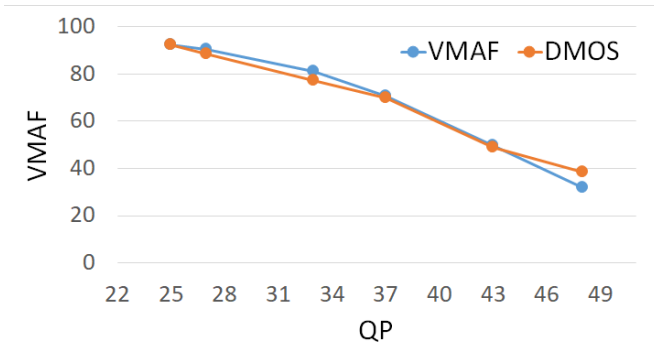
Equipment + environment



Content randomization





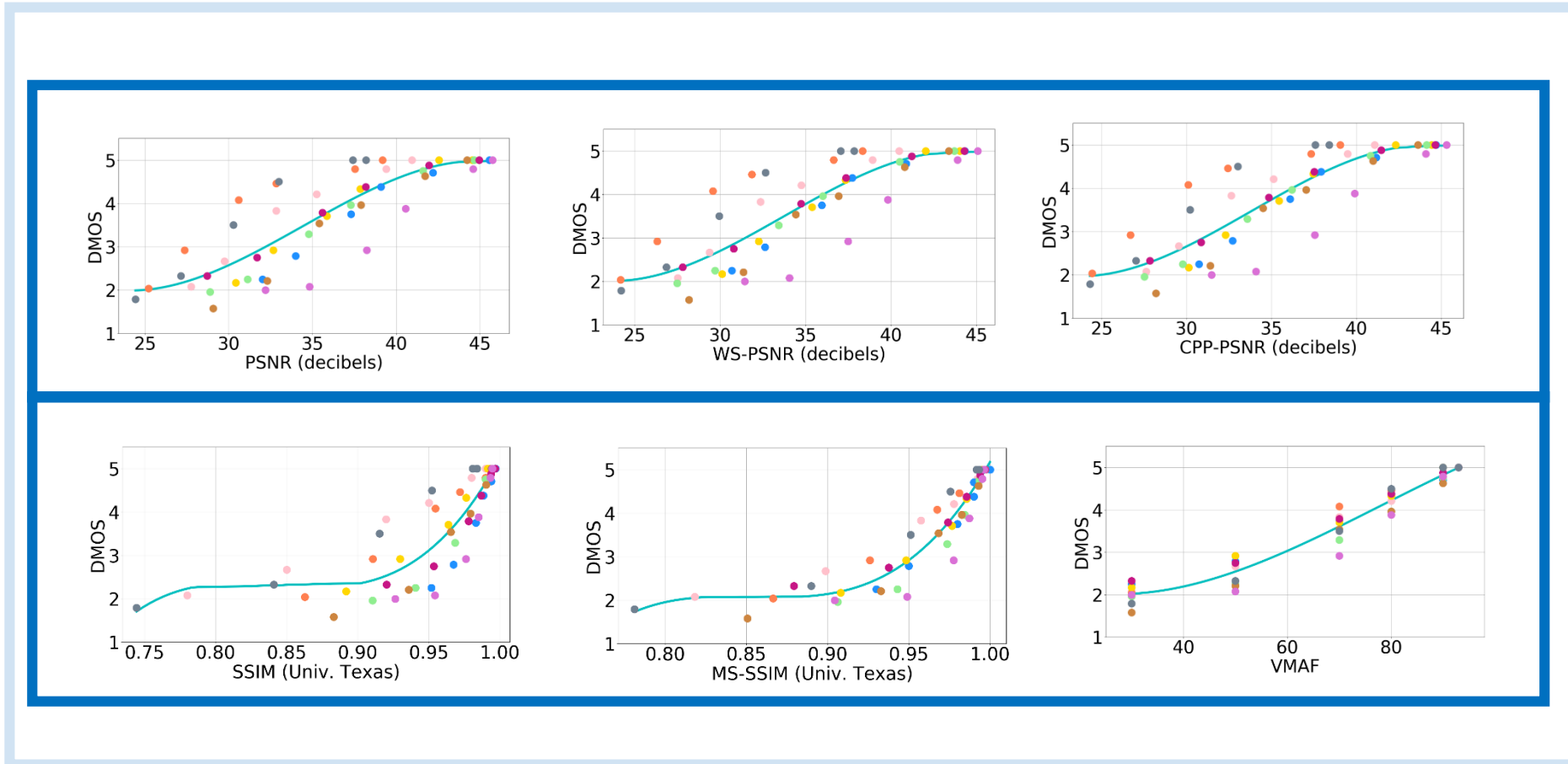


PLCC and RMSE between VMAF and DMOS



Content	PLCC	PLCC	RMSE	RMSE	SROCC
	(QB, QC, QD, QE, QF)	(QB, QC, QD, QE)	(QB, QC, QD, QE, QF)	(QB, QC, QD, QE)	(QA, QB, QC, QD, QE, QF)
<i>AbandonedBuilding</i>	0.995	0.997	0.172	0.099	1.000
<i>Alaska</i>	0.989	0.992	0.283	0.124	1.000
<i>Beach</i>	0.995	0.994	0.211	0.124	0.975
<i>CaribbeanVacation</i>	0.962	0.997	0.349	0.339	1.000
<i>FemaleBasket</i>	0.990	1.000	0.355	0.088	1.000
<i>Happyland</i>	0.955	0.981	0.467	0.500	1.000
<i>Lions</i>	0.987	0.995	0.201	0.222	0.975
<i>Sunset</i>	0.996	0.998	0.251	0.275	1.000
<i>Waterfall</i>	0.995	0.986	0.276	0.215	1.000
Overall	0.965	0.959	0.285	0.221	0.994

Mapping of DMOS ratings to objective scores



Solid line represents the best fitting by a third degree polynomial curve



PLCC, RMSE, R^2 between Fitting curves and DMOS

Metric	PLCC	RMSE	R^2
<i>PSNR (linear)</i>	0.851	0.593	0.725
<i>WS-PSNR (linear)</i>	0.860	0.577	0.740
<i>CPP-PSNR (linear)</i>	0.873	0.551	0.763
<i>PSNR (dB)</i>	0.851	0.593	0.725
<i>WS-PSNR (dB)</i>	0.861	0.576	0.741
<i>CPP-PSNR (dB)</i>	0.874	0.550	0.763
<i>SSIM</i>	0.874	0.550	0.763
<i>MS-SSIM</i>	0.956	0.333	0.914
<i>VMAF</i>	0.980	0.227	0.960

VMAF

- Exhaustive study on the feasibility of VMAF on 360VR contents
- VMAF works sufficiently correctly with omnidirectional contents, without performing any particular adjustments
- The creation of a 360VR dataset can be avoided, thus saving computing and time resources

- ✓ Orduna, M., Díaz, C., Muñoz, L., Pérez, P., Benito, I., & García, N. (2019). Video Multimethod Assessment Fusion (VMAF) on 360VR contents. arXiv preprint arXiv:1901.06279.

An Affine-Linear Intra Prediction with Memory Constraints



CONTENTS

- Introduction
- (1) Architecture of the trained predictors
- (2) Memory and complexity assessment
- (3) Training details
- (4) Experimental results
- References

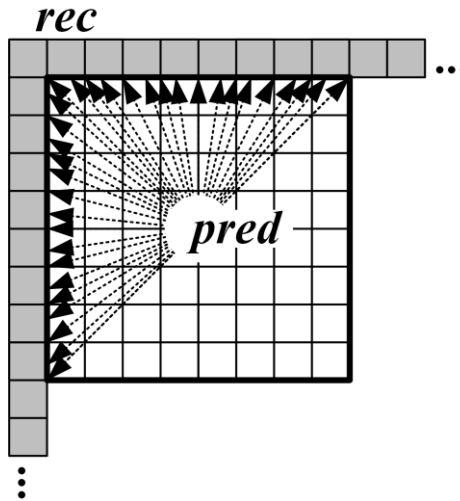
INTRODUCTION

- Modern video codecs like HEVC:
 - Recursive block-partitioning
 - Predictive Coding (**Intra-Picture Prediction**, Motion Compensation)
 - Residual Transform and Quantization
 - Entropy Coding

- The prediction residual is transmitted in the bitstream
- Hence, increased prediction accuracy leads to bit-rate savings

Intra-Picture Prediction

- Generate a prediction from reconstructed samples in the same frame
- Conventional intra modes: Angular, PLANAR and DC



Question

- Can we generate intra prediction modes as outcome of a training experiment on a large set of suitable data?

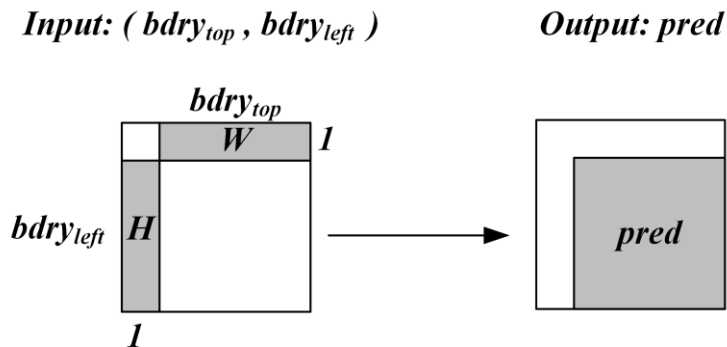
Challenges

- Narrow limits in memory and complexity for video coding applications
- Neural networks consist of multiple fully-connected or convolutional layers
- Modern video codecs support a variety of block partitions
- Loss function?

Memory vs. complexity vs. compression efficiency

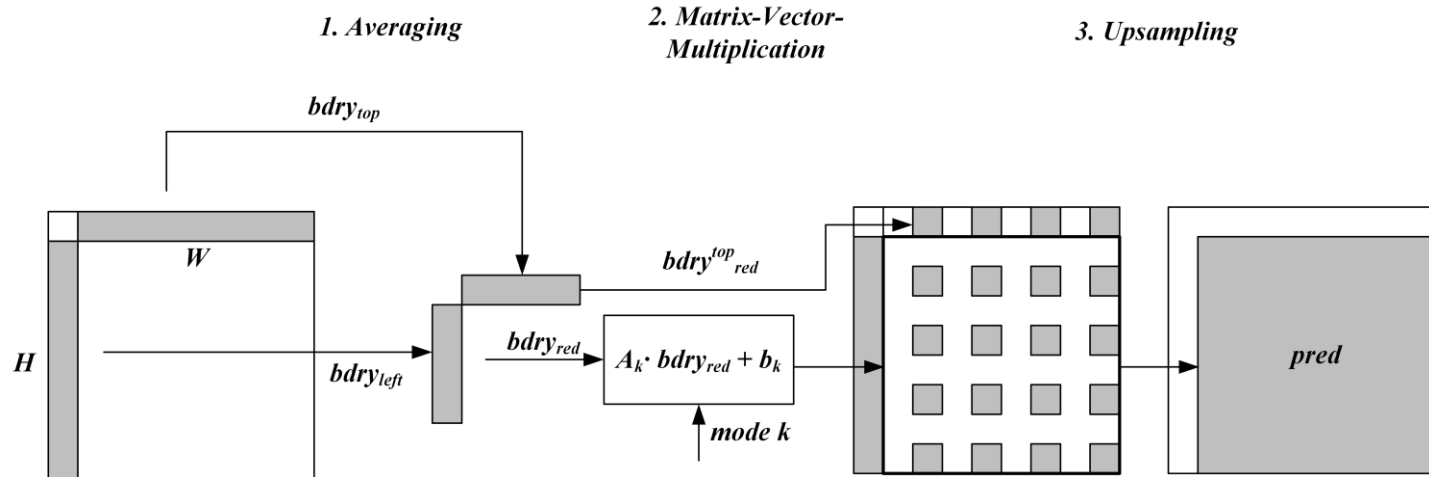
(1) ARCHITECTURE OF THE PREDICTORS

- For each luma WxH block, 19 trained intra prediction modes are provided
- These predictors are added to the list of intra modes for rate-distortion optimization
- Input for the prediction are the W samples above and the H samples left of the block



Generation of the prediction signal in three steps

- Averaging on the boundary
- Matrix vector multiplication and offset addition
- Upsampling of the result (only applied to blocks larger than 8x8)



(2) MEMORY AND COMPLEXITY

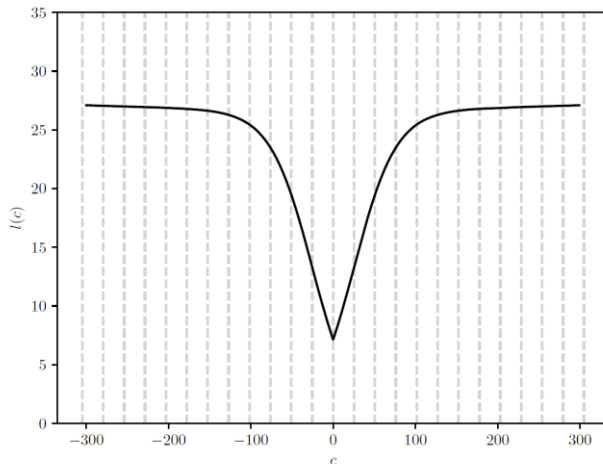
- The matrix and vector entries are stored as 10-bit values
- Consequently, the **total memory requirement is 7.2 kB**
 - 128x128 CTU, bitdepth=10, 4:2:0 sampling rate requires 30 kB of memory
- Note that the matrices A have 512 entries each
- Input and output sampling only uses additions and bitshifts
- Consequently, **not more than 8 multiplications per sample** are necessary
 - Interpolation filters for fractional angle positions require 4 mult. /sample

(3) TRAINING

- Given mode k , the DCT-transformed residuals for a $W \times H$ block are $c_k = T(\text{org} - \text{pred}_k)$
- We approximate the bitrate of the residuals by

$$L(\text{org}, k) = \sum_{i=1}^{WH} (|(c_k)_i| + \alpha \cdot \text{sig}(\beta |(c_k)_i| - \gamma))$$

- Recursive block-partitioning:
 - Start with a parent block of shape 16x16
 - Compare the cost of a parent with the accumulated costs of childs
 - Jointly train predictors for shapes 4x4, 8x8, 16x16



(4) EXPERIMENTAL RESULTS

- Reference Software: Versatile Video Coding Test Model 4.0
- Coding tools configuration according to common test condition (CTC)

All Intra	Y in %	Enc Time in %	Dec Time in %
Class A1	-1.15	142	98
Class A2	-0.67	140	99
Class B	-0.73	143	100
Class C	-0.72	142	98
Class E	-0.90	140	100
Overall	-0.82	142	99

REFERENCES

- [1] J. Pfaff, P. Helle, Dominique R. Maniry, S. Kaltenstadler, W. Samek, H. Schwarz, D. Marpe, and T. Wiegand, "Neural network based intra prediction for video coding," Fraunhofer HHI, September 2018.
- [2] P. Helle, J. Pfaff, M. Schäfer, R. Rischke, H. Schwarz, D. Marpe, and T. Wiegand, "Intra picture prediction for video coding with neural networks," 2019 Data Compression Conference, 2019, Fraunhofer HHI.
- [3] M. Schäfer, B. Stallenberger, J. Pfaff, P. Helle, H. Schwarz, D. Marpe, and T. Wiegand, "An Affine-Linear Intra Prediction with Complexity Constraints," 2019 International Conference on Image Processing (ICIP), 2019, Fraunhofer HHI.
- [4] J. Chen, Y. Ye, K. Suehring, and S. H. Kim, "Algorithm description for Versatile Video Coding and Test Model 4," JVET-M002, Joint Video Exploration Team (JVET), 2019

Dictionary Learning based Adaptive Resolution Change in Video Coding

Outline

1. Motivation and Fundamentals
2. Dictionary learning based super-resolution
3. Dynamic Resolution Change in Video Coding
4. Experimental Results

Motivation

- Dictionary Learning based super-resolution showed promising results when applied to inter-layer prediction in SHVC [1]
- The concept of dynamic resolution conversion is already known from MPEG 4 [2] and raised attention recently [3]
- The convex hull of the RD curve can be estimated by downsampling the video before coding [4]

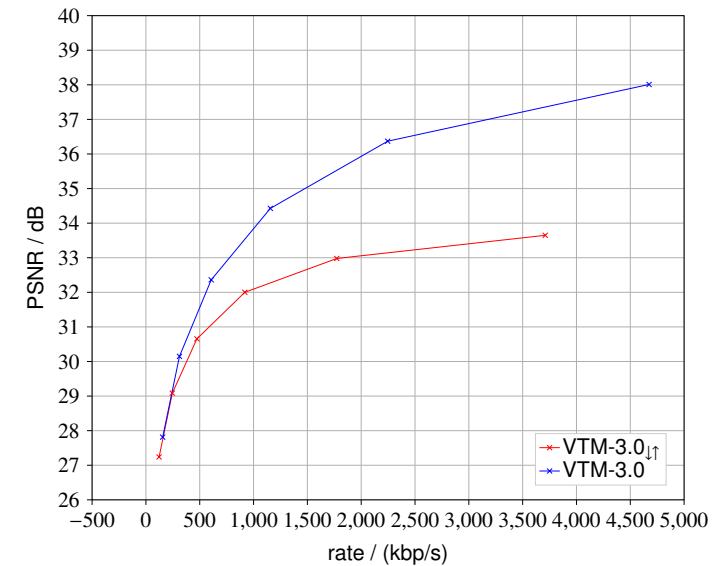
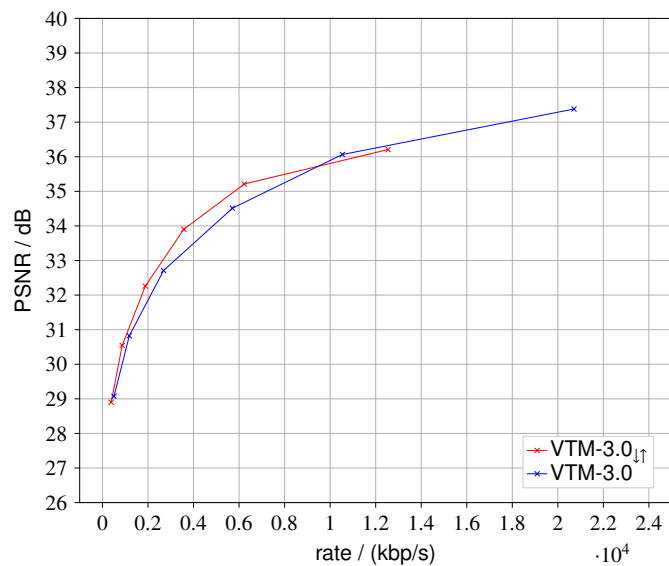
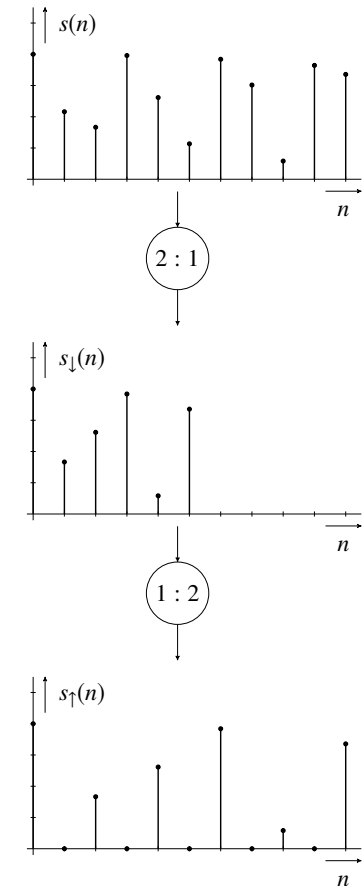


Figure: RD-curves for Campfire sequence (left) and Basketballdrive (right). First 100 frames, RA coding configuration

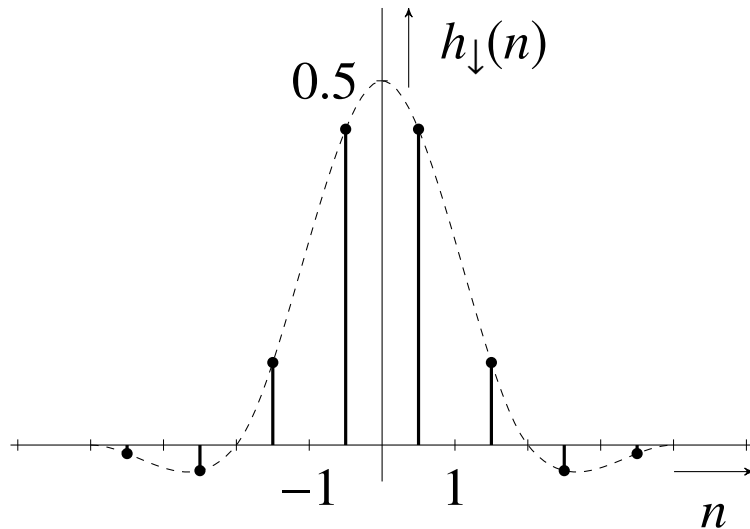
Fundamentals: Downsampling and Upsampling

- Downsampling realized by taking e.g. every second sample
- This introduces alias in general
 - The signal is filtered with a anti-aliasing filter
- Upsampling is realized by inserting zeros
 - The signal is filtered with an interpolation filter
- MATLAB's *imresize* function does not strictly follow this methodology, when using the bicubic kernel
 - samples are shifted when downsampling and shifted back when upsampling



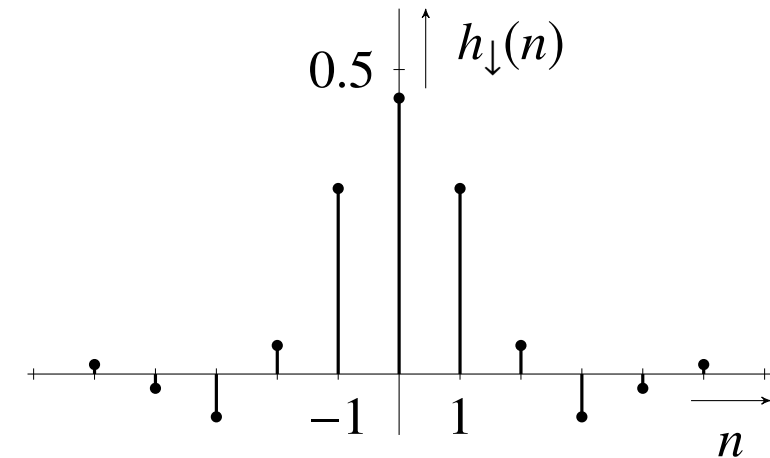
Fundamentals: Downsampling Filters

- Bicubic downsampling filter has 8 taps
 - This introduces a phase shift of the downsampled signal



(a) bicubic downsampling filter

- The downsampling filter used in SHVC has 11 taps
 - no distortion of the phase during downsampling



(b) downsampling filter used for SHVC

Figure: different downsampling filters

Fundamentals: Downsampling Filters

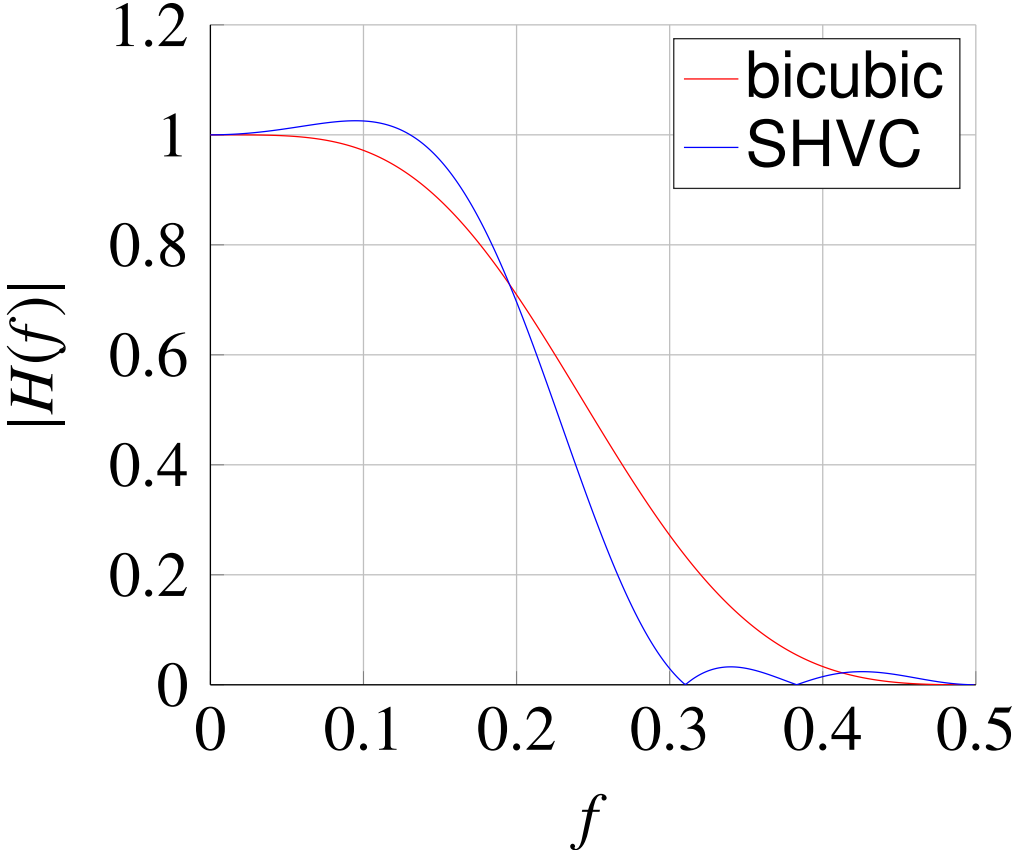
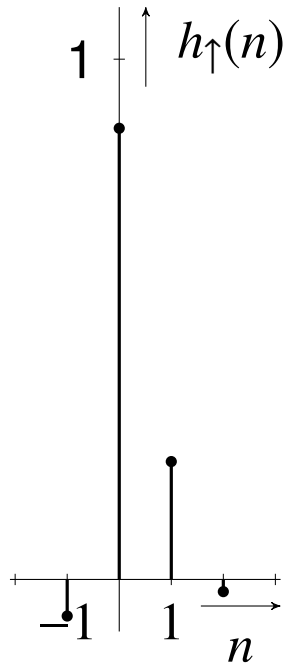


Figure: Frequency response of different downsampling filters

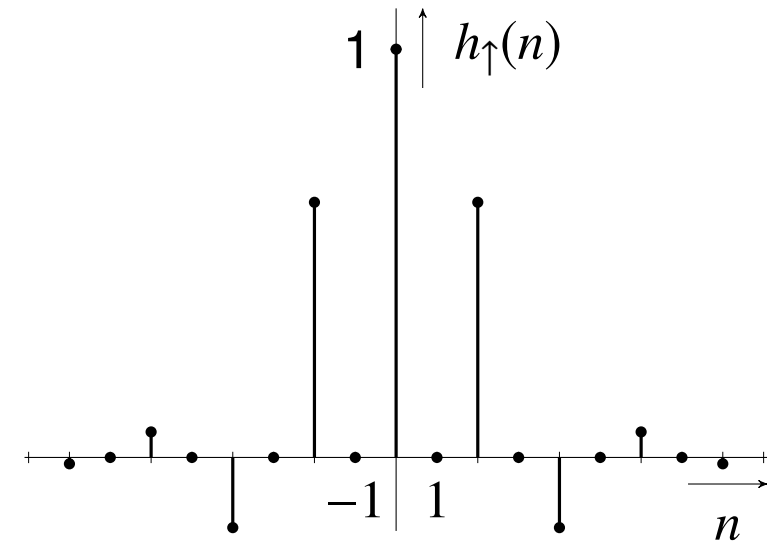
Fundamentals: Upsampling Filters

- Bicubic upsampling filter has to be applied several times since we need to “backshift” the phase



(a) bicubic upsampling filter

- The upsampling filter is derived from the half-pel interpolation filters used in HEVC
 - We need to insert a 1 at position zero and 0s at the odd sample positions



(b) upsampling filter derived from HEVC interpolation filters

Figure: different upsampling filters

Contents

1. Motivation and Fundamentals

2. Dictionary learning based super-resolution

3. Dynamic Resolution Change in Video Coding

4. Experimental Results

Dictionary Learning Fundamentals

- Dictionary is typically trained using vectorized training patches x_i of a size $s_p = 8 \times 8$
- a sparse representation of an image patch is found by sparse encoding the patch x in the dictionary D

$$D = \arg \min_D \sum_{i=1}^n \frac{1}{2} \|x_i - D\alpha_i\|_2^2 + \lambda \|\alpha_i\|_1$$

$$\alpha = \arg \min_{\alpha} \|x - D\alpha\|_2^2 + \lambda \|\alpha\|_1$$

$$x = D\alpha + \varepsilon$$

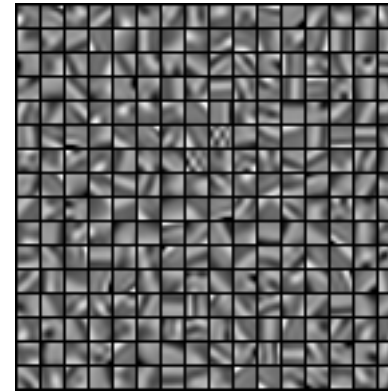
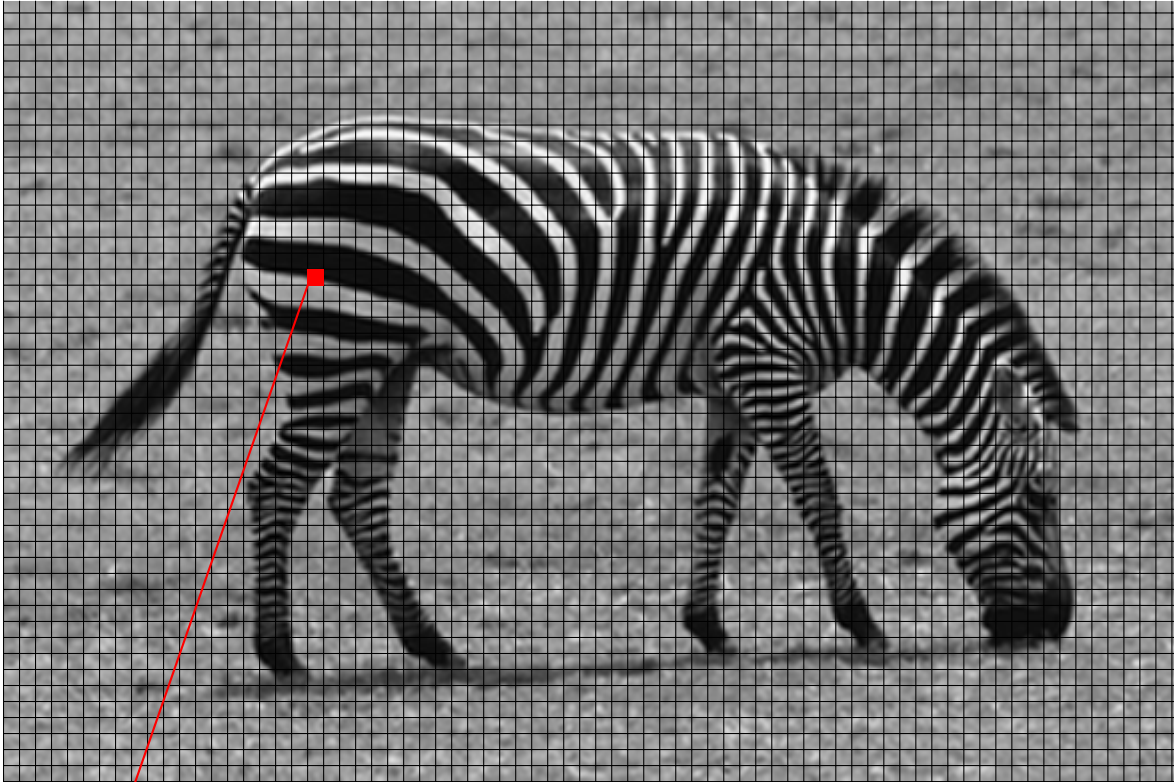


Figure: Example Dictionary

- The concept of dictionary learning can be used for super-resolution by training coupled dictionaries [5]

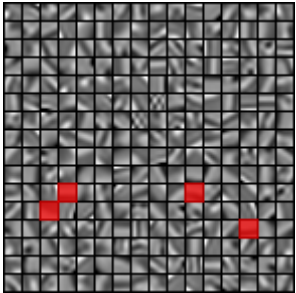
DL based SR: Coupled dictionaries approach

I_{LR}

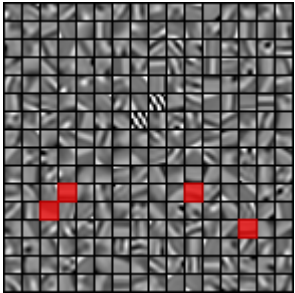


$\approx \alpha_{164} + \alpha_{171} + \alpha_{179} + \alpha_{206}$

D_{LR}

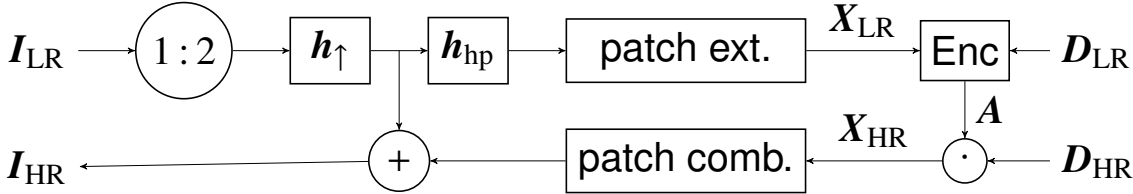


D_{HR}



$x_{LR} \approx D_{LR} \alpha$

$x_{HR} \approx D_{HR} \alpha$



Results DL based SR

	Bicubic	HEVC int.	DLSR
BQTerrace	28.2	28.8	30.3
BasketballDrive	34.7	34.9	36.1
Cactus	33.5	34.2	35.5
Campfire	37.8	38.1	38.8
CatRobot1	39.7	40.2	40.9
DaylightRoad2	37.4	37.7	38.2
FoodMarket4	48.5	48.7	48.7
MarketPlace	40.2	41.1	41.9
ParkRunning3	40.7	44.2	47.8
RitualDance	44.6	45.7	48.0
Tango2	42.4	42.6	42.6
AVG	39.0	40.0	41.4

Table: PSNR values for different downsampling and upsampling / SR algorithms: Values were measured for the Y component of the first frame of each video sequence. For DLSR: $\lambda = 0.01$ and h_{hp} was chosen to be a laplacian highpass filter.

Contents

1. Motivation and Fundamentals

2. Dictionary learning based super-resolution

3. Dynamic Resolution Change in Video Coding

4. Experimental Results

Dynamic Resolution Change with SR

- On which level of the encoding scheme should the resolution change happen?

Option	signaling cost	spacial adaptivity	temporal adaptivity	boundary issues
CU level	high	yes	yes	yes
CTU level	moderate	yes	yes	moderate
TID level	low	no	yes	almost none
Intra period level	low	no	yes	almost none
Sequence level	none	no	no	almost none

- ➔ The decision was drawn to try it at the CTU level, which seems to be a good compromise
 - Code the CTU at full and half resolution
 - upsample or apply SR to downsampled reconstructed CTU
 - decide based on RD-cost which one is coded into the bitstream

Dynamic Resolution Change with SR

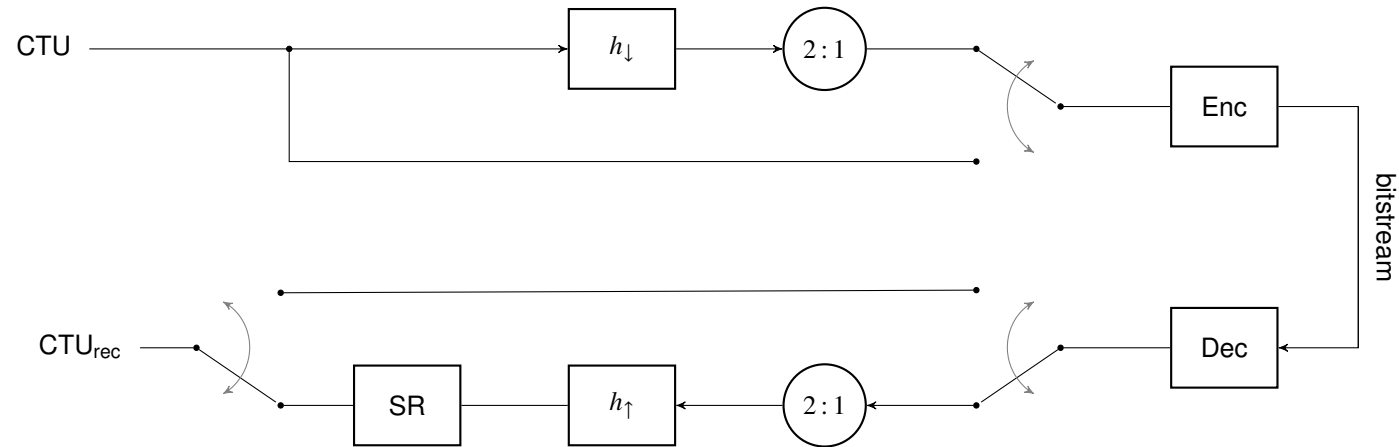


Figure: CTU level DRC scheme

- Implementation so far only for Intra-CTUs
- Implementation only for the Y-Component

Dynamic Resolution Change with SR

- The reference area needs to be downsampled in the case of prediction at the boundary of a downsampled CTU
- The downsampled CTU has to be coded at lower QP [3]:

$$QP_{LR} = QP_{HR} - 6$$

- The rate-distortion parameter λ has to be adjusted:

$$\lambda_{LR} = \frac{\lambda_{HR}}{4}$$

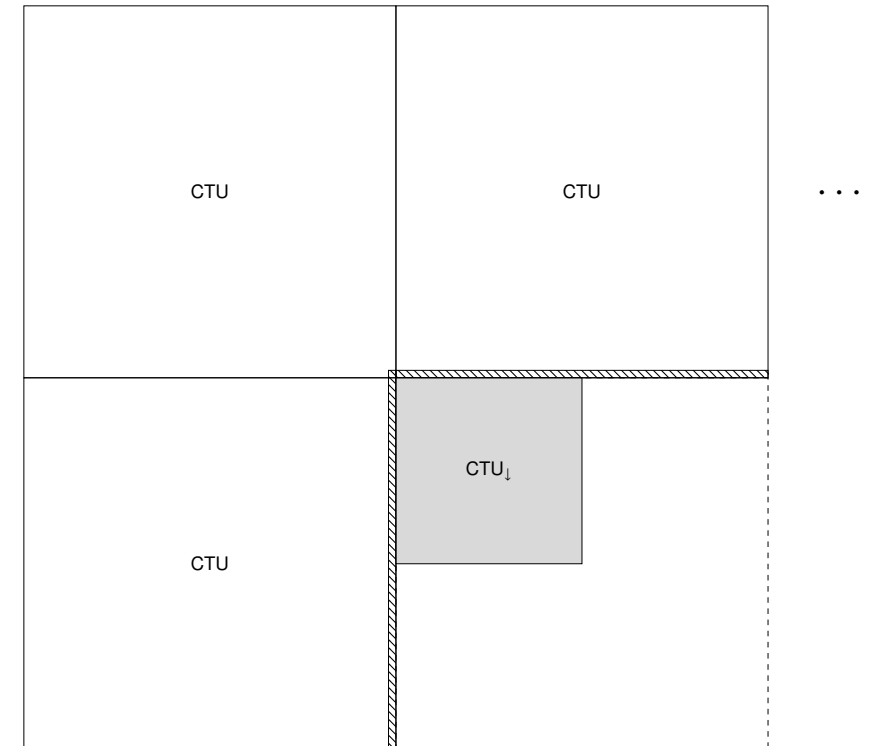


Figure: Coding of a downsampled CTU

Contents

1. Motivation and Fundamentals

2. Dictionary learning based super-resolution

3. Dynamic Resolution Change in Video Coding

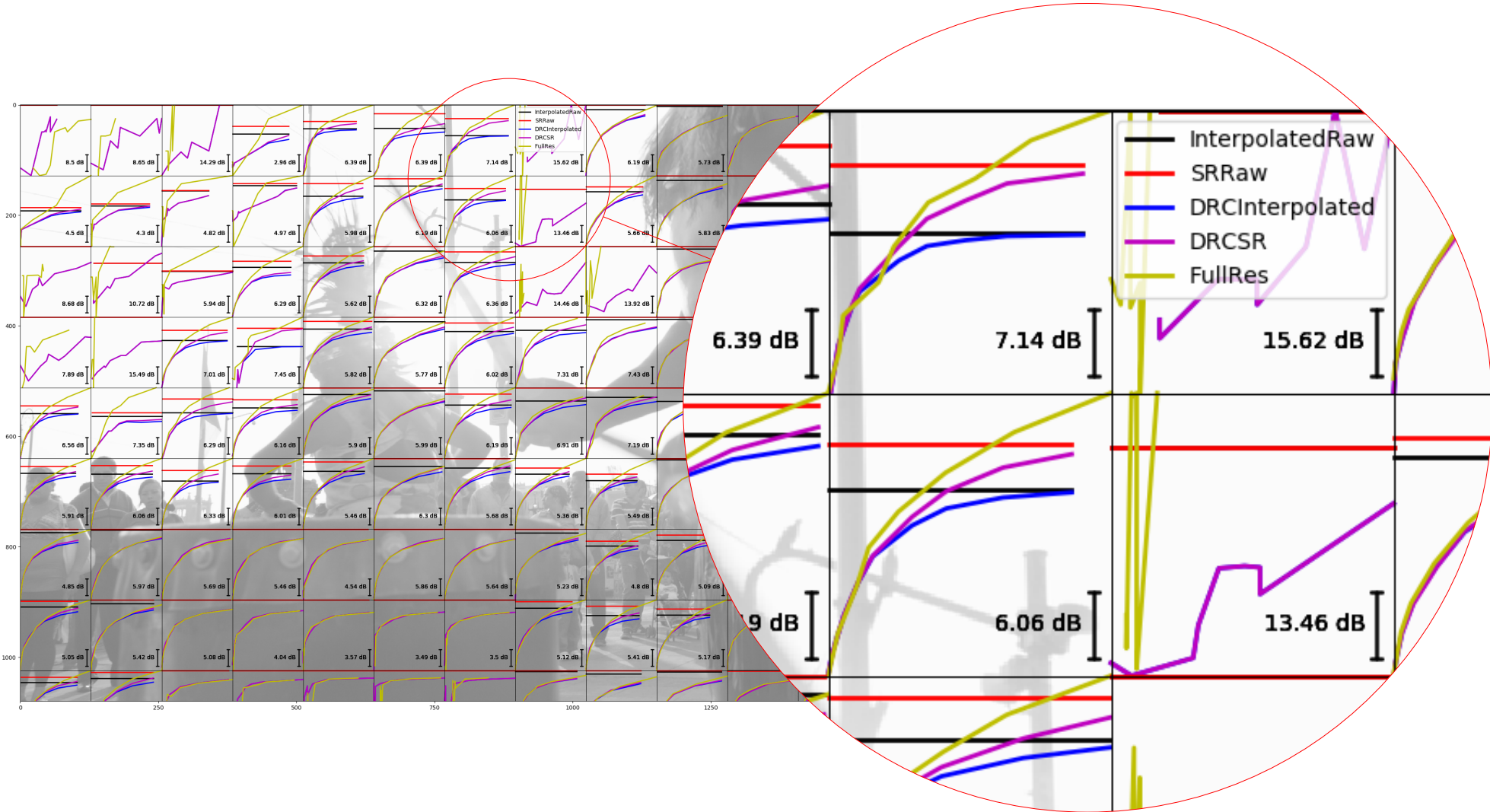
4. Experimental Results

Results

	VTM-3.0 DRC HEVC int.	VTM-3.0 DRC DLSR
BQTerrace	-0.03	-0.04
BasketballDrive	-0.27	-0.17
Cactus	-0.05	-0
Campfire	-0.19	-0.23
CatRobot1	-0.29	-0.25
DaylightRoad2	-0.05	-0.13
FoodMarket4	-3.59	-3.58
MarketPlace	-0.11	-0.01
ParkRunning3	-0.01	0.08
RitualDance	-0.48	-0.56
Tango2	-1.51	-1.43
AVG	-0.6	-0.57

Table: BD rate savings against VTM-3.0. QP \in {22, 27, 32, 37}. Only the first frame of each sequence was coded.

Results



Results

	VTM-3.0 DRC HEVC int.	VTM-3.0 DRC DLSR
BQTerrace	-1.76	-1.67
BasketballDrive	-4.72	-4.59
Cactus	-3.16	-2.73
Campfire	-6.02	-5.79
CatRobot1	-7.86	-7.48
DaylightRoad2	-8	-7.29
FoodMarket4	-6.85	-6.52
MarketPlace	-5.06	-5.29
RitualDance	-5.74	-5.35
Tango2	-5.71	-5.8
AVG	-5.33	-5.11

Table: BD rate savings against VTM-3.0. QP \in {42, 47, 52, 57}. Only the first frame of each sequence was coded.

Conclusion

- Coding gains with respect to VTM 3.0 can be achieved by performing a dynamic resolution change on the CTU level
- Dictionary Learning based super-resolution does not increase the coding gain
 - At high rates the quality gain of DLSR is too low to outperform full resolution coding
 - At low rates coding artifacts heavily influence the DLSR performance such that there is no gain over classic interpolation anymore

Thank you for your attention!

Any questions?

Jens Schneider
schneider@ient.rwth-aachen.de

Institut für Nachrichtentechnik, RWTH Aachen University
www.ient.rwth-aachen.de

References I

- [1] Jens Schneider, Johannes Sauer, and Mathias Wien. “Dictionary Learning based High Frequency Inter-Layer prediction for Scalable HEVC”. In: *Proc. of IEEE Visual Communications and Image Processing VCIP '17*. St. Petersburg, USA: IEEE, Piscataway, Dec. 2017.
- [2] F.C.N. Pereira and T. Ebrahimi. *The MPEG-4 Book*. IMSC Press multimedia series. Prentice Hall PTR, 2002.
- [3] Y. Li, D. Liu, H. Li, L. Li, F. Wu, H. Zhang, and H. Yang. “Convolutional Neural Network-Based Block Up-Sampling for Intra Frame Coding”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 28.9 (Sept. 2018), pages 2316–2330.
- [4] A Aaron, Z. Li, M. Manohara, J De Cock, and D. Ronca. *Per-Title Encode Optimization*. <https://medium.com/netflix-techblog/per-title-encode-optimization-7e99442b62a2>. [Online; accessed 2-May-2019]. 2015.
- [5] Roman Zeyde, Michael Elad, and Matan Protter. “On single image scale-up using sparse-representations”. In: *International conference on curves and surfaces*. Springer. 2010, pages 711–730.
- [6] Gisle Bjontegaard. *Calculation of average PSNR differences between RD-curves*. Technical report Doc. VCEG-M33. Austin, USA: ITU-T SG16/Q6 VCEG, 2001.
- [7] Michael Elad. *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*. 1st. Springer Publishing Company, Incorporated, 2010.
- [8] Julien Mairal, Francis Bach, Jean Ponce, and Guillermo Sapiro. “Online dictionary learning for sparse coding”. In: *Proceedings of the 26th annual international conference on machine learning*. ACM. 2009, pages 689–696.
- [9] M. Wien. *High Efficiency Video Coding*. 1st edition. Springer-Verlag Berlin Heidelberg, 2015.
- [10] Radu Timofte, Vincent De Smet, and Luc Van Gool. “A+: Adjusted Anchored Neighborhood Regression for Fast Super-Resolution”. In: volume 9006. Apr. 2015, pages 111–126.

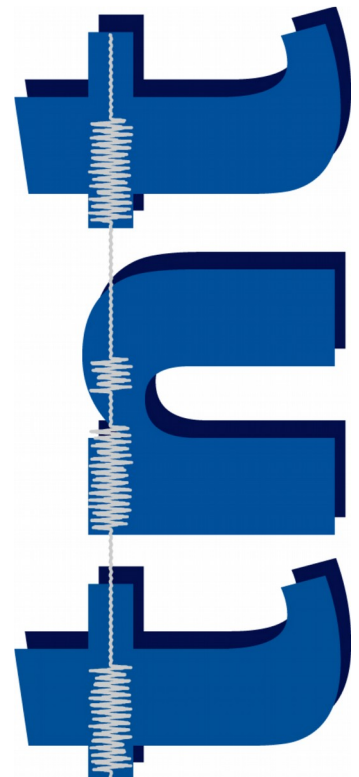
High-precision Camera Calibration for Professional Augmented-Reality Applications

SVCP 2019

Benjamin Spitschan

Institut für Informationsverarbeitung
Leibniz Universität Hannover

June 19, 2019



Motivation

- ▶ Distinct, salient **markers** are widely used in computer vision applications (also called fiducials, control points, ...)
- ▶ Black/white transitions exhibit high contrast and SNR



Motivation

- ▶ Distinct, salient **markers** are widely used in computer vision applications (also called fiducials, control points, ...)
 - ▶ Black/white transitions exhibit high contrast and SNR
-
- ▶ Camera calibration
 - ▶ Close-range photogrammetry
 - ▶ Robotics (hand-eye calibration)
 - ▶ Augmented Reality (AR)



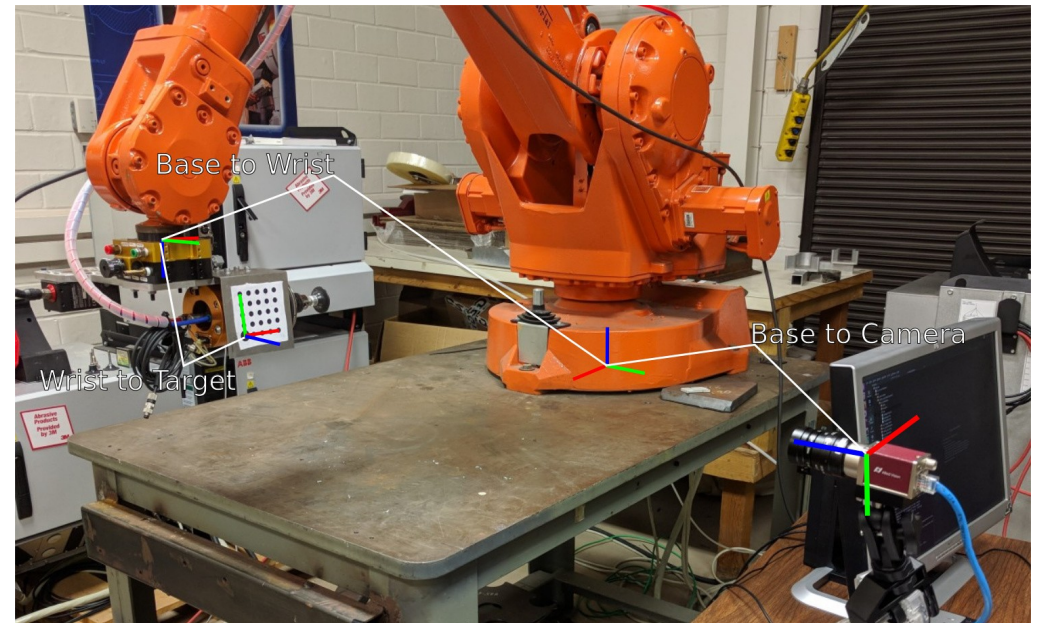
Motivation

- ▶ Distinct, salient markers are widely used in computer vision applications (also called fiducials, control points, ...)
- ▶ Black/white transitions exhibit high contrast and SNR
- ▶ Camera calibration
- ▶ Close-range photogrammetry
- ▶ Robotics (hand-eye calibration)
- ▶ Augmented Reality (AR)



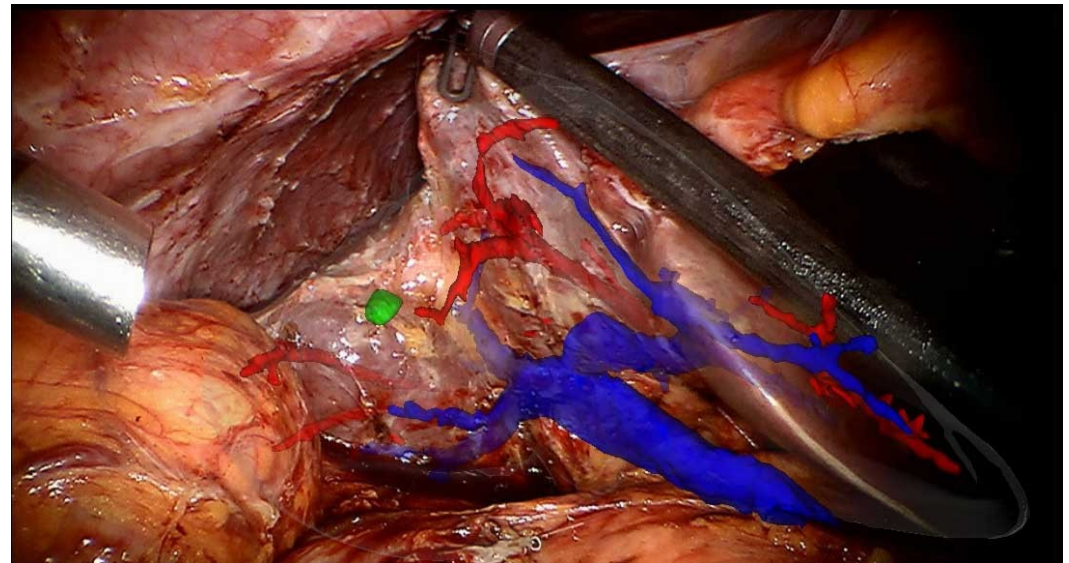
Motivation

- ▶ Distinct, salient markers are widely used in computer vision applications (also called fiducials, control points, ...)
- ▶ Black/white transitions exhibit high contrast and SNR
- ▶ Camera calibration
- ▶ Close-range photogrammetry
- ▶ Robotics (hand-eye calibration)
- ▶ Augmented Reality (AR)

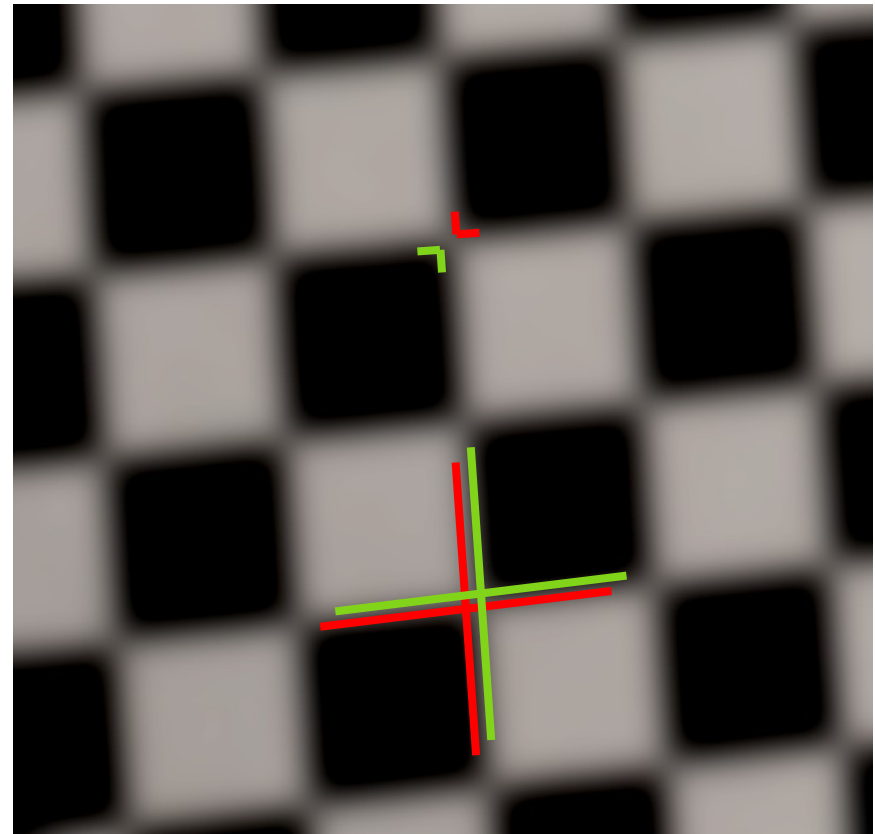
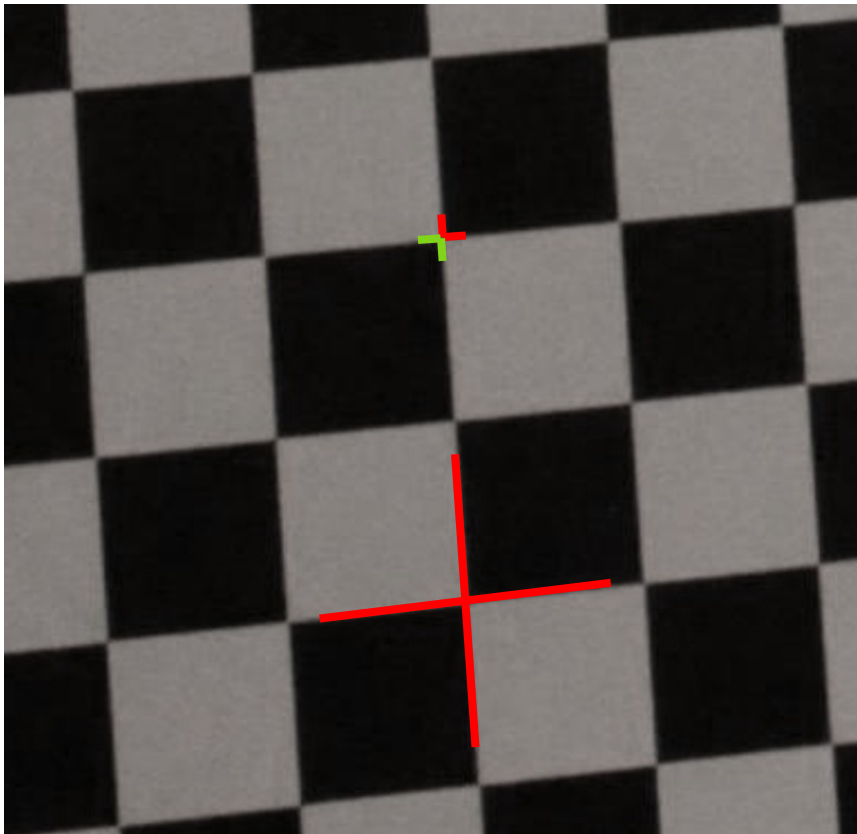


Motivation

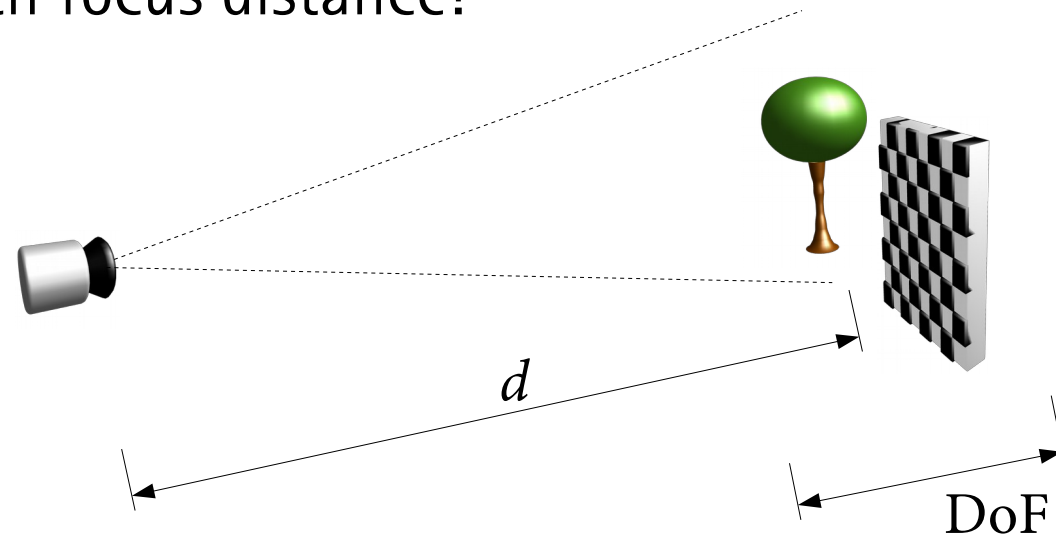
- ▶ Distinct, salient markers are widely used in computer vision applications (also called fiducials, control points, ...)
- ▶ Black/white transitions exhibit high contrast and SNR
- ▶ Camera calibration
- ▶ Close-range photogrammetry
- ▶ Robotics (hand-eye calibration)
- ▶ Augmented Reality (AR)



- ▶ **Problem:**
Marker localization difficult in blurred images
- ▶ State of the art: based upon corner or line detection
- ▶ But: neither corners nor lines are preserved under blurring

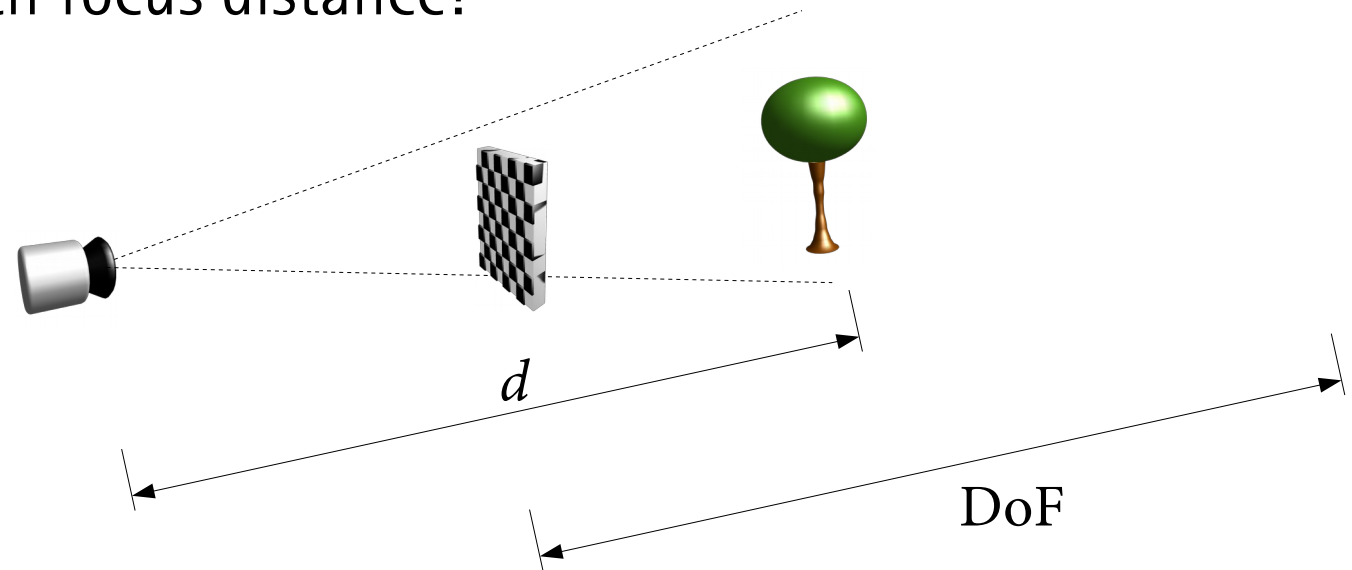


- ▶ Specific problem with camera calibration:
Calibrate at which focus distance?



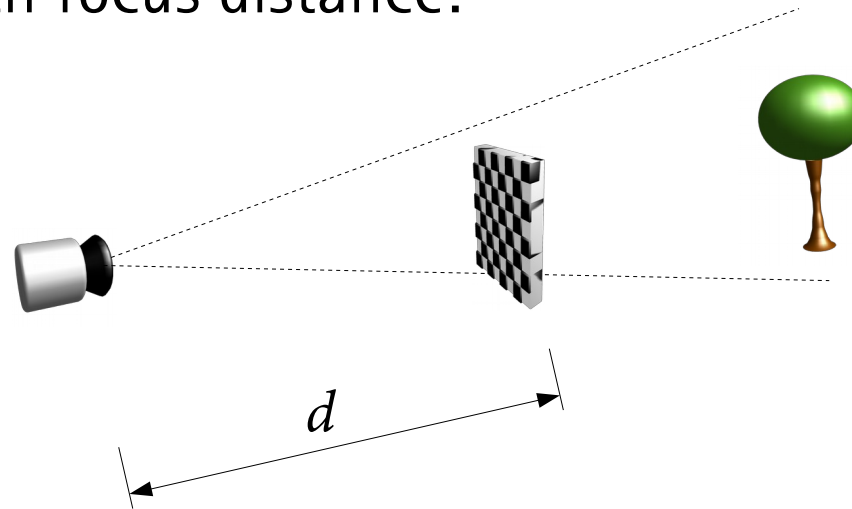
- ▶ Common photogrammetric recommendation:
 - ▶ Set focus distance d to working distance, or to infinity
 - ▶ In case of small depth of field (DoF):
 - ▶ Huge targets required
(everything outside of DoF range is blurry)

- ▶ Specific problem with camera calibration:
Calibrate at which focus distance?



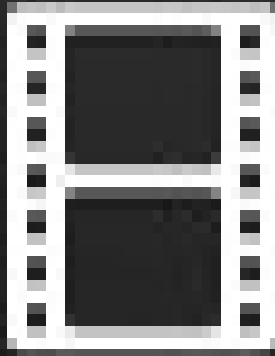
- ▶ First idea:
 - ▶ Increase DoF by stopping down
(DoF is function of d , focal length f , f-number N and acceptable blur C)
 - ▶ But: changing aperture changes camera parameters
(focal length, distortion, ...)

- ▶ **Specific problem with camera calibration:**
Calibrate at which focus distance?

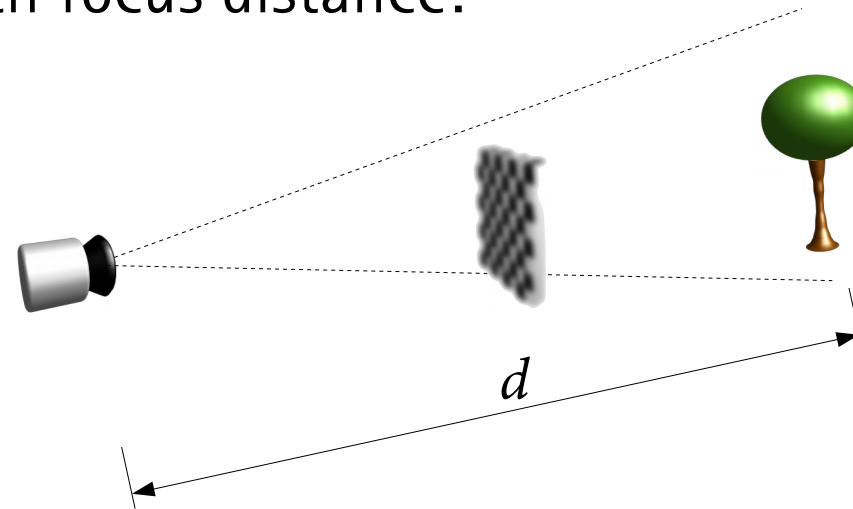


- ▶ **Second idea:**
 - ▶ Focus to target in near distance
 - ▶ **But:** changing the focus changes the camera parameters even more severely!
 - ▶ "Lens breathing"

Motivation



- ▶ Specific problem with camera calibration:
Calibrate at which focus distance?



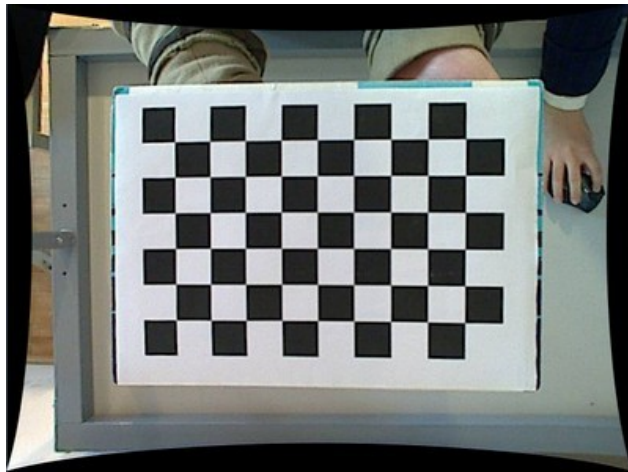
- ▶ Solution
 - ▶ Focus to original working distance d
 - ▶ Calibrate with defocused targets in near range
 - ▶ Marker detection for severely blurred markers needed

- ▶ Geometric relationship between scene and image:
Mapping \mathcal{P} from world space to the image plane,

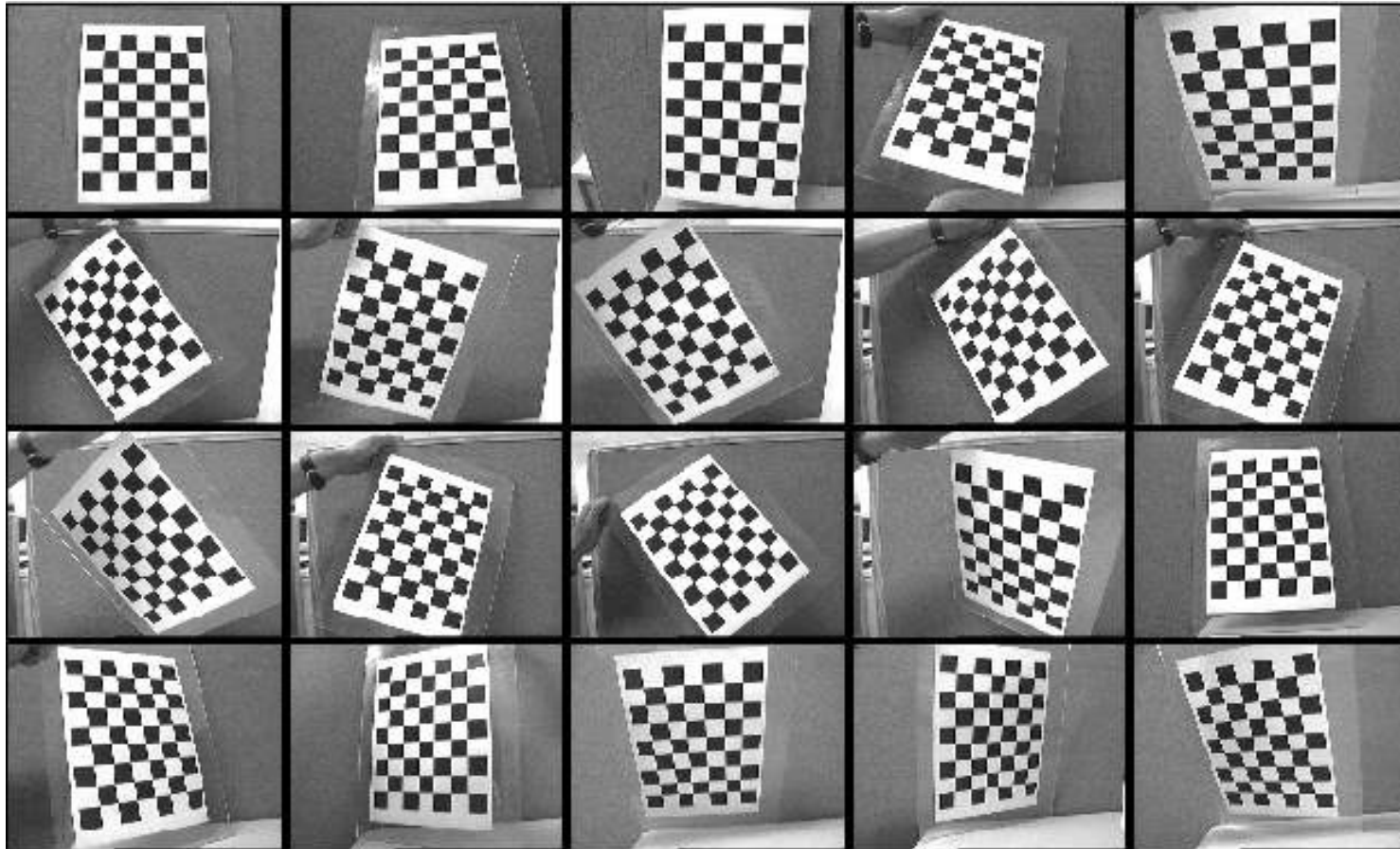
$$\mathcal{P}: \mathbb{R}^3 \rightarrow \mathbb{R}^2, (X, Y, Z) \mapsto (x, y)$$

- ▶ **(Geometric) camera calibration:**
Parameter estimation for a model of \mathcal{P}
- ▶ Estimation is carried out using
 - ▶ Point correspondences
and/or
 - ▶ Known a-priori constraints within the scene
in
 - ▶ Single or multiple images

- ▶ Self-calibration (using point correspondences within the imaged scene) methods available, *but*:
- ▶ Target-based calibration prevailing in many applications
 - ▶ Accuracy
 - ▶ Reproducibility
- ▶ Common targets in CV: Checkerboards

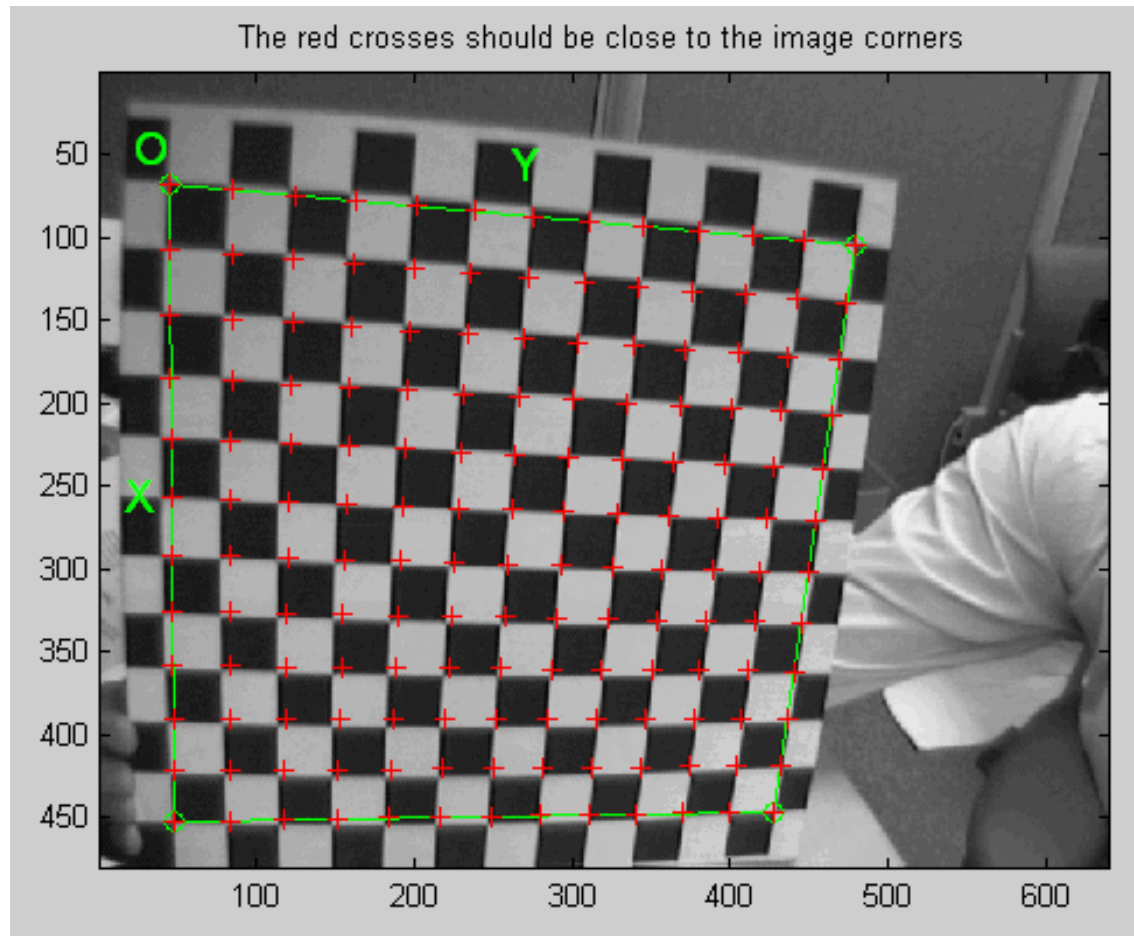


► CalTech calib toolbox¹ toolchain



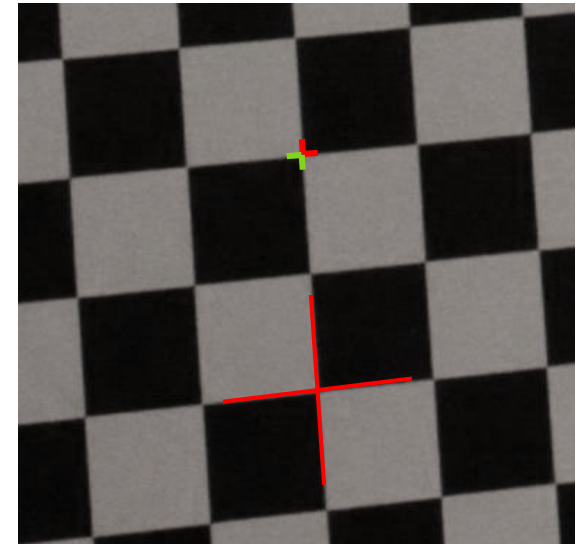
¹ J.-Y. Bouquet, "Camera Calibration Toolbox for MATLAB", version 2017-06-01, http://www.vision.caltech.edu/bouquetj/calib_doc

► CalTech calib toolbox¹ toolchain



¹ J.-Y. Bouquet, "Camera Calibration Toolbox for MATLAB", version 2017-06-01, http://www.vision.caltech.edu/bouquetj/calib_doc

- ▶ Marker localization: State of the art
 - ▶ Two-stage hierarchical approach
 1. Coarse localization
 - ▶ Harris-type corner detection
 - *or* –
 - Crossings of detected lines
 - ▶ Postprocessing to verify topology
 2. Subpixel refinement
 - ▶ Widely deployed (OpenCV¹, Geiger et. al.²):
Variant of Förstner interest point detector



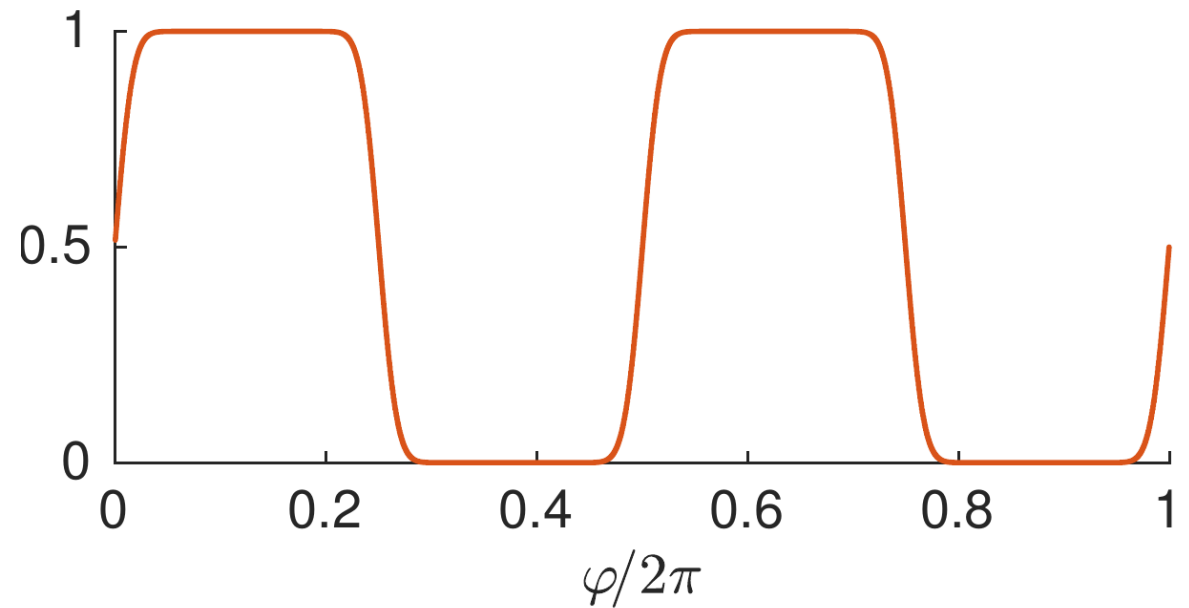
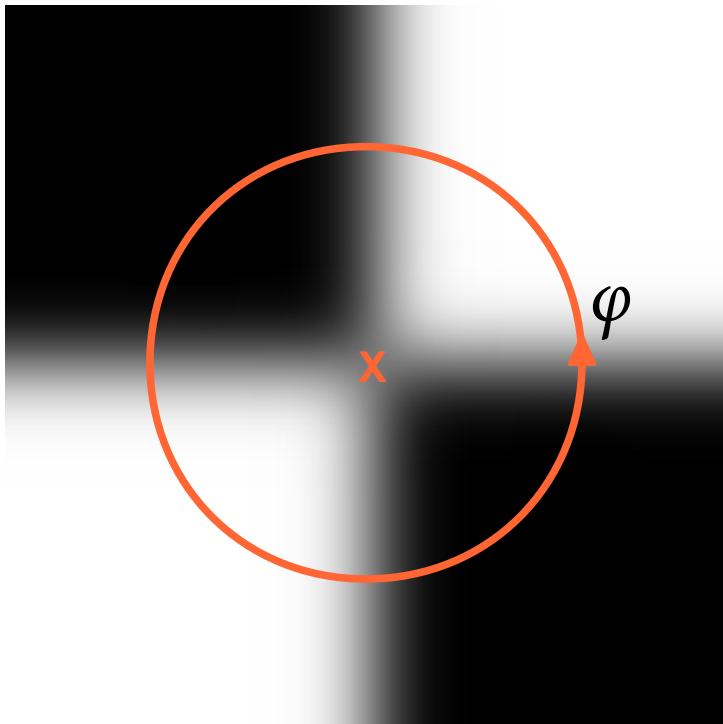
¹ OpenCV 3.4.1, `cornerSubPix()` function

² Geiger et. al., "Automatic camera and range sensor calibration using a single shot", ICRA '12

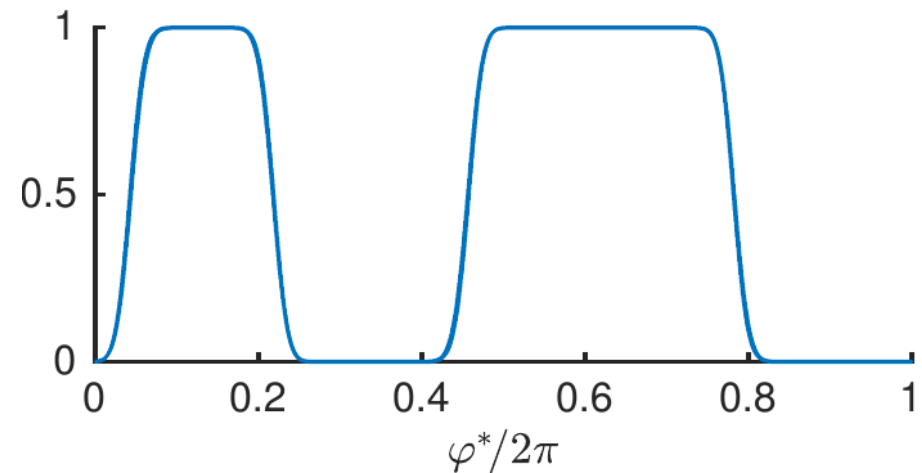
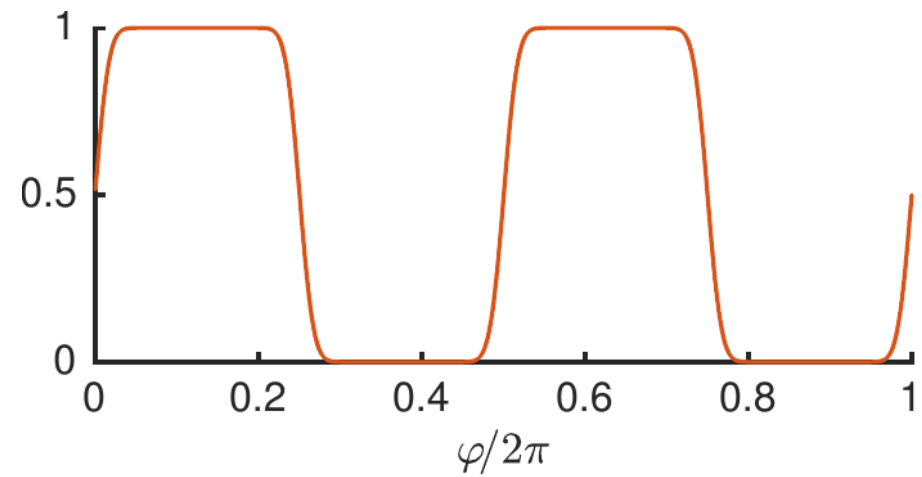
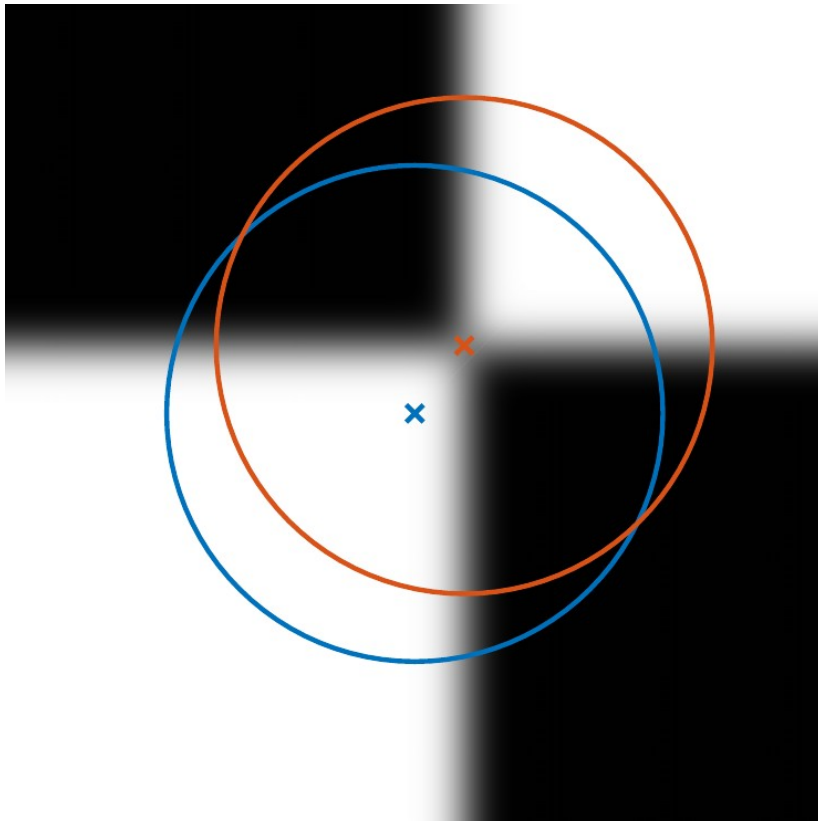
³ W. Förstner and E. Gülch, "A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features", ISPRS Conf. Proc. Ph. Data '87

- ▶ State of the art fails for:
 - ▶ Strong blur
 - ▶ High noise levels
 - ▶ Asymmetric transitions due to nonlinear response ("gamma")

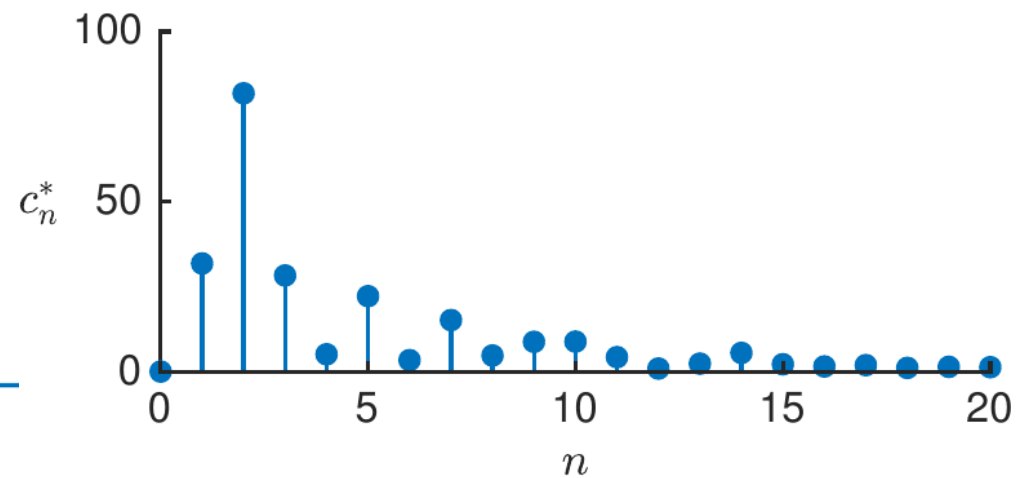
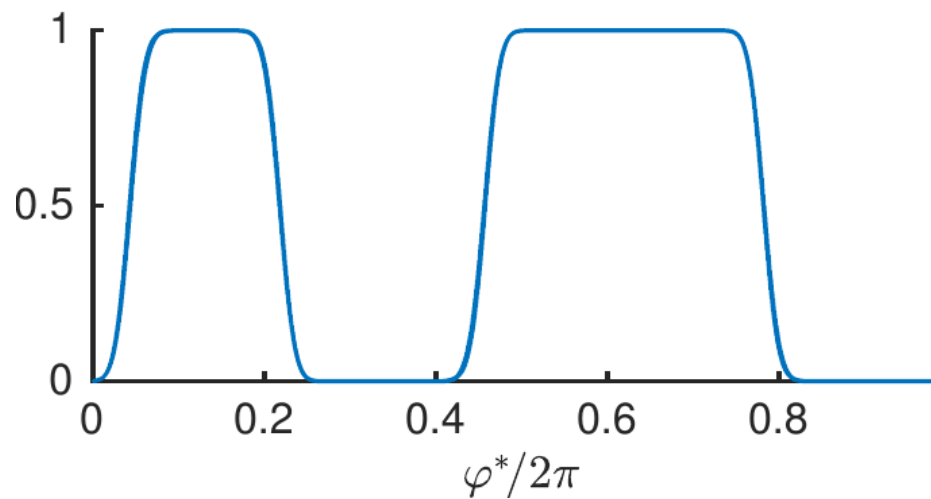
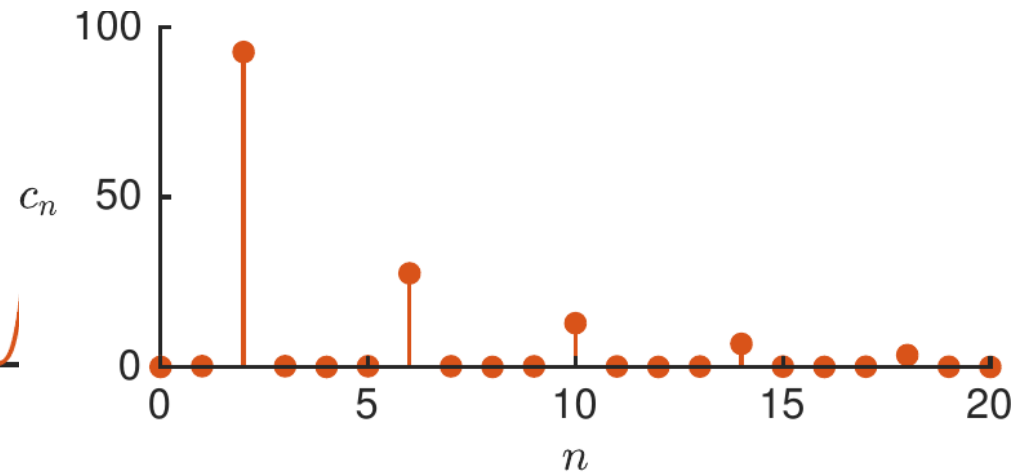
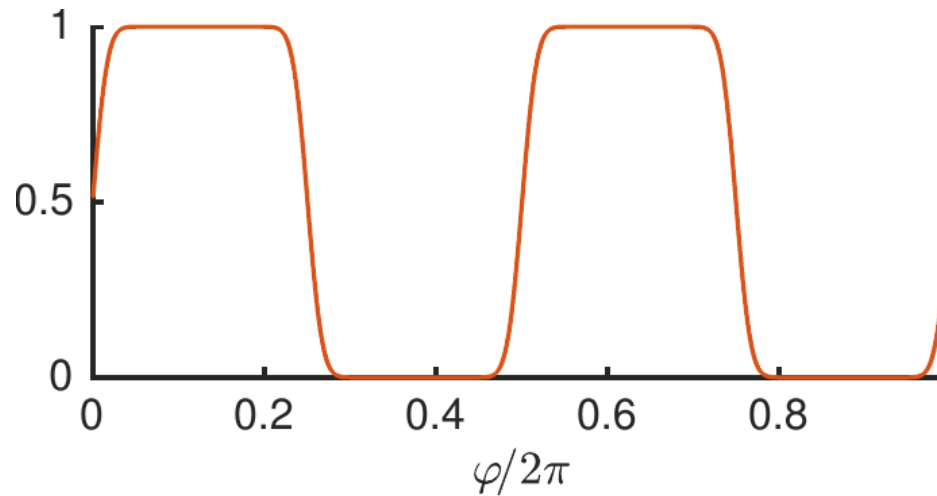
- ▶ Signal along φ is π -periodic



- ▶ Decentered signal is 2π -periodic in φ !



► Fourier analysis of angular signal



Refresher

<i>Input</i>	Periodic	Infinite
Continuous	Fourier series	Fourier transform
Discrete	DFT (Discrete FT)	DTFT (Discrete-Time FT)

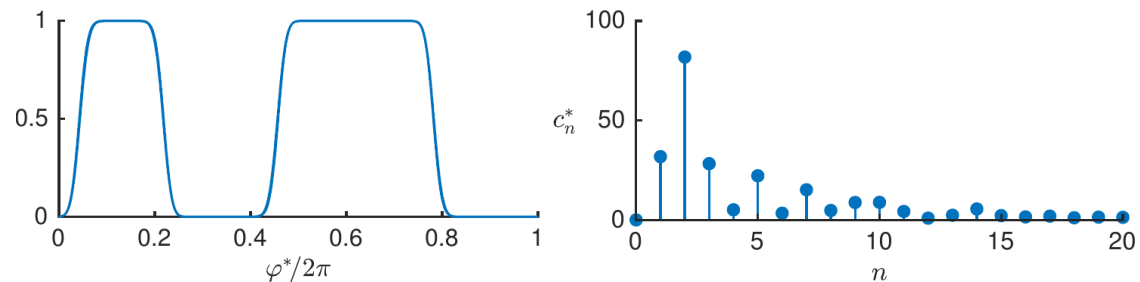
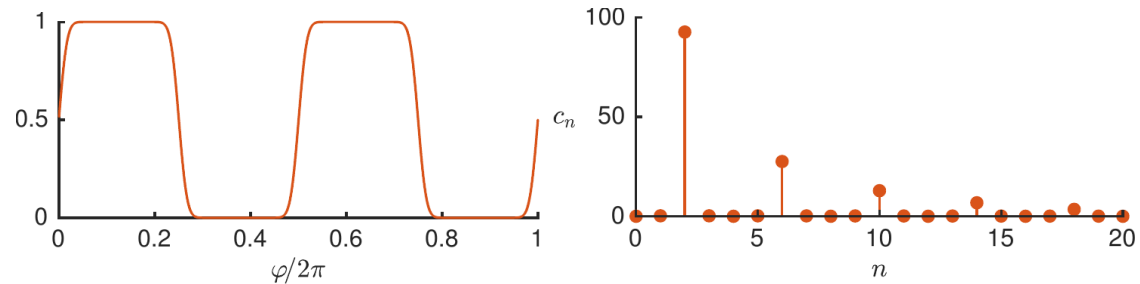
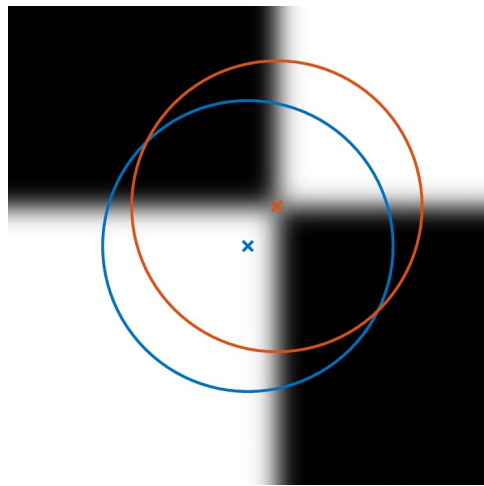
Color legend:

Discrete output

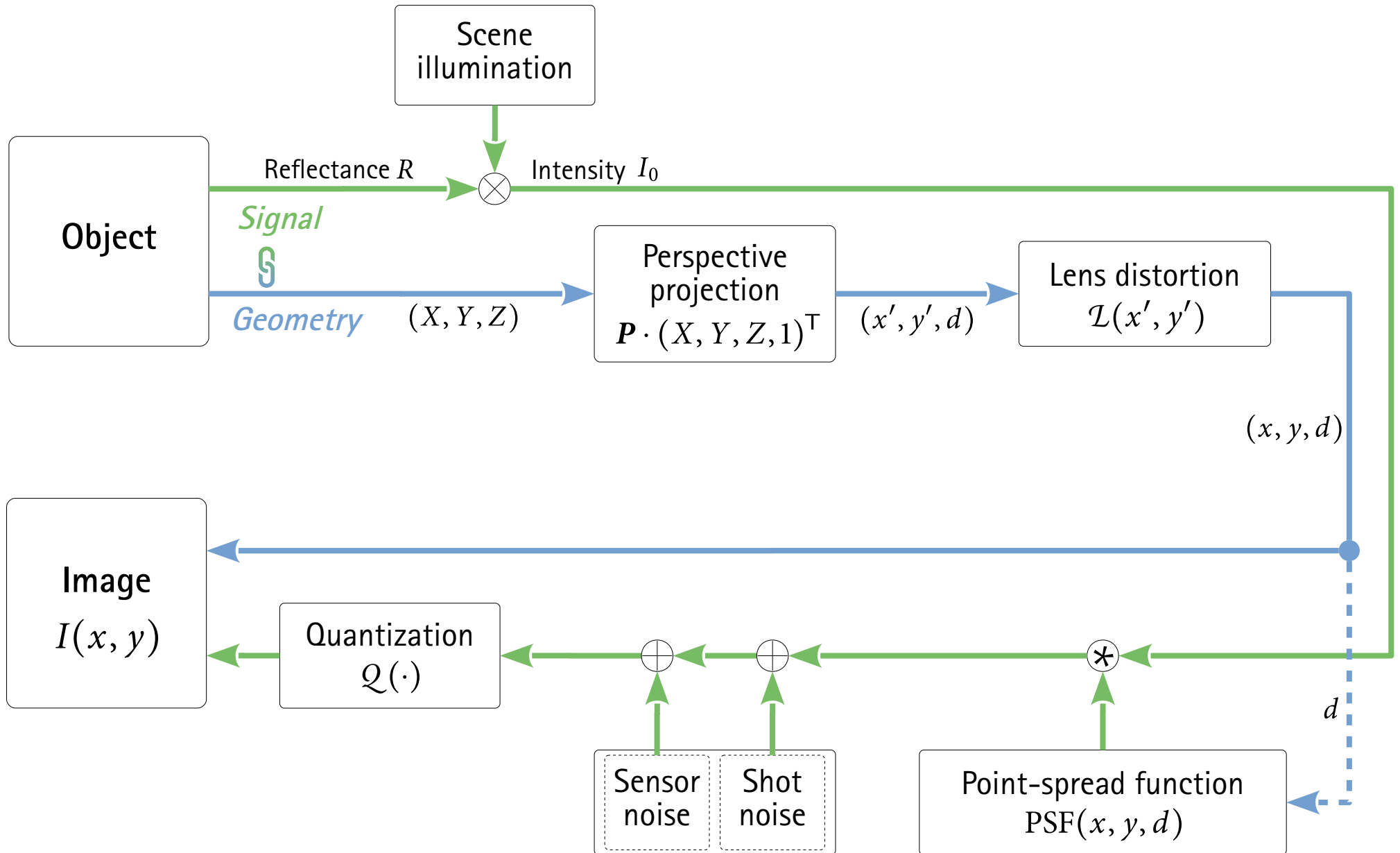
Continuous output

- ▶ Sub-pixel offset estimation:
 - ▶ Minimize the "wrong" = odd Fourier components, weighted by the ideal decay

$$\mathbf{x}_0^{\text{est}} = \arg \min_{x_0, y_0} \sum_{k=0}^{\infty} \left| b_{2k+1} \int_0^{2\pi} \hat{s}_{\text{blur}}(r^*, \varphi^*) e^{j(2k+1)\varphi^*} d\varphi^* \right|^2$$

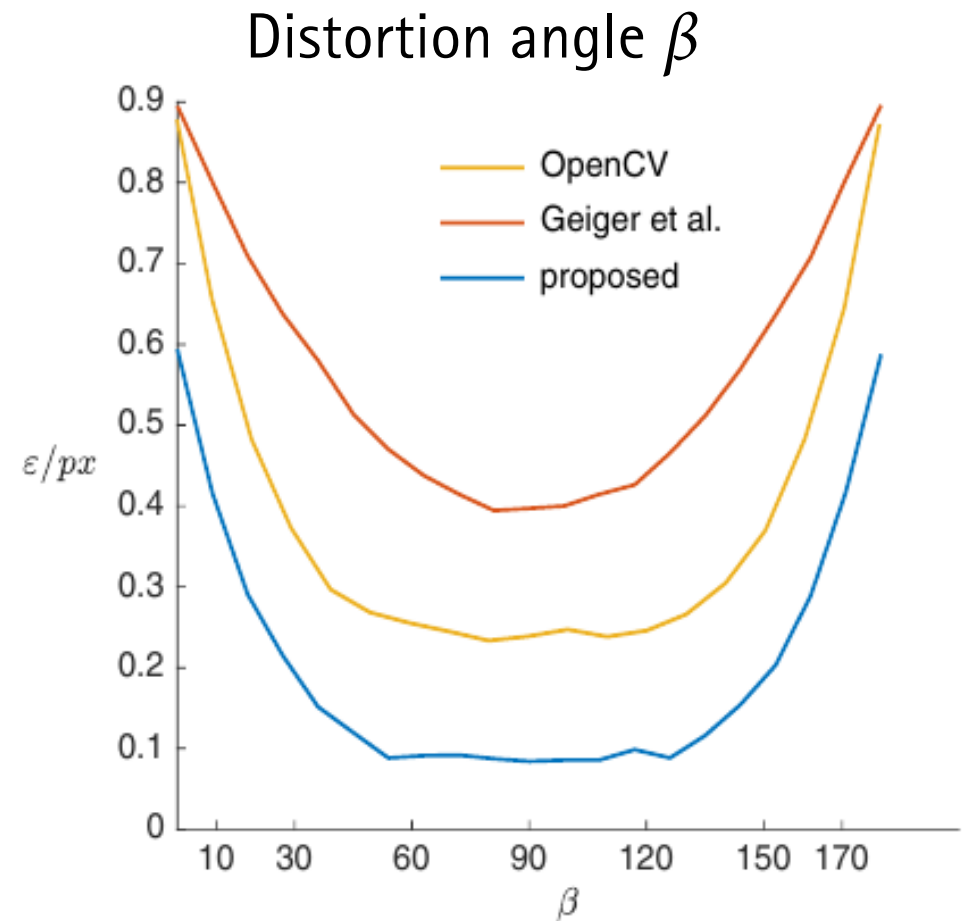
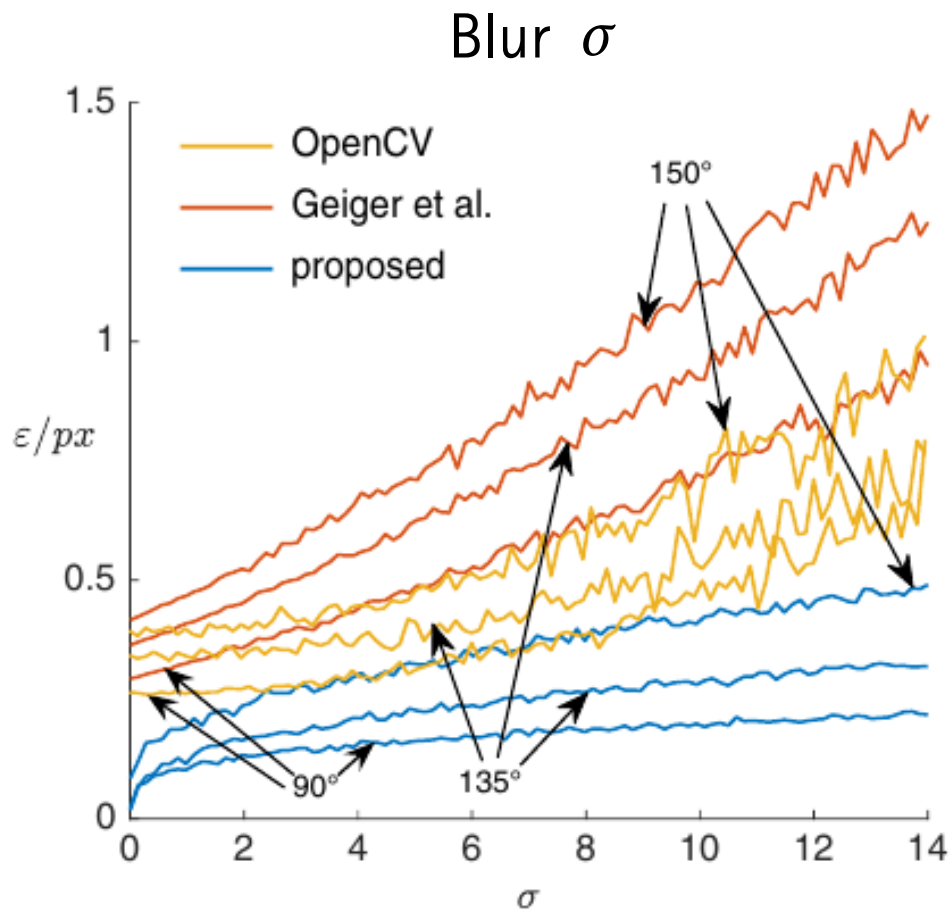


Imaging pipeline

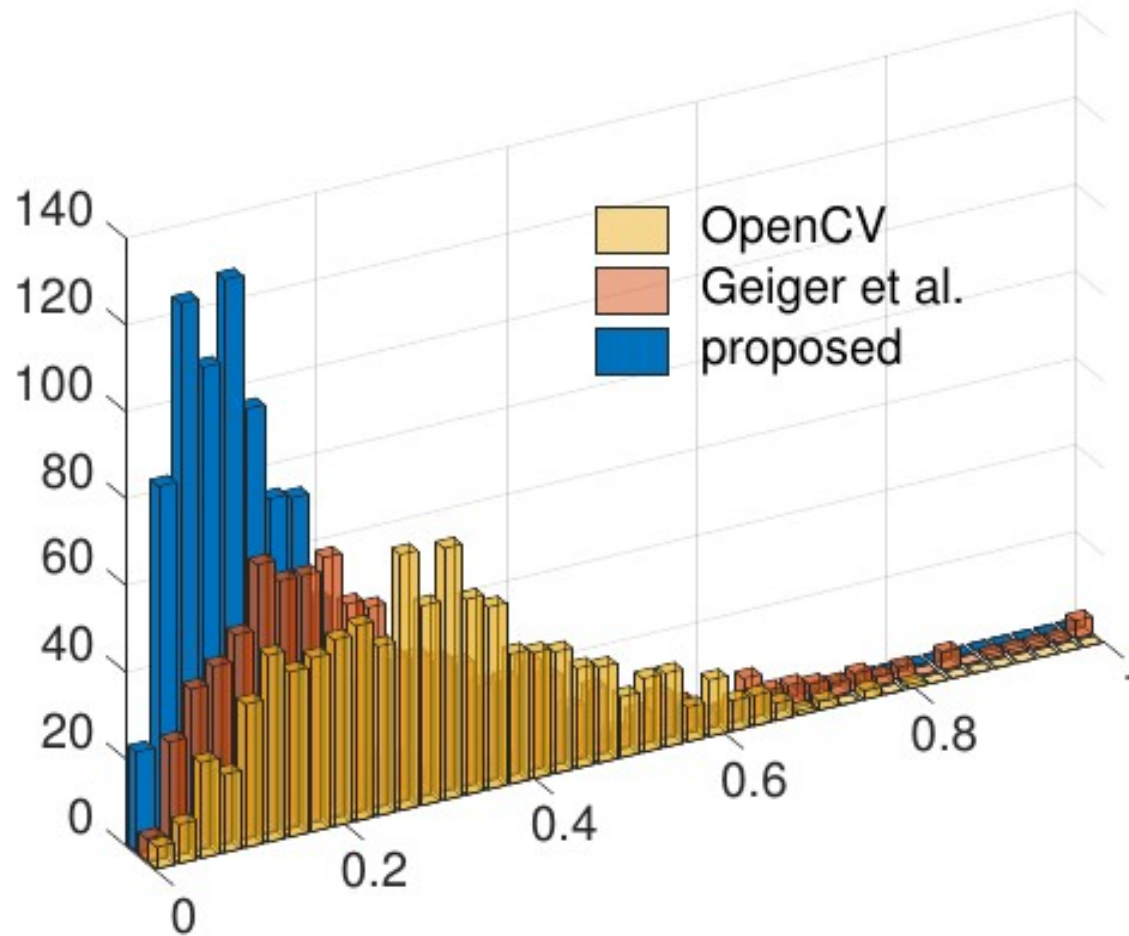


► *Synthetic images:*

Absolute localization error w.r.t.



- ▶ *Real images* (DIMA dataset):
Distribution of reprojection errors



Conclusions

- ▶ New method for marker localization exploiting angular symmetry
- ▶ Highest positional accuracy
- ▶ Robust against common perturbations during imaging process
- ▶ Highly beneficial in applications such as professional AR systems



Potential of Deep Learning in the Field of Industrial Quality Assurance

Andreas Spruck

andreas.spruck@fau.de

Chair of Multimedia Communications
and Signal Processing



FRIEDRICH-ALEXANDER
UNIVERSITÄT
ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT



Outline

- Motivation
- Automated Optical Inspection Systems
- Potential of Deep Learning
- Challenges with Deep Learning
- Implementation of Demo System
- Conclusion

Motivation

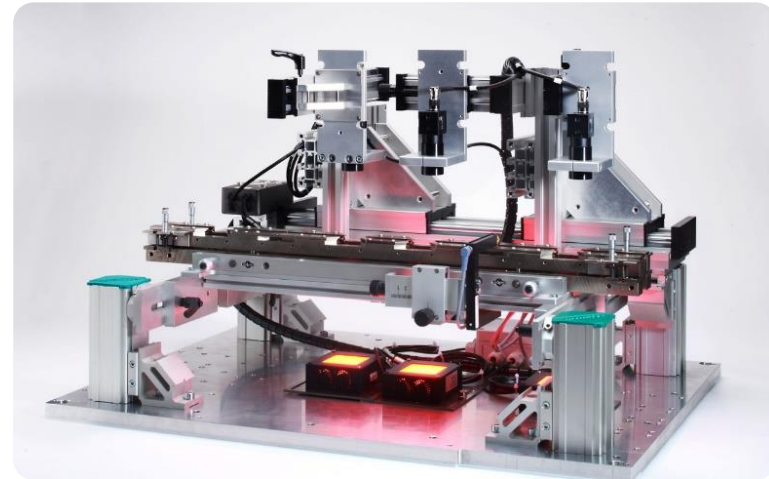
- Customers demand error-free high quality products
- Quality Assurance nowadays:
 - Mainly manual optical inspection by a worker
 - Very monotonous & tiring labor
 - Time and cost intensive process
- Automated system for certain tasks



Source: BMW

Automated Optical Inspection Systems

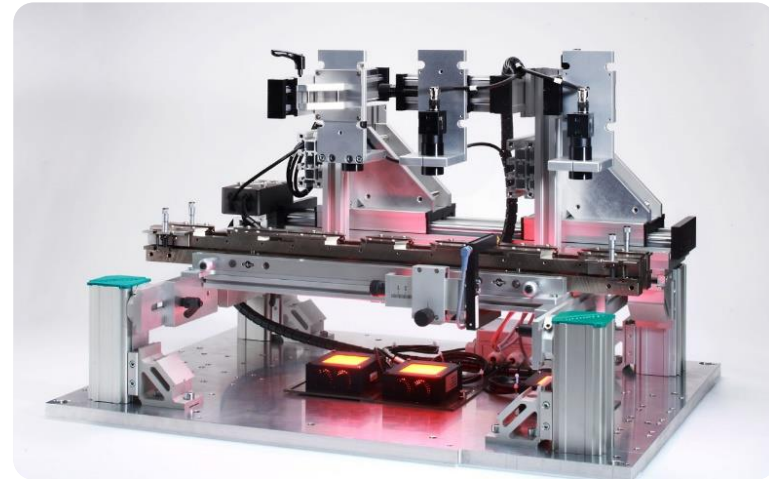
- Most common for items which are produced in large numbers
 - Screws, automotive parts, ...
- Individual configuration depending on the considered application
 - Acquisition system
 - Camera, X-Ray, ultrasound, laser triangulation
 - Lighting system
 - Item transportation
 - Sorting
 - Performable measurements



Source : www.otto-jena.de

Automated Optical Inspection Systems

- Strict parametrization of the tolerance range for every specific component necessary
- Restricted to single specified inspection task
- Elaborate reconfiguration of the test setup



Source : www.otto-jena.de

Potential of Deep Learning

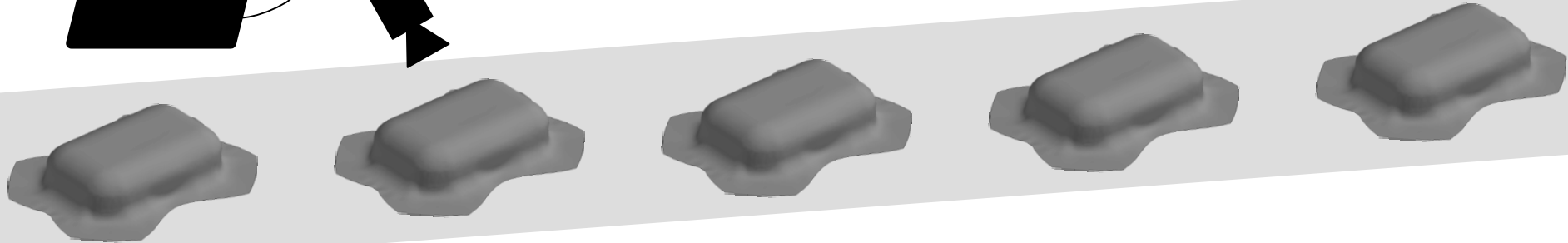
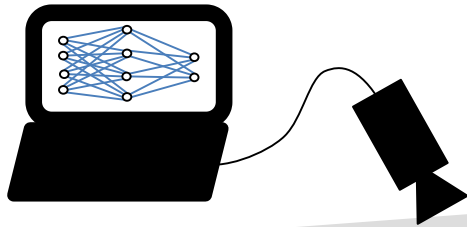
- Fast progress in the field of classification using neural networks
 - Quality Assurance can be treated as simple classification problem
 - Error recognition
 - Distinction of certain error types
- Neural networks can easily solve localization problems
 - Type of error and position of the error can be distinguished
 - More comfortable for worker to inspect the item or correct the error

Challenges with Deep Learning

- High requirements on the training data set
 - Large enough
 - High quality labeling
 - Each class evenly represented
- Works very reliable
 - Similarly high recognition rates as humans

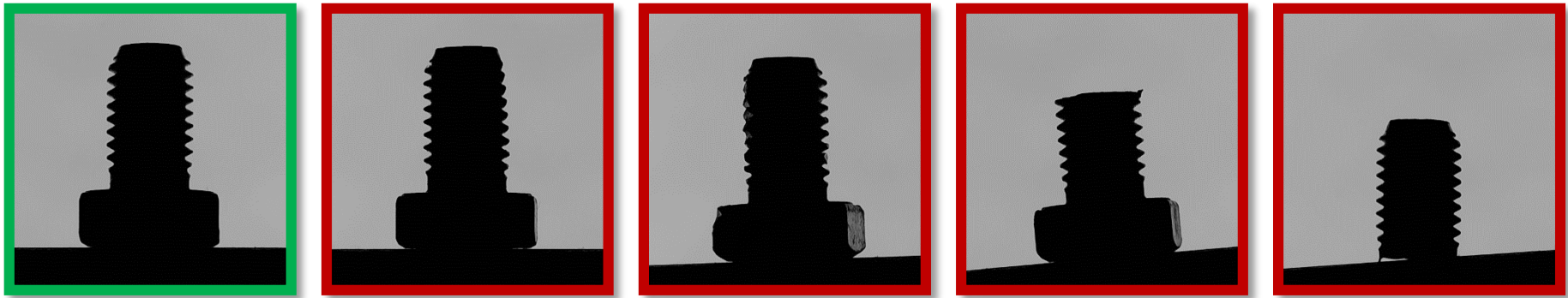
Implementation of learning based Inspection system

- Existing infrastructure may be reused
- Low roll-out costs
- Only software changes necessary



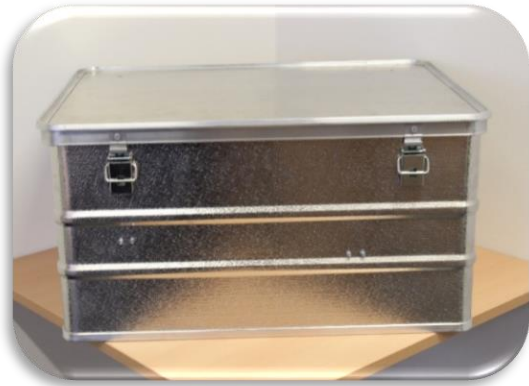
Implementation of Demo System

- Acquisition of training dataset
 - Set of 110 screws
 - 50 error-free screws
 - 60 erroneous screws with different error types

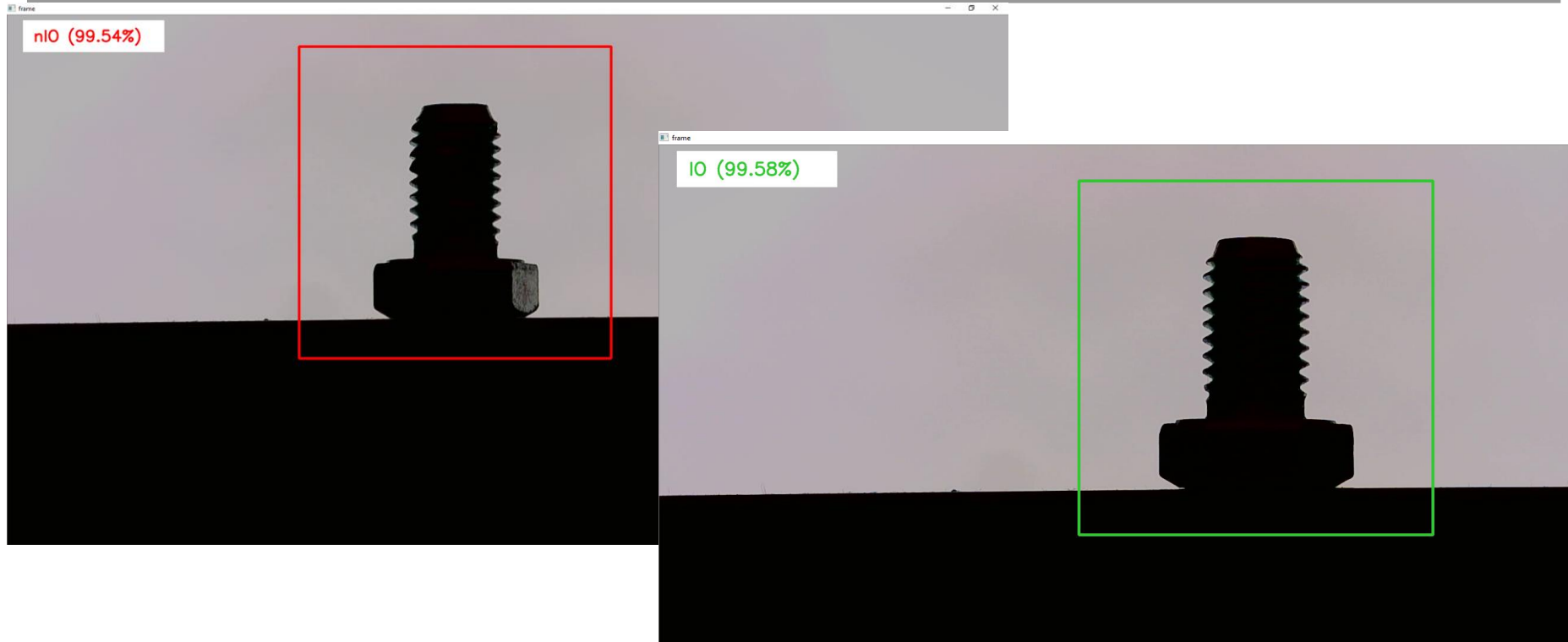


Implementation of Demo System

- Transportable system
- Inspired by real-world inspection systems



Implementation of Demo System



Conclusion

- Manual inspection of products should be automated
- Existing automated inspection systems are highly parametrized and inflexible
- Advances in the field of neural networks enable a new type of inspection systems
- Recognition rates similarly high to humans
- Overhead for training and data acquisition should be reduced

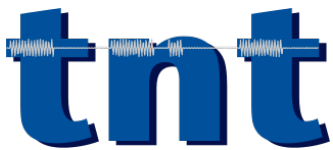
Optimization for MPEG–G compliant entropy coding

5th Summer School on Video Compression and Processing (SVCP) 2019

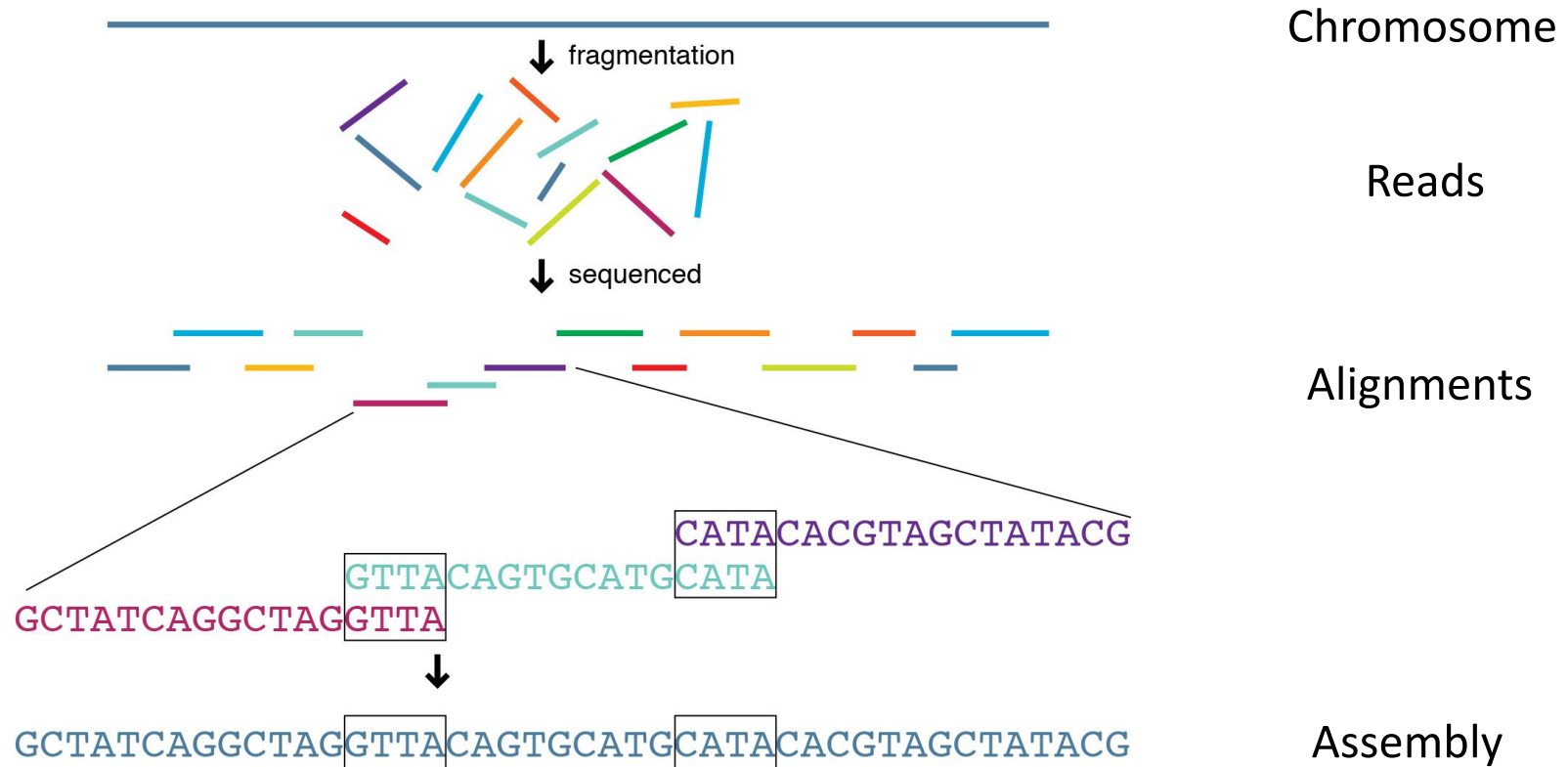
Jan Voges

Yeremia Gunawan Adhisantoso

Jörn Ostermann



Whole genome sequencing



Evolution of genome sequencing

Sequencing technology

	2009	2018/2019
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

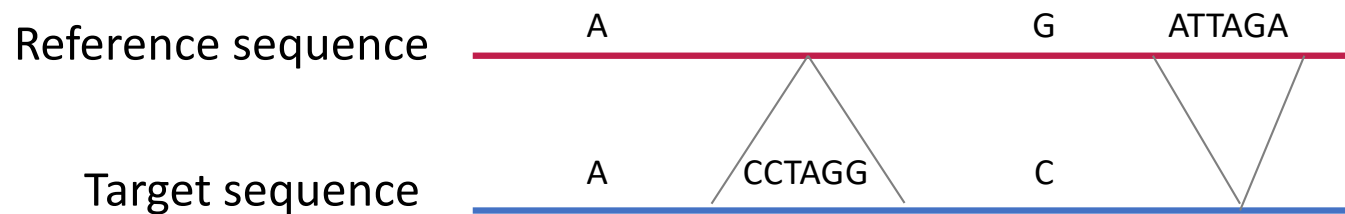
Storage & transmission infrastructure

	2009	2018/19
Cost/TB	\$100	\$50
Download speed	10 Mbps	100 Mbps

No technology is keeping with the pace of genome sequencing!

Lossless compression of DNA sequences

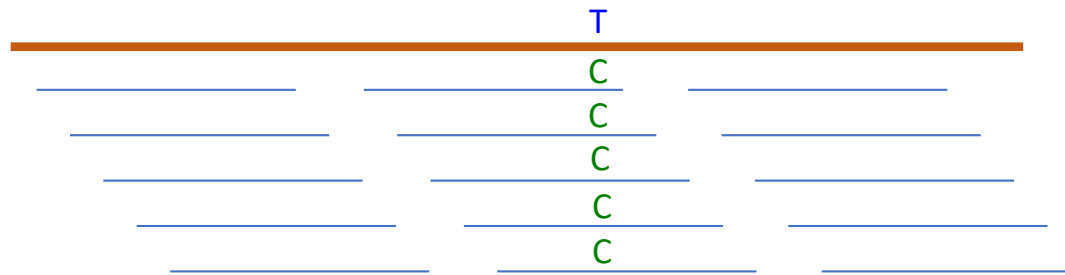
1. Find the differences between the target and the reference sequence



2. Encode those differences

$$m_i = (p_i, l_i, C_i)$$

Lossless compression of aligned reads



Approach:

- Exploit the redundancy present in the reads
- Predict variants of current read given previous ones

MPEG-G

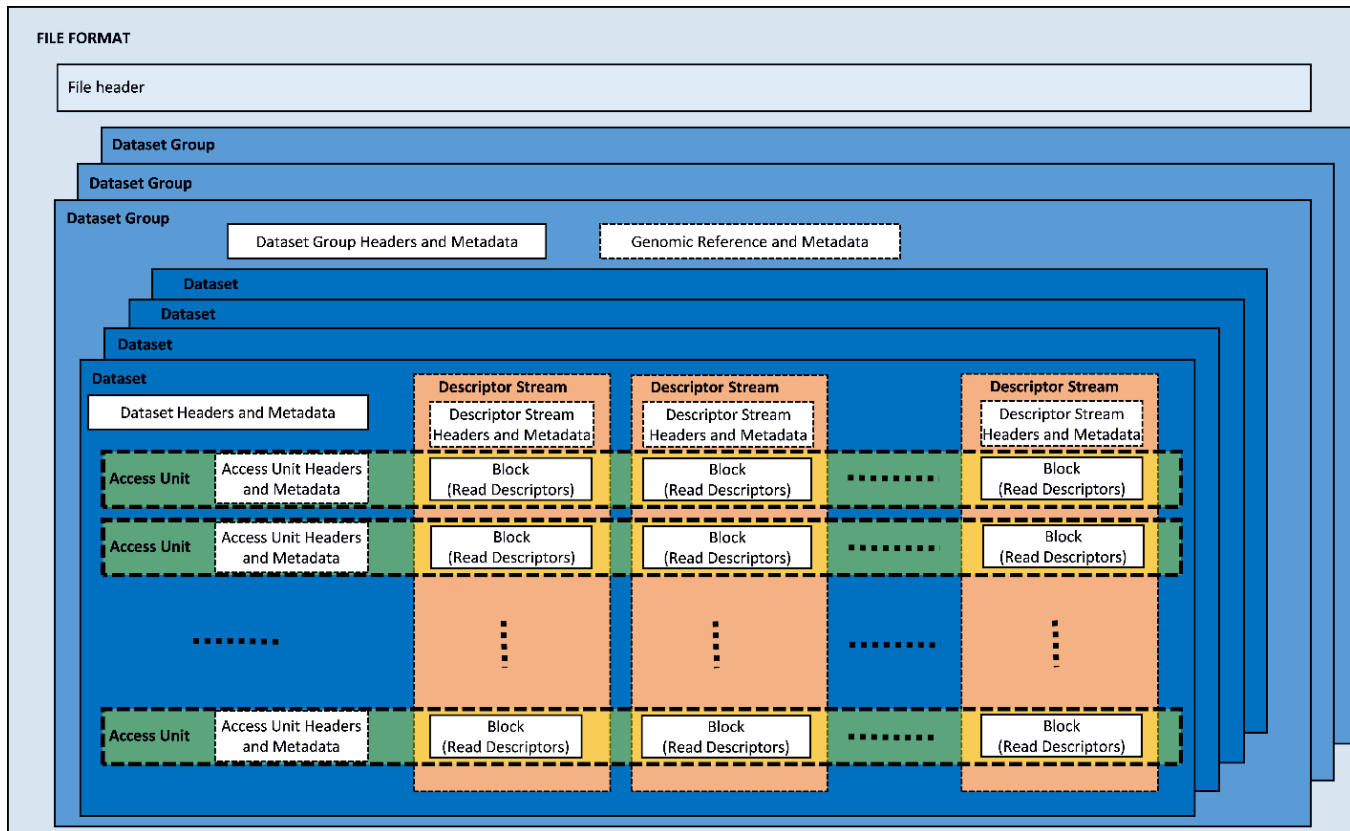
- MPEG-G = International Standard ISO/IEC 23092
- Standard = normative text
 - A set of instructions of how to retrieve genomic data from the compressed domain
 - Not tied to a particular implementation
- Largest coordinated and international effort by end users, industry and academia



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: Reference Software**
The standard support to the implementation of applications
- **Part 5: Conformance**
The methodology to test compliance with the standard

MPEG-G file format



An example:

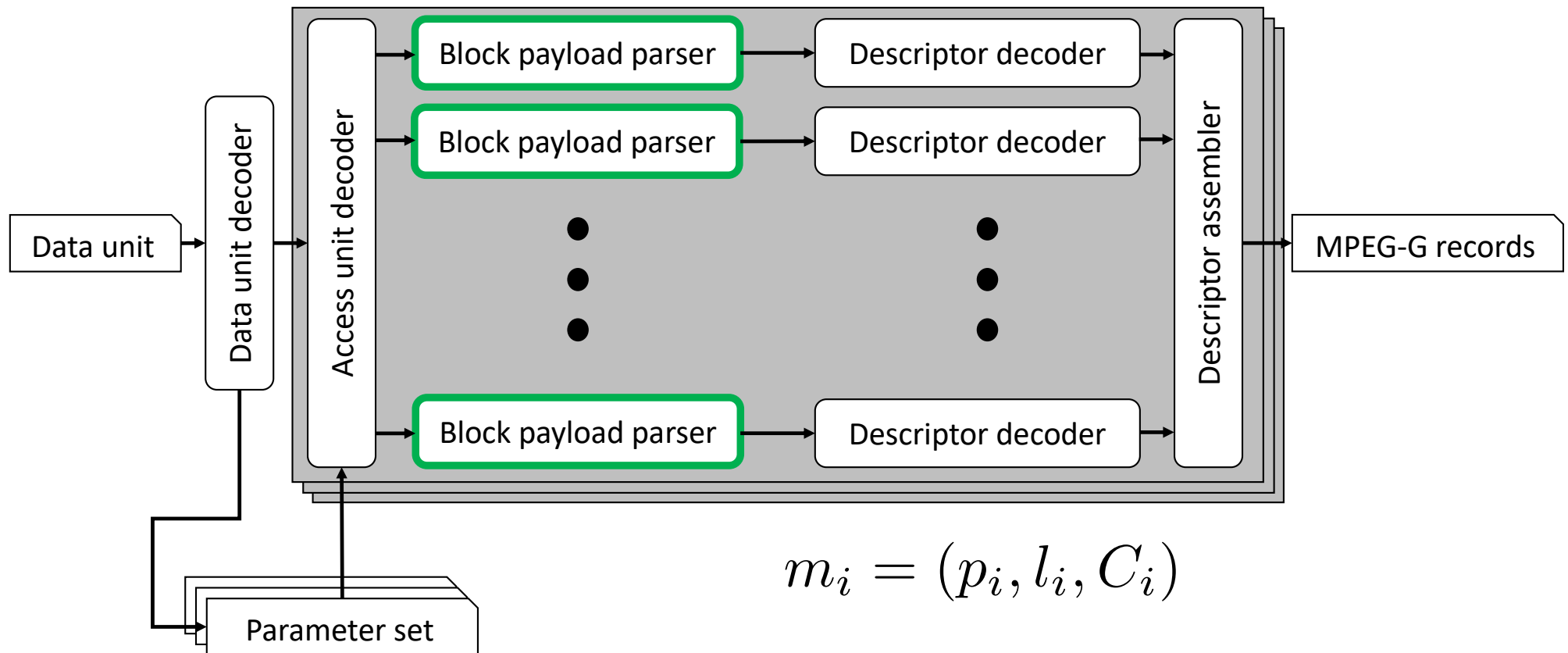
- **MPEG-G file**: sequencing data of a trio
- **File Header**: metadata related to the study
- **Dataset Group**: one per individual + metadata from the individual
- **Dataset**: sequencing data + metadata from one experiment
- **Colored structures**: this is how genomic data is represented in MPEG-G

The MPEG-G file can encapsulate the entire genomic history of one or more individuals in a unique file including the metadata describing the study, samples, etc.

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: Reference Software**
The standard support to the implementation of applications
- **Part 5: Conformance**
The methodology to test compliance with the standard

MPEG-G Part 2 – the decoder core



GABAC

- GABAC = **G**enomics-oriented context **A**daptive **B**inary **A**rithmetic **C**oding
- GABAC is part of a collaborative effort to produce a standard-compliant open source MPEG-G encoder (*genie*)



Mikel Hernaez, Idoia Ochoa, Jan Voges,
Fabian Muntefering, Liudmila S. Mainzer,
Brian Bliss, Mingyu Yang

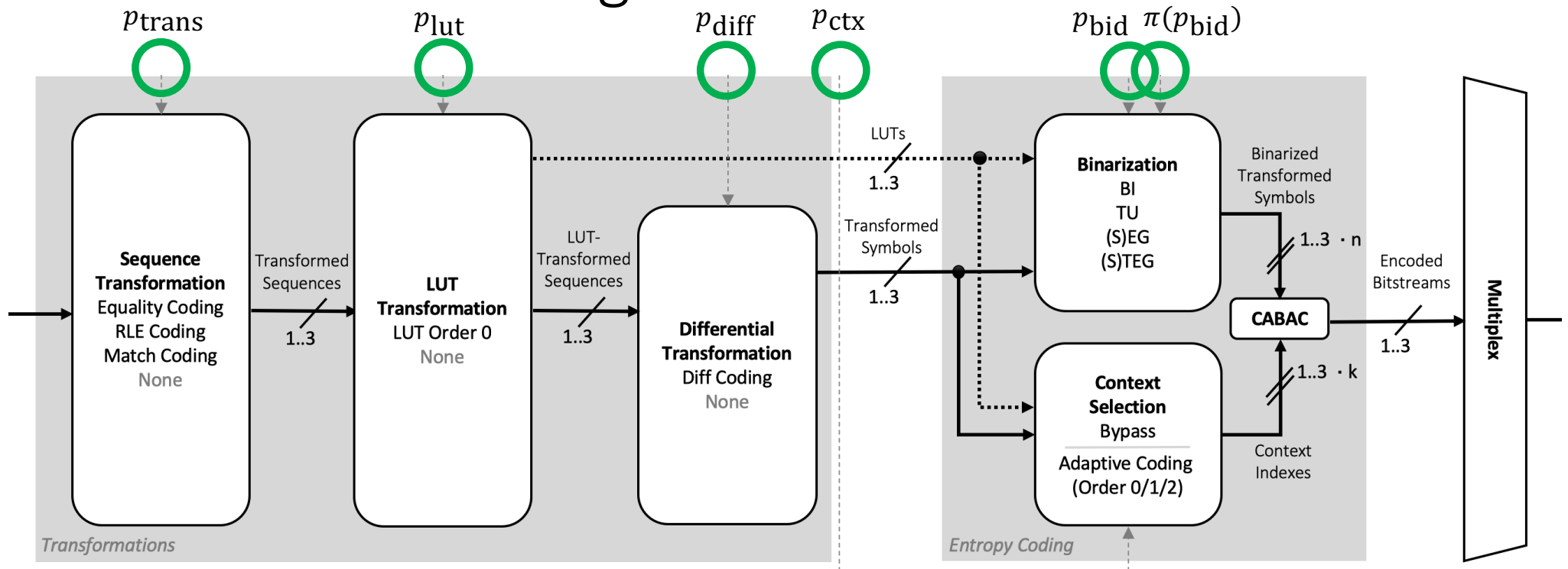


Tom Paridaens, Jan Fostier



Jan Voges, Jörn Ostermann, Fabian Muntefering

GABAC block diagram



----- Control signal

..... Only required when LUT Transformation is enabled

——— Signal

Average compression **ratio** (compressed size / uncompressed size):

0.199 %

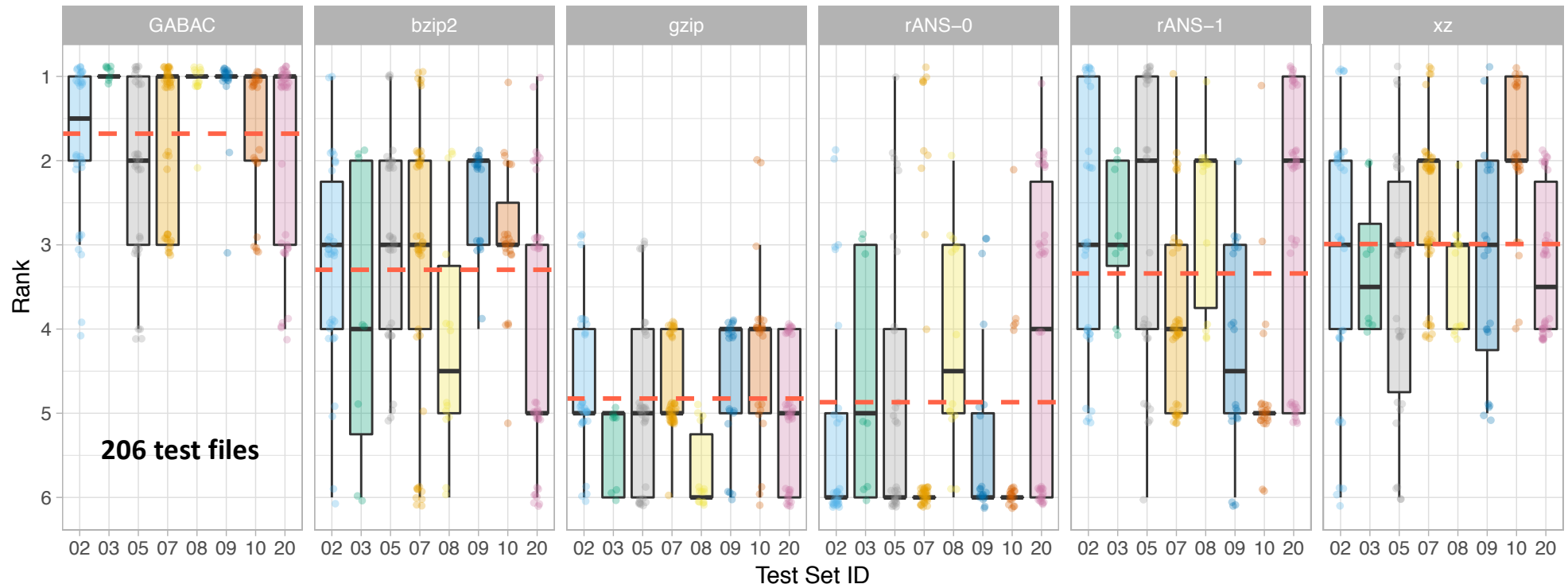
0.212 %

0.245 %

0.286 %

0.237 %

0.204 %



Average compression speed:

18.12 MiB/s

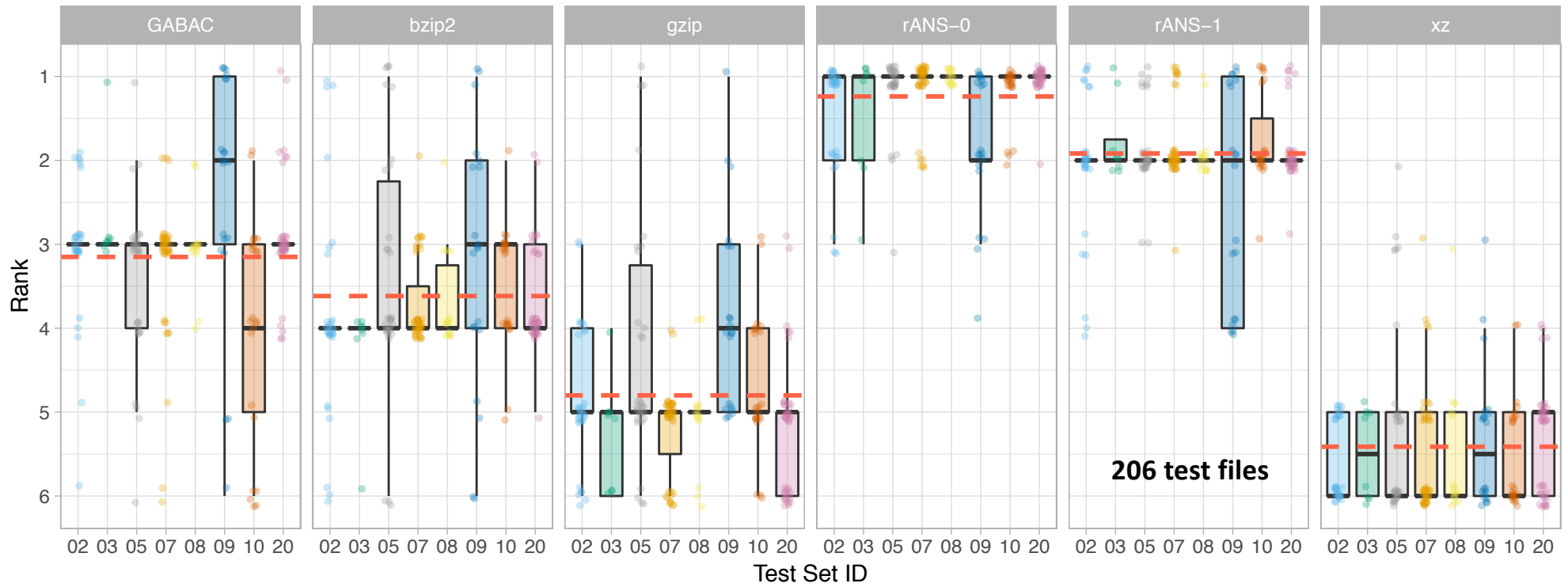
12.27 MiB/s

3.95 MiB/s

41.91 MiB/s

36.15 MiB/s

1.51 MiB/s



The GABAC configuration space

$$p_{\text{trans}} \in \mathcal{P}_{\text{trans}} = \{ \text{no_transform}, \text{equality_coding}, \text{match_coding}, \text{rle_coding} \}$$

$$n_{\text{ts}}(p_{\text{trans}}) = \begin{cases} 1, & \text{if } p_{\text{trans}} = \text{no_transform} \\ 2, & \text{if } p_{\text{trans}} = \text{equality_coding} \\ 3, & \text{if } p_{\text{trans}} = \text{match_coding} \\ 2, & \text{if } p_{\text{trans}} = \text{rle_coding} \end{cases}$$

$$p_{\text{diff}} \in \mathcal{P}_{\text{diff}} = \{ \text{disabled}, \text{enabled} \}$$

$$p_{\text{lut}} \in \mathcal{P}_{\text{lut}} = \{ \text{disabled}, \text{enabled} \}$$

The GABAC configuration space

$$p_{\text{ctx}} \in \mathcal{P}_{\text{ctx}} = \{ \text{bypass}, \text{adaptive_coding_order_0}, \text{adaptive_coding_order_1}, \text{adaptive_coding_order_2} \}$$

$$p_{\text{bid}} \in \mathcal{P}_{\text{bid}} = \{ \text{BI}, \text{TU}, \text{EG}, \text{SEG}, \text{TEG}, \text{STEG} \}$$

$$\pi(p_{\text{bid}}) = \begin{cases} 1, & \text{if } p_{\text{bid}} = \text{BI} \\ 1, & \text{if } p_{\text{bid}} = \text{TU} \\ 1, & \text{if } p_{\text{bid}} = \text{EG} \\ 1, & \text{if } p_{\text{bid}} = \text{SEG} \\ 32, & \text{if } p_{\text{bid}} = \text{TEG} \\ 32, & \text{if } p_{\text{bid}} = \text{STEG} \end{cases}$$

The GABAC configuration space

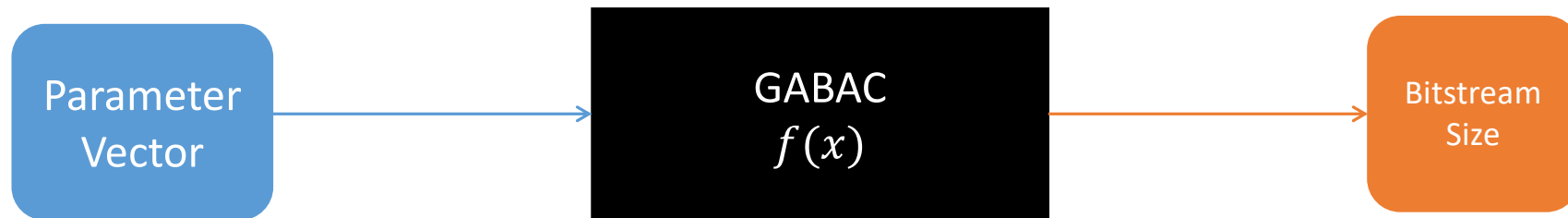
Total number of possible configurations N :

$$N \cong \sum_{\forall p_{\text{trans}} \in \mathcal{P}_{\text{trans}}} \left[\left(|\mathcal{P}_{\text{diff}}| \cdot |\mathcal{P}_{\text{lut}}| \cdot |\mathcal{P}_{\text{ctx}}| \cdot \sum_{\forall p_{\text{bid}} \in \mathcal{P}_{\text{bid}}} \pi(p_{\text{bid}}) \right) \cdot n_{\text{ts}}(p_{\text{trans}}) \right]$$

$N \cong 16,000$

Real-world implementation: $4,000 < N < 8,000$

The optimization problem



$$x^* = \operatorname{argmin}_{x \in X} f(x)$$

Optimization algorithms

Gradient
Based

Require existence of continuous first derivatives of the object function and possibly higher derivatives

Non-Gradient
Based

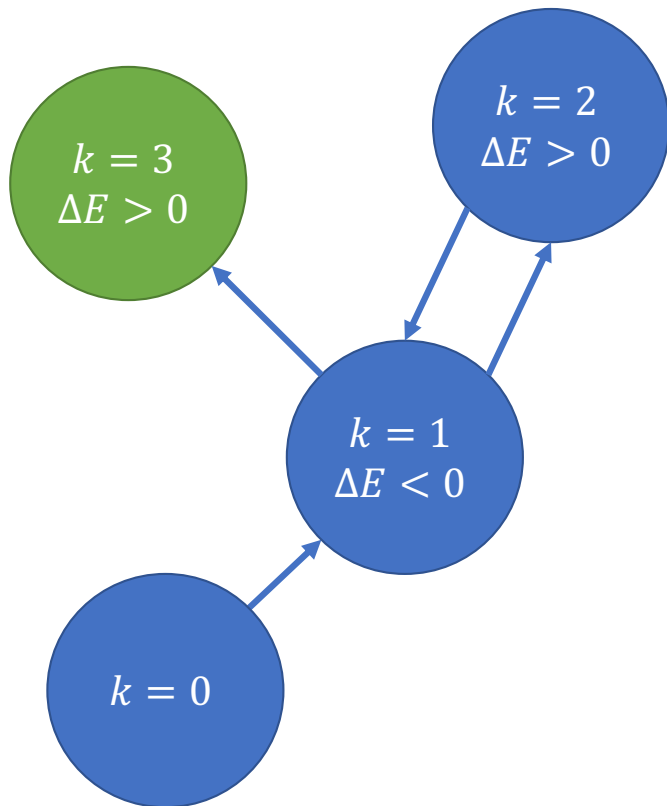
Only objective function evaluations are used to find an optimum; derivatives of the objective function are not needed

Candidates for GABAC optimization

Simulated
annealing

Genetic
algorithm

Simulated annealing



Compute $\Delta E(k)$ (i.e., the compression ratio gain)

`accept(k - 1)` // accept the old configuration

if $\Delta E(k) > 0$ // the new configuration is worse

// nevertheless accept the new solution on a random basis

if $P(k) = e^{-\Delta E/T(k)} \geq x$ // x is a random probability

`accept(k)`

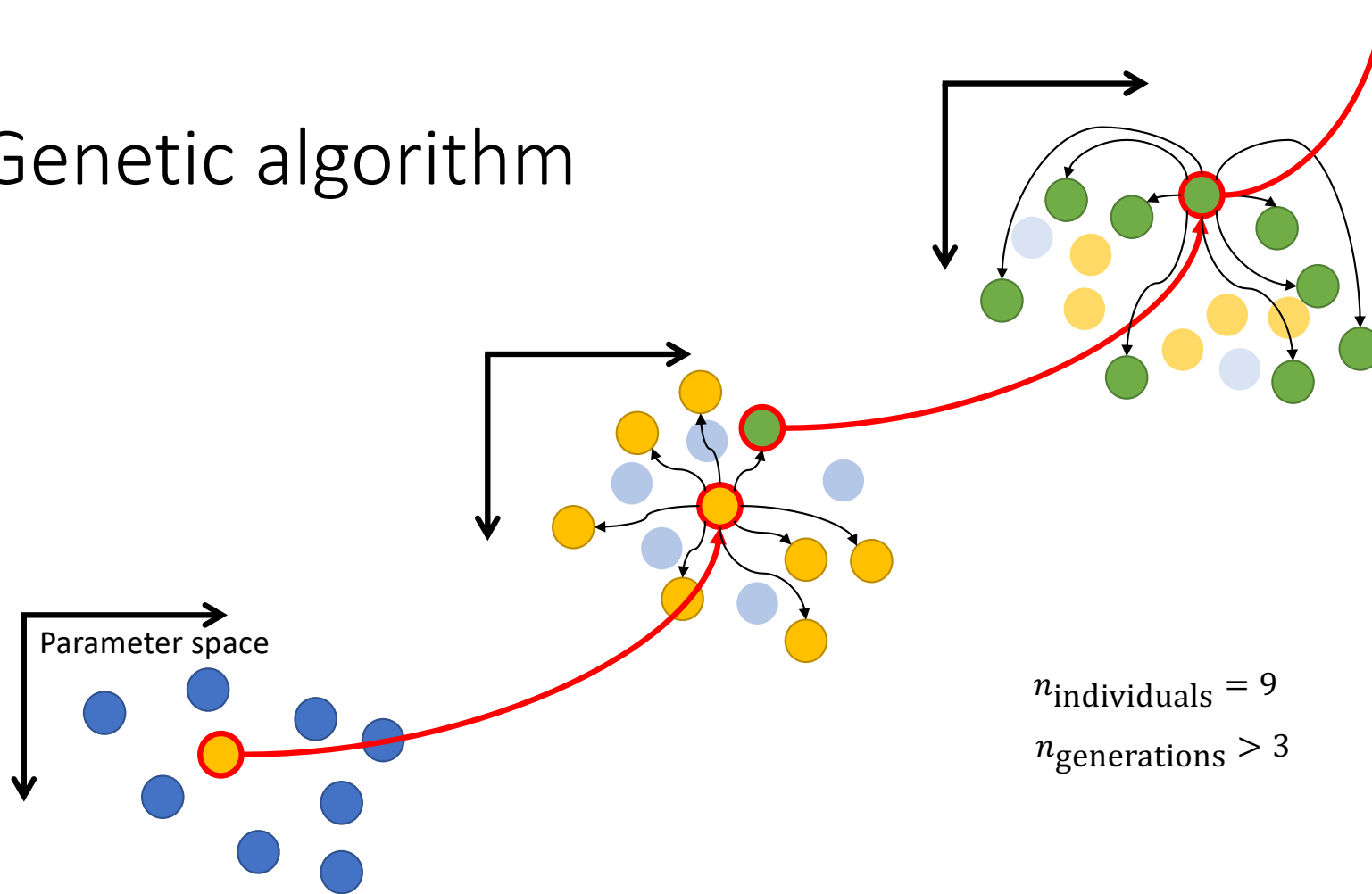
else // the new configuration is better

`accept(k)`

Simulated annealing

Parameter	Choice	Remarks
State space	GABAC parameter space	$\{p_{\text{trans}} \cdot p_{\text{diff}} \cdot p_{\text{lut}} \cdot p_{\text{bid}} \cdot p_{\text{ctx}}\}$
Energy (objective) function	Compression ratio r	$0 < r < \sim 1$
Candidate generation procedure	Random neighbor	1 random parameter is changed
Acceptance probability function	$\Delta E(k) = r(k) - r(k - 1)$ $P(k) = e^{-\Delta E / T(k)}$	“energy difference” = compression ratio gain
Annealing schedule	$T(k) = \left(1 - k / k_{\text{max}}\right) \cdot k_t$	k grows $\rightarrow T(k)$ decreases $\rightarrow P(k)$ decreases
Initialization	$T(k = 0) = 1$ $k = 0$ $k_{\text{max}} = 100$ $k_t = 1$	k : no. of iterations k_{max} : maximal no. of iterations k_t : hyper parameter k_t is large \rightarrow worse solutions are accepted more frequently

Genetic algorithm



$n_{\text{individuals}} = 9$
 $n_{\text{generations}} > 3$

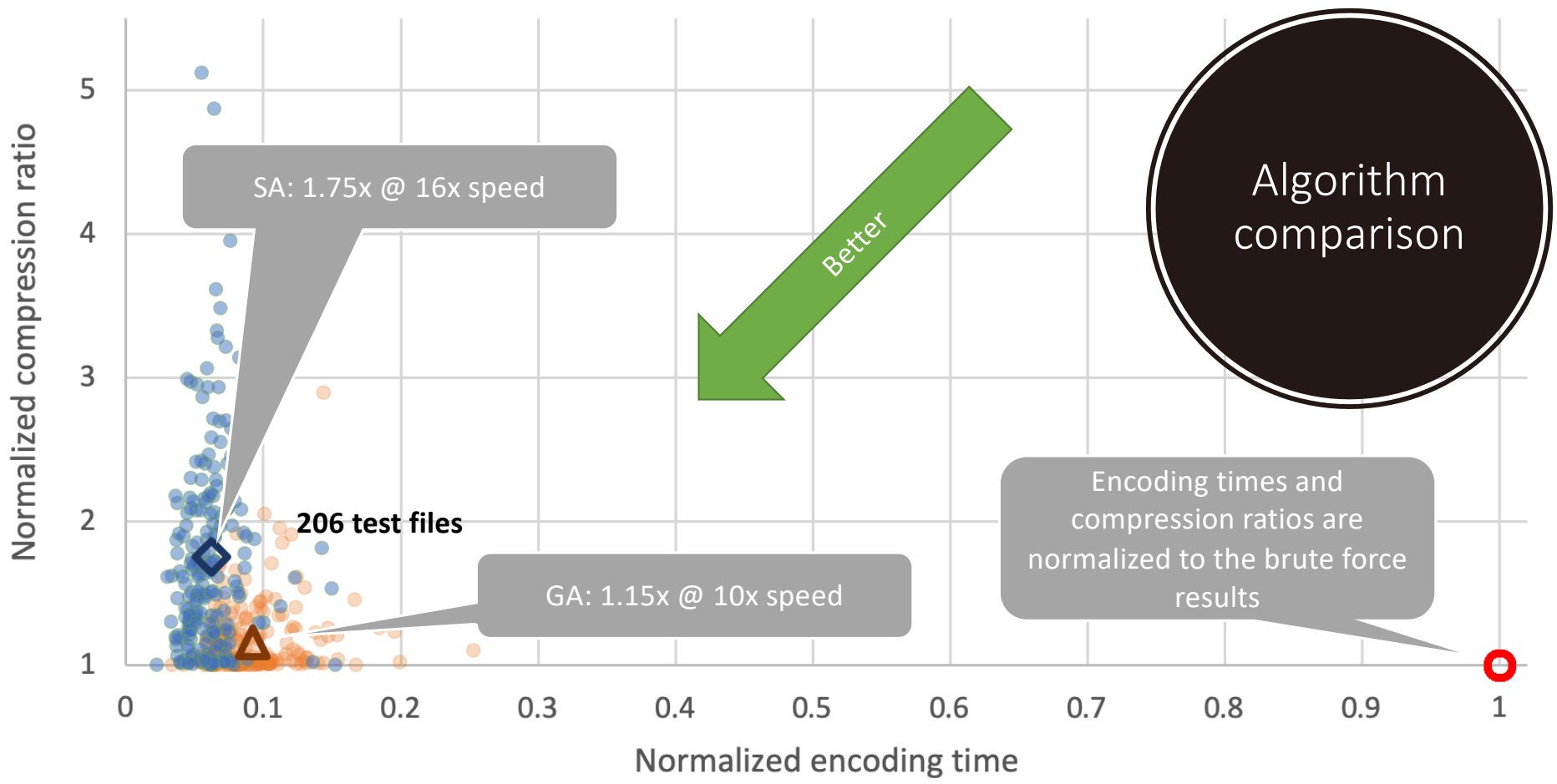
Genetic algorithm

Parameter	Choice	Remarks
State space	GABAC parameter space	$\{ p_{\text{trans}} \cdot p_{\text{diff}} \cdot p_{\text{lut}} \cdot p_{\text{bid}} \cdot p_{\text{ctx}} \}$
Objective function	Compression ratio r	$0 < r < \sim 1$
Candidate generation procedure	Random neighbor	2 random parameters are changed randomly
Acceptance probability function	$i_{\text{best}} = \underset{1 \leq i \leq n_{\text{individuals}}}{\text{argmax}} \{r(i)\}$	<ul style="list-style-type: none"> • Best individual in each generation is selected <ul style="list-style-type: none"> • No crossover • No mutation (replaced by random parameter change)
Initialization	$n_{\text{individuals}} = 10$ $n_{\text{generations}} = 10$	n/a

Results and sanity checks on artificial data

1 MiB random	Brute force	Genetic algorithm	Simulated annealing
Compression ratio	~1	~1	~1
No. of tested configurations	6,945	10 · 10	100
Encoding time	467 s	50 s	64 s

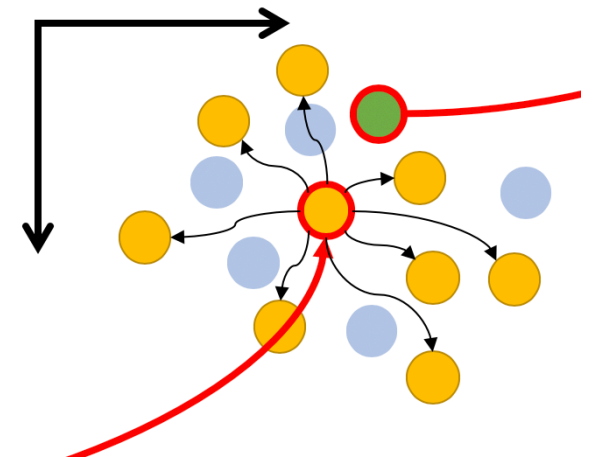
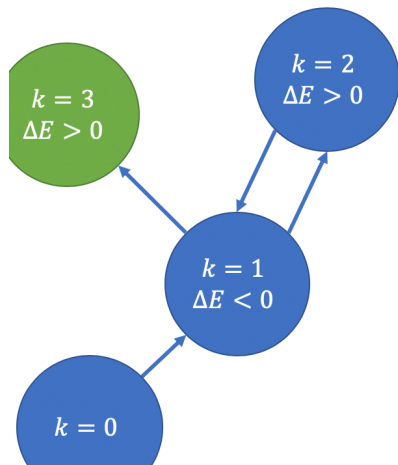
1 million 0x00	Brute force	Genetic algorithm	Simulated annealing
Compression ratio	~0.03	~0.03	~0.03
No. of tested configurations	8,481	10 · 10	100
Encoding time	41 s	6.8 s	8.6 s



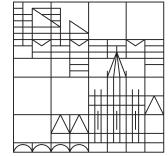
● Brute Force
 ● Genetic Algorithm (GA)
 ● Simulated Annealing (SA)
 ▲ GA Mean
 ◆ SA Mean

Conclusions

- Optimization of GABAC, an MPEG-G compliant entropy codec
- Optimization of the encoding process using a genetic algorithm increases the encoding speed by a factor of 10 at an average compression ratio of 1.15 compared to the brute force solutions



github.com/mitogen/gabac

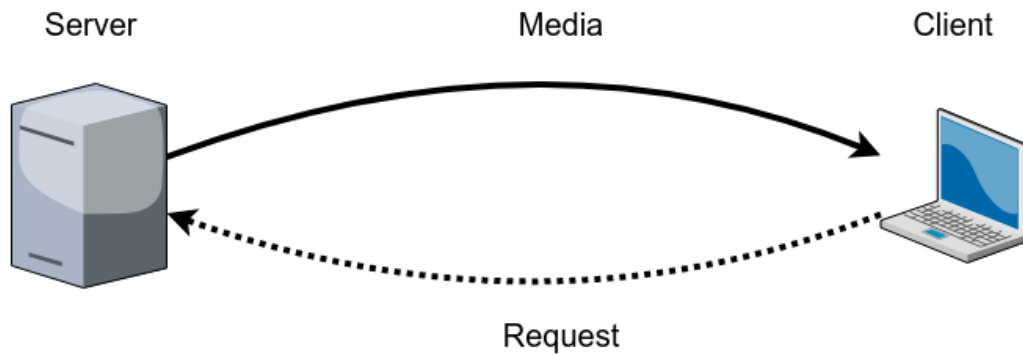


Foveated Video Coding for Real-Time Streaming Applications

Oliver Wiedemann

Universität Konstanz, 19.06.2019

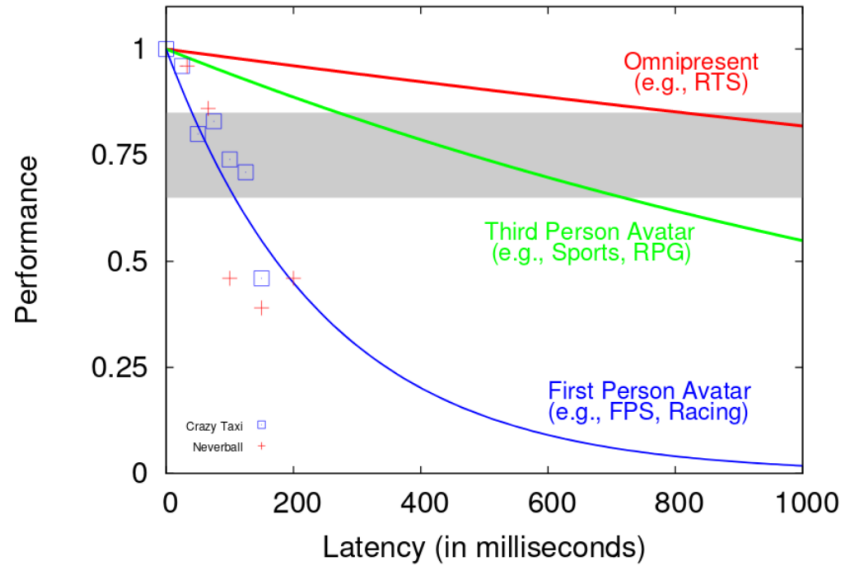
Introduction



Real-Time Video Streaming

- Prime example: cloud gaming
 - Offload the rendering / game engine to a server
 - Stream game graphics to a thin client
 - Play control feedback back to the server
- Causal video source
- Latency constraints

Latency Influence

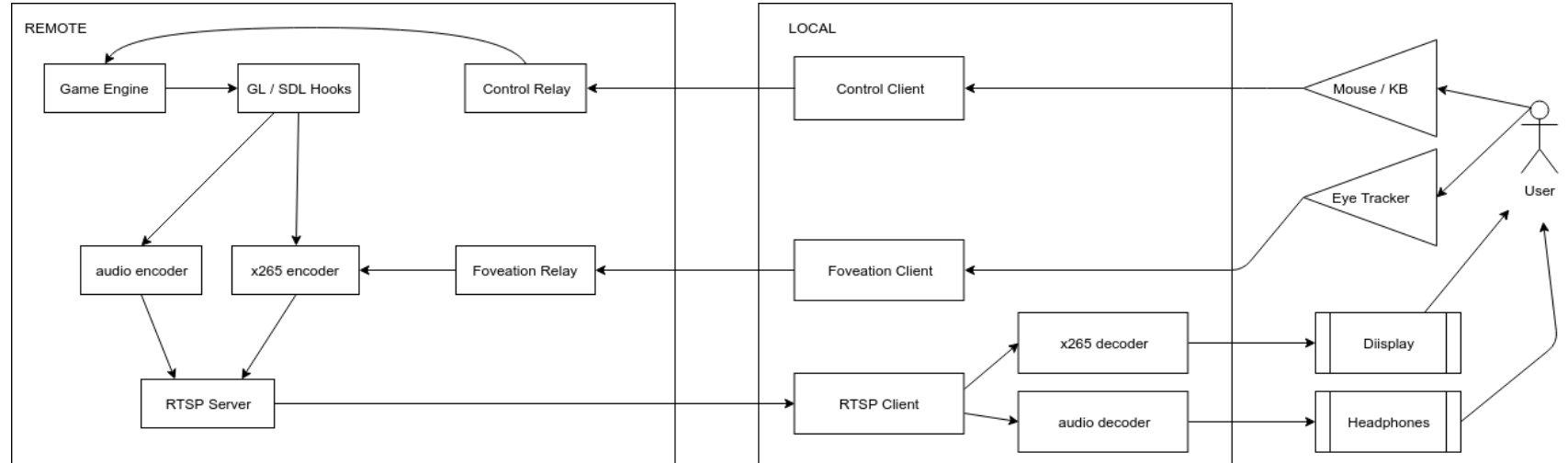


M. Claypool and D. Finkel: The Effects of Latency on Player Performance in Cloud-based Games

Encoder Choices and Parameterization

- Huang et. al: x264
 - Just one reference frame
 - Preset: At least `fast`
 - Tuning flag: `zerolatency`
 - no B-frames
 - no lookahead
- Alternative: x265
- Codecs are restricted to a basic level of operation.

Structure



Foveation

1. Send fixation coordinates (x, y) and viewer distance z
2. Server calculates an offset map
3. The next frames' macro-blocks are quantized accordingly

Macroblock offset according to a two dimensional Gaussian:

$$QO(i, j) = QO_{\max} \left(1 - \exp\left(\frac{(i-x)^2 + (j-y)^2}{2\sigma^2}\right) \right)$$

With parameters σ and QO_{\max}

Research Questions and Outlook

- How to quantify performance?
 - How to relate non-uniform quality and bitrate?
- How to choose optimal quantization parameters?
 - as a function of the network bandwidth?
- Sideproject: Try eye-tracking approximation using a notebook webcam

1. Motivation

Multi-image 3D reconstruction is a core task in machine vision and required in many applications, including augmented reality, geo-imaging, or artificial intelligence.

State of the art Multi-View Stereo:

- No explicit control over search space and resolution
- Input images assumed un-distorted to linear pin-hole camera model
- Algorithms yield polygonal meshes or sparse point clouds
- High computational complexity

Proposal:

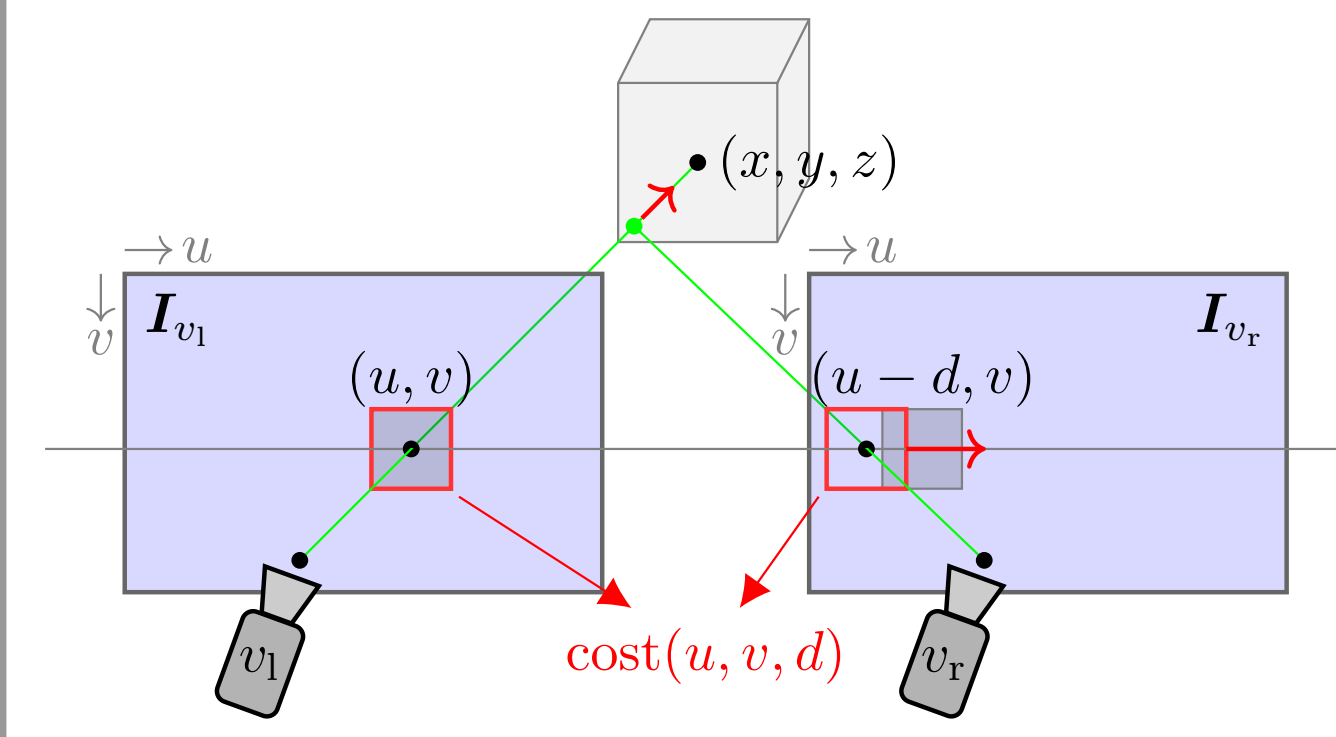
- Direct control over search space, scalable between resolution and computational effort
- Arbitrary, heterogeneous camera projection models supported
- Output: discrete 3D probability density of surfaces

2. Basics

Basic principle: find corresponding points in two or more perspectives

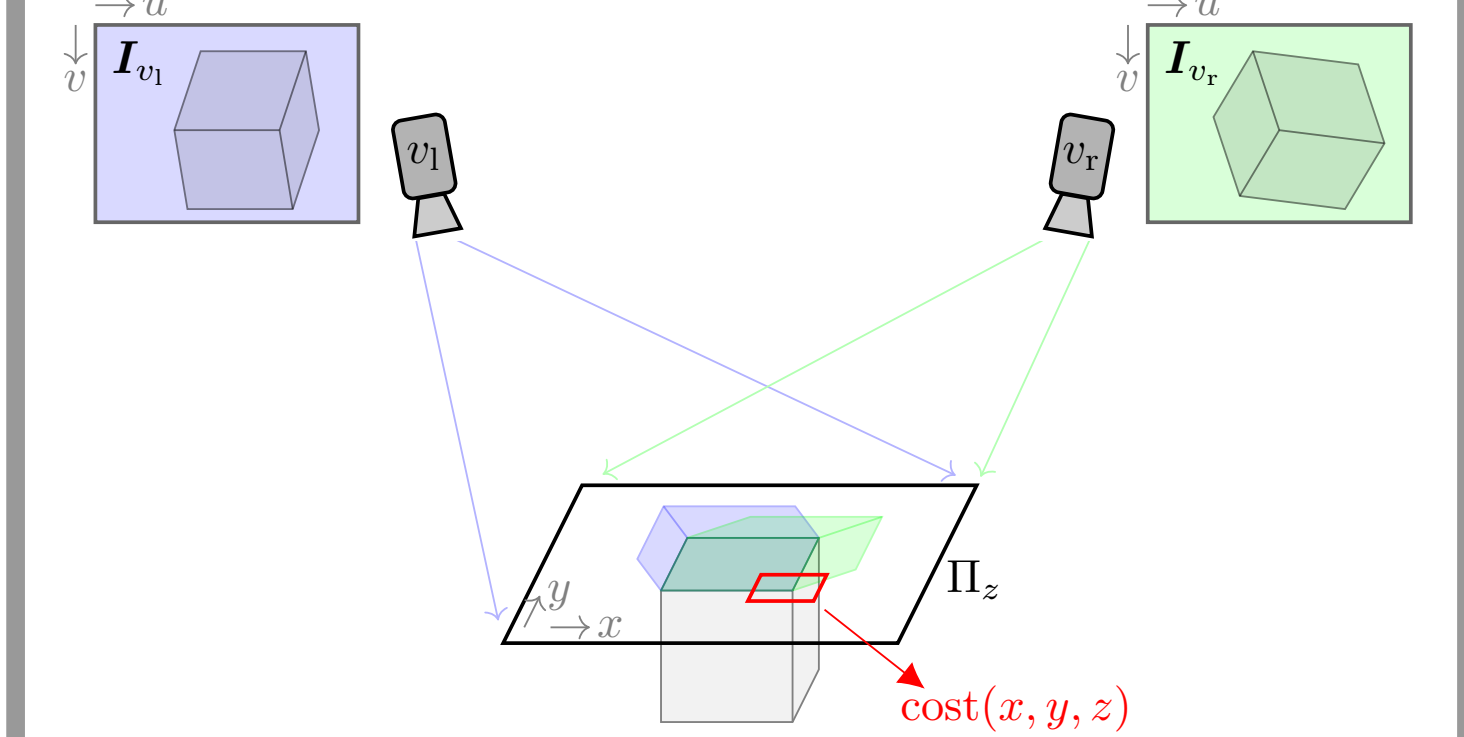
Traditional approach [1]:

- Evaluate photo-consistency along epipolar lines in rectified images
- Find optimum in disparity image space (U,V,D)



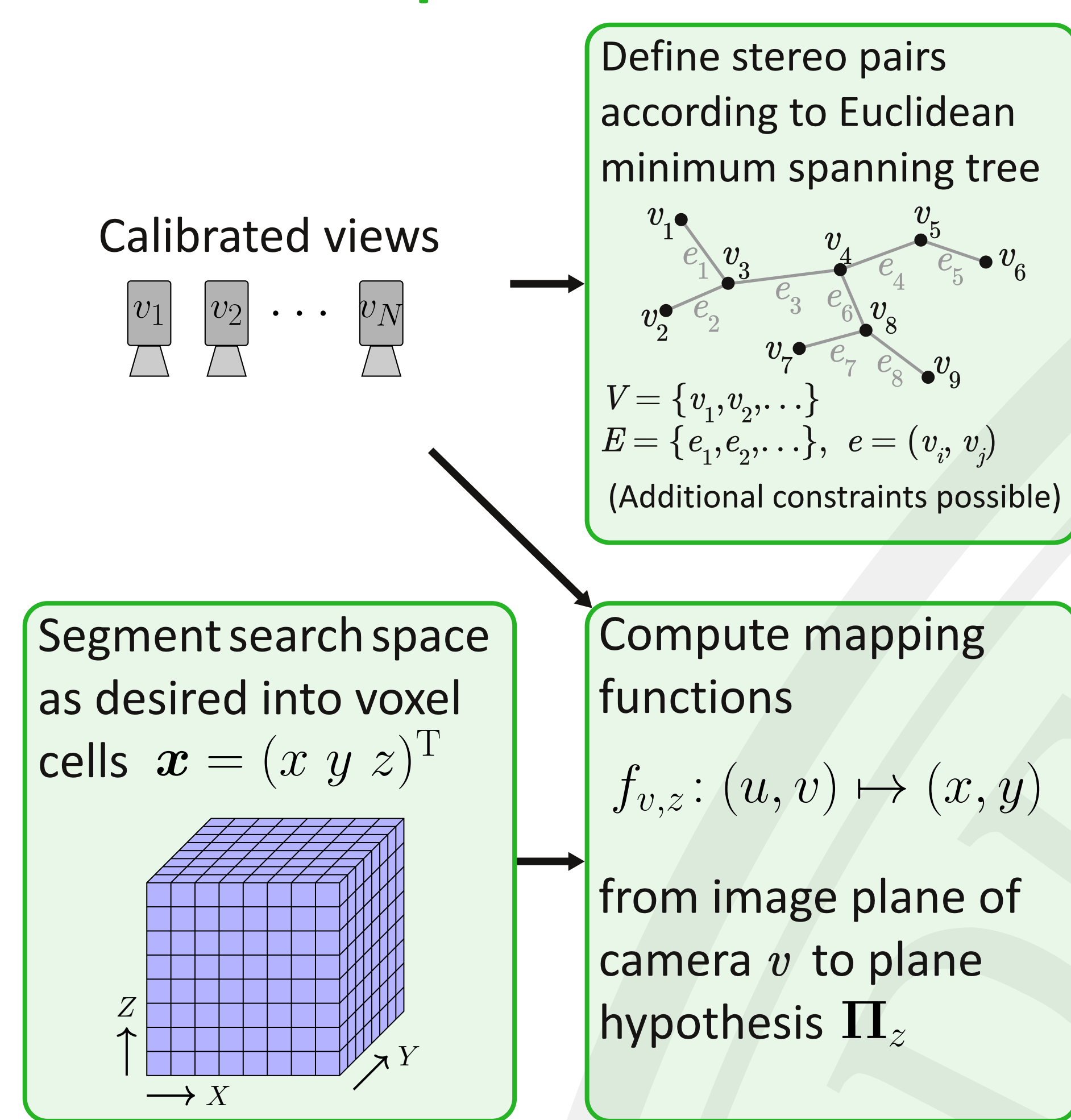
Plane sweep [2]:

- Re-project images onto plane hypotheses in object space (X,Y,Z)
- Evaluate local photo-consistency on re-projections

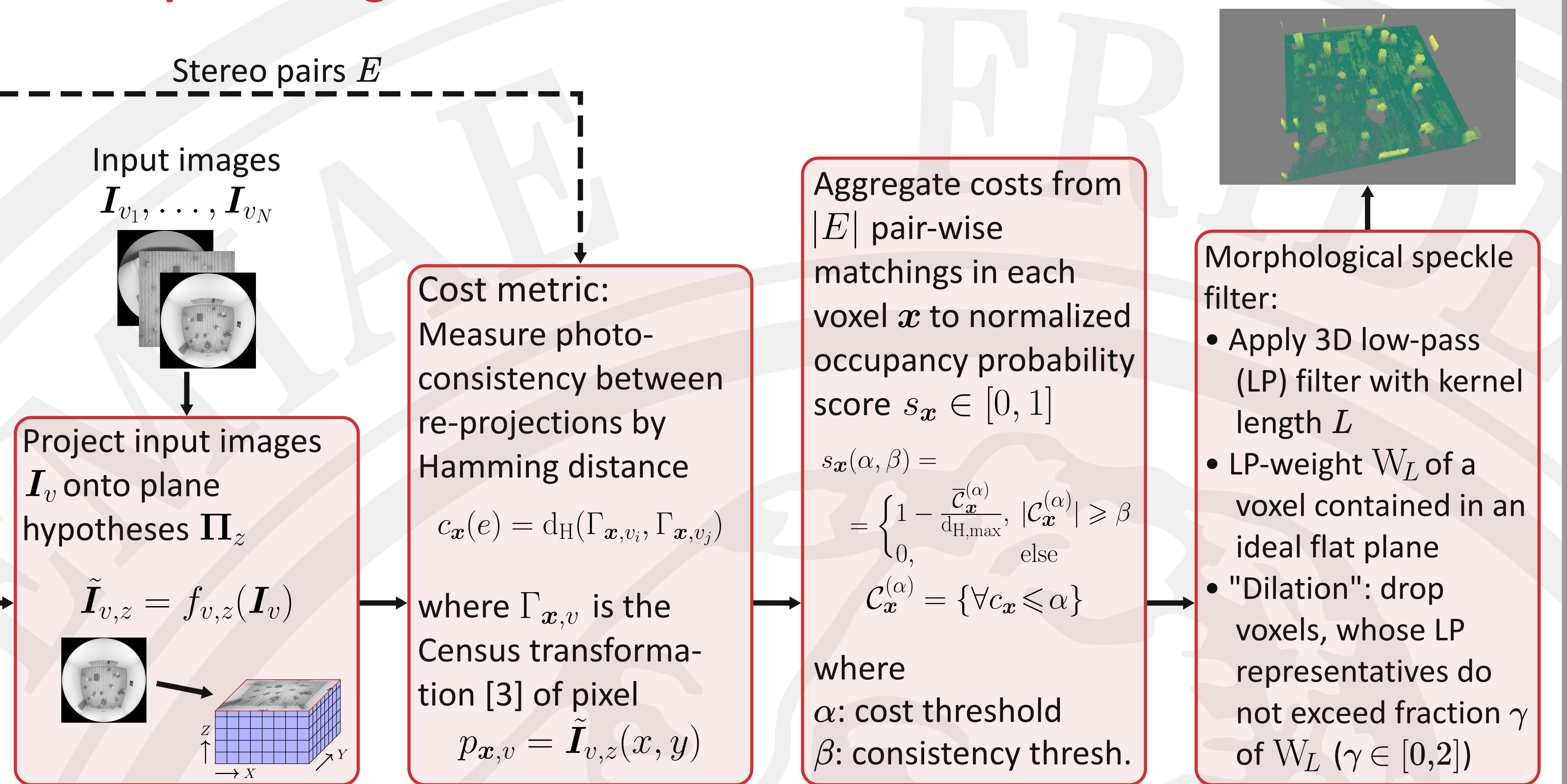


3. Proposed Processing Pipeline

Initialization phase



Online processing



4: Results

Test Scenario: 3D-modeled room, 16m x 16m

- Textured cuboids 1.4m ... 1.8m height
- Target resolution: 30 voxels/m
- Scan from 0 to 2m height, cameras at 7.5m
- Block size for Census transformation: 3x3
- Cameras: equisolid fish-eye projection
- 49 cameras roughly placed on a 7x7 grid at 2m distance

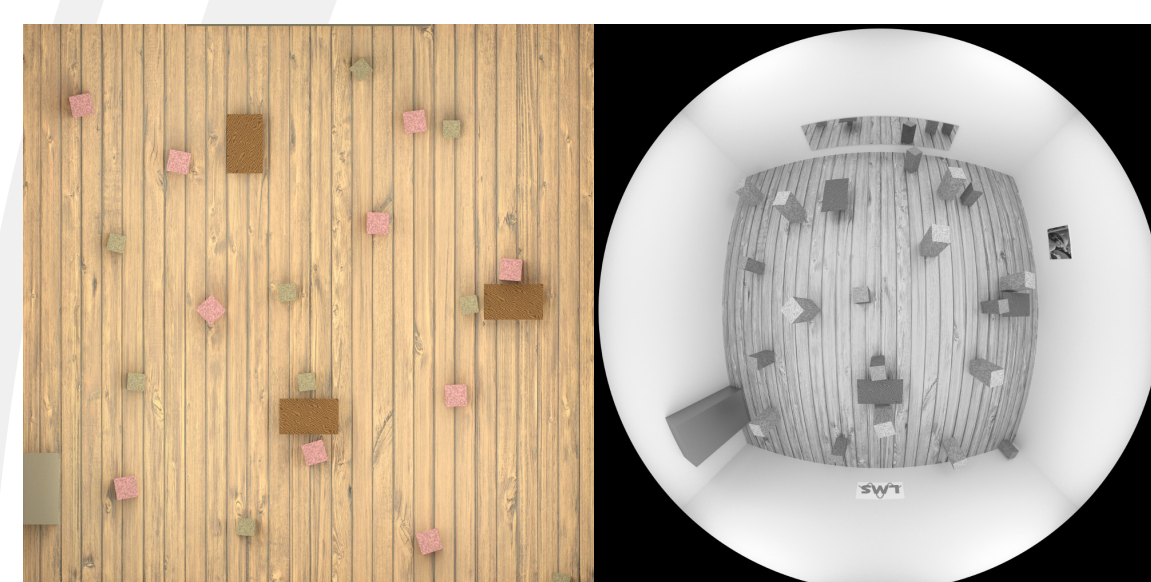


Figure: Test room outline and camera view example

Processing parameters: $\alpha=4, \beta=36, L=5, \gamma=0.6$

Run times of initialization + single frame

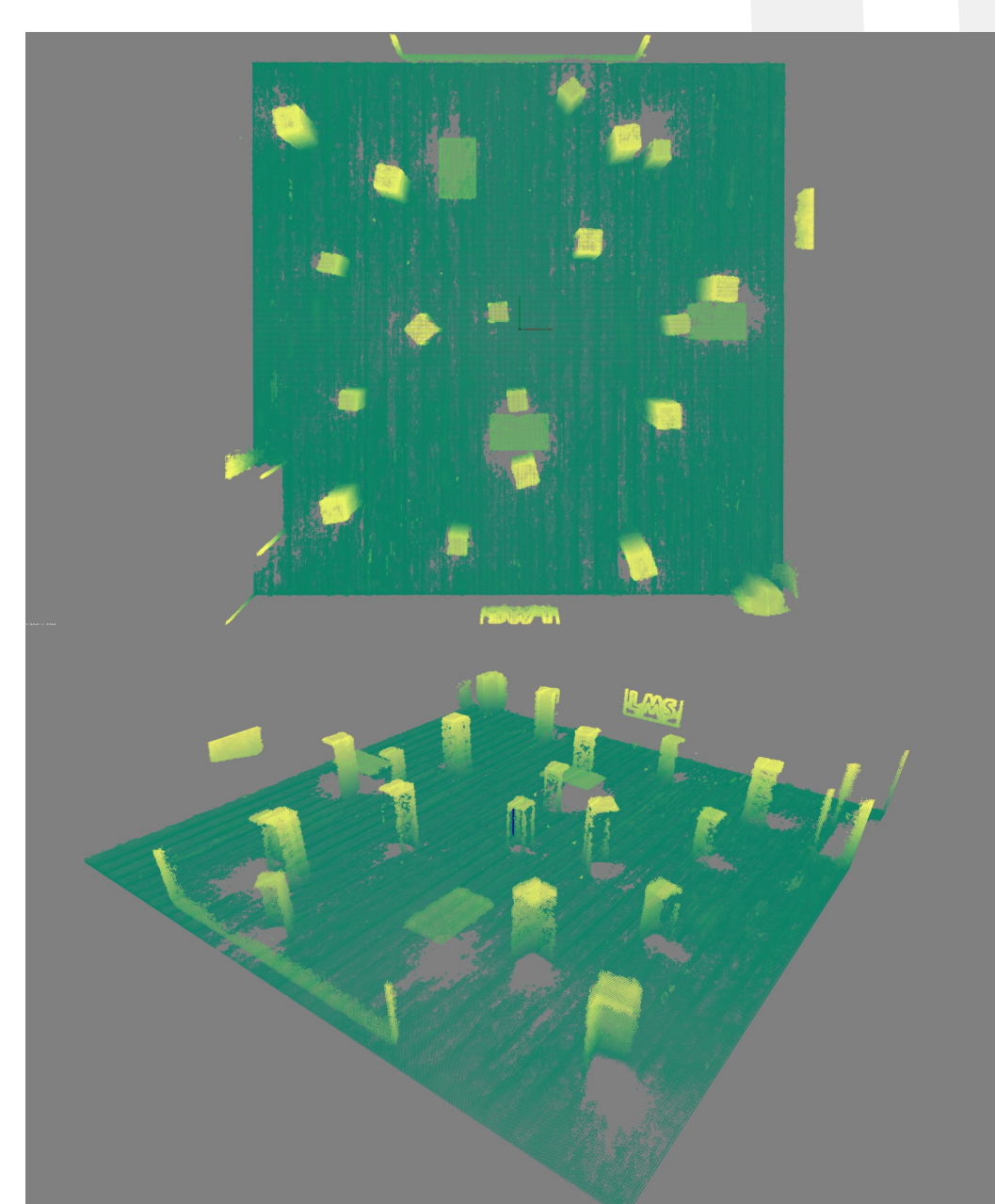
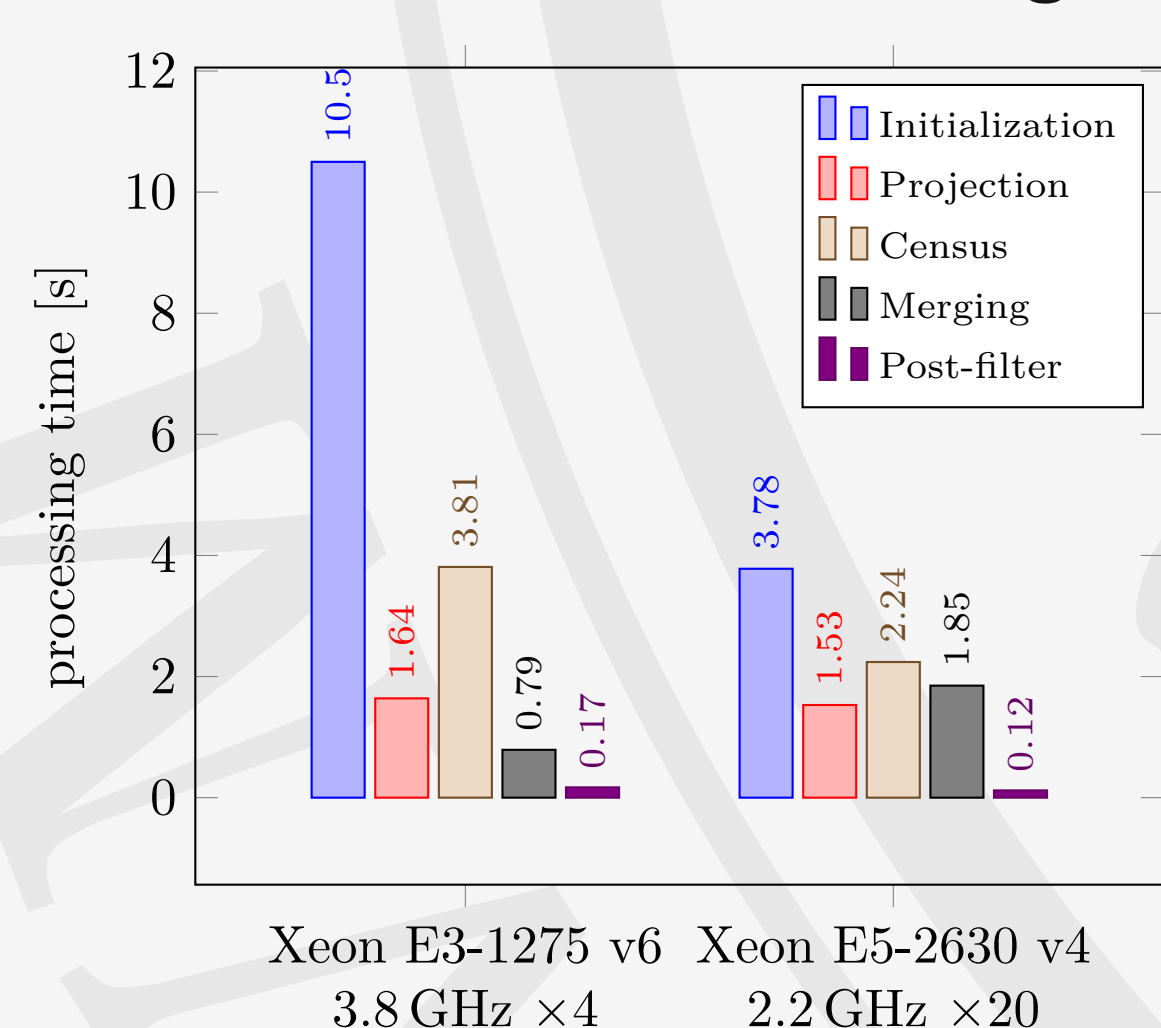


Figure: Final 3D structure

5. Conclusion

Plane sweep-based approach:

- Scalable in input size, moderate computation time at 49 input images and 30 voxels/m resolution
- Scalable search space, explicitly controllable
- Independent from used camera types
- Yields discrete 3D probability density of surfaces, defined on a voxel grid

References:

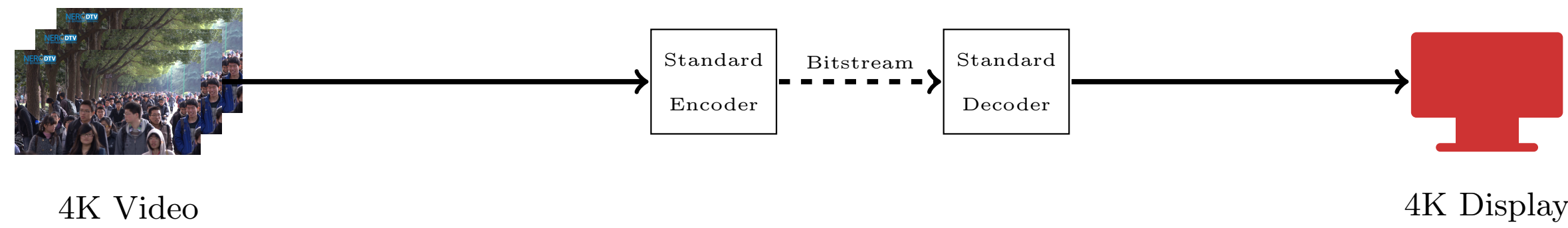
- [1] R. Hartley & A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.
- [2] R. T. Collins. "A Space-Sweep Approach to True Multi-Image Matching," *Proceedings CVPR 1996 IEEE*, San Francisco, CA, USA, pp. 358-363.
- [3] B. Fröba & A. Ernst, "Face Detection with the Modified Census Transform," *Proceedings FG 2004 IEEE*, Seoul, South Korea, pp. 91-96.

Acknowledgment:

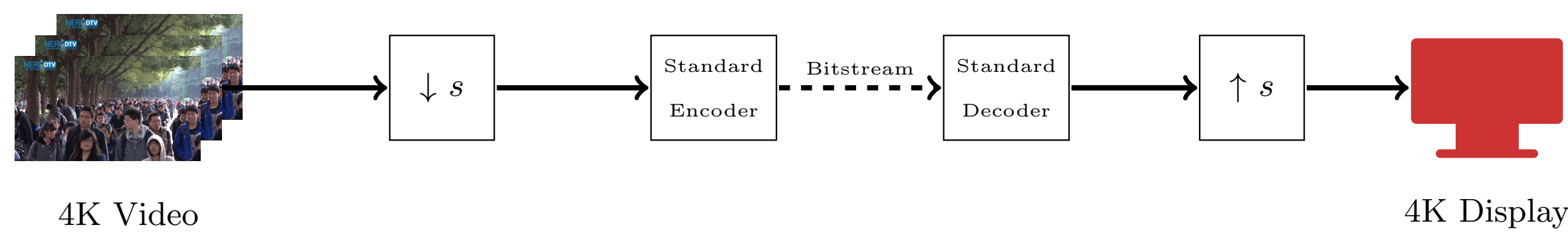
This work was funded by BOSCH Sicherheitssysteme, Nürnberg

1. Motivation

- Demand of high resolution videos has recently been increasing
- Hence, question arises how to transmit the massive amounts of high resolution data
- Conventional scenario with standard video codecs (Reference):



- Scenario with restrictions (e.g. limited bandwidth or time)



- Georgis et al. showed in [1] that there is a critical bandwidth up to which it is beneficial to encode video sequences with the second scenario that downscales the videos before encoding
- Video has to be upscalded at the decoder side
- In the following these results are verified by using the VP9 codec with several upscaling algorithms

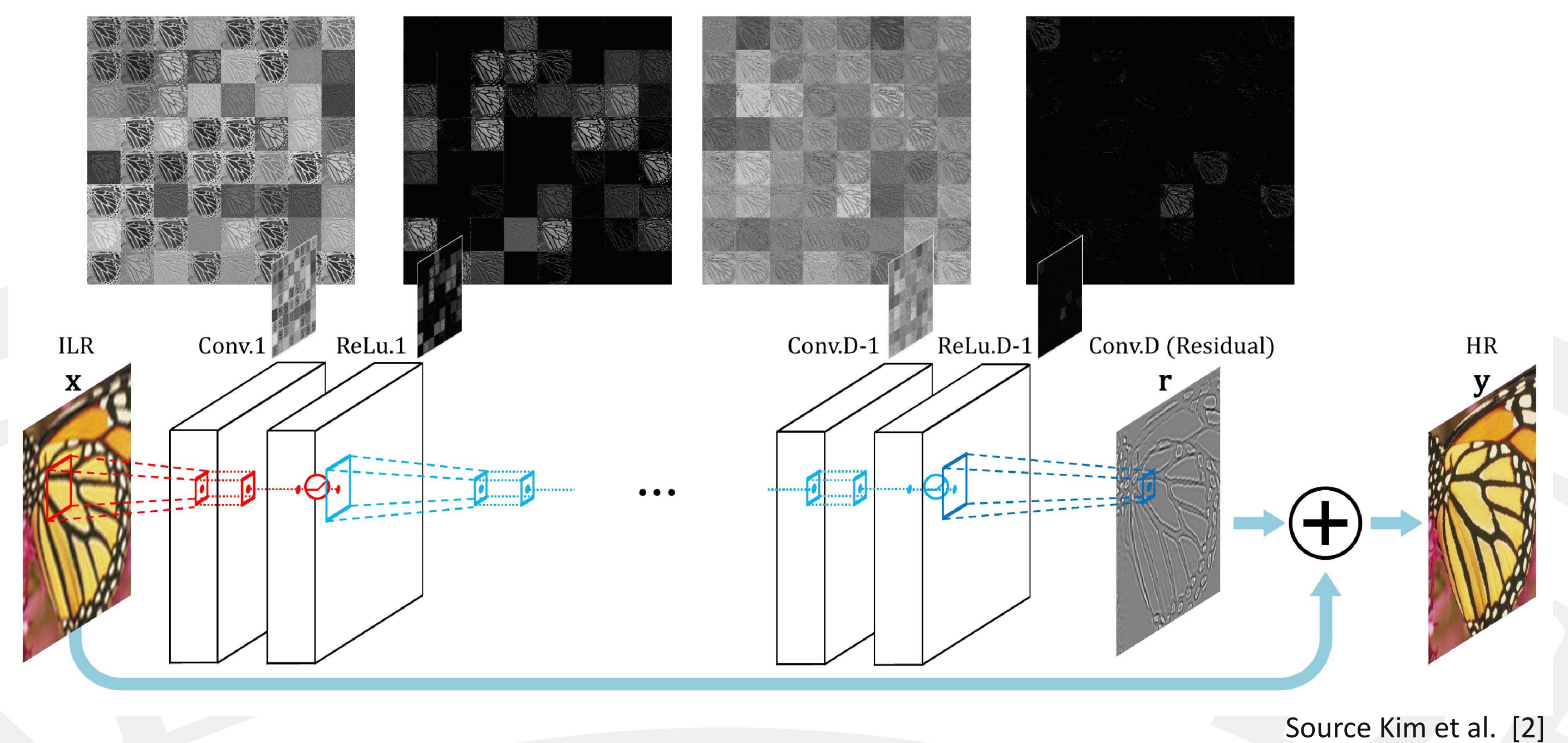
2. Evaluated Upscaling Algorithms

Low-complexity Statistical Edge-Adaptive Back-projected Interpolation (L-SEABI) [1]

- Additionally proposed upscaling algorithm by Georgis et al.
- Defines whether a sample belongs to edge area or not
- Edge areas are interpolated with bicubic, non-edge areas with bilinear interpolation
- Iterative refinement phase with adaptive back-projecting

Very-Deep Super-Resolution (VDSR) [2]

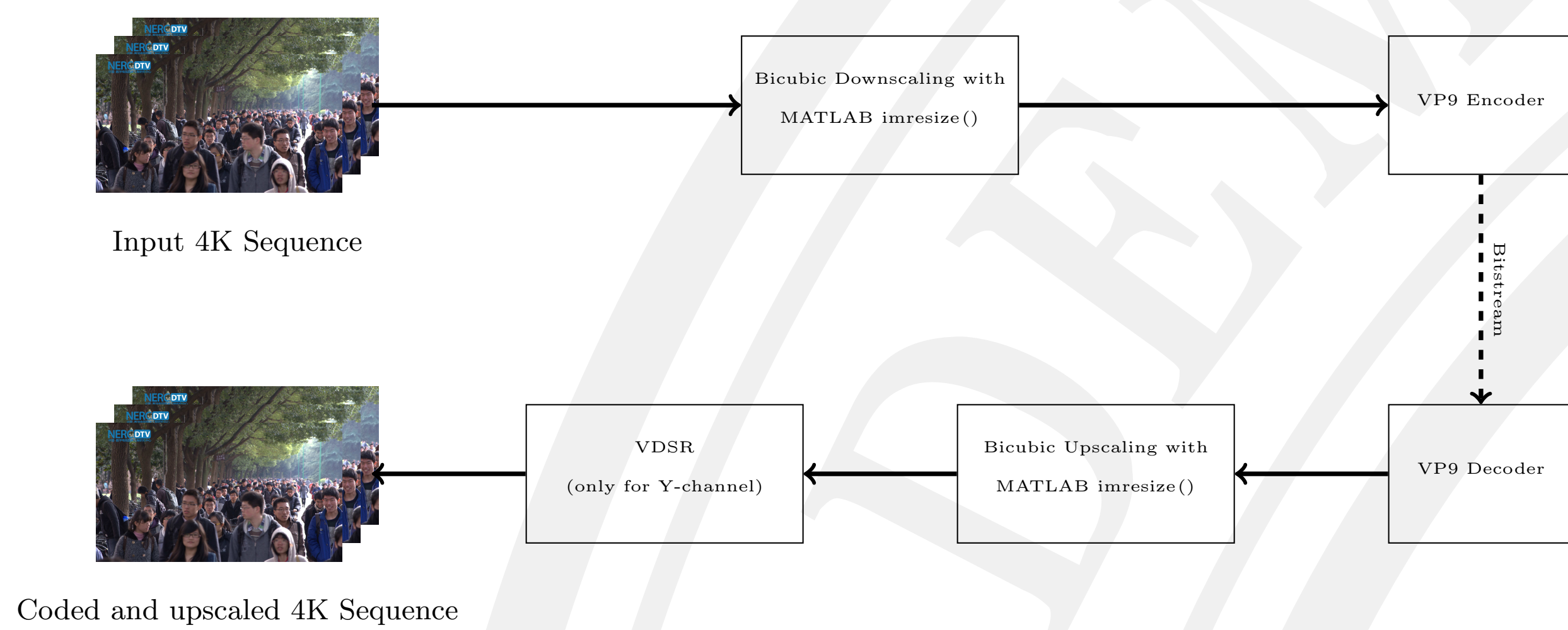
- Single image super-resolution network proposed by Kim et al.
- General Idea: Enhance the quality of bicubically interpolated image



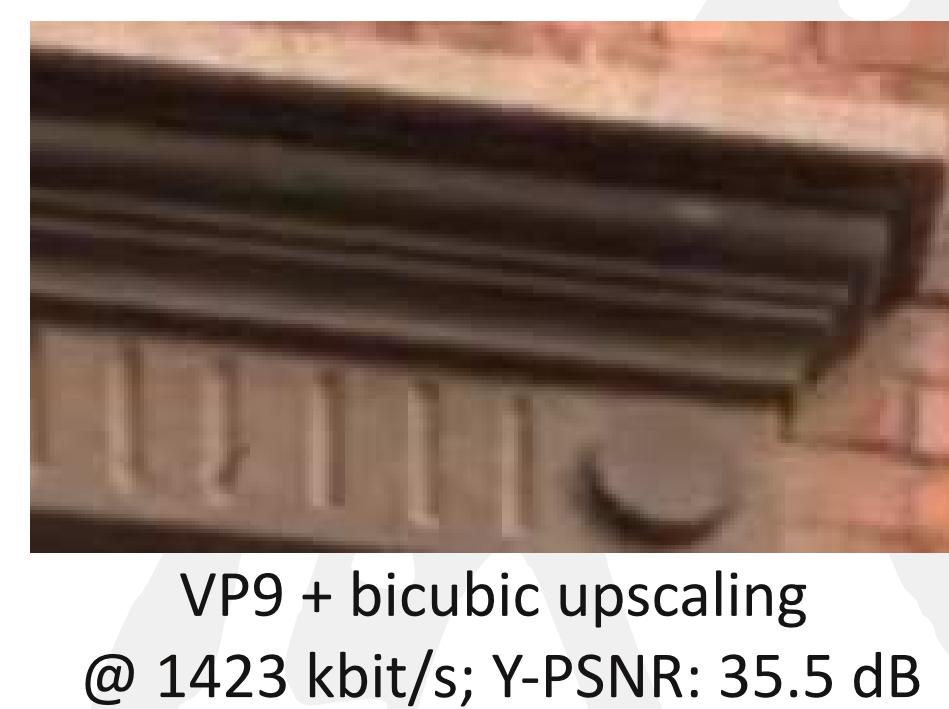
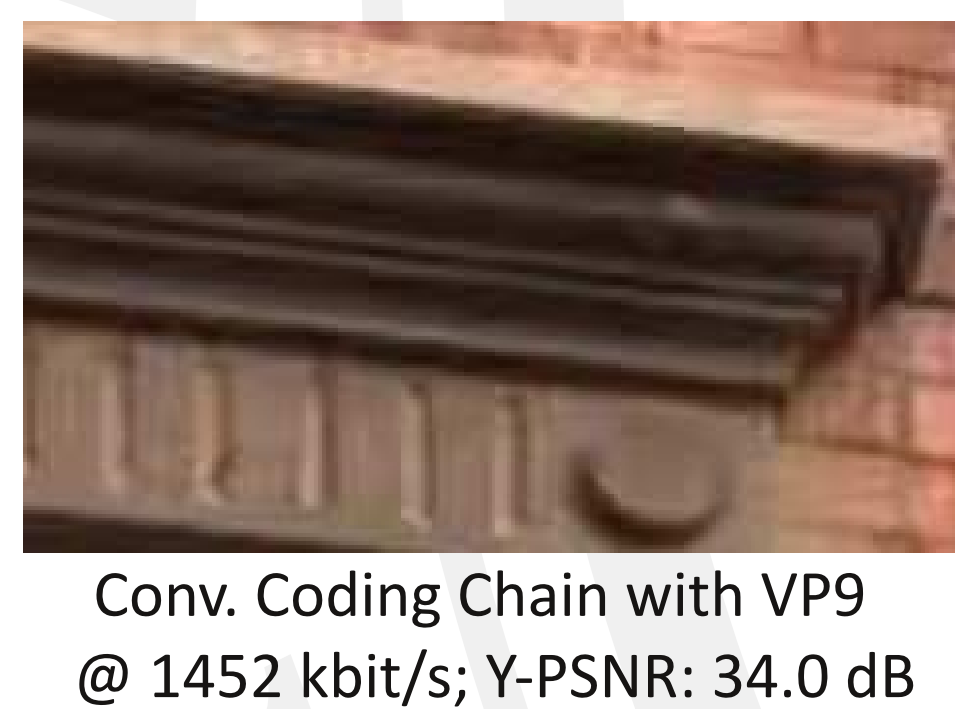
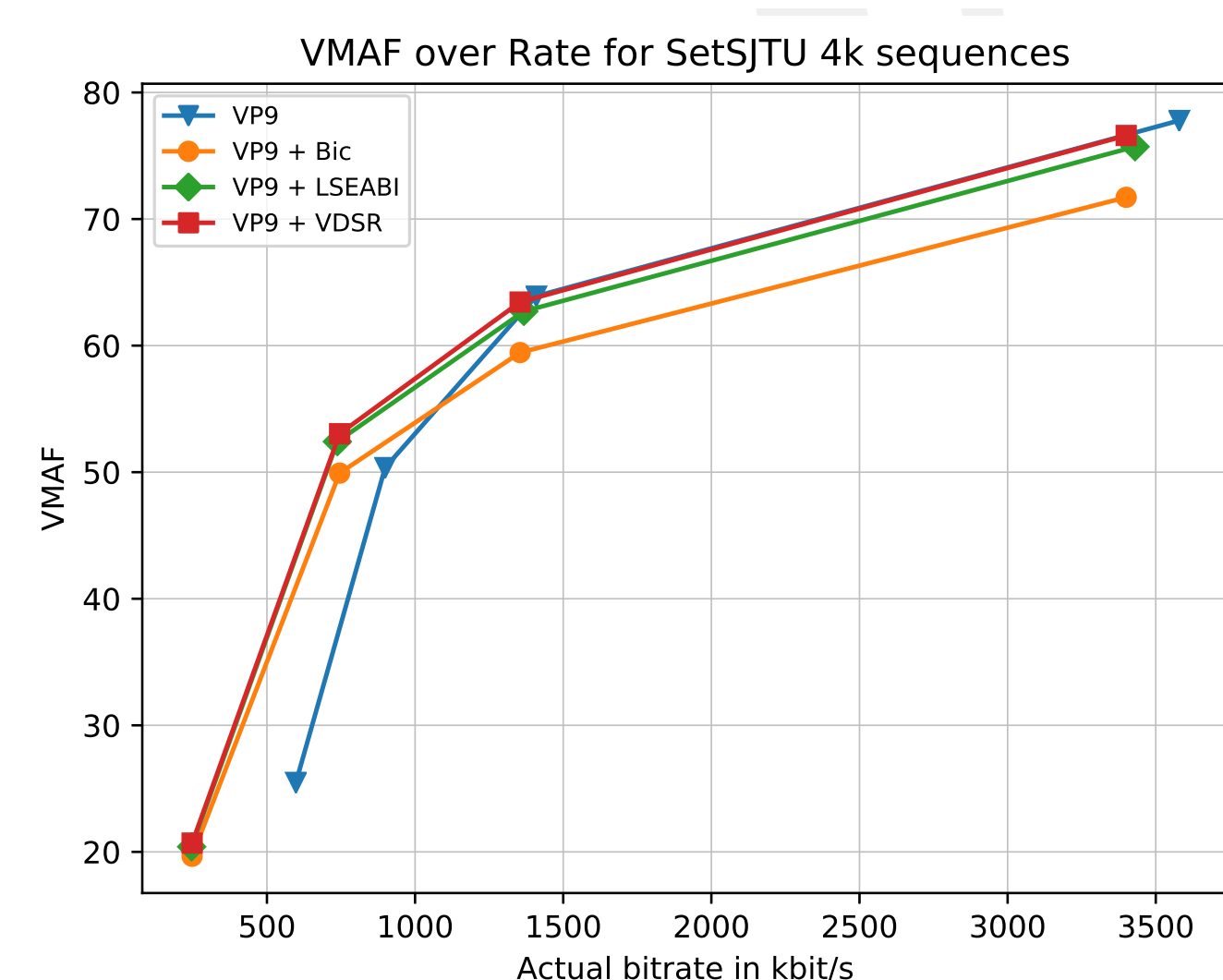
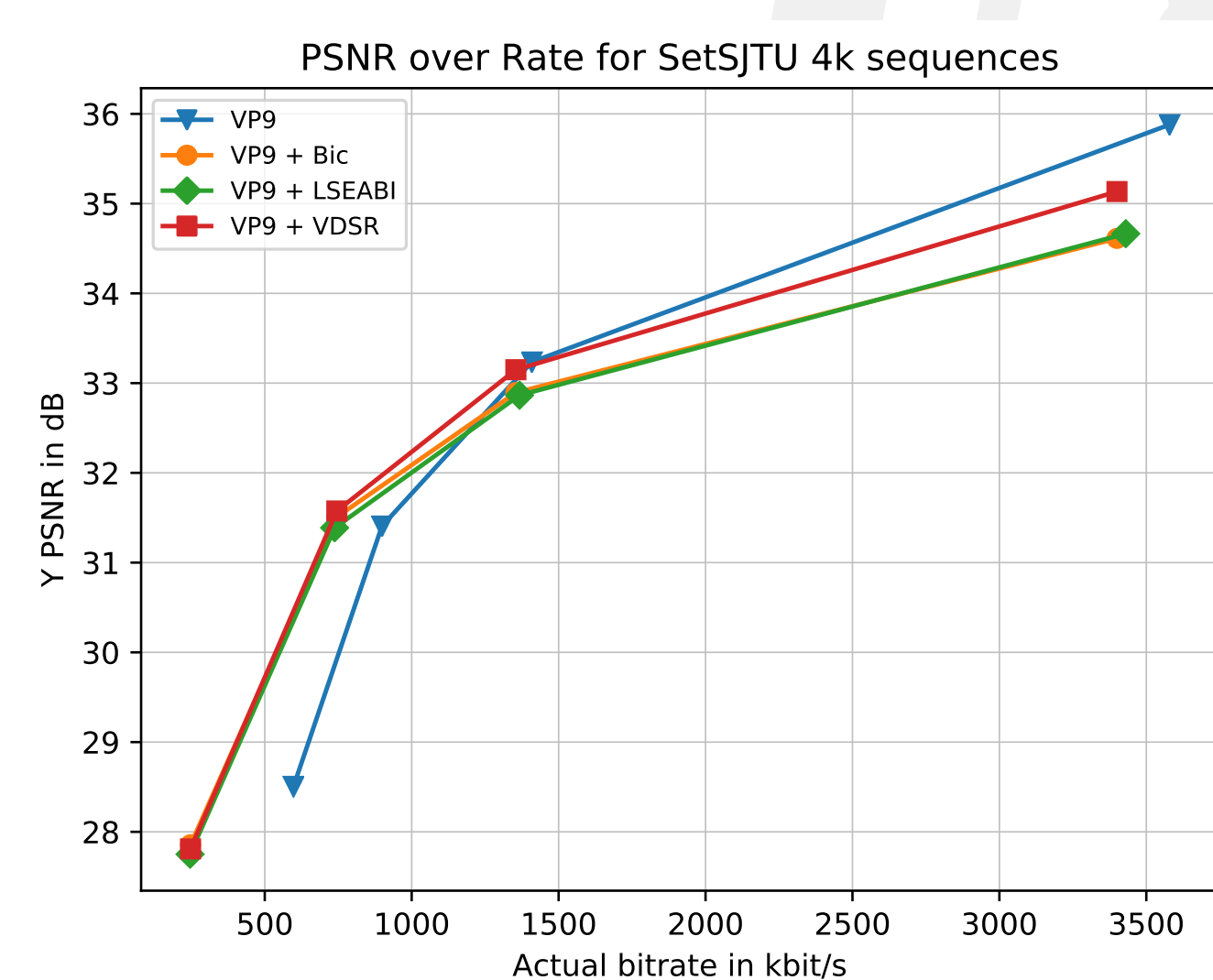
- Network learns difference between bicubically interpolated image and original image
- 20 convolution layers (3x3 filters) with 64 feature maps, each
- Subsequent Rectified Linear Unit (ReLU) layer for activation and to induce non-linearities
- Advantages:
 - VDSR is trained end-to-end on residual data
 - Faster convergence of training process
 - Can be trained for multiple scaling factors
 - Deep network has a large receptive field (41x41 pixels) and more ReLU layers

3. Simulation Framework

- First 100 frames of 16 sequences from set SJTU in 4k resolution [3]
- Libvpx VP9 codec [4], v 1.7.0 settings:
 - Constant bitrate mode (between 100 to 3000 kbit/s)
 - 1 pass coding; --good --cpu-used=5
- VDSR with weights from reference implementation
- LSEABI from reference implementation



4. Simulation Results



Conclusion

- Coding chain with spatial down-/ up-scaling is beneficial up to a certain bitrate for all upscaling algorithms
- Compression artifacts can be reduced through spatial down-/ upscaling for low bitrates
- Bicubic interpolation leads to blurry images
- VP9 + VDSR with best upscaling quality and superior quality up to 1400 kbit/s

References

- [1] G. Georgis, G. Lentaris, and D. Reisis, "Reduced complexity super-resolution for low-bitrate video compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 2, Feb 2016, pp. 332–345.
- [2] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 1646–1654.
- [3] L. Song, X. Tang, W. Zhang, X. Yang, and P. Xia, "The SJTU 4K video sequence dataset," in *Proc. Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, July 2013, pp. 34–35.
- [4] The WebM Project, "WebM: an open web media project," [Online]. Available: <https://www.webmproject.org/>

Michael Gatzen, Christian Rohlfing, Jens-Rainer Ohm
 Institut für Nachrichtentechnik, RWTH Aachen University

Introduction

DNA: Two-stranded molecule (chromosome).
 Different bases along each strand:
Adenine, Cytosine, Guanine, and Thymine

Applications of genomic data:

- Medical: Personalized cancer treatment, genetic diseases
- Non-medical: Genetic ancestry testing

Human genome: 50 GB – 2 TB in size:

- Problematic for storage and transmission
 ⇒ Compression inevitable
- **Need for specialized lossless compression algorithms**

Proposed methods exploit statistics for efficient reference-free coding of base sequences

Preliminary Signal Analysis

- Discrete signal in alphabet $\mathcal{A} = \{A, C, G, T\} \Rightarrow$ trivial binary representation with 2 bit per base (bpb)
- Markov Property: Given the current letter $X \in \mathcal{A}$ consider n past symbols $Y \in \mathcal{A}^n$

Model

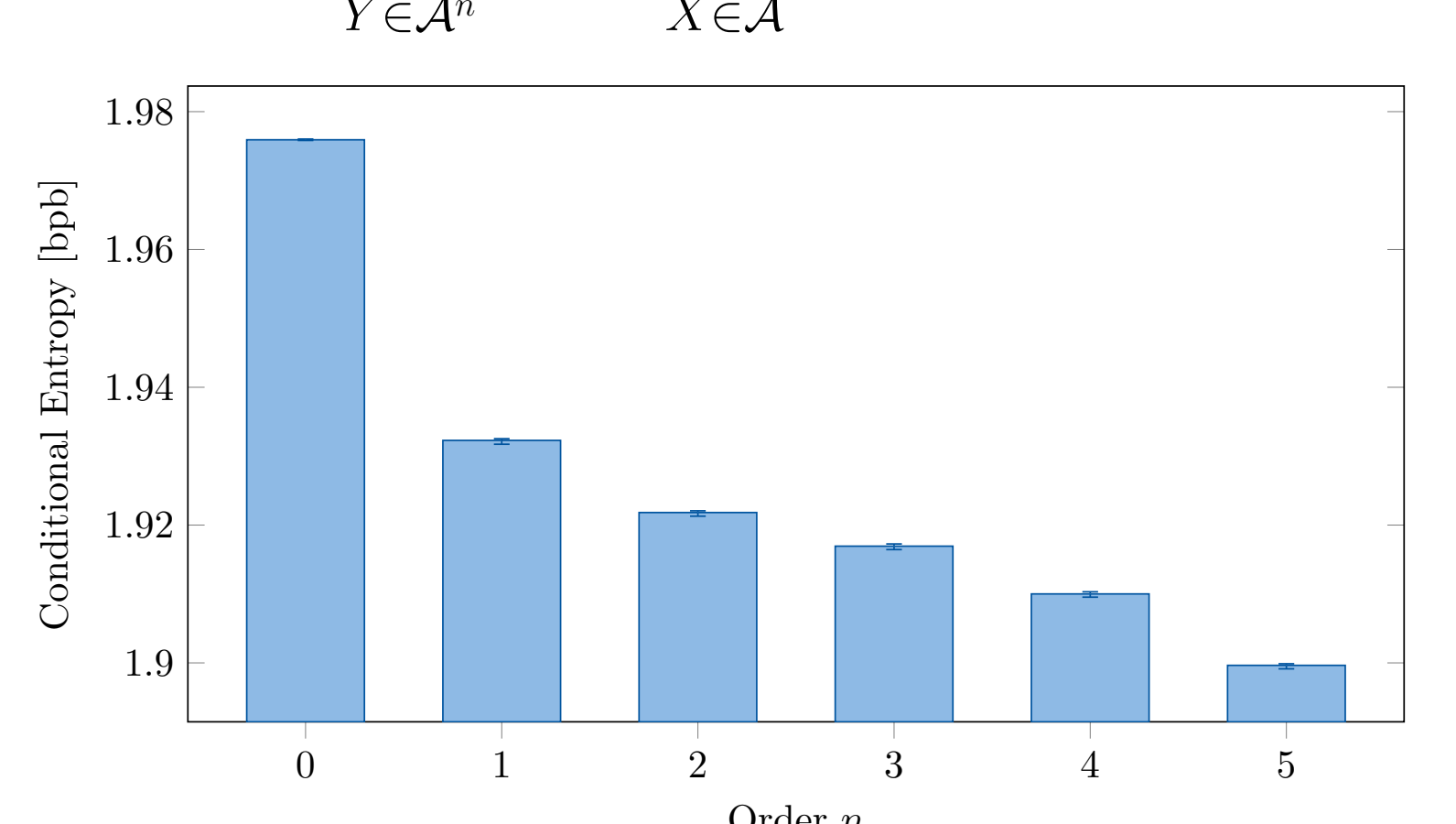
... A **C** **T** **G** **G** T ...
 ... X_{i-2} X_{i-1} X_i

Construct conditional probability matrix

$$P = [P(X | Y)]$$

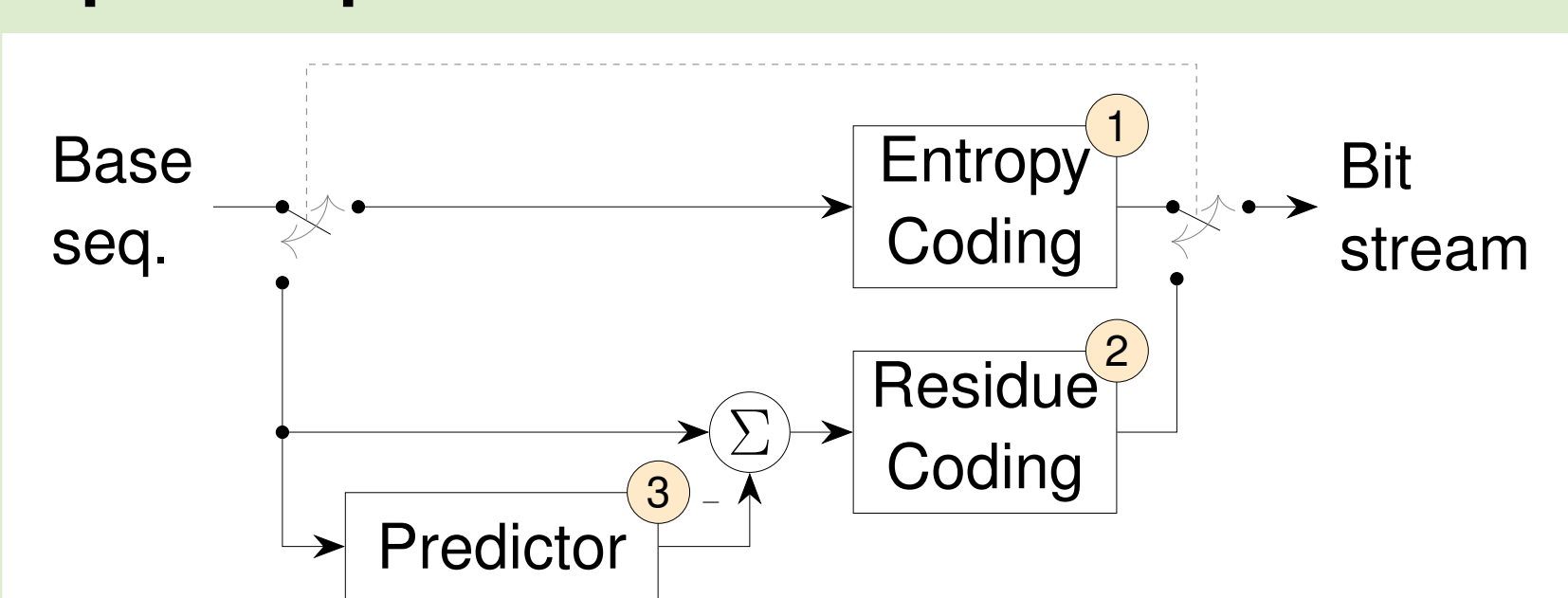
$$P = \begin{bmatrix} P(A|A) & P(C|A) & P(G|A) & P(T|A) \\ P(A|C) & P(C|C) & P(G|C) & P(T|C) \\ P(A|G) & P(C|G) & P(G|G) & P(T|G) \\ P(A|T) & P(C|T) & P(G|T) & P(T|T) \end{bmatrix} \quad (1)$$

Analysis

$$H(X|Y) = \sum_{Y \in \mathcal{A}^n} P(Y) \sum_{X \in \mathcal{A}} P(X|Y) \log_2 P(X|Y)$$


Predictive Coding

Open-loop Prediction



Entropy Coding ①

Using Context-Adaptive Arithmetic Encoder:

- Encoder keeps track of symbol occurrences and calculates probabilities on the fly
- For high probabilities, fewer bits are required

Context Model

Modeled probabilities: $P(X | X_{i-1}, \dots, X_{i-n})$

Residue Coding ②

G A C T ...
 - G T C G ...
 Residue?

- Use arithmetic coder
- Use predicted value of each position as context model

Example

Prediction	G	A	C	T	...
Actual value	G	T	C	G	...
Context	$P(G G)$	$P(T A)$	$P(C C)$	$P(G T)$...

Predictors ③

Similar to [1]. Mostly used predictors:

- *Direct* encoding (↪ skip prediction)
- *Repeat*: Copy previous block
 - No transmission of extra data
 - Many prediction errors
- *Cache* consisting of previous blocks
 - Entry with minimal Hamming distance to actual block selected
 - Fewer errors
 - Transmission of index required

Last Block: [AGCTGACAC]

Current Block: [CGATACCTA]

Residue: [AGCTGACAC]

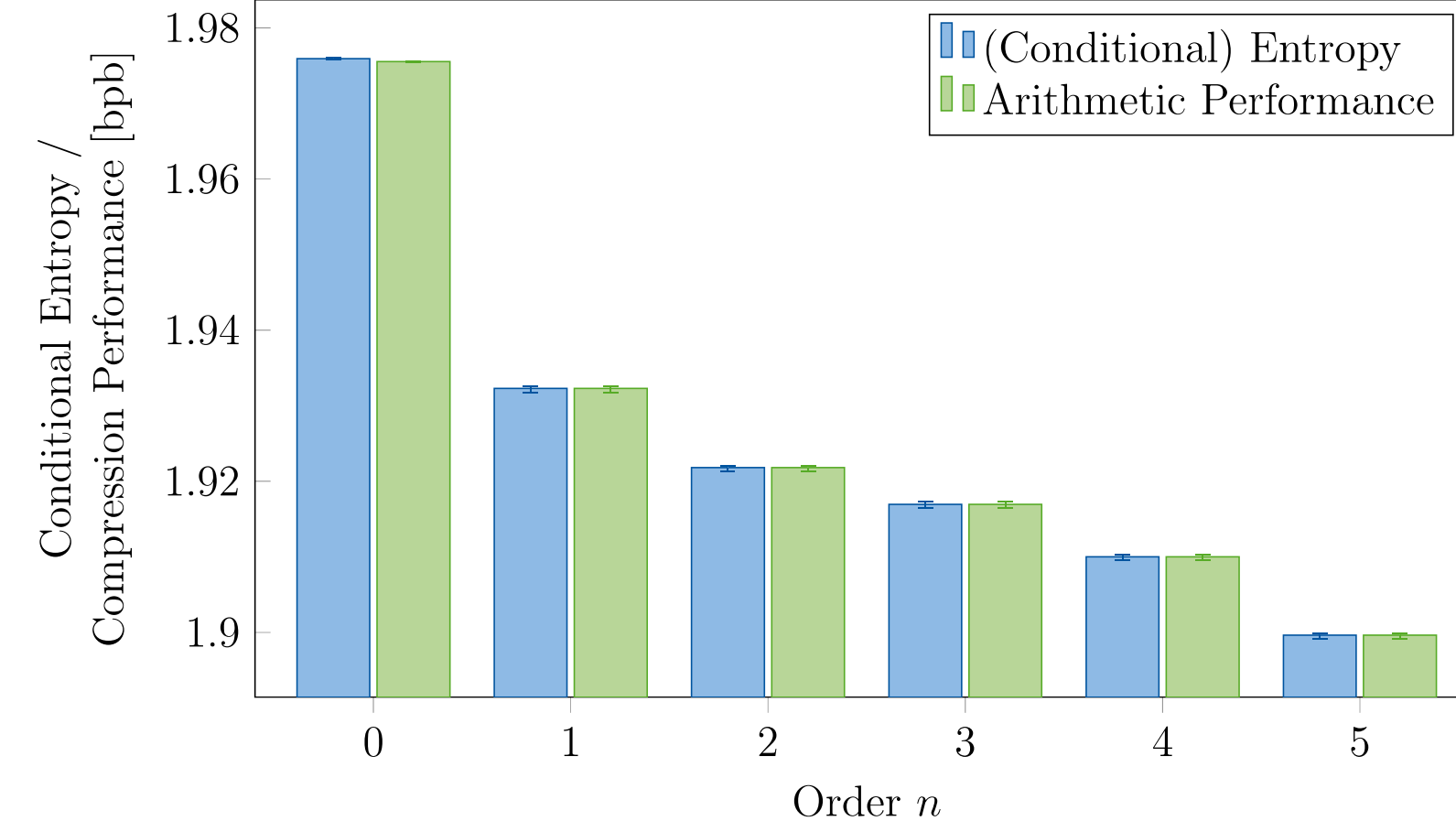
Cache:

- [AGCTGACAC]
- [CCATAGCAA]
- [GACGATCTC]
- ...
- [TACTAGCAA]

Current Block: [CGATACCTA]

Residue: [CCATAGCAA]

Preliminary Results



- Direct coding of the signal
- **Only very local patterns can be considered:**
 44% of genome: larger-scale repetitive regions
 ⇒ Accounting for large patterns necessary

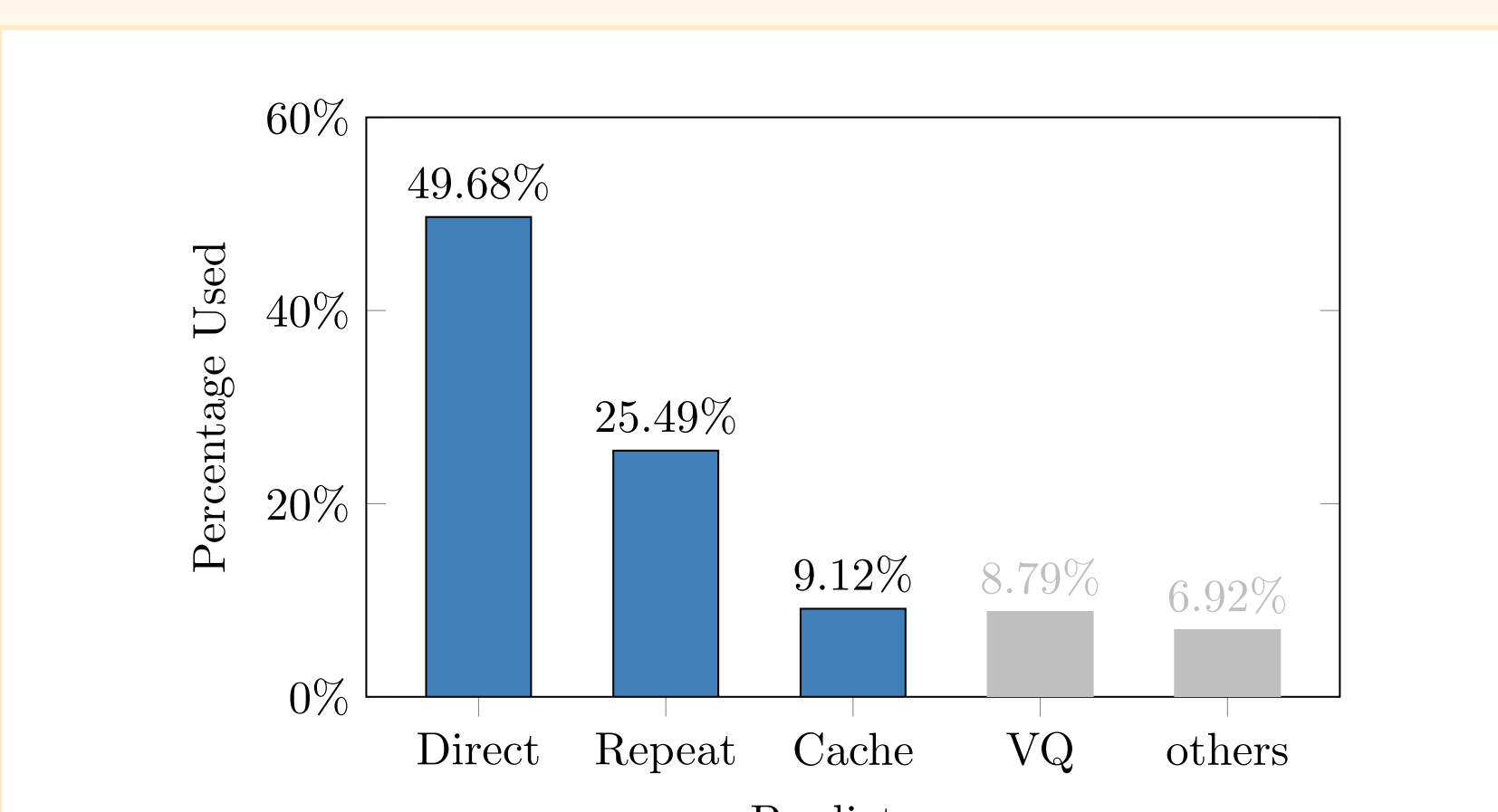
Results

Predicted value x_{pred}	A	0.69	0.07	0.14	0.10
	C	0.07	0.71	0.07	0.16
	G	0.15	0.07	0.71	0.07
	T	0.09	0.13	0.07	0.70
		A	C	G	T

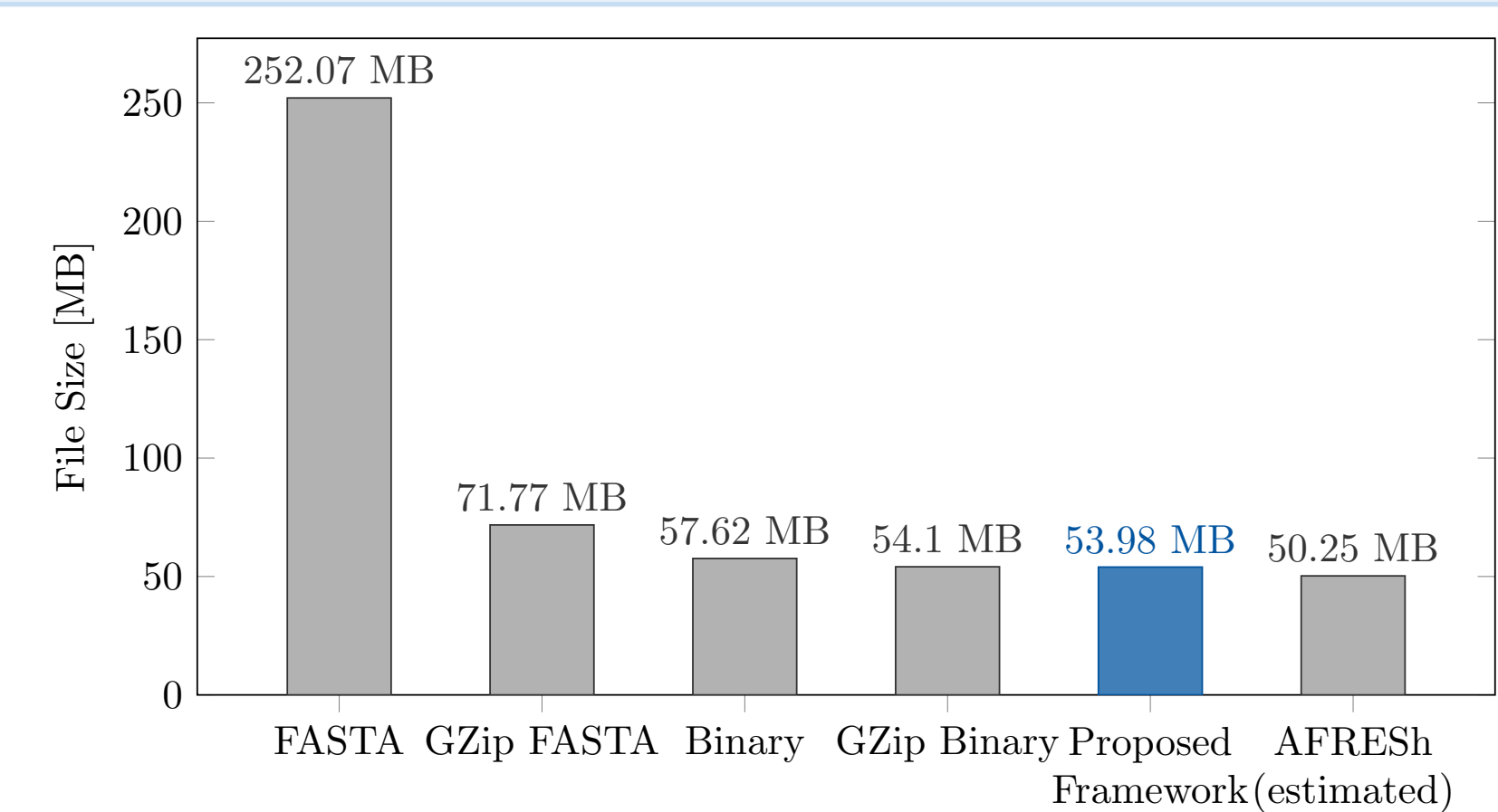
$P(X = x | Y = x_{pred})$

- **Correct prediction:** 0.51 bit per base
- **Incorrect prediction:** 3.42 bit per base

Selection of best encoding method for each block



Evaluation and Summary



- Redundancy in signal exploited by Markov model
- More redundancy exploitable due to larger-scale patterns, such as repetitions
- Use of predictors with lossless residue coding
- Comparison to AFRESH [1]: Different predictor types and error correction (bit error mask)
- Faulty prediction leads to increasing cost for residue coding
 ⇒ Better predictors required

[1] T. Paridaens, *Compression and Interoperable Representation of Genomic Information*, Ghent University, 2018.



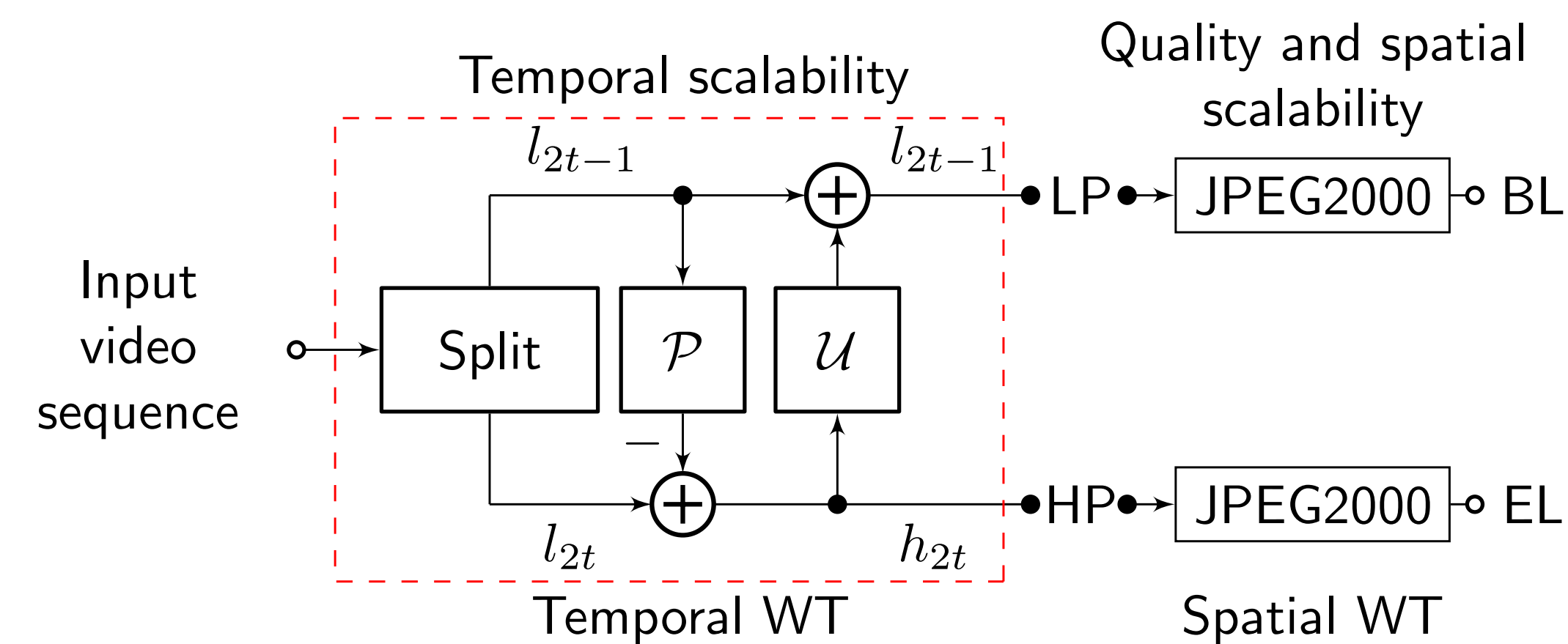
Content Adaptive Wavelet Lifting for Scalable Lossless Video Coding

Daniela Lanz, Christian Herbert, and André Kaup

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

1. Introduction

- **Task:** Professional applications often require lossless compression
- **Challenge:** Lossless compression leads to high bit rates
- **Solution:** Scalable lossless video coding based on transmitting a base layer (BL) with coarser quality and one or more enhancement layers (ELs), comprising the residual video data
- **Approach:** 3-D subband coding based on Wavelet Transforms (WT) [1]



- By realizing \mathcal{P} as the warping operator \mathcal{W} , Motion Compensated Temporal Filtering (MCTF) is achieved [2]:

$$h_{2t} = l_{2t} - \lfloor \mathcal{W}_{2t-1 \rightarrow 2t}(l_{2t-1}) \rfloor$$

$$l_{2t-1} = l_{2t-1} + \lfloor \frac{1}{2} \mathcal{W}_{2t \rightarrow 2t-1}(h_{2t}) \rfloor$$

2. Content Adaptive Wavelet Lifting (CA-WL)

- **Idea:** Adaptive temporal scaling based on significant changes among subsequent frames
- **Stopping Criterion:**
 - Haar WTs can be represented with tree structures
 - With each node a basis vector $\mathbf{b}_{i,t}$ and a wavelet coefficient vector $\mathbf{c}_{i,t}$ is associated, which is the inner product of the signal \mathbf{s} with the basis $\mathbf{b}_{i,t}$
 - If combined costs of child nodes exceed costs of parent node, i.e.

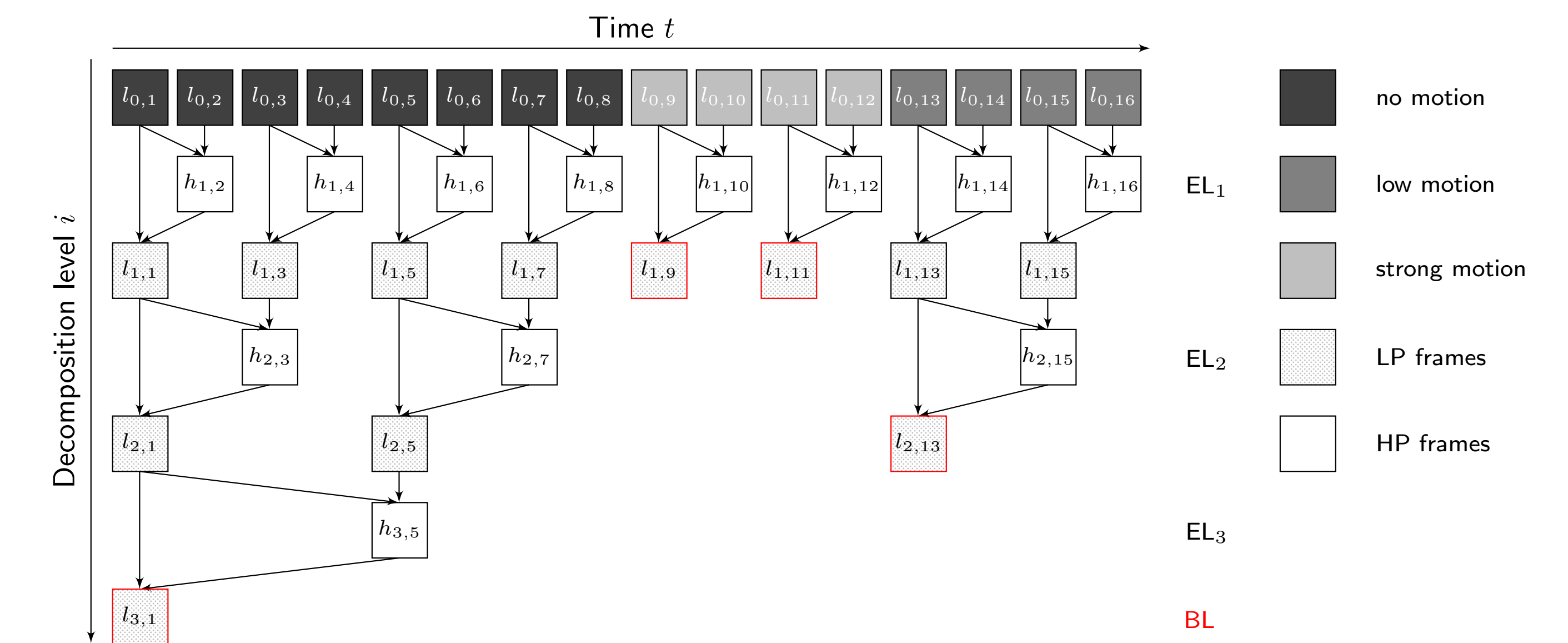
$$\mathcal{C}(\mathbf{s}, \mathbf{b}_{i,[2t-1,2t]}) \leq \mathcal{C}(\mathbf{s}, \mathbf{b}_{i+1,2t-1} \cup \mathbf{b}_{i+1,2t}),$$

the child nodes shall be pruned from the tree

- $\mathcal{C}(\cdot)$ describes a Lagrangian cost functional, which represents the coding costs:

$$\mathcal{C}(\mathbf{s}, \mathbf{b}) = D(\mathbf{s}, \mathbf{b}) + \lambda R(\mathbf{s}, \mathbf{b})$$

- Rate $R(\mathbf{s}, \mathbf{b})$ is composed of the required rate for lossless coding of the LP and HP frames and, in case of MC, the file size of the motion vectors
- Distortion $D(\mathbf{s}, \mathbf{b})$ is calculated by the MSE of the corresponding wavelet coefficients compared to the original signal according to [3]



- **Handling of the Overhead:**

- Realized by transmitting a vector \mathbf{v} , whose length equals the number of input frames:

$$\begin{aligned} \text{Initialize } \mathbf{v} : & (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0) \\ \mathbf{v} \text{ after level } i=1 : & (1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0, 1, 0) \\ \mathbf{v} \text{ after level } i=2 : & (2, 0, 0, 0, 2, 0, 0, 0, 1, 0, 1, 0, 2, 0, 0, 0) \\ \mathbf{v} \text{ after level } i=3 : & (3, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 2, 0, 0, 0) \end{aligned}$$

- Non-zero entries correspond to the number of applied decomposition levels i
- Distance d to the corresponding HP frame is given by $d=2^{i-1}$
- Encoded using multiple-context adaptive arithmetic coding [4]

3. Experimental Results

- **Simulation Setup (8 bpp):**

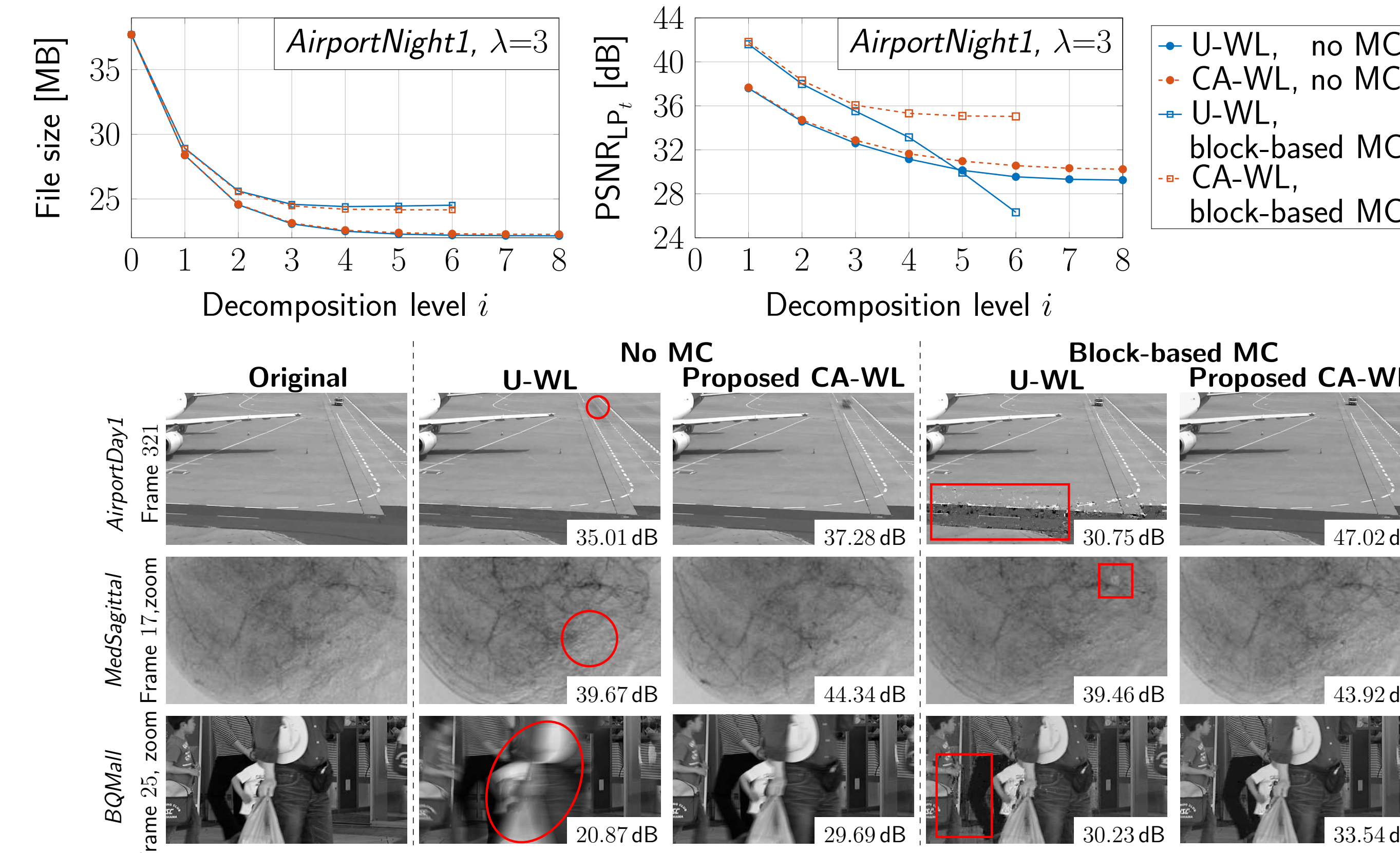
		Spatial resolution	Number of frames
Surv	AirportNight1	688 × 352	500
	AirportNight2	688 × 432	500
	AirportNight3	688 × 372	500
	AirportDay1	688 × 432	500
Med	MedFrontal	512 × 512	29
	MedSagittal	512 × 512	29
HEVC	ClassC	832 × 480	300
	ClassD	416 × 240	300

- **Coding parameters:**

- LP and HP frames are encoded by JPEG2000 [5]
- Block-based MC with block size equals 8
- Search range equals 8 and is doubled for every decomposition level until a maximum size of 64
- Motion vectors are encoded using the QccPack library [6]

Differences of our proposed CA-WL compared to the uniform WL (U-WL) with and without block-based MC.

	λ	Surv	Med	HEVC	Total average
No MC	1	4.12	5.28	15.45	8.88
	3	1.64	1.91	8.86	5.30
	5	0.97	1.16	6.31	3.67
	7	0.65	1.16	6.18	3.50
	1	5.99	0.09	10.36	6.56
	3	0.80	-0.96	4.15	2.18
	5	0.23	-1.29	2.44	1.08
Block-based MC	7	0.16	-1.29	1.66	0.67
	1	9.30	15.56	10.57	10.98
	3	8.17	13.89	10.43	10.28
	5	7.42	13.89	9.38	9.47
	7	7.27	13.89	8.68	9.02
	1	0.16	-5.58	4.44	1.34
	3	-0.52	-5.64	-0.18	-1.06
5	-0.69	-5.64	-0.66	-1.38	
7	-0.80	-5.64	-0.94	-1.57	



4. Conclusion

- Temporal resolution controlled by recursive application of WT
- Visual quality of BL is degraded by strong motion of underlying video
- CA-WL locally adapts temporal scaling by evaluating a Lagrangian cost functional
- For $\lambda=3$ and MC, PSNR_{LP_i} of BL is increased by 10.28dB and rate is reduced by 1.06%

References:
 [1] G. Karlsson and M. Vetterli, Three dimensional sub-band coding of video, ICASSP, New York City, NY, USA, vol. 2, Apr 1988.
 [2] J. R. Ohm, Three-dimensional subband coding with motion compensation, TIP, vol. 3, no. 5, Sep 1994.
 [3] D. Lanz, J. Seiler, K. Jaskolka, and A. Kaup, Compression of dynamic medical CT data using motion compensated wavelet lifting with denoised update, PCS, San Francisco, CA, USA, June 2018.
 [4] I. H. Witten, R. M. Neal, and J. G. Cleary, Arithmetic coding for data compression, Communications of the ACM, vol. 30, no. 6, June 1987.
 [5] ITU-T and ISO/IEC, JPEG 2000 Image Coding System: Core Coding System, in ITU-T Rec. T.800 and ISO/IEC 15444-1:2004, Sep 2004.
 [6] J.E. Fowler, Qccpack: An open-source software library for quantization, compression, and coding, App. of Digital Image Proc., San Diego, CA, USA, vol. 4115, Aug 2000.

Introduction

- Deep learning based IQA methods require massive amounts of data to train
- However, the current largest artificially distorted IQA database, TID2013 [1], contains only 3,000 rated images
- Unable to generate more distorted images for further subjective study as source code is not available
- Our contributions:
 - Konstanz Artificially Distorted Image quality Database (KADID-10k) and Konstanz Artificially Distorted Image quality Set (KADIS-700k)
 - Multi-Level Spatially Pooled IQA method

Dataset creation

- Reference image collection
 - Collect pristine images from Pixabay.com, free to be edited and redistributed
 - Download 654,706 images whose resolution are greater than 1500-by-1200, rescaled and cropped to 512-by-384
 - Manually select 81 reference images in KADID-10k (Fig. 1)
 - Randomly select 140,000 images as reference images in KADIS-700k
- Distorted image generation
 - 25 distortions, grouped into blurs, color distortions, compression, noise, brightness change, spatial distortions, sharpness, and contrast
 - KADID-10k: degraded by 25 distortions in 5 levels each
 - KADIS-700k: degraded by a random distortion in 5 levels each

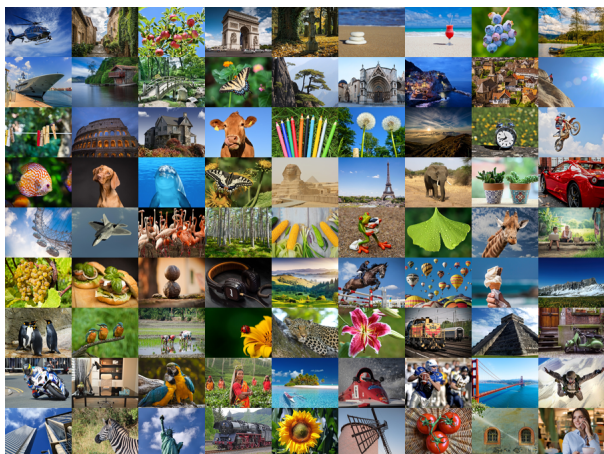


Fig.1. The 81 pristine reference images in KADID-10k.

Subjective IQA

- Performed on figure-eight.com, see interface in Fig. 2
- 5-point scale Degradation category ratings (DCR): imperceptible (5), perceptible but not annoying (4), slightly annoying (3), annoying (2), and very annoying (1)
- Test questions to control the quality of crowd workers
- 30 ratings per image, yield DMOS for each image

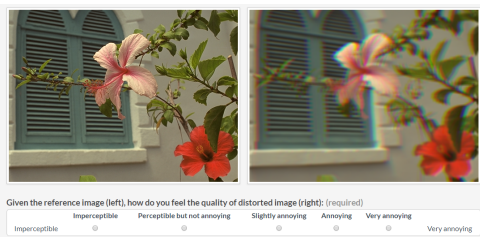
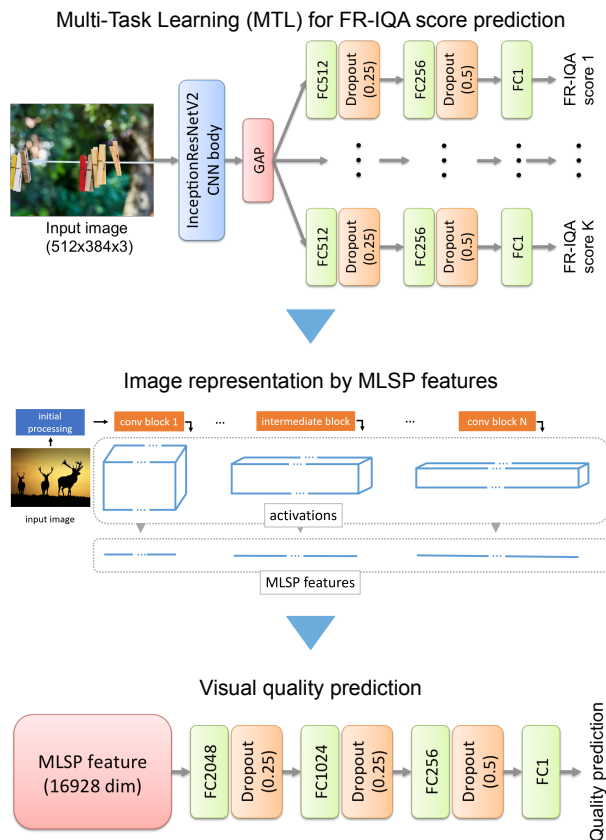


Fig.2. User interface for subjective IQA study.

References

- N. Ponomarenko, L. Jin, O. Ieremeiev, V. Lukin, K. Egiazarian, J. Astola, B. Vozel, K. Chehdi, M. Carli, F. Battisti et al., "Image database TID2013: Peculiarities, results and perspectives," *Signal Processing: Image Communication*, vol. 30, pp. 57–77, 2015.
- H. Lin, V. Hosu, and D. Saupe, "The KADID-10K Image Database," 2019, <http://database.mmsp-kn.de/kadid-10k-database.html>.

MLSP-IQA



	Method	PLCC	SROCC
FR-IQA	SSIM	0.723	0.724
	MSSSIM	0.801	0.802
	IWSSIM	0.846	0.850
	MDSI	0.873	0.872
	VSI	0.878	0.879
	FSIM	0.851	0.854
	GMSD	0.847	0.847
	SFF	0.862	0.862
	SCQI	0.853	0.854
	ADD-GSIM	0.817	0.818
NR-IQA	SR-SIM	0.834	0.839
	BIQI	0.460	0.431
	BLIINDS-II	0.559	0.527
	BRISQUE	0.554	0.519
	CORNIA	0.580	0.541
	DIIVINA	0.532	0.489
	HOSA	0.653	0.609
	SSEQ	0.463	0.424
	InceptionResNetV2 (fine-tune)	0.734	0.731
	MLSP-IQA	0.941	0.939

Table 2. IQA performance comparison on KADID-10K.

Conclusion

- Introduce two KADID-10k and KADIS-700k
 - KADID-10k contains 81 reference images and 10,125 distorted images with 30 quality ratings each
 - KADIS-700k contains 140,000 reference images and 700,000 distorted images
- Both datasets, together with the source code for the 25 distortions, are available in [2]
- Developed MSLP-IQA method, a deep learning based IQA method by weakly supervised feature learning; 0.2 SROCC improvement than fine-tuned CNN, 0.06 SROCC improvement than best FR-IQA metric

Video Transmission Optimization

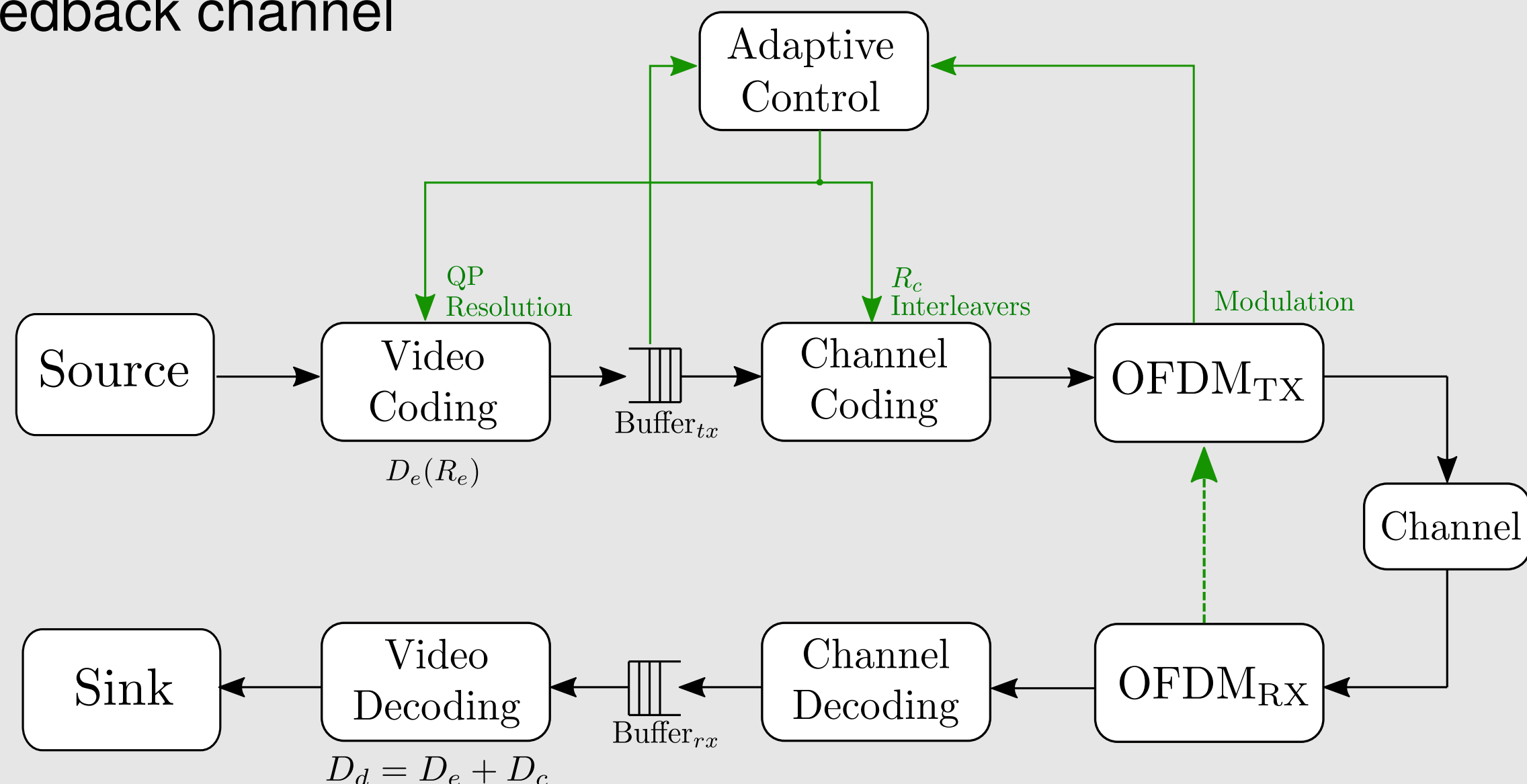
An Overview of Video Compression and Communication Systems

Yasser Samayoa

1. Motivation

Video compression and communication

- ▶ Real time data transmission
- ▶ Time- and frequency-selective channels
- ▶ Feedback channel



Classical separation principle

- ▶ Video (source) coding: operate closely to the rate-distortion bound
 - ▶ Channel coding: operate closely to the channel capacity
- Assumptions
- long block lengths for source and channel codes
 - high computational resources and associated delays

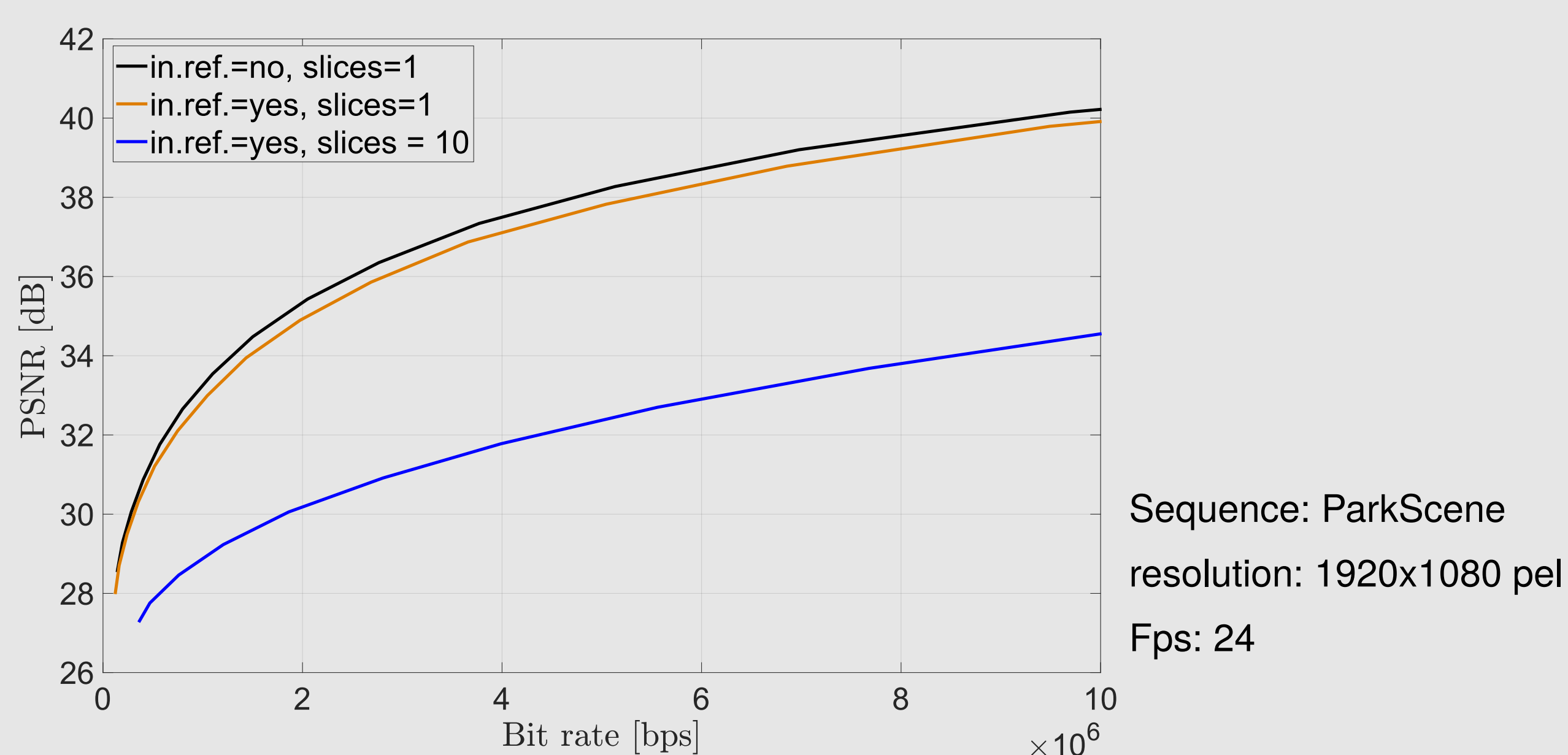
Assumptions do not hold in practice

Goal

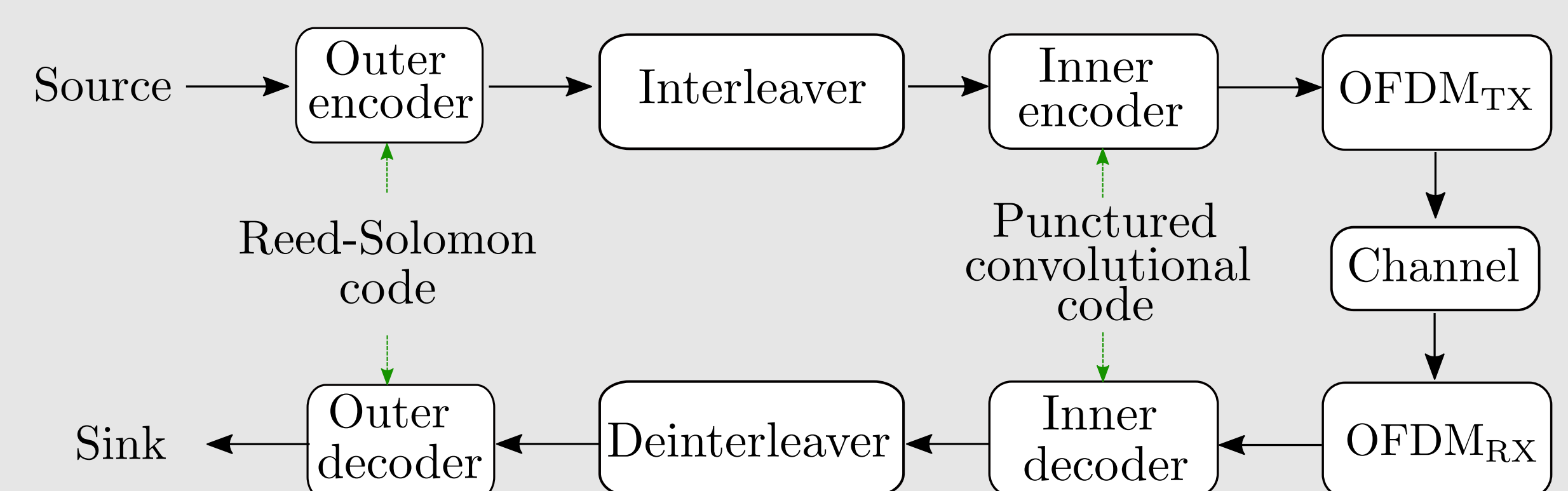
Minimize the end-to-end distortion of the delivered copy of the source under some constraints: bandwidth, transmission power or energy, delay and complexity.

2. Videocoding system: HEVC

- ▶ Adaptive parameters, e.g., space resolution and QP
- ▶ Error resilience techniques, e.g., slices and intra-refresh



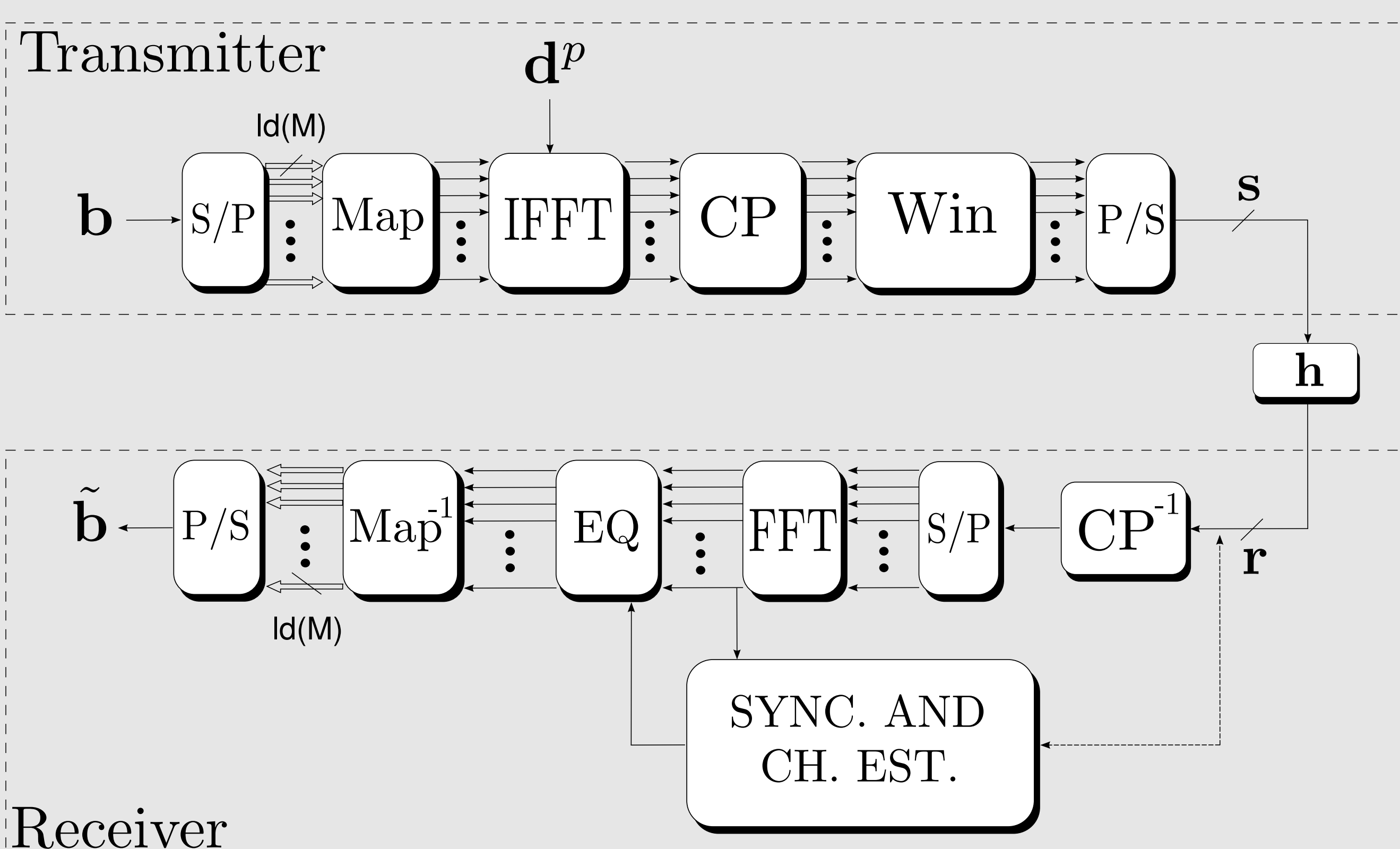
3. Channel coding: serial concatenated codes



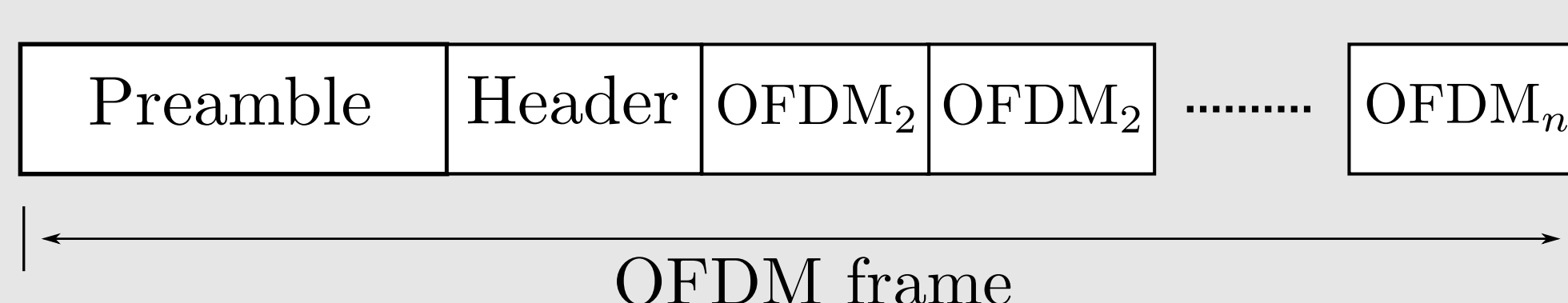
- ▶ Code rate R_c = k/n, with k information bits and n coding bits
- ▶ Punctured convolutional code drawback: burst-error
- ▶ Reed-Solomon code: against burst-error
- ▶ Performance vs. delay

4. Communication system: OFDM

- ▶ Avoids intersymbol interference
- ▶ Optimization, e.g., water filling

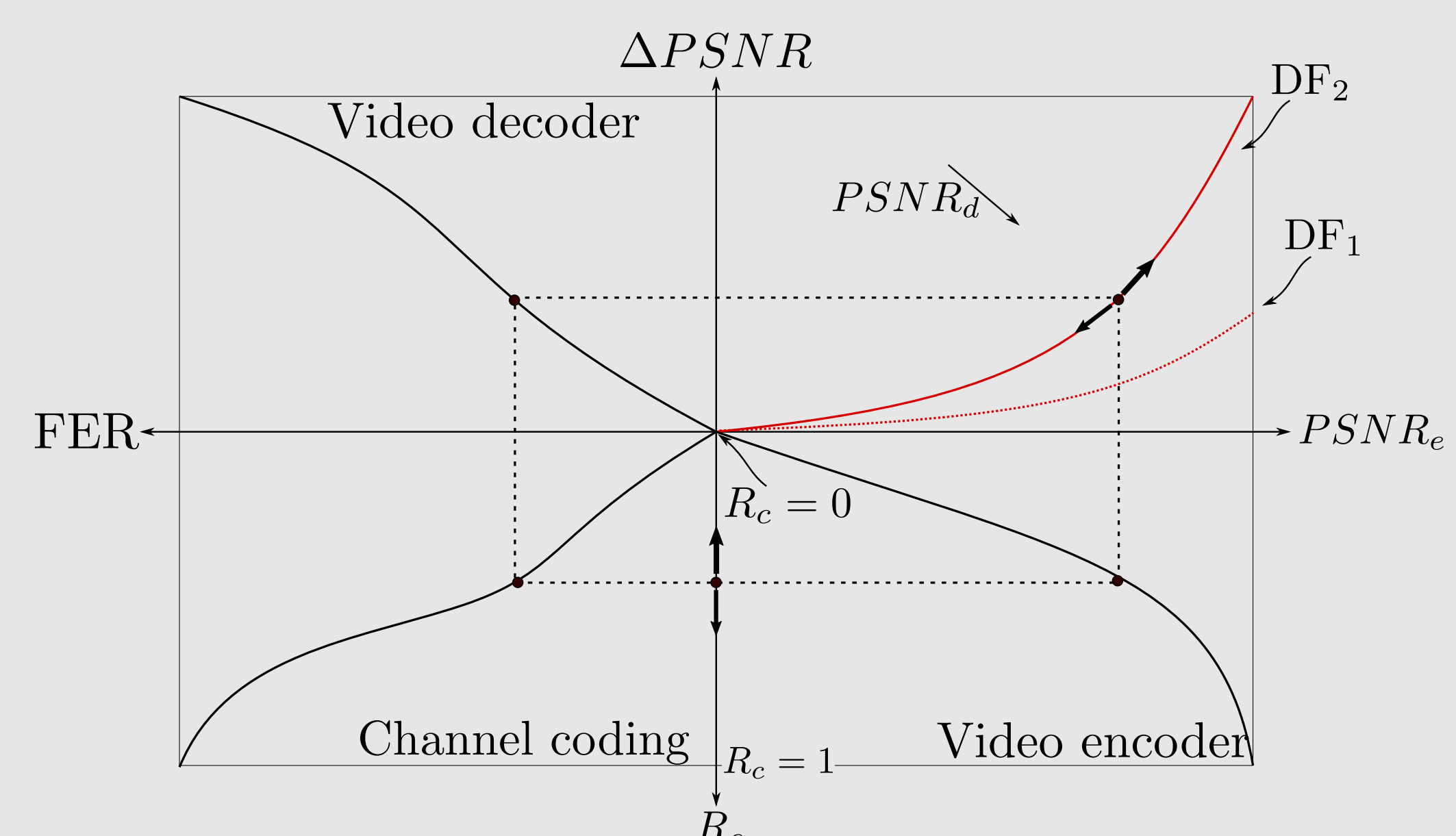


- ▶ Packet transmission mode: one OFDM frame for synchronization and adaptivity



5. Optimization procedure

- ▶ $PSNR_e = 10 \log_{10} \frac{255^2}{D_e}$ and $PSNR_d = 10 \log_{10} \frac{255^2}{D_d}$
- ▶ $\Delta PSNR = PSNR_e - PSNR_d = 10 \log_{10} \frac{D_e}{D_e + D_c}$
- ▶ Distortion Function (DF)



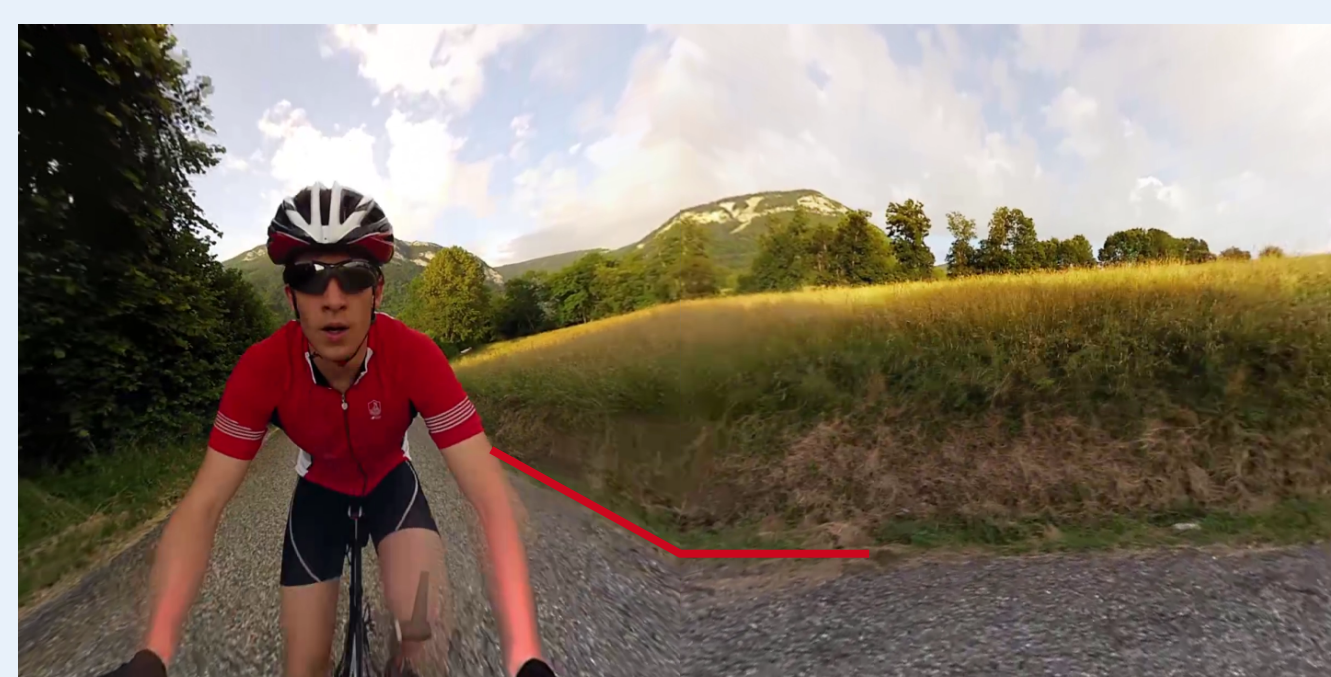
6. Conclusions and future work

- ▶ Minimize the expected video distortion at the decoder, subject to bandwidth, Tx power and delay constraints.
- ▶ Optimization with help of the Distortion Function
- ▶ The delay considerations are pending

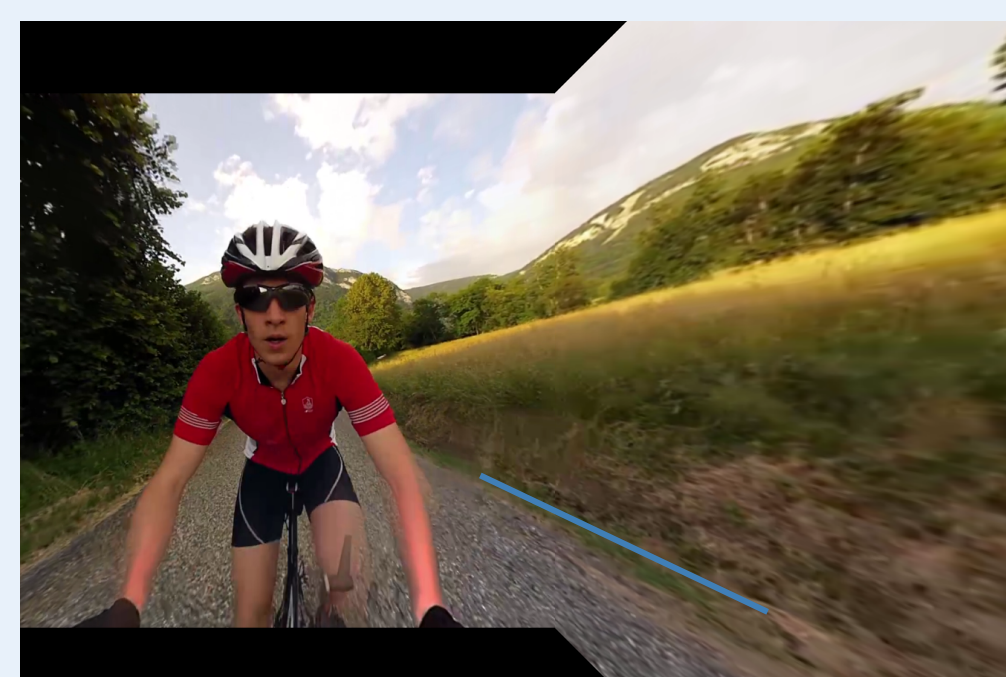
Johannes Sauer and Christian Rohlfing
 Institut für Nachrichtentechnik, RWTH Aachen University

Geometry padding corrects distortions at face boundaries in cube-based 360° video

- Improved inter prediction across face boundaries
- Objective coding gains of 2% on average [1]
- Improvement visually apparent [3]



Neighboring cube faces. Red: Geometric distortion at face boundary, straight lines appear bend.



Geometry padding of the left face. Blue: Corrected geometric distortion.

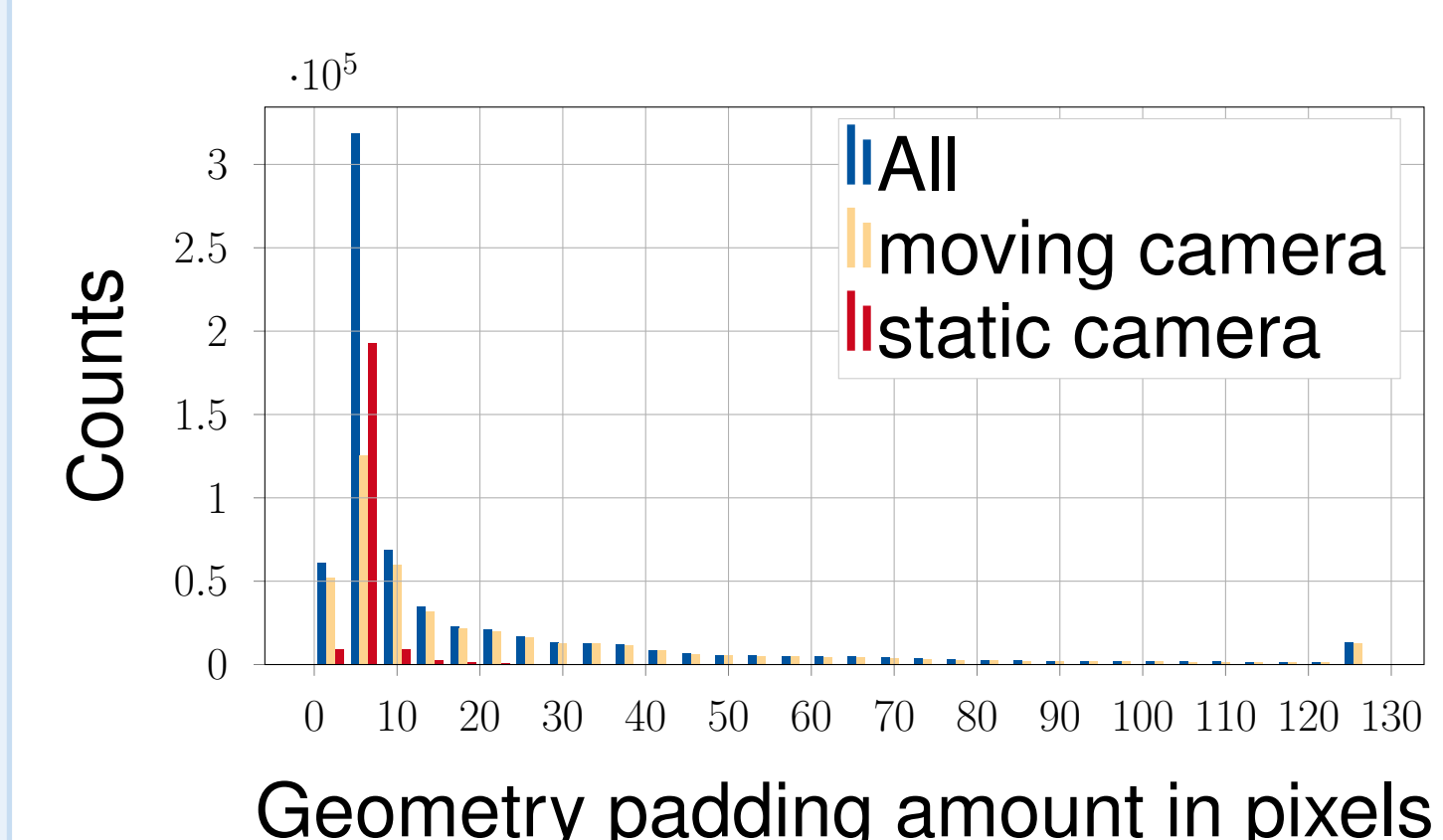
Analysis of geometry padding usage

Derivation of geometry padding usage

- Inspection of encoder motion buffer
- Derivation of required padding separately for each face boundary

Observations

- Static sequences require very little padding
- Small padding amount much more likely
- Full padding required in some cases



Integration of geometry padding into coding scheme

On-the-fly geometry padding

- + Easy to implement
- + Low complexity: 5% decoder runtime increase with nearest neighbor interpolation [2]
- Support for geometry padding at block level unlikely due to overhead for conventional video
- Higher memory access at block level. Worst case: Cube corner requires pixels from three different regions
- Padding may be re-generated several times.

Picture level geometry padding

- + Padding of reference pictures only has to be generated once
- + No modification of decoder at block level required
- Full padding of faces complex and potentially not required
- Increased storage for reference pictures
- Requires treatment of uncoded areas

Geometry padding usage signaling

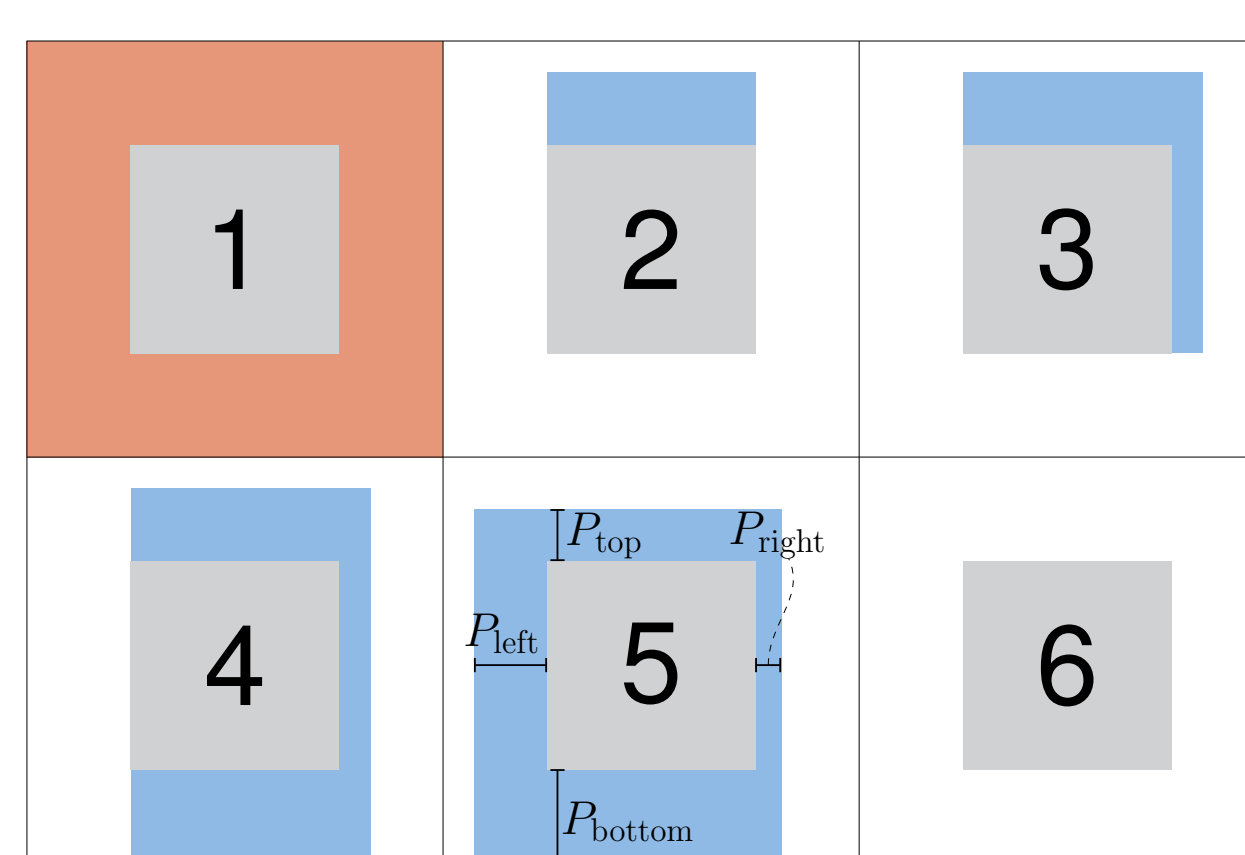
Requirements

Signaling granularity: Geometry padding can be controlled per picture, per face, and per boundary

Incremental signaling: Padding can be signaled with reference to previous padding. Previous padding can be reused

Quantization and coding: Padding is quantized into bins of 4 pixels and exponential-golomb coded

Face packing



Packing of the cube faces. Grey: Cube faces. White: Additional uncoded area reserved for geometry padding. Red: Complete padding. Blue: Partial geometry padding

Proposed supplemental enhancement information (SEI) message syntax

```

for ref_list = 0 to 2 do
  numRefIdxMinus1 ← 2 bit
  for refIdx = 0 to numRefIdxMinus1 do
    picNeedsPadding ← 1 bit
    if !picNeedsPadding then
      continue
    for r = 0 to 2 do
      for c = 0 to 3 do
        for boundary in top, bottom, left, right do
          boundary.paddingPresent ← 1 bit
        for r = 0 to 2 do
          for c = 0 to 3 do
            for boundary in top, bottom, left, right do
              if boundary.paddingPresent then
                boundary.paddingWidth ← ue
    
```

Results

	Sequence	BD-Rate in %		Decoder complexity	
		no SEI	with SEI	no SEI	with SEI
static	Gaslamp	-0.83	-0.62	302%	151%
	Harbor	-0.52	-0.34	255%	136%
	KiteFlite	-0.79	-0.72	220%	131%
	Trolley	-1.03	-0.94	262%	144%
non-static	Balboa	-3.31	-3.15	220%	113%
	BranCastle2	-3.14	-3.11	170%	98%
	Broadway	-2.55	-2.42	198%	115%
	ChairliftRide	-3.44	-3.31	223%	116%
	Landing2	-2.25	-2.20	167%	106%
SkateboardInLot	-3.11	-3.00	179%	110%	
Average	static	-0.79	-0.65	258%	140%
	non-static	-2.97	-2.86	191%	110%
	all	-2.10	-1.98	216%	121%

Configuration

- VVC Test Model (VTM) reference software, version VTM 4.2 [4]
- 360Lib extension 9.0 [5]
- Geometry padding using nearest-neighbor interpolation
- Common test conditions and evaluation procedures for 360 video (CTC360) [6],
- Distinction of static and non-static sequences

Conclusion

- Signaling of geometry padding has only small impact on coding gain
- Geometry padding can be applied efficiently at picture level
- Requires no low level modifications of the decoder

[1] Philippe Hanhart, Jian-Liang Lin, and Chirag Pujara, "CE13: Summary report on coding tools for 360° omnidirectional video," Doc. JVET-L0033, Joint Video Exploration Team (on Future Video coding) of ITU-T VCEG and ISO/IEC MPEG, Macao, CN, 12th meeting, Oct. 2018.
 [2] Philippe Hanhart, Jian-Liang Lin, and Chirag Pujara, "CE13: Summary report on coding tools for 360° omnidirectional video," Doc. JVET-M0033, Joint Video Exploration Team (on Future Video coding) of ITU-T VCEG and ISO/IEC MPEG, Marrakech, MA, 13th meeting, Jan. 2019.
 [3] Jill Boyce, "BoG Report on CE13 and CE13 related 360° video coding," Doc. JVET-M0874, Joint Video Exploration Team (on Future Video coding) of ITU-T VCEG and ISO/IEC MPEG, Marrakech, MA, 13th meeting, Jan. 2019.
 [4] "VVC Test Model, version 4.2," https://vcgit.hhi.fraunhofer.de/jvet/VVCSsoftware_VTM/tags/VTM-4.2, 2019.
 [5] "360 video conversion software, version 9.0," https://jvet.hhi.fraunhofer.de/svn/svn_360Lib/tags/360Lib-9.0, 2019.
 [6] Philippe Hanhart, Jill Boyce, Kiho Choi, and J.-L. Lin, "JVET common test conditions and evaluation procedures for 360° video," Doc. JVET-L1012, Joint Video Exploration Team (on Future Video coding) of ITU-T VCEG and ISO/IEC MPEG, Macao, CN, 12th meeting, Oct. 2018.

