Contributions

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

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

Konstanzer Online-Publikations-System (KOPS)
URL: http://nbn-resolving.de/urn:nbn:de:bsz:352-2-v37lq93m3kp2
Organizer

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 Jongebloed, 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 upcaled 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 unconstraint 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 Jongebloed  
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 downsamples 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

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 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

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)

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

Diagram:
- DC Planar
- Mode 1
- Mode 3
- Mode 5
- Mode 2
- Mode 4
- Mode 6
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

<table>
<thead>
<tr>
<th>Predictor</th>
<th># Modes</th>
<th>Prediction PSNR [dB]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Angular (HEVC)</td>
<td>35</td>
<td>25.03</td>
</tr>
<tr>
<td>NN</td>
<td>35</td>
<td>24.73</td>
</tr>
<tr>
<td>Angular</td>
<td>18</td>
<td>24.67</td>
</tr>
<tr>
<td>NN</td>
<td>18</td>
<td>24.46</td>
</tr>
<tr>
<td>Angular</td>
<td>6</td>
<td>22.71</td>
</tr>
<tr>
<td>NN</td>
<td>6</td>
<td>23.94</td>
</tr>
</tbody>
</table>
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

![Graph showing prediction PSNR (dB) vs generation for different models, including NN - 6 modes, NN - 18 modes (Baseline), Angular - 18 modes, and Angular - 35 modes.](image)
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


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)

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

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

→ Long measurement time, high data rate

→ Interpolate remaining pixels
→ Resolution limited by aliasing

→ Higher resolution after appropriate reconstruction
→ Reduced aliasing

Incremental Non-Regular Sampling

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

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**

How to choose the Sampling Pattern?

**Uniformity**
- Local density ≈ 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
Sampling Patterns (central section)

Reconstruction Quality

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

Visual Comparisons

Visual Comparisons

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

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?

- Translation
- Rotation
- Zoom
- Super-Resolution
- Frame Rate Up-Conversion
- Panorama Stitching
- Homography
- Fisheye
- Tracking
- And many more...

Source: https://testimages.org/
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]

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

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

Original Samples = Mesh

Compute Error Energy for every Basis Function

Selection of Best Fitting Basis Function

Update Model

Stopping Criteria

Cut out reconstructed pixels
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

Original Samples = Mesh

Merge

Key points

Adaptive Key Point Weighting

Compute Error Energy for every Basis Function

Selection of Best Fitting Basis Function

Update Model

Stopping Criteria

Cut out reconstructed pixels
Results with Adaptive Key Point Weighting
## Comparison

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

<table>
<thead>
<tr>
<th></th>
<th>Cubic</th>
<th>Natural Neighbor</th>
<th>Inverse Distance Weighting</th>
<th>Lanczos</th>
<th>FSMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotation</td>
<td>3.30</td>
<td>-0.11</td>
<td>0.56</td>
<td>0.32</td>
<td>9.77</td>
</tr>
<tr>
<td>Translation</td>
<td>3.23</td>
<td>-0.13</td>
<td>1.63</td>
<td>3.59</td>
<td>12.06</td>
</tr>
<tr>
<td>Zoom</td>
<td>3.65</td>
<td>-0.26</td>
<td>4.20</td>
<td>4.44</td>
<td>7.09</td>
</tr>
</tbody>
</table>

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

- **Model 1:** Polynomial with degree 2
  - Linear regression problem, solved by least squares
  - In some cases not robust enough (few contour points, volatile curvature, imprecise contour pixel location)

- **Model 2:** Slope modeling
  - More robust modeling in different space with fewer parameters
  - Smoothed slope

- **Model 3:** Outlier consideration
  - Iterative residue weightening during least squares

- **Model 4:** Awareness for evasive curvature changes
  - Distance-dependent residue weightening
Contour Combination

Problem:
Contours can be intersected by current block → Two independent contours

Solution:
1. Detect intersecting contours after extrapolation
2. Join contours to single contour
3. Interpolation between contours
Sample Value Prediction

- Reference sample continuation *along contour*
- Continuation along shifted contours

Differences to HEVC
- Multiple directions
- Non-linear directional prediction
Bit rate savings relative to JPEG at fixed quality
Coding Efficiency 2

BD rates relative to JPEG
<table>
<thead>
<tr>
<th>Coding Efficiency 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>BD rates relative to HM</td>
</tr>
<tr>
<td>All Intra</td>
</tr>
</tbody>
</table>

- Steam Locomotive
- Kimono
- Park Scene
- Basketball Drive
- BBQ Terrace
- Johnny
- Ball Under Water
- Bubbles Clear
- Calming Water
- Carpet Circle Fast
- Drops on Water
- Flowers 2
- Paper Static
- Smoke Clear
- Squirrel
- Bike 1
- Bike 2
- Bike 3
- Bike 4
- Bike 5
- Bike 6
- Bike 7
- Bike 8
- Bike 9
- Bike 10
- Bike 11
- Bike 12
- Bike 13
- Bike 14
The CoMIC Codec Today

```
 Contour detection

 Linear contour model

 Non-linear contour model 1

 Non-linear contour model 2

 Non-linear contour model 3

 Non-linear contour model 4

 Boundary Recall-based contour model selection

 Contour connection

 Reference sample continuation with diminishing

 Along-contour reference sample continuation

 Prediction error coding and signal reconstruction

 DC mode

 Rate-distortion optimization

 Novel contribution

 Previous work (CoMIC v1)

 Input

 Output
```
Current Work

Stochastic contour modeling
Machine learning-based sample value prediction
How much machine learning is good?

Conventional codecs
- HEVC
- JPEG

Encoder control
- Partitioning, mode decision, etc.
- Classification, reinforcement learning
  - Laude2016, Yu2015

Coding Tools
- Integration in hybrid codecs
- Intra Coding (Meyer2019)
- Inter Coding (Laude2018)

End-to-end coding
- Complete coding system
- Toderici2015+2016
- Johannes Ballé

no machine learning

solely machine learning
More Information


- 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
Introduction

Motion Estimation

Example: Analysis of Hamburg Taxi Sequence

Frame 1  ➔  Frame 2

Movements of the cars?

Color-coded Displacement Field

Flow Field

Flow color coding

Visual Quality Assessment for Motion Compensated Frame Interpolation
Hui Men
Introduction

Motion Estimation

- Optical Flow

![Frame 1](image1.png) ![Displacement ?](image2.png) ![Frame 2](image3.png)

Flow color coding
光学流动

定量评估
- 法向错误与端点误差之间的流动向量与ground-truth流动

地面真实流动

流动向量
Optical Flow

Quantitative Evaluation

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

MSE: 11.9

Ground-truth in-between image
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](image1)
![Interpolated Images](image2)

*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**

**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

<table>
<thead>
<tr>
<th>Running Time</th>
<th>Average</th>
<th>Mequon</th>
<th>Schefflera</th>
<th>Urban</th>
<th>Teddy</th>
<th>Backyard</th>
<th>Basketball</th>
<th>Dumptruck</th>
<th>Evergreen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hours</td>
<td>29</td>
<td>60</td>
<td>72</td>
<td>10</td>
<td>20</td>
<td>10</td>
<td>3</td>
<td>20</td>
<td>35</td>
</tr>
</tbody>
</table>
Result Reconstruction

- Adding 2 Anchors

Anchor: worst

423 pairs of images

Anchor: best
Result Reconstruction

- Adding 2 Anchors

- Reconstruction using Thurstone’s Model
- Rescale to [0,1]
- Correlations: Average SROCC = 0.598
## Re-ranking Result

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

---

Visual Quality Assessment for Motion Compensated Frame Interpolation
Hui Men
<table>
<thead>
<tr>
<th>Optrical Flow Method</th>
<th>Ranking Old / New</th>
<th>Ranking Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>CtxSyn</td>
<td>1 / 2</td>
<td>1</td>
</tr>
<tr>
<td>FGIK</td>
<td>2 / 20</td>
<td>18</td>
</tr>
<tr>
<td>MDP-Flow2</td>
<td>3 / 18</td>
<td>15</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>NN-field</td>
<td>11 / 48</td>
<td>37</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>NNF-EAC</td>
<td>18 / 54</td>
<td>36</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>F-TV-L1</td>
<td>33 / 63</td>
<td>30</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Bartels</td>
<td>77 / 11</td>
<td>66</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>TI-DOFE</td>
<td>103 / 34</td>
<td>69</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>PGAM+LK</td>
<td>135 / 50</td>
<td>85</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Pyramid LK</td>
<td>139 / 122</td>
<td>17</td>
</tr>
<tr>
<td>GroupFlow</td>
<td>140 / 139</td>
<td>1</td>
</tr>
<tr>
<td>Periodicity</td>
<td>141 / 141</td>
<td>0</td>
</tr>
</tbody>
</table>
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

[Diagram showing optical flow method and quality assessment process]

Ground-truth in-between Image → Optical Flow Method → Interpolated Image → Model → Predicted Quality Score → MOS

Ground-truth Flow Field → Optical Flow Method → Predicted Flow Field
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
  - $\Rightarrow$ High prediction error
  - $\Rightarrow$ High entropy $\Rightarrow$ 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
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 rate$^2$

---


Holger Meuel
meuel@tnt.uni-hannover.de
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:

\[
\begin{align*}
    x' &= a_{11} \cdot x + a_{12} \cdot y + a_{13} \\
y' &= a_{21} \cdot x + a_{22} \cdot y + a_{23}
\end{align*}
\]

- \(a_{11}, a_{12}, a_{21}, a_{22}\) “purely affine” parameters (rotation, scaling, shearing)
- \(a_{13}\) and \(a_{23}\) translational parameters
Affine Motion Estimation

Estimated affine motion:

\[
\begin{align*}
    x' &= a_{11} \cdot x + a_{12} \cdot y + a_{13} \\
y' &= a_{21} \cdot x + a_{22} \cdot y + a_{23}
\end{align*}
\]

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

\[
\begin{align*}
    \Delta x' &= \hat{x}' - x' = \left(\hat{a}_{11} - a_{11}\right) \cdot x + \left(\hat{a}_{12} - a_{12}\right) \cdot y + \left(\hat{a}_{13} - a_{13}\right) \\
    \Delta y' &= \hat{y}' - y' = \left(\hat{a}_{21} - a_{21}\right) \cdot x + \left(\hat{a}_{22} - a_{22}\right) \cdot y + \left(\hat{a}_{23} - a_{23}\right)
\end{align*}
\]
Efficiency Analysis of Affine MCP / Affine Motion and Error Model

Affine Error Model

Displacement estimation error in the frame:

\[
\Delta x' = e_{11} \cdot x + e_{12} \cdot y + e_{13}
\]
\[
\Delta y' = e_{21} \cdot x + e_{22} \cdot y + e_{23}
\]

- 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

  $\Rightarrow$ Joint pdf for independent $e_{ij}$:

  \[ p_{E_{11},...,E_{23}}(e_{11}, \ldots, e_{23}) = p(e_{11}) \cdot \ldots \cdot p(e_{23}) \]

- 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:

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

with

\[
\sigma_{\Delta x'}^2 = \sigma_{e_{11}}^2 x^2 + \sigma_{e_{12}}^2 y^2 + \sigma_{e_{13}}^2
\]

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
Efficiency Analysis of Affine MCP / Signal and Error PSD Modeling

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
Efficiency Analysis of Affine MCP / Rate-Distortion Analysis

**Rate-Distortion Theory**

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

\[
R(D) = \frac{1}{8\pi^2} \int \int \log_2 \left( \frac{S_{ee}(\Lambda)}{\Theta} \right) \, d\Lambda \, \text{bit}
\]

\[\Lambda: \left( S_{ss}(\Lambda) > \Theta \right) \quad \text{and} \quad S_{ee}(\Lambda) > \Theta \]

\[\Theta: \text{generating function varying distortion} \quad D \text{ and corresponding rate } R(D)\]

---

3 based on Toby Berger, “Rate Distortion Theory: A Mathematical Basis for Data Compression”, Prentice-Hall electrical eng. series, Prentice-Hall, 1971
Outline

Efficiency Analysis of Affine Motion Compensated Prediction
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
Simulations

Location Dependent Bit Rate

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

Minimum Required Bit Rate for Prediction Error Coding

Distortion SNR = 30 dB, $\sigma^2_{e_{11}} = \sigma^2_{e_{12}} = \sigma^2_{e_{21}} = \sigma^2_{e_{22}}$ and $\sigma^2_{e_{13}} = \sigma^2_{e_{23}}$, full HD resolution, 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
Experiments

Experimental Setup

- Video signal $s$ with artificially introduced motion of specific variances
- Most-trivial motion estimation always predicting “no motion”
  \[ \Rightarrow \text{Introduced motion becomes exactly prediction error } e \]

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

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

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 $2.507 \text{bit/sample}$ at maximum)
- More pronounced lower plateau in experimental data due to interpolation filter
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 $128 \times 128$
Experiments

Distinct Affine Test Sequences

ShieldsPart, frame 1

ShieldsPart, frame 100

TractorPart, frame 1

TractorPart, frame 100

Experiments

Model vs. Real-World Measurements

- Block size: \(128 \times 128\) pel as in JEM
- Translational quarter-pel, non-translational \(1/16\) pel accuracy

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ShieldsPart</td>
<td>0.398</td>
<td>0.5</td>
<td>0.71</td>
<td>Model approximates minimum bit rate</td>
</tr>
<tr>
<td>TractorPart</td>
<td>0.058</td>
<td>0.07</td>
<td>0.012</td>
<td>Isotropic assumption violation, low-contrast signal, high amount of blur</td>
</tr>
</tbody>
</table>

Conclusion:

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

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
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
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

Problem Statement: Neural Networks for Intra Prediction

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

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)
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
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
### 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](image-url)
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
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

<table>
<thead>
<tr>
<th>Loss Function</th>
<th>L1</th>
<th>SATD</th>
</tr>
</thead>
<tbody>
<tr>
<td>BQTerrace</td>
<td>-1.51%</td>
<td>-1.61%</td>
</tr>
<tr>
<td>BasketballDrive</td>
<td>-1.97%</td>
<td>-2.30%</td>
</tr>
<tr>
<td>Cactus</td>
<td>-2.08%</td>
<td>-2.30%</td>
</tr>
<tr>
<td>Kimono</td>
<td>-2.61%</td>
<td>-3.17%</td>
</tr>
<tr>
<td>ParkScene</td>
<td>-2.75%</td>
<td>-2.85%</td>
</tr>
<tr>
<td>AVG Class B</td>
<td>-2.18%</td>
<td>-2.45%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Architecture</th>
<th>C2</th>
<th>C1</th>
<th>C0</th>
</tr>
</thead>
<tbody>
<tr>
<td>BQMall</td>
<td>-2.00%</td>
<td>-1.85%</td>
<td>-1.85%</td>
</tr>
<tr>
<td>BasketballDrill</td>
<td>-1.99%</td>
<td>-1.96%</td>
<td>-1.81%</td>
</tr>
<tr>
<td>PartyScene</td>
<td>-1.46%</td>
<td>-1.39%</td>
<td>-1.34%</td>
</tr>
<tr>
<td>RaceHorses</td>
<td>-1.89%</td>
<td>-1.84%</td>
<td>-1.75%</td>
</tr>
<tr>
<td>AVG Class C</td>
<td>-1.84%</td>
<td>-1.76%</td>
<td>-1.69%</td>
</tr>
<tr>
<td>BQSquare</td>
<td>-0.98%</td>
<td>-0.88%</td>
<td>-0.79%</td>
</tr>
<tr>
<td>BasketballPass</td>
<td>-1.85%</td>
<td>-1.51%</td>
<td>-1.49%</td>
</tr>
<tr>
<td>BlowingBubbles</td>
<td>-1.70%</td>
<td>-1.74%</td>
<td>-1.63%</td>
</tr>
<tr>
<td>RaceHorses</td>
<td>-2.43%</td>
<td>-2.30%</td>
<td>-2.00%</td>
</tr>
<tr>
<td>AVG Class D</td>
<td>-1.74%</td>
<td>-1.61%</td>
<td>-1.48%</td>
</tr>
<tr>
<td>AVG All Classes</td>
<td>-2.01%</td>
<td>-1.97%</td>
<td>-1.91%</td>
</tr>
</tbody>
</table>
### 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

<table>
<thead>
<tr>
<th>Version</th>
<th>Y</th>
<th>U</th>
<th>V</th>
<th>Y</th>
<th>U</th>
<th>V</th>
<th>Y</th>
<th>U</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>BQTerrace</td>
<td>-1.79%</td>
<td>-0.84%</td>
<td>-0.36%</td>
<td>-1.66%</td>
<td>-0.75%</td>
<td>-0.79%</td>
<td>-1.57%</td>
<td>-0.26%</td>
<td>-0.03%</td>
</tr>
<tr>
<td>Basket.Drive</td>
<td>-2.33%</td>
<td>-1.64%</td>
<td>-1.97%</td>
<td>-1.83%</td>
<td>-0.15%</td>
<td>-0.88%</td>
<td>-1.34%</td>
<td>-0.05%</td>
<td>-0.56%</td>
</tr>
<tr>
<td>Cactus</td>
<td>-2.46%</td>
<td>-1.89%</td>
<td>-2.05%</td>
<td>-1.99%</td>
<td>-1.24%</td>
<td>-1.12%</td>
<td>-1.60%</td>
<td>-0.89%</td>
<td>-0.50%</td>
</tr>
<tr>
<td>Kimono</td>
<td>-2.66%</td>
<td>-2.46%</td>
<td>-1.84%</td>
<td>-1.71%</td>
<td>-1.60%</td>
<td>-1.41%</td>
<td>-1.62%</td>
<td>-1.46%</td>
<td>-1.26%</td>
</tr>
<tr>
<td>ParkScene</td>
<td>-2.55%</td>
<td>-1.75%</td>
<td>-1.91%</td>
<td>-1.87%</td>
<td>-1.27%</td>
<td>-1.82%</td>
<td>-1.88%</td>
<td>-0.79%</td>
<td>-1.15%</td>
</tr>
<tr>
<td><strong>AVG Class B</strong></td>
<td><strong>-2.36%</strong></td>
<td><strong>-1.72%</strong></td>
<td><strong>-1.63%</strong></td>
<td><strong>-1.81%</strong></td>
<td><strong>-1.00%</strong></td>
<td><strong>-1.20%</strong></td>
<td><strong>-1.59%</strong></td>
<td><strong>-0.69%</strong></td>
<td><strong>-0.62%</strong></td>
</tr>
<tr>
<td>BQMall</td>
<td>-2.00%</td>
<td>-1.62%</td>
<td>-1.57%</td>
<td>-1.71%</td>
<td>-1.22%</td>
<td>-1.05%</td>
<td>-1.58%</td>
<td>-1.37%</td>
<td>-0.36%</td>
</tr>
<tr>
<td>BasketballDrill</td>
<td>-1.99%</td>
<td>-2.42%</td>
<td>-2.26%</td>
<td>-1.21%</td>
<td>-0.38%</td>
<td>-0.74%</td>
<td>-0.63%</td>
<td>-0.17%</td>
<td>-0.21%</td>
</tr>
<tr>
<td>PartyScene</td>
<td>-1.46%</td>
<td>-0.86%</td>
<td>-0.96%</td>
<td>-1.31%</td>
<td>-0.68%</td>
<td>-0.70%</td>
<td>-1.21%</td>
<td>-0.68%</td>
<td>-0.65%</td>
</tr>
<tr>
<td>RaceHorses</td>
<td>-1.89%</td>
<td>-1.28%</td>
<td>-1.44%</td>
<td>-1.55%</td>
<td>-0.90%</td>
<td>-0.60%</td>
<td>-1.22%</td>
<td>-0.69%</td>
<td>-0.49%</td>
</tr>
<tr>
<td><strong>AVG Class C</strong></td>
<td><strong>-1.84%</strong></td>
<td><strong>-1.55%</strong></td>
<td><strong>-1.56%</strong></td>
<td><strong>-1.45%</strong></td>
<td><strong>-0.80%</strong></td>
<td><strong>-0.77%</strong></td>
<td><strong>-1.16%</strong></td>
<td><strong>-0.73%</strong></td>
<td><strong>-0.43%</strong></td>
</tr>
<tr>
<td>BQSquare</td>
<td>-0.98%</td>
<td>-0.65%</td>
<td>-0.28%</td>
<td>-1.04%</td>
<td>-0.55%</td>
<td>0.00%</td>
<td>-1.00%</td>
<td>-0.28%</td>
<td>0.20%</td>
</tr>
<tr>
<td>BasketballPass</td>
<td>-1.85%</td>
<td>-1.67%</td>
<td>-1.36%</td>
<td>-1.39%</td>
<td>-1.37%</td>
<td>-0.82%</td>
<td>-1.21%</td>
<td>-0.26%</td>
<td>-0.58%</td>
</tr>
<tr>
<td>BlowingBubbles</td>
<td>-1.70%</td>
<td>-1.54%</td>
<td>-0.97%</td>
<td>-1.40%</td>
<td>-0.60%</td>
<td>-0.43%</td>
<td>-1.32%</td>
<td>-0.58%</td>
<td>-0.37%</td>
</tr>
<tr>
<td>RaceHorses</td>
<td>-2.43%</td>
<td>-1.40%</td>
<td>-2.13%</td>
<td>-1.91%</td>
<td>-1.34%</td>
<td>-1.34%</td>
<td>-1.58%</td>
<td>-0.79%</td>
<td>-0.78%</td>
</tr>
<tr>
<td><strong>AVG Class D</strong></td>
<td><strong>-1.74%</strong></td>
<td><strong>-1.32%</strong></td>
<td><strong>-1.19%</strong></td>
<td><strong>-1.44%</strong></td>
<td><strong>-0.97%</strong></td>
<td><strong>-0.65%</strong></td>
<td><strong>-1.28%</strong></td>
<td><strong>-0.48%</strong></td>
<td><strong>-0.38%</strong></td>
</tr>
<tr>
<td><strong>AVG All Classes</strong></td>
<td><strong>-2.01%</strong></td>
<td><strong>-1.54%</strong></td>
<td><strong>-1.47%</strong></td>
<td><strong>-1.58%</strong></td>
<td><strong>-0.93%</strong></td>
<td><strong>-0.90%</strong></td>
<td><strong>-1.37%</strong></td>
<td><strong>-0.57%</strong></td>
<td><strong>-0.52%</strong></td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Version</th>
<th>END, with CRCO</th>
<th>UP, with CRCO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Y</td>
<td>U</td>
</tr>
<tr>
<td>BQTerrace</td>
<td>-1.75%</td>
<td>-0.69%</td>
</tr>
<tr>
<td>Basket.Drive</td>
<td>-2.24%</td>
<td>-1.64%</td>
</tr>
<tr>
<td>Cactus</td>
<td>-2.35%</td>
<td>-1.95%</td>
</tr>
<tr>
<td>Kimono</td>
<td>-2.42%</td>
<td>-2.33%</td>
</tr>
<tr>
<td>ParkScene</td>
<td>-2.44%</td>
<td>-1.46%</td>
</tr>
<tr>
<td>AVG Class B</td>
<td>-2.24%</td>
<td>-1.61%</td>
</tr>
<tr>
<td>BQMall</td>
<td>-1.97%</td>
<td>-1.63%</td>
</tr>
<tr>
<td>BasketballDrill</td>
<td>-2.00%</td>
<td>-2.17%</td>
</tr>
<tr>
<td>PartyScene</td>
<td>-1.46%</td>
<td>-0.83%</td>
</tr>
<tr>
<td>RaceHorses</td>
<td>-1.84%</td>
<td>-1.20%</td>
</tr>
<tr>
<td>AVG Class C</td>
<td>-1.82%</td>
<td>-1.46%</td>
</tr>
<tr>
<td>BQSquare</td>
<td>-1.00%</td>
<td>-0.56%</td>
</tr>
<tr>
<td>BasketballPass</td>
<td>-1.78%</td>
<td>-1.76%</td>
</tr>
<tr>
<td>BlowingBubbles</td>
<td>-1.69%</td>
<td>-1.74%</td>
</tr>
<tr>
<td>RaceHorses</td>
<td>-2.35%</td>
<td>-1.92%</td>
</tr>
<tr>
<td>AVG Class D</td>
<td>-1.71%</td>
<td>-1.50%</td>
</tr>
<tr>
<td>AVG All Classes</td>
<td>-1.94%</td>
<td>-1.53%</td>
</tr>
</tbody>
</table>
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)
Presentation scheme

• 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
Motivation

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

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

All these solutions require a quality metric
Review of quality metrics on 360VR contents

• Traditional metrics
  o PSNR (PSNR)
  o Structural Similarity Index (SSIM)
  o Multi-Scale SSIM (MS-SSIM)
  o VMAF

• Adaptations of quality metrics to 360VR contents
  o Weighted to Spherically - PSNR (WS-PSNR)
  o 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**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of reference videos</td>
<td>34</td>
</tr>
<tr>
<td>Duration</td>
<td>6 seconds</td>
</tr>
<tr>
<td>Encoding</td>
<td>H.264/AVC</td>
</tr>
<tr>
<td>Resolution</td>
<td>384x288 – 1920x1080</td>
</tr>
<tr>
<td>Bitrate</td>
<td>375 kbps – 20 Mbps</td>
</tr>
</tbody>
</table>

Total of distorted videos: 300

Developed by:
Work approach (I)

• **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

VMAF computation - Objective Analysis

- No temporal pooling challenge
- 4K throughout the process

<table>
<thead>
<tr>
<th>Number of reference videos</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duration</td>
<td>10 seconds</td>
</tr>
<tr>
<td>Encoding</td>
<td>H.265/HEVC</td>
</tr>
<tr>
<td>Resolution</td>
<td>4K (3840x1920)</td>
</tr>
<tr>
<td>Hypothetical Reference Circuits (HRCs)</td>
<td>QP range (1-51)</td>
</tr>
<tr>
<td>Framerate</td>
<td>25 fps</td>
</tr>
</tbody>
</table>

Total number of videos: 459

![Graph showing VMAF scores vs QP Values]
Subjective assessment – Test material

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

**DATASET CHARACTERISTICS**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of source videos</td>
<td>9</td>
</tr>
<tr>
<td>Duration</td>
<td>10 seconds</td>
</tr>
<tr>
<td>Encoding</td>
<td>H.265/HEVC</td>
</tr>
<tr>
<td>Resolution</td>
<td>4K (3840x1920)</td>
</tr>
<tr>
<td>Number of QP values</td>
<td>6</td>
</tr>
</tbody>
</table>

Total of videos: 54
Subjective assessment – Test material

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

DATASET CHARACTERISTICS

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of source videos</td>
<td>9</td>
</tr>
<tr>
<td>Duration</td>
<td>10 seconds</td>
</tr>
<tr>
<td>Encoding</td>
<td>H.265/HEVC</td>
</tr>
<tr>
<td>Resolution</td>
<td>4K (3840x1920)</td>
</tr>
<tr>
<td>Number of QP values</td>
<td>6</td>
</tr>
</tbody>
</table>

Total of videos: 54
Subjective assessment – Test session

Methodology

ACR-HR

<table>
<thead>
<tr>
<th>Grade</th>
<th>Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Excellent</td>
</tr>
<tr>
<td>4</td>
<td>Good</td>
</tr>
<tr>
<td>3</td>
<td>Fair</td>
</tr>
<tr>
<td>2</td>
<td>Poor</td>
</tr>
<tr>
<td>1</td>
<td>Bad</td>
</tr>
</tbody>
</table>

- 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
Experimental Results - MOS
Experimental Results - DMOS

[Bar chart showing quality perception for different scenes and qualities]
VMAF adjustment for 360VR contents

Good fit

Bad fit
PLCC and RMSE between VMAF and DMOS

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

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

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

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.

\[\text{Input: } (\text{bdry}_{\text{top}}, \text{bdry}_{\text{left}}) \quad \text{Output: } \text{pred}\]
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)
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)

<table>
<thead>
<tr>
<th>All Intra</th>
<th>Y in %</th>
<th>Enc Time in %</th>
<th>Dec Time in %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class A1</td>
<td>-1.15</td>
<td>142</td>
<td>98</td>
</tr>
<tr>
<td>Class A2</td>
<td>-0.67</td>
<td>140</td>
<td>99</td>
</tr>
<tr>
<td>Class B</td>
<td>-0.73</td>
<td>143</td>
<td>100</td>
</tr>
<tr>
<td>Class C</td>
<td>-0.72</td>
<td>142</td>
<td>98</td>
</tr>
<tr>
<td>Class E</td>
<td>-0.90</td>
<td>140</td>
<td>100</td>
</tr>
<tr>
<td>Overall</td>
<td>-0.82</td>
<td>142</td>
<td>99</td>
</tr>
</tbody>
</table>
REFERENCES


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 an anti-aliasing filter
- Upsampling is realized by inserting zeros
  - The signal is filtered with an interpolation filter
- MATLAB’s \textit{imresize} function does not strictly follow this methodology, when using the bicubic kernel
  - Samples are shifted when downsampling and shifted back when upsampling

\[
\begin{align*}
\text{s}(n) & \quad \downarrow(n) \quad \text{n} \\
\text{s}_1(n) & \quad \downarrow(1:2) \quad \text{n} \\
\text{s}_2(n) & \quad \downarrow(2:1) \quad \text{n}
\end{align*}
\]
Fundamentals: Downsampling Filters

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

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

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.

\[ h_\uparrow(n) \]

- 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.

\[ h_\uparrow(n) \]

Figure: different upsampling filters
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 $D$

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

\[
\alpha = \arg \min_\alpha \|x - D\alpha\|_2^2 + \lambda \|\alpha_i\|_1
\]

\[
x = D\alpha + \varepsilon
\]

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

Figure: Example Dictionary
DL based SR: Coupled dictionaries approach

\[ I_{LR} \approx \alpha_{164} + \alpha_{171} + \alpha_{179} + \alpha_{206} \]

\[ D_{LR} \]

\[ D_{HR} \]

\[ x_{LR} \approx D_{LR} \alpha \]

\[ x_{HR} \approx D_{HR} \alpha \]

\[ I_{LR} \xrightarrow{1:2} h_{hp} \]

\[ x_{LR} \rightarrow \text{Enc} \rightarrow D_{LR} \]

\[ I_{HR} \xrightarrow{+} \text{patch comb.} \rightarrow X_{HR} \]

\[ X_{HR} \xrightarrow{A} \text{patch ext.} \rightarrow D_{HR} \]
<table>
<thead>
<tr>
<th></th>
<th>Bicubic</th>
<th>HEVC int.</th>
<th>DLSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>BQTerrace</td>
<td>28.2</td>
<td>28.8</td>
<td>30.3</td>
</tr>
<tr>
<td>BasketballDrive</td>
<td>34.7</td>
<td>34.9</td>
<td>36.1</td>
</tr>
<tr>
<td>Cactus</td>
<td>33.5</td>
<td>34.2</td>
<td>35.5</td>
</tr>
<tr>
<td>Campfire</td>
<td>37.8</td>
<td>38.1</td>
<td>38.8</td>
</tr>
<tr>
<td>CatRobot1</td>
<td>39.7</td>
<td>40.2</td>
<td>40.9</td>
</tr>
<tr>
<td>DaylightRoad2</td>
<td>37.4</td>
<td>37.7</td>
<td>38.2</td>
</tr>
<tr>
<td>FoodMarket4</td>
<td>48.5</td>
<td>48.7</td>
<td>48.7</td>
</tr>
<tr>
<td>MarketPlace</td>
<td>40.2</td>
<td>41.1</td>
<td>41.9</td>
</tr>
<tr>
<td>ParkRunning3</td>
<td>40.7</td>
<td>44.2</td>
<td>47.8</td>
</tr>
<tr>
<td>RitualDance</td>
<td>44.6</td>
<td>45.7</td>
<td>48.0</td>
</tr>
<tr>
<td>Tango2</td>
<td>42.4</td>
<td>42.6</td>
<td>42.6</td>
</tr>
<tr>
<td>AVG</td>
<td>39.0</td>
<td>40.0</td>
<td>41.4</td>
</tr>
</tbody>
</table>

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?

<table>
<thead>
<tr>
<th>Option</th>
<th>signaling cost</th>
<th>spacial adaptivity</th>
<th>temporal adaptivity</th>
<th>boundary issues</th>
</tr>
</thead>
<tbody>
<tr>
<td>CU level</td>
<td>high</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>CTU level</td>
<td>moderate</td>
<td>yes</td>
<td>yes</td>
<td>moderate</td>
</tr>
<tr>
<td>TID level</td>
<td>low</td>
<td>no</td>
<td>yes</td>
<td>almost none</td>
</tr>
<tr>
<td>Intra period level</td>
<td>low</td>
<td>no</td>
<td>yes</td>
<td>almost none</td>
</tr>
<tr>
<td>Sequence level</td>
<td>none</td>
<td>no</td>
<td>no</td>
<td>almost none</td>
</tr>
</tbody>
</table>

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 downsamped reconstructed CTU
– decide based on RD-cost which one is coded into the bitstream
Dynamic Resolution Change with SR

- 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 \( \lambda \) has to be adjusted:
  \[ \lambda_{LR} = \frac{\lambda_{HR}}{4} \]
1. Motivation and Fundamentals

2. Dictionary learning based super-resolution

3. Dynamic Resolution Change in Video Coding

4. Experimental Results
<table>
<thead>
<tr>
<th></th>
<th>VTM-3.0 DRC HEVC int.</th>
<th>VTM-3.0 DRC DLSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>BQTerrace</td>
<td>-0.03</td>
<td>-0.04</td>
</tr>
<tr>
<td>BasketballDrive</td>
<td>-0.27</td>
<td>-0.17</td>
</tr>
<tr>
<td>Cactus</td>
<td>-0.05</td>
<td>-0.0</td>
</tr>
<tr>
<td>Campfire</td>
<td>-0.19</td>
<td>-0.23</td>
</tr>
<tr>
<td>CatRobot1</td>
<td>-0.29</td>
<td>-0.25</td>
</tr>
<tr>
<td>DaylightRoad2</td>
<td>-0.05</td>
<td>-0.13</td>
</tr>
<tr>
<td>FoodMarket4</td>
<td>-3.59</td>
<td>-3.58</td>
</tr>
<tr>
<td>MarketPlace</td>
<td>-0.11</td>
<td>-0.01</td>
</tr>
<tr>
<td>ParkRunning3</td>
<td>-0.01</td>
<td>0.08</td>
</tr>
<tr>
<td>RitualDance</td>
<td>-0.48</td>
<td>-0.56</td>
</tr>
<tr>
<td>Tango2</td>
<td>-1.51</td>
<td>-1.43</td>
</tr>
<tr>
<td>AVG</td>
<td>-0.6</td>
<td>-0.57</td>
</tr>
</tbody>
</table>

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

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

*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


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
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)
Motivation

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, ...)

Motivation
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 $P$ from world space to the image plane,
$P: \mathbb{R}^3 \rightarrow \mathbb{R}^2, (X, Y, Z) \mapsto (x, y)$

(Geometric) camera calibration:
Parameter estimation for a model of $P$

Estimation is carried out using
- Point correspondences
  
  and/or
- Known a-priori constraints within the scene
  
  in
- Single or multiple images
Calibration revisited

- 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 toolbox\(^1\) toolchain

Calibration revisited

- CalTech calib toolbox\(^1\) toolchain

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 (OpenCV\(^1\), 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
State of the art fails for:

- Strong blur
- High noise levels
- Asymmetric transitions due to nonlinear response ("gamma")
Novel method

Signal along $\varphi$ is $\pi$-periodic
Novel method

Decentered signal is $2\pi$-periodic in $\varphi$!
Novel method

Fourier analysis of angular signal

\[ \varphi/2\pi \]

\[ c_n \]

\[ \varphi^*/2\pi \]

\[ c_n^* \]
<table>
<thead>
<tr>
<th>Input</th>
<th>Periodic</th>
<th>Infinite</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous</td>
<td>Fourier series</td>
<td>Fourier transform</td>
</tr>
<tr>
<td>Discrete</td>
<td>DFT (Discrete FT)</td>
<td>DTFT (Discrete–Time FT)</td>
</tr>
</tbody>
</table>

Color legend:
- Discrete output
- Continuous output
Novel method

Sub-pixel offset estimation:

Minimize the “wrong” = odd Fourier components, weighted by the ideal decay

\[
x_0^{\text{est}} = \arg \min_{x_0, y_0} \sum_{k=0}^{\infty} b_{2k+1} \left| \int_0^{2\pi} \hat{s}_{\text{blur}}(r^*, \varphi^*) e^{i(2k+1)\varphi^*} \, d\varphi^* \right|^2
\]
Imaging pipeline

Object

Scene illumination

Reflectance $R$ → Intensity $I_0$

Signal $I_0$

Geometry $(X, Y, Z)$

Perspective projection $P \cdot (X, Y, Z, 1)^T$

Lens distortion $\mathcal{L}(x', y')$

Point-spread function PSF$(x, y, d)$

Shot noise

Sensor noise

Quantization $Q(\cdot)$

Image $I(x, y)$
Results

**Synthetic images:**

Absolute localization error w.r.t.

**Blur $\sigma$**

**Distortion angle $\beta$**
Results

Real images (DIMA dataset):
Distribution of reprojection errors
Conclusions

- New method for marker localization exploiting angular symmetry
- Highest positional accuracy
- Robust again 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
Outline

• Motivation

• Automated Optical Inspection Systems

• Potential of Deep Learning

• Challenges with Deep Learning

• Implementation of Demo System

• Conclusion
Spruck: Potential of Deep Learning in the Field of Quality Assurance

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

- Chromosome
- Reads
- Alignments
- Assembly
# Evolution of genome sequencing

## Sequencing technology

<table>
<thead>
<tr>
<th></th>
<th>2009</th>
<th>2018/2019</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost/genome</td>
<td>$100k</td>
<td>~$1k</td>
</tr>
<tr>
<td>Coverage</td>
<td>~30x</td>
<td>&gt; 200x</td>
</tr>
<tr>
<td>Number of reads</td>
<td>~1 billion</td>
<td>&gt; 6 billion</td>
</tr>
<tr>
<td>Size of raw sequencing files</td>
<td>~0.25 TB</td>
<td>&gt; 1.5 TB</td>
</tr>
</tbody>
</table>

## Storage & transmission infrastructure

<table>
<thead>
<tr>
<th></th>
<th>2009</th>
<th>2018/19</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost/TB</td>
<td>$100</td>
<td>$50</td>
</tr>
<tr>
<td>Download speed</td>
<td>10 Mbps</td>
<td>100 Mbps</td>
</tr>
</tbody>
</table>

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

Reference sequence

Target 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

\[ m_i = (p_i, l_i, C_i) \]
GABAC

- GABAC = Genomics-oriented context Adaptive Binary Arithmetic Coding
- GABAC is part of a collaborative effort to produce a standard-compliant open source MPEG-G encoder (genie)
GABAC block diagram

$p_{\text{trans}}$ $p_{\text{lut}}$ $p_{\text{diff}}$ $p_{\text{ctx}}$ $p_{\text{bid}}$ $\pi(p_{\text{bid}})$

Transformations

Sequence Transformation
Equality Coding
RLE Coding
Match Coding
None

LUT Transformation
LUT Order 0
None

Differential Transformation
Diff Coding
None

Binarization
BI
TU
(S)EG
(S)TEG

Context Selection
Bypass
Adaptive Coding (Order 0/1/2)

Entropy Coding

Control signal
Only required when LUT Transformation is enabled
Signal
Average compression ratio (compressed size / uncompressed size):

- GABAC: 0.199%
- bzip2: 0.212%
- gzip: 0.245%
- rANS−0: 0.286%
- rANS−1: 0.237%
- xz: 0.204%

206 test files
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

206 test files
The GABAC configuration space

\[ p_{\text{trans}} \in \mathcal{P}_{\text{trans}} = \{ \text{no\_transform, equality\_coding, match\_coding, rle\_coding} \} \]

\[ n_{\text{ts}} (p_{\text{trans}}) = \begin{cases} 
1, & \text{if } p_{\text{trans}} = \text{no\_transform} \\
2, & \text{if } p_{\text{trans}} = \text{equality\_coding} \\
3, & \text{if } p_{\text{trans}} = \text{match\_coding} \\
2, & \text{if } p_{\text{trans}} = \text{rle\_coding} 
\end{cases} \]

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

\[ p_{\text{lut}} \in \mathcal{P}_{\text{lut}} = \{ \text{disabled, enabled} \} \]
The GABAC configuration space

\[ p_{\text{ctx}} \in \mathcal{P}_{\text{ctx}} = \{ \text{bypass, adaptive\_coding\_order\_0, adaptive\_coding\_order\_1, adaptive\_coding\_order\_2} \} \]

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

\[ \pi (p_{\text{bid}}) = \begin{cases} 1, & \text{if } p_{\text{bid}} = \text{BI} \\ 1, & \text{if } p_{\text{bid}} = \text{TU} \\ 1, & \text{if } p_{\text{bid}} = \text{EG} \\ 1, & \text{if } p_{\text{bid}} = \text{SEG} \\ 32, & \text{if } p_{\text{bid}} = \text{TEG} \\ 32, & \text{if } p_{\text{bid}} = \text{STEG} \end{cases} \]
The GABAC configuration space

Total number of possible configurations $N$:

$N \cong 16{,}000$

Real-world implementation: $4{,}000 < N < 8{,}000$
The optimization problem

\[ x^* = \arg\min_{x \in X} f(x) \]
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

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Choice</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>State space</td>
<td>GABAC parameter space</td>
<td>( { p_{\text{trans}} \cdot p_{\text{diff}} \cdot p_{\text{lut}} \cdot p_{\text{bid}} \cdot p_{\text{ctx}} } )</td>
</tr>
<tr>
<td>Energy (objective) function</td>
<td>Compression ratio ( r )</td>
<td>( 0 &lt; r &lt; \sim 1 )</td>
</tr>
<tr>
<td>Candidate generation procedure</td>
<td>Random neighbor</td>
<td>1 random parameter is changed</td>
</tr>
<tr>
<td>Acceptance probability function</td>
<td>( \Delta E(k) = r(k) - r(k - 1) )</td>
<td>“energy difference” = compression ratio gain</td>
</tr>
<tr>
<td></td>
<td>( P(k) = e^{-\Delta E / T(k)} )</td>
<td></td>
</tr>
<tr>
<td>Annealing schedule</td>
<td>( T(k) = \left( 1 - \frac{k}{k_{\text{max}}} \right) \cdot k_{t} )</td>
<td>( k ) grows ( \rightarrow T(k) ) decreases ( \rightarrow P(k) ) decreases</td>
</tr>
<tr>
<td>Initialization</td>
<td>( T(k = 0) = 1 )</td>
<td>( k_{t} ): hyper parameter</td>
</tr>
<tr>
<td></td>
<td>( k = 0 )</td>
<td>( k_{t} ) is large ( \rightarrow ) worse solutions are accepted</td>
</tr>
<tr>
<td></td>
<td>( k_{\text{max}} = 100 )</td>
<td>more frequently</td>
</tr>
<tr>
<td></td>
<td>( k_{t} = 1 )</td>
<td></td>
</tr>
</tbody>
</table>
Genetic algorithm

Parameter space

\[ n_{\text{individuals}} = 9 \]
\[ n_{\text{generations}} > 3 \]
## Genetic algorithm

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Choice</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>State space</td>
<td>GABAC parameter space</td>
<td>( { p_{\text{trans}} \cdot p_{\text{diff}} \cdot p_{\text{lut}} \cdot p_{\text{bid}} \cdot p_{\text{ctx}} } )</td>
</tr>
<tr>
<td>Objective function</td>
<td>Compression ratio ( r )</td>
<td>( 0 &lt; r &lt; \sim 1 )</td>
</tr>
<tr>
<td>Candidate generation procedure</td>
<td>Random neighbor</td>
<td>2 random parameters are changed randomly</td>
</tr>
<tr>
<td>Acceptance probability function</td>
<td>( i_{\text{best}} = \arg \max_{1 \leq i \leq n_{\text{individuals}}} { r(i) } )</td>
<td>• Best individual in each generation is selected</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No crossover</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• No mutation (replaced by random parameter change)</td>
</tr>
<tr>
<td>Initialization</td>
<td>( n_{\text{individuals}} = 10 ) ( n_{\text{generations}} = 10 )</td>
<td>n/a</td>
</tr>
</tbody>
</table>
### Results and sanity checks on artificial data

<table>
<thead>
<tr>
<th></th>
<th>1 MiB random</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Brute force</td>
<td>Genetic algorithm</td>
<td>Simulated annealing</td>
<td></td>
</tr>
<tr>
<td>Compression ratio</td>
<td>~1</td>
<td>~1</td>
<td>~1</td>
<td></td>
</tr>
<tr>
<td>No. of tested configurations</td>
<td>6,945</td>
<td>10 · 10</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Encoding time</td>
<td>467 s</td>
<td>50 s</td>
<td>64 s</td>
<td></td>
</tr>
</tbody>
</table>

|                  | 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                     |                           |
Algorithm comparison

206 test files

GA: 1.15x @ 10x speed

SA: 1.75x @ 16x speed

Encoding times and compression ratios are normalized to the brute force results.

Better
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

![Graph showing the performance of different types of games with varying latencies.](image-url)
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

REMOTE

Game Engine → GL / SDL Hooks → Control Relay

audio encoder → x265 encoder → Foveation Relay

RTSP Server

LOCAL

Control Client

Foveation Client

x265 decoder

RTSP Client

audio decoder

Display

Headphones

Mouse / KB

Eye Tracker

User
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:

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

With parameters \(\sigma\) and \(QO_{\text{max}}\)
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
Scalable, Object-Centered Multi-Image 3D Reconstruction using Plane Sweep

Johannes Bauer and André Kaup

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

1. Motivation
Multi-image 3D reconstruction is a core task in machine vision and required in many applications, including augmented reality, geoinformation, 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
- Define stereo pairs according to Euclidean minimum spanning tree
- Segment search space as desired into voxel cells
- Compute mapping functions from image plane of camera to plane hypothesis 1

Online processing
- Input images I_{xy}:
- Project input images I_{xy} onto plane hypotheses 11
- Cost metric: Measure photo-consistency between re-projections by Hamming distance c_{xy}(e) = d_{xy}(I_{xy}, I_{xy})
- Aggregate costs from |E| pair-wise matchings in each voxel α to normalized occupancy probability probability score s_{xy} ∈ [0, 1]

3D probability density
- Morphological speckle filter:
- Apply 3D low-pass (LP) filter with kernel length L
- LP-weight W_{xy} of a voxel contained in an ideal flat plane
- Dilatation; drop voxels, whose LP representatives do not exceed fraction γ

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: α = 4, β = 36, L = 5, γ = 0.6

Run times of initialization + single frame

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:

Acknowledgment:
This work was funded by BOSCH Sicherheitssysteme, Nürnberg
Video Coding with Spatial Downscaling and Super-Resolution for 4K Data
Kristian Fischer and André Kaup

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

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):
  4K Video → Standard Encoder → Bitstream → Standard Decoder → 4K Display

- Scenario with restrictions (e.g. limited bandwidth or time)
  4K Video ↓ Bicubic Downscaling with MATLAB imresize() → VP9 Codec → Coded and upscaled 4K Sequence

- 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 down scales the videos before encoding
- Video has to be up scaled at the decoder side
- In the following results are verified by using the VP9 codec with several upsampling algorithms

2. Evaluated Upscaling Algorithms
- Low-complexity Statistical Edge-Adaptive Back-projected Interpolation (L-SEABI) [1]
  - Additionally proposed upsampling 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

3. Simulation Framework
- First 100 frames of 16 sequences from set SITU 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; --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 upsampling algorithms
- Compression artifacts can be reduced through spatial down-/up-scaling for low bitrates
- Bicubic interpolation leads to blurry images
- VP9 + VDSR with best upsampling quality and superior quality up to 1400 kbit/s

References
Model-Based Compression of Genomic Sequences

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

Predictive Coding
Open-loop Prediction
Base seq. \( \rightarrow \) Entropy Coding \( \rightarrow \) Bit stream
\( \rightarrow \) Residual Coding
 predictors

Predictors
Similar to [1]. Mostly used predictors:
- Direct encoding (\( \checkmark \) 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

Selection of best encoding method for each block

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

Model
\[
\text{Construct conditional probability matrix } P = [P(X|Y)]
\]

Analysis
\[
H(X|Y) = \sum_{Y \in \mathcal{A}} P(Y) \left( \sum_{X \in \mathcal{A}} P(X|Y) \log_2 P(X|Y) \right)
\]

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: \( P(X|X_{i-1}, \ldots, X_0) \)

Preliminary Results
- Direct coding of the signal
- Only very local patterns can be considered: 14% of genome: larger-scale repetitive regions
  ⇒ Accounting for large patterns necessary

Residue Coding 2
- Use arithmetic coder
- Use predicted value of each position as context model

Example
Prediction \( G \ A \ C \ T \ldots \) \( \rightarrow \) Residue
\( G \ T \ C \ G \ldots \)
Actual value \( G \ T \ C \ G \ldots \)
\( X \) \( \rightarrow \) Context
\( P(G|G) \ P(T|A) \ P(C|C) \ P(G|T) \); \( \ldots \)

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

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

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]

![Diagram of Temporal scalability](image)

- **Input video sequence**
- **Output:** Temporal WT and Spatial WT
- **Quality and spatial scalability**
- **By realizing P as the warping operator Y, Motion Compensated Temporal Filtering (MCTF) is achieved [2]:**
  \[ h_{2i} = l_{2i} - |W_{2i-1} \cdot l_{2i-1}| \]
  \[ l_{2i-1} = l_{2i-1} + \frac{1}{2} W_{2i-2} \cdot h_{2i} \]

2. Content Adaptive Wavelet Lifting (CA-WL)

- **Idea:** Adaptive temporal scaling based on significant changes among subsequent frames
- **Stopping Criterion:**
  - Haar WT can be represented with tree structures
  - With each node a basis vector \( b_i \) and a wavelet coefficient vector \( c_i \) is associated, which is the inner product of the signal \( s \) with the basis \( b_i \)
  - If combined costs of child nodes exceed costs of parent node, i.e.
    \[ C(s, b_i) = C(s, b_{i1}) + C(s, b_{i2}) \]
  - child nodes shall be pruned from the tree
  - \( C(s) \) describes a Lagrangian cost functional, which represents the coding costs:
    \[ C(s, b_i) = D(s, b_i) + \lambda R(s, b_i) \]
- **Rate** \( R(s, b_i) \) 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** \( D(s, b_i) \) is calculated by the MSE of the corresponding wavelet coefficients compared to the original signal according to [3]

3. Experimental Results

- **Simulation Setup (8 bpps):**
  - Spatial resolution | Number of frames
  - 480 x 320 | 500
  - 480 x 320 | 500
  - 480 x 320 | 500
  - 480 x 320 | 500
- **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 4
  - Motion vectors are encoded using the QcPac library [6]

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, PSNRp, of BL is increased by 10.28 dB and rate is reduced by 1.06%**

References:

MLSP-IQA: Weak supervision for deep IQA feature learning

Hanhe Lin
Department of Computer and Information Science
University of Konstanz

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

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

MLSP-IQA

- Multi-Task Learning (MTL) for FR-IQA score prediction

<table>
<thead>
<tr>
<th>Method</th>
<th>PLCC</th>
<th>SROCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSIM</td>
<td>0.723</td>
<td>0.724</td>
</tr>
<tr>
<td>MSSSIM</td>
<td>0.801</td>
<td>0.802</td>
</tr>
<tr>
<td>IWSSIM</td>
<td>0.846</td>
<td>0.850</td>
</tr>
<tr>
<td>MSII</td>
<td>0.873</td>
<td>0.872</td>
</tr>
<tr>
<td>VIF</td>
<td>0.878</td>
<td>0.879</td>
</tr>
<tr>
<td>FSIM</td>
<td>0.851</td>
<td>0.854</td>
</tr>
<tr>
<td>GMSD</td>
<td>0.847</td>
<td>0.847</td>
</tr>
<tr>
<td>SFF</td>
<td>0.862</td>
<td>0.862</td>
</tr>
<tr>
<td>SCQI</td>
<td>0.853</td>
<td>0.854</td>
</tr>
<tr>
<td>ADD-GSIM</td>
<td>0.817</td>
<td>0.818</td>
</tr>
<tr>
<td>SR-SIM</td>
<td>0.834</td>
<td>0.839</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>PLCC</th>
<th>SROCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>BIQI</td>
<td>0.460</td>
<td>0.431</td>
</tr>
<tr>
<td>BLINDS-III</td>
<td>0.559</td>
<td>0.527</td>
</tr>
<tr>
<td>BRISQUE</td>
<td>0.554</td>
<td>0.519</td>
</tr>
<tr>
<td>CORNIA</td>
<td>0.580</td>
<td>0.541</td>
</tr>
<tr>
<td>DIVINE</td>
<td>0.532</td>
<td>0.489</td>
</tr>
<tr>
<td>HOISA</td>
<td>0.653</td>
<td>0.609</td>
</tr>
<tr>
<td>SSEQ</td>
<td>0.463</td>
<td>0.424</td>
</tr>
<tr>
<td>InceptionResNetV2 (fine-tune)</td>
<td>0.734</td>
<td>0.731</td>
</tr>
<tr>
<td>MLSP-IQA</td>
<td>0.941</td>
<td>0.939</td>
</tr>
</tbody>
</table>

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 MLSP-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

References


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

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

5. Optimization procedure

- $PSNR_a = 10 \log_{10} \frac{255^2}{D_a}$ and $PSNR_d = 10 \log_{10} \frac{255^2}{D_d}$
- $\Delta PSNR = PSNR_a - PSNR_d = 10 \log_{10} \frac{D_a + D_d}{D_d}$
- 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
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
- Improvement visually apparent

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

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
  - 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 only 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 uncodable 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

#### Proposed supplemental enhancement information (SEI) message syntax

```plaintext
for r < r_max do
  for c < c_max do
    for l < l_max do
      if boundary[boundary_paddingPresent ← 1 bit]
      do
        for v ∈ top, bottom, left, right do
          if v = 0
            continue
          boundary.paddingPresent ← 0
          boundary.paddingWidth ← numRefIdxMinus1 + 1
          do
            if r = 0
              do
                for v ∈ top, bottom, left, right do
                  if v = 0
                    do
                      boundary.paddingPresent ← 1 bit
                      boundary.paddingWidth ← numRefIdxMinus1 + 1
                      Geometry padding amount in pixels
```

### Results

<table>
<thead>
<tr>
<th>BD-Rate in %</th>
<th>Decoder complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequence (no SEI)</td>
<td>no SEI, with SEI</td>
</tr>
<tr>
<td>static</td>
<td>0.85, 0.63</td>
</tr>
<tr>
<td>Gaasang</td>
<td>-0.52, 0.34</td>
</tr>
<tr>
<td>Haar</td>
<td>-0.57, 0.72</td>
</tr>
<tr>
<td>Trolley</td>
<td>0.03, 0.94</td>
</tr>
<tr>
<td>Ballosa</td>
<td>3.31, 3.15</td>
</tr>
<tr>
<td>BranCarved</td>
<td>3.31, 3.11</td>
</tr>
<tr>
<td>Broadway</td>
<td>2.55, 2.42</td>
</tr>
<tr>
<td>non-static</td>
<td>3.44, 3.31</td>
</tr>
<tr>
<td>Landing2</td>
<td>2.25, 2.20</td>
</tr>
<tr>
<td>Skateboard2</td>
<td>3.11, 3.00</td>
</tr>
<tr>
<td>static</td>
<td>-0.79, -0.65</td>
</tr>
<tr>
<td>Average</td>
<td>-2.37, -2.86</td>
</tr>
<tr>
<td>all</td>
<td>-2.70, -1.98</td>
</tr>
</tbody>
</table>

#### 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 (ITC360) [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

Observations

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

Analysis of geometry padding usage

<table>
<thead>
<tr>
<th>Derivation of geometry padding usage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inspection of encoder motion buffer</td>
</tr>
<tr>
<td>Derivation of required padding separately for each face boundary</td>
</tr>
</tbody>
</table>

References


Institut für Nachrichtentechnik, RWTH Aachen University

sauer@ient.rwth-aachen.de
www.ient.rwth-aachen.de

Institut für Nachrichtentechnik, Melatenstr. 23, 52074 Aachen

SVCP 2019