Leibniz Universität Hannover

tnt

Proceedings of the

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

4–6 July 2018

Hannover, Germany

Edited by

Jan Voges

Published by

Leibniz Universität Hannover
Institut für Informationsverarbeitung
Appelstr. 9A
30167 Hannover
Germany

Contents

1 Scientific chairs

Jörn Ostermann

Jörn Ostermann received a Dr.-Ing. degree from University of Hannover in 1994. In 1994 he joined the Visual Communications Research Department at AT&T Bell Labs. Since 2003 he is Full Professor at Leibniz Universität Hannover, Germany. Since 2008, he is the Chair of the Requirements Group of MPEG.

He is a Fellow of the IEEE (class of 2005). Jörn served as a Distinguished Lecturer of the IEEE CAS Society (2002/2003). He published more than 100 research papers and book chapters. He holds more than 30 patents.

André Kaup

André Kaup (M'96–SM'99–F'13) received the Dipl.-Ing. and Dr.-Ing. degrees in electrical engineering from Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen University, Aachen, Germany, in 1989 and 1995, respectively.

He was with the Institute for Communication Engineering, RWTH Aachen University, from 1989 to 1995. He joined the Networks and Multimedia Communications Department, Siemens Corporate Technology, Munich, Germany, in 1995 and became Head of the Mobile Applications and Services Group in 1999. Since 2001 he has been a Full Professor and the Head of the Chair of Multimedia Communications and Signal Processing, University of Erlangen-Nuremberg, Erlangen, Germany. From 1997 to 2001 he was the Head of the German MPEG delegation. He has authored around 350 journal and conference papers and has over 80 patents granted or pending.

André Kaup is a member of the IEEE Multimedia Signal Processing Technical Committee, a member of the scientific advisory board of the German VDE/ITG, and a Fellow of the IEEE.

Jens-Rainer Ohm

Jens-Rainer Ohm holds the chair position of the Institute of Communication Engineering at RWTH Aachen University, Germany since 2000. His research and teaching activities cover the areas of multimedia signal processing, analysis, compression, transmission and content description, including 3D and VR video applications, bio signal processing and communication, application of deep learning approaches in the given fields, as well as fundamental topics of signal processing and digital communication systems.

Since 1998, he participates in the work of the Moving Picture Experts Group (MPEG). He has been chairing/co-chairing various standardization activities in video coding, namely the MPEG Video Subgroup 2002 – 2018, the Joint Video Team (JVT) of MPEG and ITU-T SG 16 VCEG 2005 – 2009, the Joint Collaborative Team on Video Coding (JCT-VC) since 2010, as well as the Joint Video Experts Team (JVET) since 2015.

Prof. Ohm has authored textbooks on multimedia signal processing, analysis and coding, on communication engineering and signal transmission, as well as numerous papers from the fields mentioned above.

2 Abstracts

(listed alphabetically by last name of first author)

3D Models in Motion Compensation – Bakhshi Golestani, Hossein
RWTH Aachen University, Institut für Nachrichtentechnik

The main idea of this presentation is to employ the 3D geometry of scene to generate a more precise prediction for video coding. The proposed method first extracts the point cloud and the 3D model of the scene based on a subset of 2D key-frames and then employs it to predict the remaining frames by virtually synthesizing them. Structure from Motion (SfM) and Multi-View Reconstruction (MVR) have been employed for estimating camera parameters and constructing the 3D model of the scene. The virtual views are then synthesized based on the 3D model and Depth Image Based Rendering (DIBR) methods. This 3D model-based prediction is then offered to HEVC as an additional reference for motion compensation. It means HEVC could choose between our prediction and its own built-in prediction through a rate-distortion optimization. Our simulation results show 2.5% and 8.7% bitrate reduction for dynamic and static scenes, respectively.

Robust Super-Resolution for Mixed-Resolution Image Data Using Displacement-Compensated Projection – Bauer, Johannes
Friedrich-Alexander-Universität Erlangen-Nürnberg, Chair of Multimedia Communications and Signal Processing

Multi-view super-resolution techniques aim for the increase of spatial resolution and image quality by exploiting adjacent high-resolution perspectives. By projecting their high-frequency parts into an up-scaled low-resolution image, the level of detail can be strongly enhanced. However, the projection of information from neighboring views typically requires exact depth maps, since even small errors in depth data can lead to annoying artifacts in the resulting images.

Therefore, a displacement compensation approach is presented which aims at correcting resulting errors after projection, introduced by inaccurate depth maps. This leads to a super-resolution technique that is far more robust against noisy depth data and non-ideal depth sensor calibration.

Multiple Feature-based Classifications Adaptive Loop Filter – Erfurt, Johannes
Fraunhofer Heinrich Hertz Institute

In video coding, adaptive loop filter (ALF) has attracted attention due to its increasing coding performances. The main idea of ALF is to apply a classification to obtain multiple classes, which gives a partition of a set of all pixel locations. After that, a Wiener filter is applied for each class. Therefore, the performance of ALF essentially relies on how its classification behaves. Here, we introduce a novel classification method, Multiple feature-based Classifications ALF (MCALF) extending a classification in GALF and show that it increases coding efficiency while only marginally raising encoding complexity. The key idea is to apply more than one classifier at the encoder to group all reconstructed samples and then to select a classifier with the best RD performance to carry out the classification process.

Scalable Lossless Video Coding – Fischer, Kristian

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

Commonly, videos are coded before transmitting them over a channel in order to save bits. Thereby, some information is lost. Indeed, for several professional applications, e.g., medical Computed Tomography, a lossless video transmission is indispensable.

Therefore, an existing approach takes the lossy coded video as base layer (BL). If needed, the residual image between the original frame and the coded frame is additionally transmitted with the Sample-based Weighted Prediction for Enhancement Layer Coding (SELC) codec afterwards. This results in an on-demand lossless video transmission.

In this presentation the SELC and its components are presented first. Thereby, the focus is set on the Context-Based Adaptive Binary Arithmetic Coding (CABAC). The presentation is concluded by discussing the results when combining the SELC with different BL codec implementations like HM, JM, and VP9.

Real-time Video Conferencing – Gehlert, Alexander

LogMeIn Germany GmbH, Dresden

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

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

Demonstration of Rapid Frequency Selective Reconstruction for Image Resolution Enhancement – Genser, Nils

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

For processing, transmitting, or displaying images, the samples are required being placed on a regular grid, in general. Indeed, there are scenarios where the samples to be processed are located on non-integer positions, e.g., in image super-resolution, image warping or when utilizing random sampling sensors.

All these applications have in common that a suitable reconstruction method is required for resampling the pixels to a regular grid. A method, which achieves a very high reconstruction quality is the Frequency Selective Reconstruction, which outperforms competing state-of-the-art algorithms in reconstruction quality. As well as the competing methods, it is computational demanding as the complete images has to be processed.

In this contribution, the latest research was bundled to exploit frequency, signal, and loss pattern properties and to decrease computational complexity. Consequently, a demonstration system is shown, which reconstructs HD image content in real-time.

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

Non-regular sampling can be used instead of regular sampling to reduce aliasing and therefore increase the resolution per pixel. As in other compressed sensing problems, the missing data needs to be reconstructed on a high-resolution grid subsequent to the acquisition.

It turns out, that the actual 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 masks, i.e., an increased reconstruction quality. In terms of the reconstruction method, we highlight Frequency Selective Reconstruction being well-suited for such tasks and leading to a high reconstruction quality.

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

Head pose estimation of a person has many applications such as attention modeling, gaze estimation or fitting 3D face models for face alignment or animation. Recently, the state-of-the-art method to estimate head pose from an image has changed from using geometric key points to using convolutional neural networks (CNNs). CNN based methods provide several advantages to other methods such as robustness to partial occlusions. This work summarizes our current findings working on the problem of head pose estimation.

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

The current state-of-the-art for standardized video codecs is High Efficiency Video Coding (HEVC) which was developed jointly by ISO/IEC and ITU-T. Recently, the development of two contenders for the next generation of standardized video codecs began:

ISO/IEC and ITU-T advance the development of the Joint Exploration Model (JEM), a possible successor of HEVC, while the Alliance for Open Media forwards the video codec AV1.

It is asserted by both groups that their codecs achieve superior coding efficiency over the state-of-the-art. We discuss the distinguishing features of JEM and AV1 and evaluate their coding efficiency and complexity under well-defined and balanced test conditions.

Rate-Distortion Theory for Affine Motion Compensation in Video Coding – Meuel, Holger
Leibniz Universität Hannover, Institut für Informationsverarbeitung

In this work, the rate-distortion function for video coding using affine global motion compensation is derived. The displacement estimation error during motion estimation is modeled and the bit rate after applying the rate-distortion theory is obtained. The displacement estimation error is assumed to be caused by a perturbed affine transformation. The 6 affine transformation parameters are assumed statistically independent, with each of them having a zero-mean Gaussian distributed estimation error. Based on that, the joint probability density function (p.d.f.) of the displacement estimation errors is derived and related to the prediction error. Using the rate-distortion theory, the bit rate in

dependence of the perturbation of the affine transformation parameters is calculated. Comparing with a translational motion model in video coding standards like HEVC, accuracy boundaries for the affine transformation are determined, with which a gain can be achieved.

Computer-assisted Optical Verification of Precast and Reinforced Concrete Structures – Munderloh, Marco

Leibniz Universität Hannover, Institut für Informationsverarbeitung

Quality control and inspection as well as documentation are important issues on construction sites to ensure the stability and safety of the constructed buildings. Concerning reinforced concrete elements, the contained steel bars as well as the resulting concrete covering need to be inspected prior to filling in the concrete as an inspection would not be possible afterwards. Due to the short time frames between finishing of the construction, arrival of the concrete trucks and the availability of inspection engineers, this task might not always be handled out thoroughly. We are working on computer-assisted inspection systems based on computer vision methods to compare the constructed parts to the pre-planed models which are more and more available due to the spreading of building information modeling (BIM). We present three different methods in this field: the registration and validation of reinforcement bars in reinforced concrete elements, the registration and validation of precast concrete elements and the identification of precast concrete elements using auto-id techniques.

Motion Adapted Three-Dimensional Frequency Selective Extrapolation – Spruck, Andreas

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

High resolution images can be acquired using a low-resolution sensor with non-regular sampling. Therefore, post-processing is necessary. In terms of video data, not only the spatial neighborhood can be used to support the reconstruction, but also the temporal neighborhood. A popular algorithm for this kind of problem is the three-dimensional frequency selective extrapolation (3D-FSE) for which a novel spatial weighting function is introduced here. The proposed weighting function solves the problem of changing content within the area considered by the 3D-FSE, which is caused by motion within the sequence. Due to this motion, it may happen that the weighting function emphasizes regions that are not present in the original signal within the considered area. By that, false content is introduced into the extrapolated sequence, which affects the resulting image quality negatively. The novel weighting function incorporates motion data of the sequence in order to adapt the function accordingly, and weights the regions corresponding to their content.

MPEG-G: The Standard for Genomic Information Representation – Voges, Jan

Leibniz Universität Hannover, Institut für Informationsverarbeitung

In its 30 years of activity ISO/IEC JTC 1/SC 29/WG 11 – also known as Moving Picture Experts Group (MPEG) – has developed many generations of successful standards that have transformed the world of media from analog to digital. Four years ago, MPEG started to investigate the compression of the abundance of data produced by high-throughput sequencing machines. Today, the MPEG-G standard (ISO/IEC 23092 Genomic Information Representation) is close to conclusion. The standard will offer

high levels of compression, about one order of magnitude better than currently used formats. By utilizing the latest technologies to compress and transport sequencing data, the MPEG-G standard will provide new functionalities such as enforcement of privacy rules, selective encryption, and many more. Moreover, MPEG-G makes use of well-established MPEG technologies to provide interoperable streaming and selective data access such as those available today in the most advanced media and internet applications.