Guest Editorial
Screen Content Video Coding and Applications

I. Introduction

THIS special issue aims to present recent technical advances in screen content video coding and applications. Screen content video has recently evolved from a niche to the mainstream due to the rapid advances in mobile and cloud technologies. Real-time, low-latency transport of screen visuals between devices in the form of screen content video is becoming prevalent in many applications, e.g., wireless displays, screen mirroring, display interfaces, screen/desktop virtualization and cloud-based mobile virtual reality. Today's commonly-used video coding methods, however, have been developed primarily with camera-captured content in mind. These new applications create an urgent need for efficient coding of screen content video, especially as the support of 4k or even 8k resolution begins to achieve mass market appeal.

Screen content video coding poses numerous challenges. Such content usually features a mix of computer-generated graphics, text, and camera-captured images and video. With their distinct signal characteristics, content-adaptive coding becomes necessary. Moreover, given the varied level of the human's visual sensitivity to distortion in different types of content; visually or mathematically lossless quality may be required for all or part of the video.

Recognizing the demand for an industry standard for coding of screen content, standardization organizations started to define standards in their respective domains. In early 2016, the ISO/IEC Moving Picture Experts Group and ITU-T Video Coding Experts Group concluded their two-year joint standardization project, adding screen content coding extensions to HEVC/H.265 (referred hereafter to as HEVC-SCC). Likewise, the Video Electronics Standards Association (VESA) recently updated their Display Stream Compression (DSC) standard published in mid-2014 to include more features for next-generation television and cinema display interfaces. At the time of writing, the ISO/IEC Joint Picture Experts Group was working on JPEG XS, a new standard aimed at defining a low-latency lightweight coding system with potential use in screen content coding and other applications. The development of these standards is expected to inspire more future innovations and benefit the varied applications of screen content coding.

In addition to compression efficiency, the whole spectrum of screen content video coding and applications has many additional challenging aspects to be addressed. For instance, obtaining objective quality metrics for predicting the perceptual quality of compressed screen content video has been found to be extremely difficult due to the varied content characteristics. As such, how to optimize the encoding process with regard to the varying rate-distortion characteristics arises naturally as a challenging issue. Another challenge is that for some applications, e.g., cloud-based mobile virtual reality, the system complexity could grow exponentially due to the need to optimize over many adjustable parameters and constraints, such as the graphics rendering quality, compression level for individual stereo views, and network conditions. For other applications, hardware implementation of screen content codecs could be the only viable option. The needs for striking a balance among throughput, latency, power consumption, and cost often impose harsh constraints on the design. The pre-/post-processing of screen content, such as segmentation among different content types in a picture, is another area full of research opportunities. There are more challenges to be added. The list could never be exhaustive.

Addressing these issues requires research from multiple disciplines. The intent of this special issue is to present the latest developments in standards, algorithms, and system implementations related to the coding and processing of screen content video.

II. Organization and Overview

This special issue accepted 15 papers in several key areas, namely 1) screen content coding standards, 2) application-specific screen content coding, 3) mathematically lossless screen content coding, 4) machine learning for mode decision and quality assessment, 5) cloud-based mobile streaming, 6) string matching hardware, and 7) screen content segmentation.

We begin with a tutorial-style paper, "Overview of Screen Content Video Coding: Technologies, Standards, and Beyond," by the Guest Editors and other experts, which aims to prepare the reader for delving deeply into individual papers in this issue through a comprehensive introduction to relevant technology developments over the past decade. This paper includes an exclusive account of two recently published international standards, i.e., HEVC-SCC and DSC, and the upcoming JPEG XS standard from a comparative perspective that attempts to distinguish between their uses, design requirements, building elements, and performance. It then concludes with an outlook for future research and developments, some of which are nicely echoed and responded with solutions in the other accepted papers.

A. Screen Content Coding Standards

Following this high level overview, several papers present in greater detail the key elements in the aforementioned standards. For example, "Intra Block Copy in HEVC Screen Content Coding Extensions" by Xu *et al.* details how the redundancy of repetitive patterns in screen content is utilized in HEVC-SCC via a block-based prediction from the reconstructed current picture. Part of its central theme is around the

Digital Object Identifier 10.1109/JETCAS.2016.2631958

efficient reuse of existing syntax for inter-picture prediction. In addition, to address the discrete-tone nature of screen content that a local area often contains few distinct colors, "Palette Mode Coding in HEVC Screen Content Coding Extension" by Pu *et al.* introduces the palette mode design in HEVC-SCC, with a particular focus on the generation of representative colors for a block of pixels and the coding of these chosen colors and palette indices used to associate them with pixel values. On top of these standardized approaches, Sun *et al.* introduce, among others, improved merge candidates for intra block copy and a transition copying technique for palette index coding in "Improvements of HEVC SCC Palette Mode and Intra Block Copy," demonstrating the potential for further enhancements in coding efficiency. Another key aspect of the HEVC-SCC is its ability to process native RGB inputs, where different color components often exhibit strong correlations. The paper "Adaptive Color-space Transform in HEVC Screen Content Coding" by Li *et al.* presents the design details of HEVC-SCC's in-loop adaptive color-space transform. Extensive experiments are conducted to show its interaction with another tool, cross-component prediction, which addresses similar inter-component redundancy and was adopted into an earlier version of HEVC/H.265.

Besides the papers dedicated to HEVC-SCC, this section gains broader scope by also including papers related to DSC and JPEG XS. Specifically, "VESA Display Stream Compression for Television and Cinema Applications" by Walls and MacInnis overviews the DSC standard and the newly added features in its latest release (version 1.2). The other paper "Quality and Error Robustness Assessment of Low-Latency Lightweight Intra-Frame Codecs for Screen Content Compression" by Willeme *et al.* discloses the objective quality assessment methods for use in JPEG XS's standardization. A few baseline/anchor schemes, e.g., DSC, VC-2, JPEG-2000, and HEVC-SCC, are thus analyzed.

B. Application-Specific Screen Content Coding

The applications for lightweight screen content coding as targeted by DSC and JPEG XS appear in many varieties. "Fixed-Ratio Compression of an RGBW Image and Its Hardware Implementation" by Kim *et al.* is one of the few early attempts to address lightweight RGBW image compression for the increasingly popular RGBW displays. The paper points out the unique design constraints associated with this application, and it proposes a second-order spatial prediction and fixed-length Golomb-Rice coding along with their hardware implementations.

C. Mathematically Lossless Screen Content Coding

Instead of allowing for some distortion while being lightweight, compression achieving mathematically lossless quality is often more desirable in professional contexts, e.g., digital operating rooms and post-production in studios. With this observation, "DPCM-based Edge Prediction for Lossless Screen Content Coding in HEVC" by Sanchez *et al.* enhances HEVC-SCC's capability for lossless coding by introducing a sample-based edge and angular prediction technique. The idea is to generalize the conventional intra-picture prediction to allow a better predictor to be selected for each individual pixel depending on the signal pattern in its causal neighborhood. Exploiting similar signal correlations yet in a different way, "Lossless Intra Compression of Screen Content based on Soft Context Formation" by Strutz departs from the framework of HEVC-SCC to propose a standalone pixel-based arithmetic codec for screen content compression. The soft context formation merges the probabilities of a pixel conditioned on similar rather than exact causal surrounding patterns, in order to better estimate its contextual probability for arithmetic coding.

D. Machine Learning for Mode Decision and Quality Assessment

The use of machine learning in video coding has great potential. Its application to screen content coding is no exception. As an early attempt to speed up the mode decision process in the HEVC-SCC reference software, "Fast Mode and Partition Decision Using Machine Learning for Intra-Frame Coding in HEVC Screen Content Coding Extension" by Duanmu *et al.* formulates the mode and block partition decisions as typical classification problems. They adopt decision tree classifiers along with hand-crafted features for their low implementation complexity and interpretable results.

As well as being useful for fast mode decision, machine learning also sees some success in developing objective metrics for image quality assessment. "Subjective and Objective Quality Assessment of Compressed Screen Content Images" by Wang *et al.* introduces a reduced-reference image quality model for HEVC and HEVC-SCC. The model builds on a support vector regressor to fit the divergence between the compressed and reference screen images in the wavelet feature domain to the perceptual quality. Their dataset has been made publicly accessible to promote further developments.

E. Cloud-Based Mobile Streaming

Cloud-based mobile streaming is opening up a new paradigm for enjoying realistic, immersive 3-D content on mobile devices. This new architecture involves graphics rendering on remote servers with resulting images streamed to the client in video form. To ensure the user's experience, "JAVRE: A Joint Asymmetric Video Rendering and Encoding Approach to Enable Optimized Cloud Mobile 3D Virtual Immersive User Experience" by Lu and Dey discusses a three-pronged joint optimization approach, where the quality of graphics rendering and video encoding can both be adapted in an asymmetric manner to network conditions for two stereo views. The adaptation is guided by an analytical user experience model that reflects the perceptual quality in response to the parameter settings.

F. String Matching Hardware

The notion of string matching lays the foundation of intra block copy and palette index coding in HEVC-SCC. Whether this technique can be effectively implemented in hardware has been a long-debated question. In attempting to provide an answer, "A Hardware Decoder Architecture for General

String Matching Technique" by Zhou *et al.* presents a low-cost hardware decoder for general string matching. The design has the remarkable features of supporting real-time 4K video decoding and being able to perform intra block copy and palette decoding in two special modes.

G. Screen Content Segmentation

Image segmentation often accompanies screen content applications as an essential pre-/post-processing step for text extraction, region adaptive coding, etc. "Screen Content Image Segmentation Using Robust Regression and Sparse Decomposition" by Minaee and Wang tackles the segmentation of a screen image into its background and foreground parts. The process is done on a block-by-block basis with two distinct approaches: one performing hard segmentation of a block by fitting its background pixels to a smooth function using the random sample consensus algorithm; and the other decomposing the block of pixels in a soft manner into a smooth (background) and a sparse (foreground) component through solving a convex optimization problem with sparsity constraints.

III. Paper Selection Process and Acknowledgement

This special issue has a highly selective review process, in order to comply with the high-quality standards of the IEEE Journal on Emerging and Selected Topics in Circuits and Systems. In total, we received 42 submissions, of which 22 survived the first round of review and 15 were accepted after the subsequent rounds. All accepted papers went through two revisions, and some even more with consent from the Editor-in-Chief in order to meet the quality requirements. The review of each paper involved at least 3 independent reviewers, with criteria emphasizing equally on quality and fitness to the scope of this issue. We regret that some good papers had to be rejected due to limited space and the needs for balanced coverage of different areas.

With this relentless pursuit of quality, we are very grateful to still have a sizeable number of papers accepted. Notably, all the papers hit closely the page limit 14, the average being 12.8. The Guest Editors would like to thank everyone who made submissions to this issue and committed much time to revisions to meet our quality expectations. We are also deeply indebted to anonymous reviewers for their time and effort to help us make difficult decisions, especially because they had to work under high pressure to ensure the timely publication of this issue.

Last but not least, we are grateful to the Editor-in-Chief, Dr. Yen-Kuang Chen, the Deputy Editor-in-Chief, Prof. Eduard Alarcon, as well as the Senior Editorial Board of this journal for their support and advice in difficult situations. We also want to thank Alison Larkin and other IEEE Publishing Operations personnel for helping us work through many problems during the migration to the new submission system. We hope you enjoy the contents of this special issue and will follow up with contributing more to advance this field.

WEN-HSIAO PENG, *Guest Editor*
Department of Computer Science
National Chiao Tung University
Hsinchu 30010, Taiwan

JIZHENG XU, *Guest Editor*
Microsoft Research Asia
Beijing 100080, China

ROBERT A. COHEN, *Guest Editor*
Mitsubishi Electric Research Laboratories
Cambridge, MA 02139 USA

JÖRN OSTERMANN, *Guest Editor*
Institut für Informationsverarbeitung
Leibniz Universität Hannover
D-30167 Hannover, Germany

Wen-Hsiao Peng (M'09–SM'13) received the B.S., M.S., and Ph.D. degrees from National Chiao Tung University (NCTU), Hsinchu, Taiwan, in 1997, 1999, and 2005, respectively, all in electronics engineering.

He was with the Intel Microprocessor Research Laboratory, Santa Clara, CA, USA, from 2000 to 2001, where he was involved in the International Organization for Standardization (ISO) Moving Picture Experts Group (MPEG)-4 fine granularity scalability and demonstrated its application in 3-D peer-to-peer video conferencing. Since 2003, he has actively participated in the ISO MPEG digital video coding standardization process and contributed to the development of the High Efficiency Video Coding (HEVC) standard and MPEG-4 Part 10 Advanced Video Coding Amd.3 Scalable Video Coding standard. His research group at NCTU is one of the few university teams around the world that participated in the Call-for-Proposals on HEVC and its Screen Content Coding extensions. He is currently an Associate Professor with the Computer Science Department, NCTU. He was a Visiting Scholar with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA, from 2015 to 2016. He has authored over 60 technical papers in the field of video/image processing and communications and over 50 standard contributions. His research interests include video/image coding, multi-modality data analytics, and machine learning.

Dr. Peng is a Technical Committee Member of the Visual Signal Processing and Communications and Multimedia Systems and Application tracks of the IEEE Circuits and Systems Society. He organized several special sessions on HEVC and related topics in prestigious conferences and was a Technical Program Co-Chair for the Conference on Visual Communications and Image Processing in 2011.

Jizheng Xu (M'07–SM'10) received the B.S. and M. S. degrees in computer science from the University of Science and Technology of China (USTC), Hefei, China, and the Ph.D. degree in electrical engineering from Shanghai Jiaotong University, Shanghai, China.

He joined Microsoft Research Asia (MSRA) in 2003 and currently he is a Lead Researcher. He has authored and co-authored over 100 conference and journal refereed papers. He has over 30 U.S. patents granted or pending in image and video coding. His research interests include image and video representation, media compression, and communication. He has been an active contributor to ISO/MPEG and ITU-T video coding standards. He has over 30 technical proposals adopted by H.264/AVC, H.264/AVC scalable extension, High Efficiency Video Coding, HEVC range extension and HEVC screen content coding standards.

Dr. Xu chaired and co-chaired the ad-hoc group of exploration on wavelet video coding in MPEG, and various technical ad-hoc groups in JCT-VC, e.g., on screen content coding, on parsing robustness, on lossless coding. He co-organized and co-chaired special sessions on scalable video coding, directional transform, high quality video coding at various conferences. He also served as special session co-Chair of IEEE International Conference on Multimedia and Expo 2014.

Robert A. Cohen (S'85–M'90–SM'12) received the B.S. (summa cum laude) and M.Eng. degrees in computer and systems engineering, and the Ph.D. degree in electrical engineering from Rensselaer Polytechnic Institute, Troy, NY, USA.

He held positions with IBM, San Jose, CA, USA, and Harris RF Communications, Rochester, NY, USA. From 1991 to 2001, he was a Senior Member of the Research Staff with Philips Research, Briarcliff Manor, NY, USA, where he performed research in areas related to the Grand Alliance HDTV decoder, rapid prototyping for VLSI video processing, statistical multiplexing, scalable MPEG-4 video streaming, and next-generation video surveillance systems. He has been a Principal Research Scientist with Mitsubishi Electric Research Laboratories, Cambridge, MA, USA, since 2007, where he performs research, publishes, and generates patents related to next-generation video coding, screen-content coding, perceptual image/video coding and processing, and point cloud compression. His research interests include video coding and communications, video, image, and signal processing, and 3D point cloud compression.

Dr. Cohen organized the Special Session on Screen Content Coding in PCS 2013. He was a Guest Editor of Signal Processing: Image Communication of the Special Issue on Advances in High Dynamic Range Video Research. He also actively participates in ISO/MPEG and ITU standardization activities, including chairing several ad hoc groups and core experiments, contributing to High Efficiency Video Coding-related call for proposals and drafting the Joint Call for Proposals for coding of screen content in JCT-VC.

Jörn Ostermann (SS'86–M'88–SM'00–F'05) studied electrical engineering and communications engineering at the University of Hannover and Imperial College London, respectively. He received Dipl.-Ing. and Dr.-Ing. from the University of Hannover, in 1988 and 1994, respectively.

In 1994 and 1995, he worked at AT&T Bell Labs. He was with AT&T Labs - Research from 1996 to 2003. Since 2003 he is Full Professor and Head of the Institut für Informationsverarbeitung at the Leibniz Universität Hannover, Germany. From 2011 till 2013, he served as Dean of the Department of Electrical Engineering and Computer Science. He published more than 100 research papers and book chapters. He is coauthor of a graduate level textbook on video communications. He holds more than 30 patents. His current research interests are video coding and streaming, computer vision, 3D modelling, face animation, and computer–human interfaces.

Dr. Ostermann was a scholar of the German National Foundation. In 1998, he received the AT&T Standards Recognition Award and the ISO award. He served as a Distinguished Lecturer of the IEEE CAS Society. Since 2008, he is the Chair of the Requirements Subgroup of MPEG (ISO/IEC JTC1 SC29 WG11).