Introduction to the Special Issue on Scalable Video Coding—Standardization and Beyond

A FEW YEARS ago, the subject of video coding received a large amount of skepticism. The basic argument was that the 1990s generation of international standards was close to the asymptotic limit of compression capability—or at least was good enough that there would be little market interest in something else. Moreover, there was the fear that a new video codec design created by an open committee process would inevitably contain so many compromises and take so long to be completed that it would fail to meet real industry needs or to be state-of-the-art in capability when finished.

The Joint Video Team (JVT) then proved those views to be wrong and revitalized the field of video coding with the 2003 creation of the H.264/AVC standard (ITU-T Rec. H.264 and ISO/IEC 14496-10 Advanced Video Coding). The JVT was formed in late 2001 as a partnership between the two preeminent relevant standards bodies—the ITU-T Video Coding Experts Group (VCEG) and the ISO/IEC Moving Picture Experts Group (MPEG). A Special Issue of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (TCSVT) devoted to the H.264/AVC standard was published just a few months after its standardization was completed. Since that time, H.264/AVC has been very strongly embraced by industry and is now being deployed in virtually all new and existing applications of digital video technology.

This new Special Issue is devoted to the next major challenge undertaken by the JVT. That challenge was to build upon the strong compression foundation of H.264/AVC to create a *Scalable Video Coding* (SVC) standard that would extend the capabilities of the H.264/AVC design to address the needs of applications to make video coding more flexible for use in highly heterogeneous and time-varying environments.

The fundamental concept of SVC is to enable the creation of a compressed bit stream that can be rapidly and easily adapted to fit with the bit rate of various transmission channels and with the display capabilities and computational resource constraints of various receivers. This is accomplished by structuring the data of a compressed video bit stream into *layers*. The *base layer* is decodable by an ordinary H.264/AVC decoder, while one or more *enhancement layers* provides improved quality for those decoders that are capable of using it.

Three fundamental types of scalability were enabled. The first is *temporal scalability*, in which the enhancement layer provides an increase of the frame rate of the base layer. To a large extent this was already supported in the original H.264/AVC standard—but new supplemental information has been designed to make such uses more powerful. The next is *spatial scalability* (or resolution scalability), in which the enhancement layer offers increased picture resolution for receivers with greater display capabilities. Among the three types of scalability, this one is perhaps the most difficult to design in a way that is both effective (from a compression capability point of view) and practical for implementation in terms of computational resource requirements. Finally, there is *quality scalability*, sometimes also referred to as SNR or fidelity scalability, in which the enhancement layer provides an increase in video quality without changing the picture resolution.

For example, 1080p50/60 enhancement of today's 720p50/60 or 1080i25/30 HDTV can be deployed as a premium service while retaining full compatibility with deployed receivers, or HDTV can be deployed as an enhancement of SDTV video services. Another example is low-delay adaptation for multipoint video conferencing applications, including enabling the compositing of multiple source video streams and channel bit rate adaptation. Yet another example is supporting highly time-varying channel rates in Internet protocol networks and mobile communication systems.

The new SVC design was kept as "clean" as possible by the JVT—several features that appeared promising in early work were removed from the draft design after further study failed to show a sufficient need, merit or maturity for those features. For example, since only the syntax format and decoding process was standardized, there was no need to standardize a feature if comparable capability could be provided using pre-processing or encoder-only techniques. In some cases it was also found that key information for enabling effective scalability features could be provided as supplemental information without affecting the core coding technology and its syntax or decoding processes. Nevertheless, SVC is indeed a major addition to the H.264/AVC standard, which was lengthened by about 240 pages in the effort, with the vast majority of that being in a single new 200-page annex—Annex G, which is entitled "Scalable Video Coding." This represents a 2/3 increase in size for the standard, which was previously about 360 pages long (which was already about 100 pages longer than the first version approved in 2003).

The astute reader may be aware that this is not the world's first attempt at creating a SVC standard. In fact it is at least the fifth, as there were prior major scalability efforts for MPEG-2 and H.263, and two generations of such work for MPEG-4 Part 2. Unfortunately, those prior designs were basically not successful from an industry adoption perspective.

The JVT approach to address the shortcomings of the prior scalability designs had three key elements.

1) **Excellent coding efficiency**—With only about a 10% maximal bit rate penalty in compression quality for either the base layer or enhancement layer (when using sufficiently-optimized encoding techniques in typical application scenarios).
2) **Minimal computational complexity**—Avoiding excessive computational resources for decoders in particular—including a novel approach referred to as *single-loop*

Digital Object Identifier 10.1109/TCSVT.2007.905355

decoding, which allows any layer of the video to be decoded when performing the temporal inter-picture prediction processes only within a single decoding layer.

3) **Maximal design consistency**—Keeping maximal similarity between the new and prior H.264/AVC coding tools—including the enabling of a novel capability known as *bit stream rewriting* (lossless translation of a scalable multilayer bit stream into an ordinary H.264/AVC single-layer bit stream without performing a full decoding process).

Hence, there are some good indications that the present effort towards a SVC capability will sufficiently overcome the shortcomings of the prior efforts to achieve widespread industry adoption. Moreover, the new SVC standard represents a significant milestone in video compression technology and its standardization.

This Special Issue opens with an overview paper by Schwarz *et al.* As its title suggests, it introduces the basic concept of the SVC design, provides a concise description of the most important tools in the standard and gives some performance results.

We then dive deeper into the design with the two papers on spatial scalability and interlaced-scan scalable coding. In the first paper, Segall and Sullivan further detail the SVC approach to spatial scalability, which is a particularly challenging aspect of SVC. In the second paper, the concept of interlaced-scan video coding in SVC, including the mixing of interlaced and progressive video into one scalable bit stream, is described by Francois *et al.*

One of the challenges for an efficient scalable codec is its high-level syntax design and the related system-level support. This Special Issue devotes three papers to addressing aspects of this design issue. The first paper in this category is entitled "System and Transport Interface of SVC" authored by Wang *et al.*, and it describes the high-level syntax that was added to H.264/AVC to enable an efficient network adaptation interface for SVC. The following paper by Wenger *et al.*, entitled as "Transport and Signaling of SVC in IP Networks" provides an overview of the forthcoming SVC-related specification work of the Internet Engineering Task Force (IETF). The third paper by Amon *et al.* describes the file format design for SVC that is created as an extension of the AVC file format (which itself is a use of the ISO base media file format).

A critical issue with all scalable video codecs has been rate-distortion performance. While the fundamental rate-distortion performance that is achievable with a video codec is determined by the design of the bit stream and the decoder, in practice a great deal of attention needs to be paid to encoder optimization for actually achieving good rate-distortion performance. The paper by Amonou *et al.* introduces the concepts of quality layers for operational encoder control of SVC. Wien *et al.* then provide a concise overview of the rate-distortion performance that can be obtained using the reference software for SVC for a large number of different cases of scalability in their paper.

The next set of papers introduces application scenarios that can be enhanced with the use of SVC. Schierl *et al.* provide an overview of the potential that SVC can achieve in mobile video transmission applications and provide experimental results for example use cases. The paper by Song and Chen even further explores the coupling of transmission network concepts with the structure of SVC when investigating "Scalable H.264/AVC Video Transmission Over MIMO Wireless Systems with Adaptive Channel Selection Based on Partial Channel Information." A system for video streaming with SVC is then described in the paper by Wien *et al.*, entitled "Real-Time System for Adaptive Video Streaming based on SVC."

Hybrid video coding as found in H.264/AVC is based on temporal "differential pulse code modulation" (DPCM). DPCM is characterized by prediction loops that are operated in a synchronous fashion in the encoder and decoder. Differences between these prediction loops will create a *drift error* within a feedback loop that can accumulate over time and produce annoying artifacts. However, the scalability operation, i.e., the removal of parts of the video bit stream can produce these differences. While in the SVC design, a solution to the drift problem was found, the alternative approach of subband or transform coding inherently does not suffer from the drift property of DPCM. Therefore, video coding based on motion-compensated 3-D wavelet/subband transforms has been studied extensively for use in SVC. Such an alternative approach to SVC is described in two papers. While the paper by Adami *et al.* provides an overview of SVC using wavelet-based approaches, the paper by Xiong *et al.* describes a particularly efficient design, which is a "Barbell-lifting-based 3D Wavelet coding scheme."

As final remarks, we also would like to mention some considerations in regard to the environment in which prior SVC standards where created. That environment was characterized by the primary video use cases being broadcast and digital versatile disc (DVD). In such systems, graceful degradation was not particularly necessary, i.e., the transmission channel typically works or it does not. Also, rapid bit rate adaptation was not especially needed since the transmission channel ordinarily had a known, fixed throughput. Power adaptation was not particularly needed since the receivers were not hand-held devices; and finally, format adaptation was not especially needed, as there was just one single video format in common use: standard definition.

Today, the situation has changed. Internet and mobile transmission are becoming primary distribution mechanisms for video applications. These are shared resource systems with varying throughput and errors, requiring some mechanisms for graceful degradation, bit rate adaptation, and (for mobile applications) power adaptation. Moreover, we see the introduction of a variety of terminals and displays ranging from QCIF, QVGA, VGA, and SD to HD resolution and beyond. The efficient introduction of new formats may be done through a backwards compatible extension, providing format adaptation, e.g., for mobile TV from QVGA to VGA, for HDTV from 720p/1080i to 1080p, enhancing sample bit-depth from 8 to 10 bit, or enhancing the chroma sampling format from 4:2:0 to 4:2:2 or 4:4:4. The latter two cases are being considered for a possible second phase of the SVC project.

THOMAS WIEGAND
Fraunhofer-Institute for Telecommunications
Heinrich-Hertz Institute
10587 Berlin, Germany

GARY J. SULLIVAN
Entertainment and Devices Division
Microsoft Corporation
Redmond, WA 98052 USA

JENS-RAINER OHM
Institute for Communications Engineering
RWTH Aachen University
D-52056 Aachen, Germany

AJAY K. LUTHRA
Connected Home Solutions
Motorola, Inc.
San Diego, CA 92121 USA

Thomas Wiegand (M'05) received the Dipl.-Ing. degree in electrical engineering from the Technical University of Hamburg-Harburg, Germany, in 1995 and the Dr.-Ing. degree from the University of Erlangen-Nuremberg, Germany, in 2000.

He is the Head of the Image Communication Group, Image Processing Department, Fraunhofer Institute for Telecommunications—Heinrich Hertz Institute Berlin, Germany. From 1993 to 1994, he was a Visiting Researcher at Kobe University, Kobe, Japan. In 1995, he was a Visiting Scholar at the University of California at Santa Barbara. From 1997 to 1998, he was a Visiting Researcher at Stanford University, Stanford, CA, and served as a consultant to 8x8, Inc., Santa Clara, CA. He is currently a member of the technical advisory boards of the two start-up companies: Layered Media, Inc., Rochelle Park, NJ, and Stream Processors, Inc., Sunnyvale, CA. Since 1995, he has been an active participant in standardization for multimedia, with successful submissions to ITU-T VCEG, ISO/IEC MPEG, 3GPP, DVB, and IETF. In October 2000, he was appointed as the Associated Rapporteur of ITU-T VCEG. In December 2001, he was appointed as the Associated Rapporteur/Co-Chair of the JVT. In February 2002, he was appointed as the Editor of the H.264/AVC video coding standard and its extensions (FRExt and SVC). In January 2005, he was appointed as Associated Chair of MPEG Video. His research interests include video processing and coding, multimedia transmission, semantic image representation, and computer vision and graphics.

Dr. Wiegand received the SPIE VCIP Best Student Paper Award in 1998. In 2004, he received the Fraunhofer Award for Outstanding Scientific Achievements and the ITG Award of the German Society for Information Technology. Since January 2006, he has been an Associate Editor of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (TCSVT). In 2003, he was a Guest Editor of IEEE TCSVT Special Issue on the H.264/AVC Video Coding Standard in July 2003.

Gary J. Sullivan (S'83–M'91–SM'01–F'06) received the B.S. and M.Eng. degrees in electrical engineering from the University of Louisville J.B. Speed School of Engineering, Louisville, KY, in 1982 and 1983, respectively, and the Ph.D. and Engineer degrees in electrical engineering from the University of California, Los Angeles, in 1991.

He is the Rapporteur/Chairman of the ITU-T Video Coding Experts Group (VCEG), a Rapporteur/Co-Chairman of the ISO/IEC Moving Picture Experts Group (MPEG), and a Rapporteur/Co-Chairman of the Joint Video Team (JVT), which is a joint project between the VCEG and MPEG organizations. He is also the ITU-T video and image coding liaison representative to MPEG and JPEG/JBIG. Prior to joining Microsoft in 1999, he was the Manager of Communications Core Research at PictureTel Corporation (now Polycom), the world leader in videoconferencing communication at the time. He was previously a Howard Hughes Fellow and Member of the Technical Staff in the Advanced Systems Division of Hughes Aircraft Corporation and was a Terrain-Following Radar (TFR) System Software Engineer for Texas Instruments. He holds the position of Video Architect in the Core Media Processing Team of the Entertainment and Devices Division of Microsoft Corporation, Redmond, WA. At Microsoft he designed and remains lead engineer for the DirectX Video Acceleration (DXVA) API/DDI video decoding feature of the Microsoft Windows operating system. His research interests and areas of publication include image and video compression, rate-distortion optimization, motion estimation and compensation, scalar and vector quantization, and scalable and error/packet-loss resilient video coding.

Dr. Sullivan received the Technical Achievement award of the International Committee on Technology Standards (INCITS) in 2005 for his work on H.264/MPEG-4 AVC and other video standardization topics. He was a Guest Editor of the TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY for the Special Issue on the H.264/AVC Video Coding Standard in July 2003.

Jens-Rainer Ohm (M'92) received the Dipl.-Ing., Dr.-Ing., and the Habil. degrees from the Technical University of Berlin (TUB), Germany, Berlin, Germany, in 1985, 1990, and 1997, respectively.

From 1985 to 1990, he was a Research and Teaching Assistant with the Institute of Telecommunications, TUB, and from 1990 to 1995, he performed work within research projects on image and video coding. Between 1992 and 2000, he also served as Lecturer on topics of digital image processing, coding, and transmission at TUB.

From 1996 to 2000, he was Project Manager/Coordinator at the Image Processing Department of Heinrich Hertz Institute (HHI), Berlin, Germany. He was involved in research projects on motion-compensated, stereoscopic, and 3-D image processing, image/video, and content description for image/video database retrieval. Since 1998, he has participated in the work of the Moving Picture Experts Group (MPEG), where he has been active in the development of MPEG-4 and MPEG-7 standards. In 2000, he was appointed Full Professor and since then holds the chair position of the Institute of Communication Engineering, RWTH Aachen University, Aachen, Germany. His present research and teaching activities are in the areas of multimedia communication, multimedia signal processing/coding and services for mobile networks, with emphasis on video signals, also including fundamentals of digital communication systems. He served as a Chair of the MPEG Video Subgroup from May 2002 to January 2005; since then, he has been a Co-Chairman of the group. Since January 2005, he has also been co-chairing the JVT, which has been primarily focused on the development of the SVC extensions of the AVC standard. He has authored textbooks on multimedia signal processing, analysis and coding, on communications engineering and signal transmission, as well as numerous papers in the various fields mentioned above.

Ajay K. Luthra (S'79–M'81–SM'89) received the B.E. (hons.) degree from BITS, Pilani, India, in 1975, the M.Tech. degree in communications engineering from IIT Delhi, Delhi, India, in 1977, and the Ph.D. degree from Moore School of Electrical Engineering, University of Pennsylvania, Philadelphia, in 1981.

From 1981 to 1984, he was a Senior Engineer at Interspec Inc., Philadelphia, PA, where he was involved in applications of digital signal and image processing for biomedical applications. From 1984 to 1995, he was with Tektronix, Beaverton, OR, where he was Manager of the Digital Signal and Picture Processing Group during 1985–1990 and then Director of the Communications/Video Systems Research Lab from 1990–1995. He is currently a Senior Director in Advanced Technology Group at Motorola, Connected Home Solutions (formerly General Instrument), San Diego, CA, where he is involved in advanced development work in the areas of digital video compression and processing, streaming video, interactive TV, cable head-end system design, and advanced set-top box architectures. He has been an active member of the MPEG committee for the last fifteen years where he has chaired several technical subgroups. He is an Associate Rapporteur/Co-Chair of the Joint Video Team consisting of ISO/IEC MPEG and ITU-T/VCEG experts working video coding standardization. He is also an Associate Chair of the MPEG video subcommittee and the chairman of the INCITS L3.1 technical advisory group and the United States national delegation to MPEG.

Dr. Luthra was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY (2000–2002) and also a Guest Editor for its Special Issues on Streaming Video, March 2001, and the H.264/AVC Video Coding Standard in July 2003.