Introduction to the Issue on Emerging Technologies for Video Compression

I. Introduction

THE topic "Emerging Technologies for Video Compression" was chosen for this special issue simply because it is currently a very exciting and challenging time for this area of research. First, we have ongoing standards activity in HEVC. HEVC has driven the video community to challenge existing methods, with the aim of delivering drammatic quality improvements. While much of the research that will underpin HEVC has already been reported, some is presented here for the first time. As with previous standards, it will no doubt create opportunities for new research to provide further improvements and differentiation of offerings.

Second, the predicted growth in demand for bandwidth, driven largely by video applications, is probably greater now than it has ever been. There are four primary drivers for this: 1) Recently introduced formats such as 3-D and multiview, coupled with pressures for increased dynamic range, spatial resolution and framerate, all require increased bit-rate to deliver improved levels of immersion or interactivity. 2) Video-based web traffic continues to grow and dominate the internet through social networking and catch up TV. In recent years, Youtube has accounted for 27% of all video traffic and, by 2015, it is predicted that there will be 700 billion minutes of video downloaded.[1] That represents a full-length movie for every person on the planet. 3) User expectations continue to drive flexibility and quality, with a move from linear to nonlinear delivery. Users are demanding My-Time rather than Prime-Time viewing. 4) Finally new services, in particular mobile delivery through 4G/LTE to smart phones. Some mobile network operators are predicting the demand for bandwidth to double every year for the next 10 years!

While these demands can to some extent be addressed through efficiency improvements in network and physical layer technology, because video is the prime driver, the role of video compression is also enormously important.

II. State of the Art

All major video coding standards since H.261 (created in the late 1980s) have been based on incremental improvements to the hybrid motion-compensated block transform coding model. While this approach has produced impressive rate-distortion improvements over the past two decades, the question arises whether more disruptive techniques can provide substantial gains.

Digital Object Identifier 10.1109/JSTSP.2011.2170331

[1]Opportunities and impact of video on LTE networks, Motorola. http://business.motorola.com/experiencelte/pdf/LTEVideoImpactWhitePaper.pdf.

The current H.264/AVC compression standard is based on the picture-wise processing and waveform-based coding of video signals. The technology now being considered for the new standardization project on high-efficiency video coding (HEVC) is a generalization of this approach which promises significant gains through innovations such as improved intra-prediction, larger block sizes, more flexible ways of decomposing blocks for inter- and intra-coding and better exploitation of long-term correlations and picture dependencies. It will support a wide range of encoder modes, which are typically optimized using mean-squared-error-based or related distortion measures.

New and alternative frameworks for video compression are also beginning to emerge, where prediction and signal representation are based on a parametric or data-driven model of scene content. These often invoke a combination of waveform coding and texture replacement, where computer graphic models are employed to replace target textures at the decoder. Such approaches can also be combined with higher order motion models for texture warping and mosaicing. Work to date has demonstrated the potential for dramatic rate-quality improvements with such methods.

III. The Future for Video Compression

HEVC clearly represents the immediate future for compression and will undoubtedly produce impressive rate distortion gains. In the longer term, alternative approaches such as those described above have the potential to create a new content-driven rate-quality optimization framework for video compression. However, it is equally clear that a huge amount of research work needs to be done in order to fully exploit their potential and to yield stable and efficient solutions. For example, in such cases, mean square error is no longer a valid objective function or measure of quality and emphasis must shift from rate distortion to rate-quality optimization, demanding new embedded perceptually driven quality metrics. The choice of texture analysis and synthesis models, alongside meaningful quality metrics and the exploitation of long-term picture dependencies will be key if an effective and reliable system is to result.

Key research topics include: quality assessment and artifact detection methods for perceptual video coding; rate quality optimization frameworks; static and dynamic texture models for analysis, classification, and synthesis; advanced motion models; methods for exploiting long-term picture and non-local picture dependencies; new sampling methods; and finally exploitation of increased spatio-temporal resolutions and higher dynamic range. Many of these topics are addressed by the contributions in this special issue.

IV. Scanning the Special Issue

This special issue has contributions related to four key research areas: 1) the emerging HEVC standardization process; 2)

1932-4553/$26.00 © 2011 IEEE

advances in 3-D coding; 3) quality assessment for next generation codecs; and 4) new compression methods exploiting spatio-temporal content analysis and synthesis.

We have four papers addressing aspects of with potential relevance to HEVC. First, Mrak *et al.* address the important area of advanced intra coding. The paper presents a novel approach referred to as Combined Intra Prediction (CIP), based on spatial closed- and open-loop predictions. This exploits redundancy between neighboring blocks and also within a coding block to enable excellent rate-distortion performance with low-complexity. The approach is flexible in the context of different coding settings and gains of up to 4.5% YUV BD-rate for HEVC test sequences. Asai *et al.* address the topic of compressing high-resolution video content, something central to HEVC. With the aims of enhancing compression performance and lowering complexity, the approach exploits larger and hierarchically decomposable blocks for motion compensation and transformation. The paper demonstrates significant savings, achieving approximately 26% bit-rate saving compared to H.264/AVC high profile. It also discusses the important tradeoffs between complexity and performance. Lan *et al.* present a new approach to exploitation of spatial and temporal correlations. They prove that strong correlations exist in non-local spatial regions. To exploit these they propose a signal dependent transform (SDT), which is derived from decoded non-local blocks that are selected by neighborhood matching. The proposed transform has been implemented into the Key Technology Area (KTA) software to exploit both spatial and temporal non-local correlations. Results presented indicate coding gains up to 1.4 dB for intra-, and up to 1.0 dB for inter-frame coding, compared to KTA. Finally, Krutz *et al.* revisit the concept of object-based video coding that was originally introduced in MPEG-4 Visual. Background Sprites with automatic segmentation and background subtraction are shown to outperform H.264/AVC in cases where there is significant global motion. The authors introduce a rate-distortion optimization approach for this type of coder. They focus on joint optimization of quantization parameters for foreground and background regions and show superior performance to H.264/AVC for appropriate source material.

As stated above, quality assessment will be key to the success of future video codecs as we move away from PSNR-like metrics, particularly in the case of new compression strategies. In this context, Lee *et al.* introduce an interesting and alternative way of providing coding enhancements using audio visual focus of attention. The authors identify this as a basis for foveated coding, applying varying quality levels according to the distance from an identified sound source. The method has been evaluated through extensive subjective experiments and demonstrates the potential for considerable coding gain without significant quality degradation. Several issues are raised by the authors including the impact of uneven distributions of coding artifacts, content dependence, memory effects and the difference in quality perception for different frame sizes. The next paper by Bosc *et al.* examines the important area of quality assessment in the context of 3-D and multiview coding, in particular considering the question of synthesized view evaluation. They discuss the issue of the geometric distortion artifacts introduced through depth image-based rendering (DIBR) where the impact on quality is related to context. The authors compare a range of objective and subjective assessment methods alongside seven different view synthesis algorithms. Their results highlight the shortcomings of conventional metrics showing that they do not relate well to human judgment. The study proposes a new objective measure which addresses some of these issues. Oh *et al.* introduce a new rendering view distortion function to provide efficient depth map coding for 3-D applications. Rather than use just depth map coding error, they also invoke co-located color information to estimate rendered view quality. The proposed area-based scheme mimics the view-rendering process accurately. Coding performance is further improved by using the color coding information to inform the block SKIP mode. The authors present simulation results that show bit rate savings of 30% for depth data, and 10% overall for multi-view data.

The final group of papers look at new approaches to video compression based on content analysis and synthesis. By identifying contextual textures and by using varying combinations of segmentation, region warping and texture synthesis, in conjunction with more conventional coding, they demonstrate the potential for this approach as a basis for moving beyond the block based coding framework. Balle *et al.* propose a hybrid approach where static and dynamic texture models are integrated alongside conventional transform coding. Gaussian Markov random field (GMRF) methods are used for analysis/synthesis of static textures where optimal methods for classification, analysis, quantization, and synthesis are presented. In the case of dynamic textures, a linear dynamic model is derived from frames encoded in a conventional fashion. Benefits are obtained if frames are synthesized after camera motion compensation. Interestingly, the authors integrate synthesized frames into long term memory so that they are available for future prediction. This yields significant bitrate savings while preserving PSNR. Bosch *et al.* show using an integrated system, based on H.264/AVC, that coding efficiency can be improved using texture and motion models of the content in a scene. Based on these models they force block skipping in the host coder. Transmitted side information is then used to reconstruct the missing regions at the decoder. The authors propose and evaluate a number of spatial-texture models, segmentation strategies, and texture features with which textured regions can be detected. A further extension is proposed where an HVS inspired motion classification model is used to separate moving foreground objects from the background. The results presented show potential for significant improvements in coding efficiency compared to spatial texture-based methods. Importantly, the authors present results from perceptual experiments and show gains up to 25% compared to H.264/AVC for a given perceptual quality. Finally, the paper by Zhang and Bull presents an efficient parametric region based framework to identify, characterize and code both static and dynamic texture regions. It exploits properties of the dual tree complex wavelet transform to perform segmentation and classification, with warping provided by a hierarchical perspective model and dynamic synthesis based on Doretto's approach with the addition of frame pre-warping. The paper introduces a new video metric (AVM) which is used in-loop as part of the rate-quality optimization process. This metric was validated on VQEG and, importantly, is the first to perform well on both conventionally coded and synthesized content. Results show content dependent bitrate savings, of up to 60% compared with H.264/AVC at the same objective quality.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of Patrick Ndjiki-Nya from HHI Berlin.

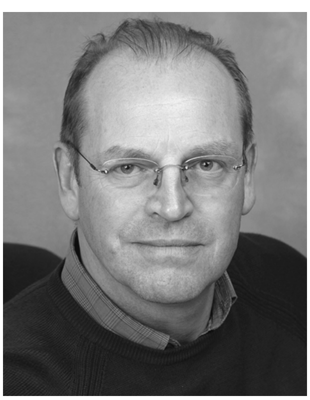
DAVID R. BULL, *Lead Guest Editor*
Department of Electrical and Electronic Engineering
University of Bristol
Bristol BS8 1UB, U.K.

EDWARD J. DELP, *Guest Editor*
School of Electrical and Computer Engineering and the School of Biomedical Engineering
Purdue University
West Lafayette, IN 47907 USA

SEISHI TAKAMURA, *Guest Editor*
NTT Cyber Space Laboratories
Yokosuka, 239-0847 Japan

THOMAS WIEGAND, *Guest Editor*
Image Processing Department
Heinrich Hertz Institute
10587 Berlin, Germany

FENG WU, *Guest Editor*
Microsoft Research Asia
Beijing 100080, China



David R. Bull (M'94–SM'07) PhD, FIET, SMIEEE, CEng received the B.Sc. degree from the University of Exeter, Exeter, U.K., in 1980, the M.Sc. degree from the University of Manchester, Manchester, U.K., in 1983, and the Ph.D. degree from the University of Cardiff, Cardiff, U.K., in 1988.

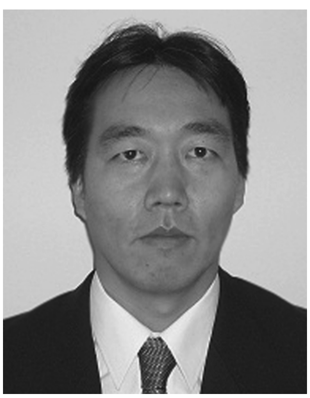
He currently holds the Chair in Signal Processing at the University of Bristol, Bristol, U.K., and is Head of the Visual Information Laboratory. His previous roles include: Lecturer at the University of Wales (Cardiff) and Systems Engineer for Rolls Royce. He was Head of the Electrical and Electronic Engineering Department at Bristol between 2001 and 2006 and is now Director of the Bristol Vision Institute (BVI). David has worked widely in the fields of 1–D and 2-D signal processing. He has published numerous patents, several of which have been exploited commercially. His current activities are focused on the problems of image and video communications and analysis for low bit-rate wireless, Internet, military, and broadcast applications. He has published some 400 academic papers, various articles and two books and has given numerous invited/keynote lectures and tutorials. In 1996, he helped to establish the U.K. DTI Virtual Centre of Excellence in Digital Broadcasting and Multimedia Technology and was one of its Directors from 1997 to 2000. He was also a founder member of the U.K.'s £50M MoD Defence Technology Centre in Data and Information Fusion, serving on its Science and Technology Board. He was appointed as an independent member of U.K. Government's Defence Scientific Advisory Council (DSAC) in 2002, contributing to the U.K. (MoD)-U.S. (DoD) Working Group on Persistent Surveillance (2004). He has also advised Government through membership of the U.K. Foresight Panel and he is a panel member for the U.K. 2014 Research Evaluation Framework. Throughout his career, he has acted as an expert advisor to many companies. In 2004, he helped to specify video coding formats for the U.K. satellite HDTV launch. In 2007, he cofounded the Bristol Vision Institute (BVI), bringing together engineers and scientists from a range of academic disciplines to work on vision research.

Dr. Bull was awarded the IEE Ambrose Fleming Premium for his work on Primitive Operator Digital Filters (1992) and a second IEE Premium for his work on 3-D Scanning Confocal Near-Infrared Microscopy in 1997. He received a Best Paper Award at IEEE SVTC in 2003 for "Link Adaptation for Video Transmission over COFDM based WLANs." He is a member of the Steering Committee for the U.K. Technology Strategy Board's Special Interest Group on Imaging. He is a past member of the Editorial Board for the Academic Press Series on Digital Signal Processing and its Applications and was Guest Editor of the *EURASIP Journal on Advances in Signal Processing* (Wireless Video—2008). He was selected as DTI Distinguished Professor in 2004. In 2001, he cofounded ProVision Communication Technologies, Ltd. ProVision launched the world's first robust multisource wireless HD sender for consumer use at CES in 2009 and was awarded a European Innovation Award for Wireless Multimedia Streaming in 2010.

Edward J. Delp (S'70–M'79–SM'86–F'97) was born in Cincinnati, OH. He received the B.S.E.E. (*cum laude*) and M.S. degrees from the University of Cincinnati, Cincinnati, OH, and the Ph.D. degree from Purdue University, West Lafayette, IN. In May 2002, he received an Honorary Doctor of Technology from Tampere University of Technology, Tampere, Finland.

From 1980 to 1984, he was with the Department of Electrical and Computer Engineering, The University of Michigan, Ann Arbor. Since August 1984, he has been with the School of Electrical and Computer Engineering and the School of Biomedical Engineering, Purdue University. He is currently The Charles William Harrison Distinguished Professor of Electrical and Computer Engineering and Professor of Biomedical Engineering. His research interests include image and video compression, multimedia security, medical imaging, multimedia systems, communication, and information theory.

Dr. Delp is a Fellow of the SPIE, a Fellow of the Society for Imaging Science and Technology (IS&T), and a Fellow of the American Institute of Medical and Biological Engineering. In 2000, he was selected a Distinguished Lecturer of the IEEE Signal Processing Society. He received the Honeywell Award in 1990, the D. D. Ewing Award in 1992 and the Wilfred Hesselberth Award in 2004 all for excellence in teaching. In 2001, he received the Raymond C. Bowman Award for fostering education in imaging science from the Society for Imaging Science and Technology (IS&T). In 2004, he received the Technical Achievement Award from the IEEE Signal Processing Society for his work in image and video compression and multimedia security. In 2008, he received the Society Award from IEEE Signal Processing Society (SPS). This is the highest award given by SPS and it cited his work in multimedia security and image and video compression. In 2009, he received the Purdue College of Engineering Faculty Excellence Award for Research.



Seishi Takamura (M'05–SM'07) received the B.E., M.E., and Ph.D. degrees from the Department of Electronic Engineering, Faculty of Engineering, University of Tokyo, Tokyo, Japan, in 1991, 1993, and 1996, respectively.

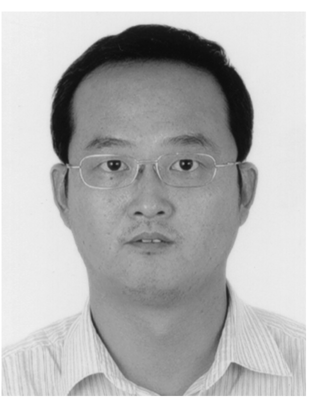
In 1996, he joined Nippon Telegraph and Telephone (NTT) Corporation, Yokosuka, Japan, where he has been engaged in research on efficient video coding and ultra-high-quality video coding (e.g., high frame rate, multi-spectral and super-high definition video coding) and evolutive video coding. He is currently a Senior Research Engineer, Supervisor in the Video Coding Group, NTT Cyber Space Laboratories. In 2009, he was elevated to Distinguished Technical Member of NTT. From 2005 to 2006, he had been a Visiting Scientist at Stanford University, Stanford, CA. He has more than 100 patents (granted or pending), authored four books, and 13 commentaries.

Dr. Takamura is a recipient of 17 academic awards including Young Researcher Award from IIEEJ in 2001, ITE Niwa-Takayanagi Memorial Award (Paper Award) in 2002, PCSJ Best Poster Award in 2002, 2003, and 2009, TAF Telecom System Technology Award in 2004 and 2008, PCSJ Frontier Award in 2004 and 2008, IPSJ Nagao Special Researcher Award in 2006, IPSJ Yamashita SIG Research Award in 2008, and FIT Funai Best Paper Award in 2008. He has served as a council member of ITE, Chair of IEEE Tokyo Section Technical Program Committee, Technical Program Committee Member of APSIPA ASC2009, Organizing Committee member of PCSJ2010, Guest Editor of Hindawi Advances in Multimedia in 2008, a reviewer for the IEEE TRANSACTION ON IMAGE PROCESSING, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEM FOR VIDEO TECHNOLOGY, and many international conferences including ICIP2008–2011, ICME2010–2012. Since 2006, he has been serving as an Associate Editor for the TRANSACTIONS ON CIRCUITS AND SYSTEM FOR VIDEO TECHNOLOGY, through which he received Certificate of Appreciation from IEEE Circuits and Systems Society. Dr. Takamura is a member of MENSA, IPSJ, and IIEEJ.

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

He is a Professor of electrical engineering at the Technical University of Berlin, Berlin, Germany. He chairs the Image Communication laboratory and is joint Heading of the Image Processing Department at the Fraunhofer Institute for Telecommunications—Heinrich Hertz Institute, Berlin. He joined the Heinrich Hertz Institute in 2000 as the head of the Image Communication Group in the Image Processing Department and remains active in this role. His research interests include video processing and coding, multimedia transmission, semantic image representation, as well as computer vision and graphics. 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 8×8, Inc., Santa Clara, CA. From 2006 to 2008, he was a consultant to Stream Processors, Inc., Sunnyvale, CA. From 2007 to 2009, he was a consultant to Skyfire, Inc., Mountain View, CA. Since 2005, he has been a member of the technical advisory boards of Vidyo, Inc., Hackensack, NJ. 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 and 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). From 2005 to 2009, he was Co-Chair of MPEG Video.

Prof. Wiegand received the SPIE VCIP Best Student Paper Award in 1998. In 2004, he received the Fraunhofer Award for outstanding scientific achievements in solving application-related problems and the ITG Award of the German Society for Information Technology. In 2009, he received the Innovations Award of the Vodafone Foundation, the EURASIP Group Technical Achievement Award, and the Best Paper Award of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY. Since January 2006, he has been an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY.



Feng Wu (M'99–SM'06) received the B.S. degree in electrical engineering from Xidian University, Xi'an, China, in 1992 and the M.S. and Ph.D. degrees in computer science from the Harbin Institute of Technology, Harbin, China, in 1996 and 1999, respectively.

He joined in Microsoft Research Asia, formerly Microsoft Research China, as an Associate Researcher in 1999. He has been a Researcher with Microsoft Research Asia since 2001 and is now a Senior Researcher/Research Manager. His research interests include image and video representation, media compression and communication, as well as computer vision and graphics. He has authored or coauthored over 200 papers published in journals like the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE TRANSACTIONS ON MULTIMEDIA, and some other International Conferences and Forums, e.g., MOBICOM, INFOCOM, CVPR, DCC, ICIP, ICME, and ISCAS. He has over 50 U.S. patents granted or pending in video and image coding.

Dr. Wu serves as an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEM FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON MULTIMEDIA, and several other International journals. He also serves as a member of the technical program committee of a number of International conferences, including ICME, ISCAS, PCS, and VCIP. He has been an active contributor to ISO/MPEG and ITU-T standards and some of his contributions have been adopted by MPEG-4 FGS, H.264/MPEG-4 AVC, and H.264 SVC standards. He served as the chairman of the China AVS Video Group between 2002 and 2004 and led the efforts on developing China's AVS video standard 1.0. As a coauthor, he received the best paper award in IEEE T-CSVT 2009, PCM 2008 and SPIE VCIP 2007.