IMAGE AND VIDEO OVERVIEW SESSION

Image transformation for privacy-preserving machine learning and adversarially robust defense

Abstract:
Although the rapid development of machine learning algorithms has led to major breakthroughs in computer vision, for a wide range of applications, where safety and security are critical, there is concern about the reliability of machine learning systems. Such systems in general suffer from untrusted environments and various attacks such as adversarial attacks, unauthorized use of data, model inversion attacks, and membership inference attacks.

Accordingly, various image transformation methods have been proposed for privacy-preserving machine learning and adversarially robust defense. In contrast, attack methods on such image transformation methods have been investigated simultaneously. In my talk, we first address compressible image encryption schemes, which have been proposed for encryption-then-compression (EtC) systems, and their application to traditional machine leaning algorithms such as support vector machine. Next, image transformation methods including perceptual encryption ones that generate images without visual information are summarized for privacy-preserving DNNs. Finally, block-wise image transformation methods with a secret key is demonstrated to be able to maintain a high classification accuracy under the use of both clean images and adversarial examples.

Biography:

HITOSHI KIYA received the B.E. and M.E. degrees from the Nagaoka University of Technology, Japan in 1980 and 1982, respectively, and the Dr.Eng. degree from Tokyo Metropolitan University, in 1987. In 1982, he joined Tokyo Metropolitan University, where he became a Full Professor, in 2000. From 1995 to 1996, he was a Visiting Fellow with The University of Sydney, Australia. He served as the Inaugural Vice President (Technical Activities) of APSIPA, from 2009 to 2013, and the Regional Director-at-Large for Region ten of the IEEE Signal Processing Society, from 2016 to 2017. He was also the President of the IEICE Engineering Sciences Society, from 2011 to 2012. He currently serves as the President of APSIPA. He is a Fellow of IEEE, IEICE and ITE. He has received numerous awards, including ten best paper awards. He has organized a lot of international conferences in such roles as the TPC Chair of the IEEE ICASSP 2012 and the General Co-Chair of the IEEE ISCAS 2019. He was an Editorial Board Member of eight journals, including the IEEE Transactions on Signal Processing, Image Processing, and Information Forensics and Security.

Recent Issues in Real-World Image Restoration Using Deep Convolutional Neural Networks

Abstract:
Deep convolutional neural networks (CNNs) are shown to provide very high performances in image restoration problems such as image denoising and single-image super-resolution. Researchers performed restoration with unrealistic assumptions on image degradation in earlier works, such as corruption by Gaussian noise and/or degradation by bicubic decimation. In other words, researchers used plausible but unrealistic models because the exact noise model and degradation model are unknown and spatially varying in the case of real-world images. By assuming known unrealistic degradation models, we can generate an unlimited number of training image pairs and gain high performances for the synthetic test images from the same domain. However, the CNNs overfitted to such synthetic data do not work well for the real-world images. Hence, now there are also many methods that attempt to reduce the gap between the performances for synthetic and real-world image degradations. For example, more realistic noise models are being developed for generating noisy training images closer to real-world ones. Also, real-world noises are being captured with careful camera settings. In the case of super-resolution, the methods for estimating degradation models are developed, which helps to modify the methods depending on the degradation model. More importantly, transfer learning from synthetic to real-world degradation models are also developed. This talk will overview these efforts, which significantly reduced the discrepancy between the unrealistic degradation models and real-world ones. We first briefly overview the methods on real-world image denoising and then give more attention to the super-resolution of real-world scenes.

Biography:

Nam Ik Cho received the BS, MS, and Ph.D. degrees in Control and Instrumentation Engineering from Seoul National University, Seoul, Korea, in 1986, 1988, and 1992, respectively. From 1991 to 1993, he was a research associate with the Engineering Research Center for Advanced Control and Instrumentation, Seoul National University. From 1994 to 1998, he was with the University of Seoul, as an assistant professor of Electrical Engineering. In 1999, he joined the Department of Electrical and Computer Engineering, Seoul National University, where he is currently a professor. His research interests include image processing, adaptive filtering, digital filter design, and computer vision.

Machine Learning for Analytics Architecture: AI to Design AI

Abstract:

Niklaus Emil Wirth introduced the innovative idea that *Programming = Algorithm + Data Structure.* Inspired by this, we advance the concept to the next level by stating that *Design = Algorithm + Architecture.* With concurrent exploration of algorithm and architecture entitled Algorithm/Architecture Co-exploration (AAC), this methodology introduces a leading paradigm shift in advanced system design from System-on-a-Chip to Cloud and Edge.

As algorithms with *high accuracy* become exceedingly more complex and Edge/IoT generated data becomes increasingly bigger, *flexible* parallel/reconfigurable processing are crucial in the design of *efficient* signal processing systems having *low power.* Hence the analysis of algorithms and data for potential computing in parallel, efficient data storage and data transfer is crucial. With extension of AAC for SoC system designs to even more versatile platforms based on analytics architecture, system scope is readily extensible to cognitive cloud and reconfigurable edge computing for multimedia and mobile health, a cross-level-of abstraction topic which will be introduced in this tutorial together with case studies.

Biography:

Chris Gwo Giun Lee is an investigator in signal processing systems for multimedia and bioinformatics. His work on analytics of algorithm concurrently with architecture, Algorithm/Architecture Co-Design (AAC), has made possible accurate and efficient computations on SoC, cloud and edge.

His work has contributed to 130+ original research and technical publications with invention of 50+ patents worldwide. His AAC work resulted in industry deployment of 60+ million LCD panels worldwide. Two patents were licensed by US health industry for development of analytics platform based precision medicine products (Boston, MA, June 1, 2015, GLOBE NEWSWIRE). This AAC work has been pivotal in delivering international standards, e.g. and Reconfigurable Video Coding 3D extension of HEVC in MPEG.

Chris was system architect in Philips Semiconductor and also project leader in the Silicon Valley. He was recruited to NCKU in 2003. He received his BSEE from National Taiwan University, MSEE and PH.DEE from the University of Massachusetts. Chris serves as the AE for IEEE TSP and JSPS. He was formerly the AE for IEEE TCSVT for which he received the Best Associate Editor's Award in 2011. He is the Distinguished Lecturer for IEEE CASS from 2019 ~ 2020.

Semantic Image Scene Segmentation by Deep Machine Learning

Abstract: Scene segmentation is a challenging task as it need classify every pixel in the image. It is crucial to exploit discriminative context and aggregate multi-scale features to achieve better segmentation. Context is essential for semantic segmentation. Due to the diverse shapes of objects and their complex layout in various scene images, the spatial scales and shapes of contexts for different objects have very large variation. It is thus ineffective or inefficient to aggregate various context information from a predefined fixed region. In this talk, I will first present a novel context contrasted local feature that not only leverages the informative context but also spotlights the local information in contrast to the context. The proposed context contrasted local feature greatly improves the parsing performance, especially for inconspicuous objects and background stuff. Furthermore, I will present a scheme of gated sum to selectively aggregate multi-scale features for each spatial position. The gates in this scheme control the information flow of different scale features. Their values are generated from the testing image by the proposed network learnt from the training data so that they are adaptive not only to the training data, but also to the specific testing image. Finally, I will present a scale- and shape-variant semantic mask for each pixel to confine its contextual region. To this end, a novel paired convolution is proposed to infer the semantic correlation of the pair and based on that to generate a shape mask. Using the inferred spatial scope of the contextual region, a shape-variant convolution is controlled by the shape mask that varies with the appearance of input. In this way, the proposed network aggregates the context information of a pixel from its semantic-correlated region instead of a predefined fixed region. In addition, this work also proposes a labeling denoising model to reduce wrong predictions caused by the noisy low-level features. This talk is based on two papers: H. Ding, X. Jiang, et al, "Context contrasted feature and gated multi-scale aggregation for scene segmentation," *CVPR'2018 Oral*, and H. Ding, X. Jiang, et al, "Semantic Correlation Promoted Shape-Variant Context for Segmentation," *CVPR'2019 Oral*.

Biography:

Xudong Jiang received the B.Eng. and M.Eng. from the University of Electronic Science and Technology of China (UESTC), and the Ph.D. degree from Helmut Schmidt University, Hamburg, Germany, all in electrical engineering. From 1986 to 1993, he was a Lecturer with UESTC, where he received two Science and Technology Awards from the Ministry for Electronic Industry of China. From 1998 to 2004, he was with the Institute for Infocomm Research, A-Star, Singapore, as a Lead Scientist and the Head of the Biometrics Laboratory, where he developed a system that achieved the most efficiency and the second most accuracy at the International Fingerprint Verification Competition in 2000. He joined Nanyang Technological University (NTU), Singapore, as a Faculty Member, in 2004, and served as the Director of the Centre for Information Security from 2005 to 2011. Currently, he is a Tenured Associate Professor with the School of EEE, NTU. Dr Jiang holds 7 patents and has authored over 150 papers with 40 papers in the IEEE journals, including 11 papers in *IEEE T-IP* and 6 papers in *IEEE T-PAMI*. Two of his first authored papers have been listed as top 1% highly cited papers in the academic field of Engineering by Essential Science Indicators. He served as IFS Technical Committee Member of the IEEE Signal Processing Society from 2015 to 2017, Associate Editor for *IEEE SPL* for 2 terms from 2014 to 2018, Associate Editor for *IEEE T-IP* for 2 terms from 2016 to 2019 and the founding editorial board member for *IET Biometrics* form 2012 to 2019. Dr Jiang is currently a Senior Area Editor for *IEEE T-IP* and Editor-in-Chief for *IET Biometrics.* His current research interests include image processing, pattern recognition, computer vision, machine learning, and biometrics.

Towards Comprehensive Understanding of 3D: Geometric Computer Vision Meets Deep Learning

Abstract: Computer vision is dedicated in to enabling machines/robots the visual perception ability as human. Geometric computer vision aims at reconstructing and understanding the three-dimensional geometric structure of the observed scene from images and videos, which has important applications in unmanned systems, autonomous driving, robotics, virtual reality/augmented reality and scene analysis. Deep learning, especially deep convolutional neural networks, has great advantages in feature learning and semantic information extraction. How to effectively combine this data-driven model with multi-view geometric model has become an active research area in computer vision. In this talk, I will present a series of recent work from our group in this direction, including how to achieve monocular depth estimation, binocular depth estimation and multi-view stereo under the framework of supervised learning, and how to construct unsupervised learning frameworks for the tasks of self-adaptive stereo, multi-view stereo, optical flow estimation and stereo-Lidar fusion. The talk will finish with discussions on future research directions, e.g., how to speed up the implementation, how to exploit recently developed NAS.

Biography:

Yuchao Dai is currently a Professor with School of Electronics and Information at the Northwestern Polytechnical University (NPU). He received the B.E. degree, M.E degree and Ph.D. degree all in signal and information processing from Northwestern Polytechnical University, Xi'an, China, in 2005, 2008 and 2012, respectively. He was an ARC DECRA Fellow with the Research School of Engineering at the Australian National University, Canberra, Australia. His research interests include structure from motion, multi-view geometry, low-level computer vision, deep learning, compressive sensing and optimization. He won the Best Paper Award in IEEE CVPR 2012, the DSTO Best Fundamental Contribution to Image Processing Paper Prize at DICTA 2014, the Best Algorithm Prize in NRSFM Challenge at CVPR 2017, the Best Student Paper Prize at DICTA 2017, the Best Deep/Machine Learning Paper Prize at APSIPA ASC 2017, the Best Paper Award Nominee at IEEE CVPR 2020. He served as Area Chair in CVPR, ACM MM, ACCV, WACV and etc.

SPEECH AND LANGUAGE OVERVIEW SESSION

Title: Co-prime Microphone Arrays and their Application to Finding the Direction of Arrival of Sound Sources

Abstract:

Co-prime microphone arrays consist of two co-incident sub-arrays of microphones, where the number of elements in the two sub-arrays are coprime with each other. This arrangement results in a type of sparse sampling of the sound field and results in an overall beampattern that is much narrower than can be obtained with an array consisting of the same number of microphones that are uniformly spaced. This is important for ensuring accurate recording of sound when there is a practical limitation on the number of microphones. This overview talk will describe the theory behind the design of coprime arrays for both linear and circular arrangements of microphones. It will include a comparison of the performance of coprime arrays, in terms of the beampattern and array gain, compared to traditional uniform microphone arrays with the same number of elements. The talk will then describe how to find the directional of arrival (DOA) of sound sources based on the Steered Response Power (SRP) derived for the coprime microphone array. This will include SRP approaches utilising the Phase Transform weighting (SRP-PHAT), known to provide a robust estimation of the DOA of speech sources within noisy and reverberant acoustic environments. The talk will also include a review of recent research into improving the performance of coprime microphone arrays for sound source DOA estimation.

Biography:

Christian Ritz graduated with a Bachelor of Electrical Engineering and a Bachelor of Mathematics (both in 1999) and a PhD in Electrical Engineering (in 2003) all from the University of Wollongong (UOW), Australia. His PhD research focused on very low bit rate coding of wideband speech signals. Since 2003, Christian has held a position within the School of Electrical, Computer and Telecommunications Engineering at UOW where he is currently a Professor. Concurrently, he is also the Associate Dean (International) for UOW's Faculty of Engineering and Information Sciences, with responsibility for managing the Faculty's international strategy including significant transnational programs and partnerships in China, Dubai, Singapore and Malaysia. Christian is the deputy director of the Centre for Signal and Information Processing (CSIP) and leads the audio, speech and acoustics signal processing research of the centre. He is actively involved in several projects, some funded from the Australian government and industry, including microphone array signal processing for the directional sound enhancement, acoustic scene classification, loudspeaker-based sound field reproduction and control and visual object classification using machine learning. He is currently a Distinguished Lecturer (2019 to 2020) of the Asia Pacific Signal and Information Processing Association (APSIPA). For more information see: https://scholars.uow.edu.au/display/christian_ritz.

Enabling *in-to-life* Emotion-AI technology: robustness, scalability, and trustworthiness

Abstract:

Emotions play a critical role in our daily life such that it affects all aspect of individual's behaviors, personality, actions, and his/her social connections. Computing emotion using measurable signals (speech, language, physiology) to understand humans scientifically and derive next-generation human-machine interface in multiple contexts, e.g., dialog systems, call centers, affective media content, mental health, etc., has sparked a tremendous interest in the signal processing/machine learning community together with fields of behavior science (further give rise to the term of 'Emotion-AI' with the recent use of deep learning techniques). In this overview talk, we lay out three key components that would take existing emotion AI solution *in-to-life* at scale, i.e., robustness, scalability, and trustworthiness. Specifically, we will highlight several recent efforts in computing emotion through multimodally learning from diverse signals, i.e., speech, language, physiology, to achieve robust accuracy in multiple contexts, to enable scalability with ease-of-adaptation by learning through unsupervised transfer between corpora, and to move toward trustworthiness with privacy-aware emotion recognition technology.

Biography:

Chi-Chun Lee (Jeremy) is an Associate Professor at the Department of Electrical Engineering with joint appointment at the Institute of Communication Engineering of the National Tsing Hua University (NTHU), Taiwan. He received his B.S. degree and Ph.D. degree both in Electrical Engineering under supervision of Prof Shri Narayanan from the University of Southern California (USC), USA in 2007 and 2012. He was a data scientist at id:a lab at the ID Analytics in 2013. His research interests are in speech and language, affective multimedia, health analytics, and behavior computing.

He is an IEEE senior member. He is awarded with Taiwan's Foundation of Outstanding Scholar's Young Innovator Award. He receives the Taiwan's Ministry of Science and Technology (MOST) 2018 and 2019 Futuretek Breakthrough Award. He is an associate editor for the IEEE Transaction on Multimedia (2019-2020), and a TPC member for APSIPA IVM and ML committee. He serves as an area chair for Interspeech 2016, 2018, 2019, senior program committee for ACII 2017,2019, publicity chair for ACM ICMI 2018, sponsorship and special session chair for ISCSLP 2018, 2020, and a guest editor in Journal of Computer Speech and Language on special issue of Speech and Language Processing for Behavioral and Mental Health. He led a team to the 1st place in Emotion Challenge in Interspeech 2009, and with his students won the 1st place in Styrian Dialect and Baby Sound subchallenge in Interspeech 2019. He is a coauthor on the best paper award/finalist in Interspeech 2008, 2010, 2018, IEEE EMBC 2018, 2019, 2020, APSIPA ASC 2019, and the most cited paper published in 2013 in Journal of Speech Communication.

He is a PI of the MOST AI Innovation Grant, is involved in multiple industry and collaborative research projects, e.g., medical centers (NTUH, VGHTC, CMGH), Allianz Insurance, E.SUN Commercial Bank, Gamania Group, C-Media Electronics, etc., and cofounds the startup AHEAD Medicine. His team’s research has been featured in Discovery, Business Today, Technews, and several major news outlets in Taiwan.

Multichannel audio source separation based on unsupervised and semi-supervised learning

Abstract:

Blind source separation (BSS) is an unsupervised learning approach for estimating original source signals using only mixed signals observed in multichannel inputs. In particular, BSS algorithms based on independent component analysis (ICA) and independent vector analysis (IVA), in which the independence among source signals is mainly used for the separation, have been studied actively in the past decade. In this talk, looking back their history from ICA to IVA, we focus our attention on the new extension to low-rank spectrogram modeling and sparse representation, and introduce independent low-rank matrix analysis (ILRMA) and multichannel nonnegative matrix factorization (MNMF). In ILRMA, several source models based on complex heavy-tailed distributions are explained with the discussion on fruitful relation between non-Gaussianity and low-rankness. Finally, thanks to audio big data capability, ILRMA (MNMF) and deep learning are combined, resulting in the sophisticated hybrid method "independent deeply learned matrix analysis (IDLMA)." In addition to the theoretical basis of the algorithms, some applications combining BSS and real-world audio system will be reviewed, e.g., binaural hearing-aid system and distributed microphone array system for speech detection.

Biography:

Hiroshi Saruwatari received the B.E., M.E., and Ph.D. degrees from Nagoya University, Japan, in 1991, 1993, and 2000, respectively. He joined SECOM IS Laboratory, Japan, in 1993, and Nara Institute of Science and Technology, Japan, in 2000. From 2014, he is currently a Professor of The University of Tokyo, Japan. His research interests include statistical speech signal processing, blind source separation (BSS), audio enhancement, and robot audition. He has successfully achieved his carrier, especially on BSS researches including theoretical bridge between unsupervised learning and spatial signal processing, and development of the real-time algorithm. He has put his research into the world's first commercially available Independent-Component-Analysis-based BSS microphone in 2007. He published 105 refereed original papers of international journals and 330 conference papers, getting more than 7300 citations. He received paper awards from IEICE in 2001 and 2006, from TAF in 2004, 2009 and 2012, from IEEE-IROS2005 in 2006, and from APSIPA in 2013 and 2018. He received DOCOMO Mobile Science Award in 2011, Ichimura Award in 2013, The Commendation for Science and Technology by the Minister of Education in 2015, and Achievement Award from IEICE in 2017. He won the first prize in IEEE MLSP2007 BSS Competition. He has been professionally involved in various volunteer works for IEEE, EURASIP, IEICE, and ASJ, including chair posts of international conferences and associate editor of journals.

Spoofing Attacks in Automatic Speaker Verification (ASV): Analysis and Countermeasures (This talk is also an ISCA Distinguished Lecture)

Abstract:

Speech is most natural way of communication between humans and it carries various levels of information, such as linguistic content, emotion, acoustic environment, language, speaker's identity and health conditions, etc. Speaker recognition verifies or identifies a speaker via his/her voice. Automatic Speaker Verification (ASV) involves verifying the claimed speaker's identity. In practice, we would like a speaker verification system to be *robust* against variations, such as microphone and transmission channel, intersession, acoustic noise, speaker ageing, etc. This robustness makes ASV system to be *vulnerable* to various spoofing attacks as it tries to nullify these effects and make spoofed speech more close to the natural speech. Hence, we would like the system to be secure against spoofing attacks. In this talk, difference issues concerning the robustness and security of a speaker verification system were discussed. We also discuss the latest progress and the research activities in anti-spoofing countermeasures against voice conversion (VC), speech synthesis (SS), replay, twins and professional mimics. In particular, brief details of risk and technological challenges associated with each of these attacks were discussed. The talk will also gave brief overview of three international challenge campaigns, namely, ASV Spoof 2015, ASV Spoof 2017 and ASV Spoof 2019 organized during INTERSPEECH 2015, INTERSPEECH 2017, and INTERSPEECH 2019, respectively. This talk also bring out connection of spoofing research with recent ongoing research on attackers perspective for ASV and first Voice Privacy Challenge 2020 during INTERSPEECH 2020. Finally, the talk concludes with overall summary of current state-of-the-art in this field and discusses future research directions.

Biography:

Hemant A. Patil received Ph.D. degree from the Indian Institute of Technology (IIT), Kharagpur, India, in July 2006. Since 2007, he has been a faculty member at DA-IICT Gandhinagar, India and developed Speech Research Lab recognized as ISCA speech labs at DA-IICT. Dr. Patil is member of ISCA, IEEE, IEEE Signal Processing Society, IEEE Circuits and Systems Society, EURASIP, APSIPA and an affiliate member of IEEE SLTC. He is regular reviewer for ICASSP and INTERSPEECH, Speech Communication, Elsevier, Computer Speech and Language, Elsevier and Int. J. Speech Tech, Springer, Circuits, Systems and Signal Processing, Springer. He has published around **250+** research publications in national and international conferences/journals/book chapters. He visited department of ECE, University of Minnesota, Minneapolis, USA (May-July, 2009) as short term scholar. He has been associated (as PI) with three MeitY sponsored projects in ASR, TTS and QbESTD. He was co-PI for DST sponsored project on India-Digital Heritage (IDH)-Hampi. His research interests include speech and speaker recognition, analysis of spoofing attacks, TTS, and infant cry analysis. He has received DST Fast Track Award for Young Scientists for infant cry analysis. He has coedited four books with Dr. Amy Neustein (EIC, IJST Springer) with titles, Forensic Speaker Recognition (Springer, 2011), Signal and Acoustic Modeling for Speech and Communication Disorders (DE GRUYTER, 2018), Voice Technologies for Speech Reconstruction and Enhancement (DE GRUYTER, 2020), and Acoustic Analysis of Pathologies from Infant to Young Adulthood (DE GRUYTER, 2020).

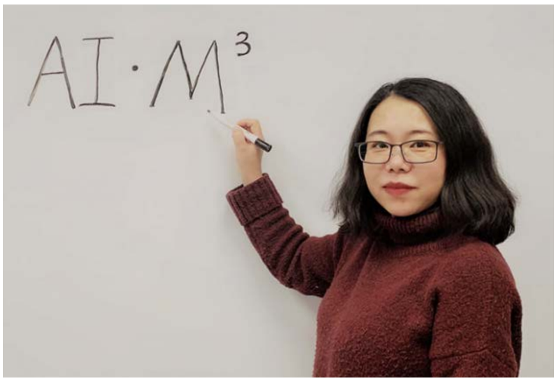
Dr. Patil has taken a lead role in organizing several ISCA supported events, such as summer/winter schools/CEP workshops (on theme as speaker and language recognition, speech source modeling, text-to-speech synthesis, speech production-perception link, advances in speech processing) and progress review meetings for two MeitY consortia projects all at DA-IICT Gandhinagar. Dr. Patil has supervised 05 doctoral and 42 M.Tech. theses (all in speech processing area). Presently, he is supervising 03 doctoral and 03 masters students. Recently, he offered a joint tutorial with Prof. Haizhou Li during Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC) 2017, and INTERSPEECH 2018. He offered a joint tutorial with H. Kawahara on the topic, “Voice Conversion: Challenges and Opportunities,” during APSIPA ASC 2018, Honolulu, USA. He has been selected as APSIPA Distinguished Lecturer (DL) for 2018-2019 and he has 20 APSIPA DLs in four countries, namely, India, Singapore, China, and Canada. Recently, he is selected as **ISCA Distinguished Lecturer** (DL) for 2020-2021 and delivered 05 ISCA DLs in India.

Natural Language Description of Image and Video Content

Abstract:

Describing an image or a video in natural language is a challenging task that bridges the gap between vision and language. Automatic image/video content description requires 1) understanding of many entities including background scene, humans, objects, actions, and events etc based on computer vision techniques; and 2) expressing the content using natural language sentences based on natural language processing techniques. In this talk, professor Jin will present the recent works from AI.M^3 lab on image and video description from different aspects, such as from coarse level to fine-grained level, from single sentence description to paragraph description, and from single modality to multiple modalities.

Biography:



Qin Jin is a professor in School of Information at Renmin University of China, who is leading the Multi-level Multi-aspect Multimedia Analysis (AI.M^3) research group. She received her Ph.D. degree in Language and Information Technologies from Carnegie Mellon University in 2007 and her B.Sc. and M.S. degrees in computer science and technologies from Tsinghua University, Beijing, China in 1996, 1999, respectively. She has research interest in multimedia computing and human computer interaction. Her recent works on image/video captioning and emotion detection have won awards in various international challenge evaluations. She was an APSIPA Distinguished Lecturer for term 2015-2016. She serves an Associate Editor of ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM).

Signal Processing an AI for Spatial Augmented Reality Audio (SARA): ***Sensing, Control and Rendering***

Abstract:

With the advent of highly interactive virtual and augmented reality head gears, spatial audio processing is becoming an important component to create a highly realistic and immersive audio experience. The demands of these new audio realities bring exciting challenges in rendering immersive audio in earphones and headphones on-the-fly. We survey some of the current signal processing and AI/machine learning techniques that are used in achieving plausible immersive sound perception and active hear through listening for different types of hearing devices (headphones, earphones, and micro-speakers). We will also share our adaptive signal processing approach in capturing fast and continuous individualized Head-Related-Transfer Function (HRTF) cues for spatial audio rendering in headphones

Biography:

Woon-Seng Gan received his BEng (1st Class Hons) and PhD degrees, both in Electrical and Electronic Engineering from the University of Strathclyde, UK in 1989 and 1993 respectively. He is currently a Professor of Audio Engineering and Director of the Smart Nation Lab in the School of Electrical and Electronic Engineering in Nanyang Technological University. He is also the Programme Director for the Singtel-NTU Cognitive and Artificial Intelligence Joint Lab. From 2011-2014, he also served as the Head of the Information Engineering Division in the School of Electrical and Electronic Engineering in Nanyang Technological University. His research has been concerned with the connections between the physical world, signal processing and sound control, which resulted in the practical demonstration and licensing of spatial audio algorithms, directional sound beam, and active noise control for headphones and window.

He has published more than 350 international refereed journals and conferences and has translated his research into 6 granted patents and has founded a company, Immersive Sound Technology Pte Ltd in 2015. He had co-authored three books on Subband Adaptive Filtering: Theory and Implementation (John Wiley, 2009); Embedded Signal Processing with the Micro Signal Architecture, (Wiley-IEEE, 2007); and Digital Signal Processors: Architectures, Implementations, and Applications (Prentice Hall, 2005). He has also been invited to give keynotes, plenary talks, and tutorials in Audio Engineering Society conference, American Society of Acoustics, IEEE International conference on Acoustic, Speech and Signal Processing, and Asia Pacific Signal and Information Processing Association.

He is a Fellow of the Audio Engineering Society(AES), a Fellow of the Institute of Engineering and Technology(IET), and a Senior Member of the IEEE. He served as an Associate Editor of the IEEE/ACM Transaction on Audio, Speech, and Language Processing (TASLP; 2012-15) and was presented with an Outstanding TASLP Editorial Board Service Award in 2016. He also served as the Associate Editor for the IEEE Signal Processing Letters (2015-2019) and the IEICE transaction (2014-2016) on Fundamentals of Electronics, Communications and Computer Sciences(Japan). He is currently serving as Senior Area Editor in the IEEE Signal Processing Letters (2019-); Associate Technical Editor of the Journal of Audio Engineering Society (JAES; 2013-); Editorial member of the Asia Pacific Signal and Information Processing Association (APSIPA; 2011-) Transaction on Signal and Information Processing; Associate Editor of the EURASIP Journal on Audio, Speech and Music Processing (2007-).

He was elected as Board of Governor and Vice President – Institutional Relations and Education Program for the Asia Pacific Signal and Information Processing Association (APSIPA) in 2014 and 2017, respectively. He also participates in the IEEE Signal Processing Industry DSP committee and IEEE IoT special interest group. He also served as member in the IEEE Signal Processing Design and Implementation of Signal Processing System from 2012-2015; IEEE Signal Processing Education Technical Committee from 2009-2015. Additionally, he is also the General Conference Chair of APSIPA Annual Summit and Conference in 2017.