The 2023 IEEE International Conference on

Image Processing (IEEE ICIP 2023)

30th
ICIP 2023
Kuala Lumpur
IEEE Signal Processing Society
CELEBRATING 75 YEARS
International Conference on Image Processing
OCTOBER 8 -11, 2023

INDUSTRY PROGRAM

Kuala Lumpur Convention Center (KLCC)
October 8-11, 2023 ■ Kuala Lumpur, Malaysia

Organized by

Organizing Committee of IEEE ICIP 2023

Coordinated by

Industry Co-Chairs of IEEE ICIP 2023:

(Lead) Jianquan Liu, NEC Corporation, Japan
(Local Affairs) Hezerul Abdul Karim, Multimedia University, Malaysia
(Industry Innovation Forums) Guan-Ming Su, Dolby Labs, USA
(Industry Expert Sessions) Kong Aik Lee, A*STAR, Singapore
(Industry Demonstrations) Yeqing Li, Google, USA
(Industry Exhibitions) Masato Ishii, Sony Research Inc., Japan

Patrons

Google
YouTube
Meta
MITSUBISHI ELECTRIC Changes for the Better
DATAFORCE
Springer
SURFING TECH
Meet in Malaysia
Malaysia Truly Asia
Malaysia Convention & Exhibition Bureau

TABLE OF CONTENTS

MESSAGE FROM INDUSTRY CHAIRS

Dear IEEE ICIP 2023 Attendees,

As the Industry Co-chairs, we are thrilled to present the comprehensive and exciting Industry Program for IEEE ICIP 2023. This year, we have dedicated ourselves to achieving the following mission:

1. Enriching industry content
2. Increasing the impact of industry
3. Attracting more attendees from industry
4. Improving the engagement between industry and academia

Our mission is driven by the desire to create an unparalleled industry program that will elevate IEEE ICIP to new heights of innovation, collaboration, and impact.

The foundation of this year's industry program is built on rethinking and reshaping key factors that will enhance the industry's impact at the conference. We have focused on fostering collaboration by establishing new components within the program. These include industry keynotes from SVP/VP-level executives, industry innovation forums, industry expert sessions, industry workshops, and industry seminars, all aimed at attracting high-profile industry researchers to share their practical insights and experiences from the industrial perspective.

This year's industry program features a diverse set of rich contents. We are honored to host three industry keynote speeches delivered by Dr. Shuicheng Yan (Kunlun 2050 Research & Skywork AI), Dr. Shriram Revankar (Dolby Labs), and Dr. Keiji Yamada (NEC Corporation). Moreover, the program will include an industry innovation forum, seven expert talks from distinguished industry researchers, an industry workshop jointly organized by Google and Meta, two industry seminars organized by NEC and NTT, and six industry demonstrations showcasing their cutting-edge technologies.

In addition to these remarkable sessions, we are grateful for the significant contributions from over 20 well-known companies, including Google, Meta, Microsoft, Dolby Labs, NEC, NTT, Samsung, Mitsubishi, Sony, and more. These partnerships are essential to creating a truly innovative and comprehensive industry program.

We would like to express our gratitude to the General Chairs of ICIP 2023 for their unwavering support and commitment to the industry program. Furthermore, we are incredibly thankful for the invaluable contributions from all industry speakers, as their expertise and practical insights are the pillars of the program’s success.

We invite you to join us in this exceptional industry program and believe that it will provide a wealth of knowledge, networking opportunities, and industry-academic engagement. We are confident that IEEE ICIP 2023 will pave the way for future collaborations and innovations that will ultimately lead to sustained growth and advancement in the field of image processing.

Warm regards,

Industry Co-chairs of IEEE ICIP 2023:
(Lead) Jianquan Liu, NEC Corporation, Japan
(Local Affairs) Hezerul Abdul Karim, Multimedia University, Malaysia
(Industry Innovation Forums) Guan-Ming Su, Dolby Labs, USA
(Industry Expert Sessions) Kong Aik Lee, A*STAR, Singapore
(Industry Demonstrations) Yeqing Li, Google, USA
(Industry Exhibitions) Masato Ishii, Sony Research Inc., Japan

INDUSTRY PROGRAM AT A GLANCE (INDUSTRY PROGRAM - ONLY)

Date Time	Monday, 9 October (Plenary Hall)		Tuesday, 10 October (Plenary Hall)		Wednesday, 11 October (Plenary Hall)	
08:30 - 09:00	Opening Ceremony					
09:00 - 10:30	Plenary Talk 1		Plenary Talk 2		Plenary Talk 3	
10:30 - 11:00	Coffee Break (Exhibition Hall 1-3)					
11:00 - 12:30	**Industry Demo** (Day 1)	**Industry Expert Session 1**	**Industry Demo** (Day 2)	**Industry Expert Session 2**	**Industry Demo** (Day 3)	**Industry Expert Session 3**
12:30 - 13:30	Lunch Break (Exhibition Hall 1-3)					
13:30 - 14:30	**Industry Keynote 1** Dr. Shuicheng Yan		**Industry Keynote 2** Dr. Shriram Revankar		**Industry Keynote 3** Dr. Keiji Yamada	
14:30 - 16:00	**Industry Demo** (Day 1)	**Meta & Google Workshop**	**Industry Demo** (Day 2)	**Industry Innovation Forum**	**Industry Demo** (Day 3)	**Industry Seminar (NTT)**
16:00 - 16:30	Coffee Break (Exhibition Hall 1-3)					
16:30 - 18:00	**Industry Demo** (Day 1)	**Meta & Google Workshop**	**Industry Demo** (Day 2)	**Industry Seminar (NEC)**	**Industry Demo** (Day 3)	
18:00 - 22:30			Conference Banquet (Majestic Hotel)		Closing Ceremony	

INDUSTRY KEYNOTE SPEECH I

Foundations of Foundation Models

Dr. Shuicheng Yan
Managing Director of Kunlun 2050 Research & Co-CEO of Skywork AI, Singapore

ABSTRACT Foundation models play a crucial role in the AI community and are considered essential for the success of AGI. In this talk, I will discuss three types of foundations for foundation models. 1) Parameter Optimizer: I will introduce a new optimizer called Adan from a theoretical perspective and demonstrate its effectiveness in improving the convergence rate of various foundation models by 1.5-2x. 2) Network Architecture: I will showcase how insights from neuroscience have influenced and will continue to drive the development of new deep learning network architectures. 3) Representative/Barrier Data: I will showcase the exemplar prototypes of Skywork AI to demonstrate the importance of representative data.

BIOSKETCH Dr. Shuicheng Yan is currently Managing Director of Kunlun 2050 Research and Co-CEO of Skywork AI, and former Group Chief Scientist of Sea. He is Fellow of Singapore's Academy of Engineering, AAAI, ACM, IEEE, and IAPR. His research areas include computer vision, machine learning, and multimedia analysis. Till now, he has published over 800 papers at top international journals and conferences, with an H-index of 140+. He has also been named among the annual World's Highly Cited Researchers eight times. His team received ten-time winners or honorable-mention prizes at two core competitions, Pascal VOC and ImageNet (ILSVRC), deemed the “World Cup” in the computer vision community. Besides, his team won more than ten best papers and best student paper awards, particularly a grand slam at the ACM Multimedia, the top-tiered conference in multimedia, including the Best Paper Awards three times, Best Student Paper Awards twice, and Best Demo Award once.

INDUSTRY KEYNOTE SPEECH II

Next-Generation Media: Content-Aware Processing and Personalization

Dr. Shriram Revankar
Senior Vice President of Dolby Labs, USA

ABSTRACT Over the course of history, media has undergone a series of transformative changes, while significantly influencing society, culture, and politics. From many revolutions to the world wars, media has been instrumental in shaping our perceptions and interactions with the world. Today, the convergence of high-speed connectivity, big-data, cloud computing, and generative AI has provided us with the ingredients needed to transform the end-to-end media workflow from creation, distribution, to rendering and consumption. This talk aims to introduce the defining characteristics of next-generation media. Through demos and examples, I will make a compelling case for a central role for content-awareness in shaping the emerging media technologies. From the inception of media content, through its formats, encoding, delivery, and rendering, content-aware algorithms are set to revolutionize every facet of the media ecosystem. Historically, optimization for human perceptions and broadcast architectures have been central to shaping of media technology. However, content awareness will not only integrate current media optimizations but also account for quantum leaps in network, big-data, and cloud infrastructures. Enabling immersive and engaging consumption through deep personalization will be the key guiding principle of this imminent transformation of the media.

BIOSKETCH As Senior Vice President of the Advanced Technology Group, Dr. Shriram Revankar oversees teams creating and delivering innovations that transform the sight and sound of immersive entertainment experiences, including the office of the Chief Technology Officer, image and sound research and development (R&D), prototyping, and technical operations. Shriram previously served as Vice President and Fellow at Adobe where he focused on delivering high impact technologies to Adobe’s Digital Experience and Document Cloud Businesses. Through his global lab leadership across the US and India, Shriram expanded on the world-class research competencies in AI, data science, social network analysis, data mining, big data analytics, natural language processing, knowledge management, machine learning and related areas at Adobe Research. Prior to joining Adobe, Shriram was a Xerox Fellow and the head of Smart Systems Lab and also the Chief Architect of the Production Solutions business at Xerox Corporation. He has a master’s and Ph.D. in computer science from SUNY Buffalo. His research interests include business intelligence and automation, smart and adaptive systems, AI and data science, computer vision and image processing, and analysis of social networks.

INDUSTRY KEYNOTE SPEECH III

Safer and more productive cities with multi-media processing

Dr. Keiji Yamada
Executive Professional of NEC Corporation, Japan

ABSTRACT Urbanization is progressing around the world. The urban population ratio is already 50% today and is expected to reach around 70% by 2050. However, each city has its own unique problems that cannot be solved by specific technologies alone. For example, London and Tokyo, for example, face aging urban infrastructure problems behind their redevelopment. Rapidly developing cities such as Bangalore are experiencing traffic congestion and logistical stagnation due to inadequate infrastructure. All cities also have problems such as chronic hospital congestion due to diverse causes. Multimedia processing, including image processing, can contribute to solving these various urban problems. In this presentation, we will discuss applications and how they can be solved through a combination of technologies.

BIOSKETCH Keiji YAMADA, Ph.D. is an Executive Professional of NEC R&D division after serving as Vice President of NEC Corporation, Senior Vice President (R&D) of NEC Asia Pacific, and Senior Vice President (R&D) of NEC India. His research covers pattern recognition, machine learning, knowledge acquisition and processing, signal processing, and intelligent system design as their applications. He has a long experience in smart city system development in ASEAN countries including Singapore, India, Japan, the UK, the US, and so on. He carries out R&D on digital transformation in several industrial domains like manufacturing, health and medical care, logistics, mobility, and city infrastructure management. He was an adjunct professor of Nara Advanced Institute of Science and Technology and contributed to several academic societies as a governing board member of IAPR, a board member of the IEICE, Japan, etc.

INDUSTRY PROGRAM (MONDAY, OCTOBER 9)

08:30 - 09:00 Opening Ceremony Plenary Hall

09:00 - 10:30 Plenary Talk I Plenary Hall

10:30 - 11:00 Coffee Break (Exhibition Hall 1-3)

11:00 - 12:00 Industry Expert Session 1: AR/VR and MOQ Plenary Hall
Session Chair: Prof. Zhiping Lin, *Nanyang Technological University, Singapore*

IES #1 Near-eye light field AR/VR display
Homer Chen, PetaRay Inc, Taiwan
(w/ Q&A, 30 minutes)

IES #2 Need for Low Latency: Media over QUIC
Ali C. Begen, Comcast NBCUniversal, USA
(w/ Q&A, 30 minutes)

12:30 - 13:30 Lunch Break (Exhibition Hall 1-3)

13:30 - 14:30 Industry Keynote Speech I Plenary Hall
Session Chair: Dr. Hezerul Abdul Karim, *Multimedia University, Malaysia*

Foundations of Foundation Models
Dr. Shuicheng Yan, *Managing Director of Kunlun 2050 Research & Co-CEO of Skywork AI, Singapore*

14:30 - 18:00 Meta & Google Workshop: Alliance for Open Media (AOMedia) Plenary Hall
Session Chair: Dr. Ioannis Katsavounidis, Research Scientist, Meta

#1 Introduction to the Alliance for Open Media
Dr. Ioannis Katsavounidis, *Research Scientist, Meta*

#2 SIWG / SVT-AV1: 15-Month Milestones and Achievements
Mr. Hassene Tmar, *Meta*

#3 AV1 Deployment at Meta
Dr. Ryan Lei, *Video Codec Specialist, Meta*

#4 AV1 Deployment @ Netflix: Past, Present & Future
Mr. Anush Moorthy, *Netflix*

#5 YouTube's AV1 Activities
Dr. In Suk Chong, *Google*

#6 UVQ in Video Compression
Dr. Yilin Wang (*YouTube/Google*), Dr. Balu Adsumili (*YouTube/Google*)

#7 Towards a Next-Gen Video Codec
Dr. Debargha Mukherjee, *Principal Engineer, Google*

#8 AOM Common Test Condition Design and Latest Result
Dr. Yeping Su (*Google*), Dr. Ryan Lei (*Meta*)

#9 Overview of Coding Tools Under Consideration in AVM
Dr. Debargha Mukherjee, Mr. Xin Zhao, Dr. Onur Guleryuz, Mr. Joe Young (*Google*)

16:00-16:30 Coffee Break (Exhibition Hall 1-3)

11:00 - 12:30 Industry Demonstration (Day 1) Exhibition Hall 1-3
14:30 - 16:00 Session Chair: Dr. Hezerul Abdul Karim, *Multimedia University, Malaysia*
16:30 - 18:00

Demo #1 JPEG XS Low latency error robust low latency video transport
Thomas Richter, Siegfried Fößel (Fraunhofer IIS)

Demo #6 Media over QUIC: Initial Testing, Findings and Results
Zafer Gurel, Ozyegin University (Turkiye), Tugce Erkilic Civelek, Ozyegin University (Turkiye), Ali C. Begen, Ozyegin University (Turkiye) and Comcast (USA), Alex Giladi, Comcast (USA)

INDUSTRY PROGRAM (TUESDAY, OCTOBER 10)

09:00 - 10:30 Plenary Talk II — Plenary Hall

10:30 - 11:00 Coffee Break (Exhibition Hall 1-3)

11:00 - 12:30 Industry Expert Session 2: AI and Neuromorphic — Plenary Hall
Session Chair: Dr. Yuangen Wang, *Guangzhou University, China*

IES #3 JPEG AI the new image compression standard entirely based on neural networks
Elena Alshina, Huawei Technologies Dusseldorf GmbH, Germany
(w/ Q&A, 30 minutes)

IES #4 Adaptive Camera Adjustment with AI
Hui Lam Ong, NEC Laboratories Singapore, Singapore
(w/ Q&A, 30 minutes)

IES #5 Low Power Image Processing on Constrained Devices Using Tiny ML and Neuromorphic Approach
Arpan Pal, TCS Research, Tata Consultancy Services, India
(w/ Q&A, 30 minutes)

12:30 - 13:30 Lunch Break (Exhibition Hall 1-3)

13:30 - 14:30 Industry Keynote Speech II — Plenary Hall
Session Chair: Dr. Jianquan Liu, *NEC Corporation, Japan*

Next-Generation Media: Content-Aware Processing and Personalization
Dr. Shriram Revankar, *Senior Vice President of Dolby Labs, USA*

14:30 - 16:00 Industry Innovation Forum: Emerging AI Trends in Image and Video Industry
Moderator: Dr. Ning Xu, *Fellow of Advanced R&D at Adeia Inc, USA*
Plenary Hall

Learning-based models in Sensing Applications
Dr. Petros T. Boufounos, Deputy Director at Mitsubishi Electric Research Laboratories (MERL), USA

Recent trends for On-device AI-Camera
Dr. Mostafa El-Khamy, Senior Principal Engineer at Samsung SOC Research and Development, USA

Robust Face Anti-Spoofing in Unseen Domains via Geometry-Aware Networks
Dr. Kai-Lung Hua, CTO of Microsoft Taiwan, Taiwan

Adaptive Camera Adjustment with AI
Mr. Hui-Lam Ong, Principal Solution Architect at NEC Laboratories Singapore, Singapore

Panel discussions

16:00 - 16:30 Coffee Break (Exhibition Hall 1-3)

16:30 - 18:00 Industry Seminar: Exploring NEC's Recognition AIs: Biometrics, Fairness, and Behavior Analysis
Moderator: Dr. Jianquan Liu, *NEC Corporation, Japan*

Plenary Hall

#1 Toward Real-time End-to-End Multi-Object Tracking Model for Biometric Solution
Dr. Hiroshi Fukui, NEC Corporation, Japan

#2 Toward Fair Face Analysis Systems: Metric Learning Loss with Adaptive Margins for Fair Multi-label Face Attribute Recognition
Dr. Masashi Usami, NEC Corporation, Japan

#3 Towards annotation-free image recognition AIs
Dr. Tomokazu Kaneko, NEC Corporation, Japan

#4 Harnessing Domain Knowledge of Intra-Class Variations to Mitigate Label Scarcity Bias (in Satellite Imagery)
Ms. Tsenjung Tai, NEC Corporation, Japan

#5 The Power of Retrieval for Video Analysis on Human Behavior Understanding
Dr. Jianquan Liu, NEC Corporation, Japan

11:00 - 12:30 Industry Demonstrations (Day 2) Exhibition Hall 1-3
14:30 - 16:00 Session Chair: Dr. Jianquan Liu, *NEC Corporation, Japan*
16:30 - 18:00

Demo #7 Camera-based Gaze Tracker Driven Robotic and Assisted Living/Hospital Bed Use-cases
Mithun B S,Tince Varghese, Aditya Choudhary, Rahul Dasharath Gavas, Ramesh Kumar Ramakrishnan, Arpan Pal (TCS Research, India)

Demo #8 VVC in a large-scale streaming environment
Kevin Rocard (Bitmovin), Jacob Arends (Bitmovin), Adam Wieckowski (Fraunhofer HHI), Benjamin Bross (Fraunhofer HHI)

INDUSTRY PROGRAM (WEDNESDAY, OCTOBER 11)

09:00 - 10:30 Plenary Talk III Plenary Hall

10:30 - 11:00 Coffee Break (Exhibition Hall 1-3)

11:00 - 12:00 Industry Expert Session 3: OCR and Video Plenary Hall
Session Chair: Dr. Lu Wang, *A*STAR Institute for Infocomm Research, Singapore*

IES #6 arTXTract – Extracting Text from Challenging Paper Documents in a FinST
Oliver Giudice, Banca D'Italia, Italy
(w/ Q&A, 30 minutes)

IES #7 Enhancing Content Experiences with Contextual Data
Viswanathan (Vishy) Swaminathan, Adobe Research, Adobe, USA
(w/ Q&A, 30 minutes)

12:30 - 13:30 Lunch Break (Exhibition Hall 1-3)

13:30 - 14:30 Industry Keynote Speech III Plenary Hall
Session Chair: Prof. Mohan Kankanhalli, *National University of Singapore*

Safer and more productive cities with multi-media processing
Dr. Keiji Yamada, *Executive Professional of NEC Corporation, Japan*

14:30 - 16:00 Industry Seminar: NTT's Media Processing AI and Its Industrial Applications
Moderator: Dr. Satoshi Suzuki, *NTT Corporation, Japan*
Plenary Hall

#1 MediaGnosis: the next-generation media processing artificial intelligence
Dr. Ryo Masumura, NTT Corporation, Japan

#2 geoNebula: Elemental Technologies for Supporting the Integration of Real Space and Cyberspace
Dr. Satoshi Suzuki, NTT Corporation, Japan

#3 Conversational system that talks about the scenery seen from vehicles
Dr. Hiroaki Sugiyama, NTT Corporation, Japan

#4 Distributed AI Video Analytics
Ms. Monikka Roslianna Busto, NTT Corporation, Japan

11:00 - 12:30 Industry Demonstration (Day 3) Exhibition Hall 1-3
14:30 - 16:00 Session Chair: Dr. Masato Ishii, *Sony Research Inc, Japan*
16:30 - 18:00

Demo #9 A Real-time Chinese Food Auto Billing System based on Synthetic images
Qiushi Guo, Yifan Chen, Jin Ma, Tengteng Zhang (China Merchants Bank)

Demo #10 Instant Object Registration System for Image Recognition of Retail Products
Tomokazu Kaneko, Soma Shiraishi, Makoto Terao (NEC Corporation)

DETAILS OF INDUSTRY INNOVATION FORUMS

Forum Scheme: Emerging AI Trends in Image and Video Industry
Moderator: Dr. Ning Xu, Fellow of Advanced R&D at Adeia Inc, USA
Panelists:

- ✧ Dr. Petros T. Boufounos, Deputy Director at Mitsubishi Electric Research Laboratories (MERL), USA
- ✧ Dr. Mostafa El-Khamy, Senior Principal Engineer at Samsung SOC Research and Development, USA
- ✧ Dr. Kai-Lung Hua, CTO of Microsoft Taiwan, Taiwan
- ✧ Mr. Hui-Lam Ong, Principal Solution Architect at NEC Laboratories Singapore, Singapore

Dr. Ning Xu currently serves as Fellow, Advanced R&D at Adeia Inc, pioneering innovations that enhance the way we live, work, and play. Before joining Adeia, Dr. Xu was the Chief Scientist of Video Algorithms at Kuaishou Technology, and before that, he held various positions at Amazon, Snap Research, Dolby Laboratories, and Samsung Research America. He earned his Ph.D. in Electrical Engineering from the University of Illinois at Urbana-Champaign (UIUC) in 2005, and his Master’s and Bachelor's degrees from the University of Science and Technology of China (USTC). Dr. Xu has co-authored over 200 journal articles, conference papers, patents, and patent applications. His research interests encompass machine learning, computer vision, video technology, and other related areas. He is a Senior Member of IEEE.

Petros T. Boufounos is a Distinguished Research Scientist and a Deputy Director at Mitsubishi Electric Research Laboratories (MERL), also leading the Computational Sensing Team. Dr. Boufounos completed his undergraduate and graduate studies at MIT. He received the S.B. degree in Economics in 2000, the S.B. and M.Eng. degrees in Electrical Engineering and Computer Science (EECS) in 2002, and the Sc.D. degree in EECS in 2006. Between September 2006 and December 2008, he was a postdoctoral associate with the Digital Signal Processing Group at Rice University. Dr. Boufounos joined MERL in January 2009, where he has been heading the Computational Sensing Team since 2016. Dr. Boufounos' immediate research focus includes signal acquisition and processing, computational sensing, inverse problems, quantization, and data representations. He is also interested in how signal acquisition interacts with other fields that use sensing extensively, such as machine learning, robotics, and dynamical system theory. He has over 40 patents granted and more than 10 pending, and more that 100 peer reviewed journal and conference publications in these topics. Dr. Boufounos has served as an Area Editor and a Senior Area Editor for the IEEE Signal Processing Letters, and as a member of the SigPort editorial board and the IEEE Signal Processing Society Theory and Methods technical committee. He is currently an Associate Editor at IEEE Transactions on Computational Imaging and the general co-chair of the ICASSP 2023 organizing committee. Dr. Boufounos is an IEEE Fellow and an IEEE SPS Distinguished Lecturer for 2019-2020.

Talk: Learning-based models in Sensing Applications

Abstract: Learning and data-driven approaches are becoming increasingly important in sensing and imaging applications. Learning-based models promise to capture characteristics of signals that are difficult to capture analytically and better describe physical processes, compared to analytical models. They are also versatile, as they can be used to refine analytical models and improve their modeling accuracy, reduce the computational complexity of using the model, or describe physical systems and processes for which no good analytical models exist. On the other hand, analytical models do not require training, which can be expensive both in computation and data requirements. In addition, analytical models are more amenable to theoretical analysis and may offer theoretical performance guarantees. This talk will explore how learning-based models enable a number of imaging applications and how combining them with analytical models can significantly reduce the training burden, improve performance, and help with theoretical analysis. We will show a range of approaches, either purely data-driven or combining analytical and learned models to various degrees. We will discuss the tradeoffs in each approach and the benefits and drawbacks, as related to their application in radar imaging, infrastructure monitoring and imaging of dynamical systems, among others.

Mostafa El-Khamy (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Alexandria University, Egypt, the M.S. and Ph.D. degrees in electrical engineering from the California Institute of Technology (Caltech), USA, and the M.B.A. degree from the Edinburgh Business School, U.K. He is currently a Senior Principal Engineer with Samsung SOC Research and Development, CA, USA. He is also an Adjunct Professor at the Faculty of Engineering, Alexandria University. He was a Founding Faculty Member of Egypt-Japan University for Science and Technology (E-JUST) and was at Qualcomm Research and Development, San Diego. His research interests include the theory and practice of artificial intelligence in multimedia and communication systems. He was a recipient of the URSI Young Scientist Award, the Caltech Atwood Fellowship, the Alexandria University Scientific Incentive Award, the Samsung Best Paper Award, and the Samsung Distinguished Inventor Award.

Talk: Recent trends for On-device AI-Camera

Abstract: This talk outlines recent trends for on-device AI. We discuss recent developments for AI-acceleration on mobile devices and how they enable AI-camera. We discuss solutions to enable content aware camera. We dive deeper into AI-methods behind the success of AI-camera, such as scene understanding and accurate scene segmentation for content-aware cameras.

Kai-Lung Hua obtained his B.S. degree in electrical engineering from National Tsing Hua University in 2000 and later achieved an M.S. degree in communication engineering from National Chiao Tung University in 2002, both situated in Hsinchu, Taiwan. He completed his Ph.D. at the School of Electrical and Computer Engineering at Purdue University in West Lafayette, IN, USA in 2010. Starting from 2010, Dr. Hua has been affiliated with the National Taiwan University of Science and Technology. Throughout this time, he has held a range of significant positions, which include being a professor in the CSIE department, serving as the director of the AI research center, taking on the role of vice dean for the EECS College, and leading as the dean of the industry-academia collaboration office. From 2022 onwards, he transitioned to Microsoft Taiwan, where he has taken on the pivotal role of Chief Technology Officer. In this role, he leads the effort to empower Taiwan's industry by harnessing the potential of cloud and AI capabilities to facilitate profound transformative change. Dr. Hua is a distinguished member of Eta Kappa Nu and Phi Tau Phi, and he has been honored as a recipient of the MediaTek Doctoral Fellowship. His research pursuits encompass digital image and video processing, computer vision, and machine learning. His accomplishments include a series of esteemed research awards, notably the 2022 K. T. Li Cornerstone Award from the Institute of Information & Computing Machinery, the 2020 Outstanding Research Award from the National Taiwan University of Science and Technology, the Top Performance Award from the 2017 ACM Multimedia Grand Challenges, the Top 10% Paper Award from the 2015 IEEE International Workshop on Multimedia Signal Processing, and the Second Prize Award from the 2014 ACM Multimedia Grand Challenge, among others.

Talk: Robust Face Anti-Spoofing in Unseen Domains via Geometry-Aware Networks

Abstract: Effective face anti-spoofing (FAS) is vital for a robust face recognition system. Despite the development of numerous texture-based countermeasures to thwart presentation attacks (PAs), their performance against unseen domains or novel spoofing methods remains unsatisfactory. Rather than attempting to comprehensively catalog all potential spoofing variations and rendering binary live/spoof determinations, we present a novel approach to the FAS challenge. Our approach focuses on discerning between normal and abnormal movements within live and spoof presentations. Introducing the Geometry-Aware Interaction Network (GAIN), we harness the power of dense facial landmarks by utilizing a spatio-temporal graph convolutional network (ST-GCN). This technique not only yields a more interpretable and modularized FAS model but also incorporates a cross-attention feature interaction mechanism. This mechanism seamlessly integrates with existing methods, resulting in a notable performance enhancement. Empirical evaluations underscore our approach's state-of-the-art performance across both standard intra-dataset assessments and cross-dataset evaluations. Particularly impressive, our model exhibits a significant performance margin over state-of-the-art methods in the cross-dataset cross-type protocol, as demonstrated on the CASIA-SURF 3DMask DATABASE. This accomplishment highlights our model's robustness against domain shifts and previously unseen forms of spoofing.

Hui Lam Ong specializes in high-performance video analytic solutions. His decades of professional experience in cyber security and software architecture design enables him to understand the complexity of video analytics deployment challenges. His current focus aims to reduce manpower and maintenance cost of large-scale video analytics solutions.

Talk: Bridging the Gaps for the Adoption of Large-Scale Video Analytics Solutions

Abstract: In recent years, city governments worldwide have recognized the benefits of implementing video analytics to enhance and maintain safety and security. Consequently, an increasing number of public areas, such as housing estates, multi-story car parks, and high-traffic transit locations like train stations and bus interchanges, are being equipped with surveillance cameras. The initial deployment process requires a significant financial investment to upgrade the necessary infrastructure, procure high-resolution surveillance cameras, and obtain modern hardware and software for data storage and analysis. Moreover, the employment of highly skilled IT professionals is essential for installing, managing, troubleshooting, and maintaining these systems. This presents a substantial resource challenge for city governments. It's also crucial not to forget the public's privacy concerns. The speaker for this session possesses extensive experience in addressing these challenges. He and his team will provide a comprehensive exploration of these industry-wide issues. Their mission to streamline the process of deploying video analytics solutions has led them to develop a semi-automated, AI-aided system. This advanced system aims to alleviate these challenges by ensuring the protection of individual privacy and simplifying the implementation of these solutions.

DETAILS OF INDUSTRY EXPERT SESSIONS

IES #1

Title: Near-eye light field AR/VR display

Speaker: Homer Chen, PetaRay Inc, Taiwan

Abstract: Human eyes have evolved over millions of years to have consistent vergence and accommodation. However, most AR/VR displays available today can easily cause vergence-accommodation conflict (VAC) that is the root cause of visual fatigue for viewers. At PetaRay, we aim to revolutionize the fundamentals of AR/VR glasses from “showing images to each eye” to “projecting light field to retina.” In this talk, I will show how a light field display with continuous focal plane can provide the most natural visual experiences for users.

Biosktech: Homer H. Chen received the Ph.D. degree in Electrical and Computer Engineering from University of Illinois at Urbana-Champaign. Dr. Chen's professional career has spanned industry and academia. Since August 2003, he has been with the College of Electrical Engineering and Computer Science, National Taiwan University, where he is Distinguished Professor. Prior to that, he held various R&D management and engineering positions with U.S. companies over a period of 17 years, including AT&T Bell Labs, Rockwell Science Center, iVast, and Digital Island (acquired by Cable & Wireless). He was a U.S. delegate for ISO and ITU standards committees and contributed to the development of many new interactive multimedia technologies that are now part of the MPEG-4 and JPEG-2000 standards. His recent research is related to AR/VR, multimedia signal processing, computational photography and display, and music data mining. Dr. Chen is an IEEE Life Fellow. Currently, he serves on the IEEE James H. Mulligan, Jr. Education Medal Committee. Previously, he served on the Senior Editorial Board of the IEEE Journal on Selected Topics in Signal Processing from 2020 to 2022, the Awards Committee from 2019 to 2021, the Conferences Board from 2020 to 2021, the Fourier Award Committee from 2015 to 2017, and the Fellow Reference Committee from 2015 to 2017, all of the IEEE Signal Processing Society. He was a General Co-chair of the 2019 IEEE International Conference on Image Processing and a Distinguished Lecturer of the IEEE Circuits and Systems Society from 2012 to 2013. He was an Associate Editor of the IEEE Transactions on Circuits and Systems for Video Technology from 2004 to 2010, the IEEE Transactions on Image Processing from 1992 to 1994, and Pattern Recognition from 1989 to 1999. He served as a Guest Editor for the IEEE Transactions on Circuits and Systems for Video Technology in 1999, the IEEE Transactions on Multimedia in 2011, the IEEE Journal of Selected Topics in Signal Processing in 2014, and Springer Multimedia Tools and Applications in 2015.

IES #2

Title: Need for Low Latency: Media over QUIC

Speaker: Ali C. Begen, Comcast NBCUniversal, USA

Abstract: This session overviews developing a low-latency solution for media ingest and distribution, the work undertaken by the IETF's new Media over QUIC (moq) working group and many industry-leading companies. It summarizes the motivation, goals, current work and potential improvements.

With the advent of QUIC, initially developed by Google and then standardized by the IETF (RFC 9000), the latest version of HTTP, H3 (RFC 9114), was built upon this new low-latency transport protocol to eliminate the known issues in the previous versions of HTTP (e.g., head-of-line blocking, HoL, due to underlying TCP protocol) and to benefit from other features such as improved congestion control, prioritized delivery and multiplexing. Although H3 outperforms its predecessors in most cases, existing adaptive streaming methods that have been highly tuned for HTTP/1.1 and 2 running on top of TCP do not give remarkably better results, with H3 running over QUIC. If timely delivery is critical, QUIC may perform better than TCP in congested environments. However, we still need a custom application-layer protocol to reap all the benefits QUIC provides at the transport layer. The new IETF working is chartered to study how to use QUIC for large-scale media transmission in one-to-one, one-to-many and many-to-one applications that might require interactivity (hence, low latency). The working group is still exploring the problem space and potential solutions, and a few different proposals are already being considered. MOQ is envisioned to be a common media protocol stack that will support (i) live streaming of events, news and sports with interactivity features, and (ii) scaling real-time collaboration applications to large audiences.

Rationale: Streaming has become increasingly important, effectively replacing most older media delivery models. Given that this is a popular research and industry topic, the audience would benefit from understanding the industry needs, constraints, pain points, and unsolved problems.

Biosktech: Ali C. Begen is currently a computer science professor at Ozyegin University and a technical consultant in Comcast's Advanced Technology and Standards Group. Previously, he was a research and development engineer at Cisco. Begen received his PhD in electrical and computer engineering from Georgia Tech in 2006. To date, he received several academic and industry awards (including an Emmy® Award for Technology and Engineering), and was granted 30+ US patents. In 2020 and 2021, he was listed among the world's most influential scientists in the subfield of networking and telecommunications. More details are at https://ali.begen.net.

IES #3

Title: JPEG AI the new image compression standard entirely based on neural networks

Speaker: Elena Alshina, Huawei Technologies Dusseldorf GmbH, Germany

Abstract: Reflecting significant progress of AI-based algorithms for image compression JPEG launched the JPEG AI standardization project. This is the first-ever international standard entirely based on AI technologies. JPEG AI is a multi-task codec targeting not only superior image reconstruction for humans but is also capable to solve computer vision and image enhancement tasks from the same latent representation. The first version of JPEG AI will be finalized at the beginning of 2024. It is expected that JPEG AI will have two profiles: base profile with a target complexity 20 kMAC/pxl (which is acceptable for mobile devices) providing 10-15% compression gain over VVC Intra coding, and high profile which is ~10 times more complex but provides 30% compression gain over VVC anchor. The talk to be presented in Industry Expert Session will focus on an overview of JPEG AI standard key design elements, major challenges of AI-based codec standardization *such as device interoperability), and deployment perspective. Also, a JPEG AI demo on mobile devices is planned.

Biosktech: Dr. Elena Alshina graduated from Moscow State University (majored in Physics) and received PhD in mathematical modeling from the Russian Academy of Science in 1998. For a series of publications on computational math, together with Alexander Alshin she was awarded the Gold Medal of the Russian Academy of Science. In 2006 she joined Samsung and start working on video codec standard development, actively participating in HEVC and VVC standard development, authoring 1000+ proposals. In 2018 she joined Huawei Technologies as Chief Video Scientist, alter became also Lab Director for Audiovisual Technology Lab and Media Codec Lab. Since 2020 she is co-chair and editor of the JPEG AI standard. JPEG AI CfP response submitted by a team led by her demonstrated thy highest objective performance, outperforming VVC anchor by 32%.

IES #4

Title: Adaptive Camera Adjustment with AI

Speaker: Hui Lam Ong, NEC Laboratories Singapore, Singapore

Abstract: In recent years, city governments across the globe have recognized the benefits of implementing video analytics to enhance and maintain safety and security. As a result, an increasing number of public areas, including housing estates, multi-storey car parks, and well-trafficked transit areas like train stations and bus interchanges, are being safeguarded with surveillance cameras. The initial deployment process involves a considerable financial investment to upgrade the necessary infrastructure, procure high-resolution surveillance cameras, and modern hardware and software for data storage and analysis. Furthermore, the employment of highly skilled IT professionals is mandatory to install, manage, troubleshoot, and maintain these systems, presenting another resource challenge for city governments. The speaker for this session is equipped with extensive experience dealing with these challenges. He, along with his team, will provide a comprehensive exploration of these industry-wide issues. Their mission to streamline the process of video analytics solutions deployment has led them to develop a semi-automated, AI-aided system. This advanced system aims to alleviate these challenges and make the implementation of these solutions more accessible and manageable.

Biosktech: Hui Lam Ong specializes in high-performance video analytic solutions. His decades of professional experience in cyber security and software architecture design enables him to understand the complexity of video analytics deployment challenges. His current focus aims to reduce manpower and maintenance cost of large-scale video analytics solutions.

IES #5

Title: Low Power Image Processing on Constrained Devices Using Tiny ML and Neuromorphic Approach

Speaker: Arpan Pal, TCS Research, Tata Consultancy Services, India

Abstract: More and more of IoT based intelligent systems demand embedding the analytics and AI on the edge device. This is driven by the needs of low latency, network unavailability/unreliability and a need for inherent privacy/security. However, edge devices are usually constrained in terms of compute, memory and power which poses challenges for such deployments.

In this presentation we will first introduce a Tiny Edge Wizard that can take large Deep Neural Network models and try to reduce their size / improve their latency automatically using an innovative integrated approach that uses both reduction using pruning using Lottery Ticket Hypothesis (LTH) [1] and synthesis using Neural Architecture Search (NAS) [2]. The automation helps in reducing development time, helps reducing over-parameterised models and minimises the human-expert time. We will take example use cases of medical image processing and manufacturing shop-floor inspection to demonstrate the efficacy of our proposed system.

Next we will introduce Neuromorphic Computing and Spiking Neural Networks (SNN) as a means towards ultra-low power intelligent processing at edge. We cover design and implementation of both efficient spike encoders and spiking neural models on neuromorphic chipsets. We will present application use cases around gesture recognition for human-robot interaction [3] and lossless image compression onboard nano satellites [4] and present some interesting results.

Finally we will conclude with technology trends we see in this area that also relates to sustainable analytics to negate the ever-increasing energy consumption trends in computation.

[1] Ishan Sahu, Arijit Ukil, Sundeep Khandelwal, and Arpan Pal, "LTH-ECG: Lottery Ticket Hypothesis-based Deep Learning Model Compression for Atrial Fibrillation Detection from Single Lead ECG On Wearable and Implantable Devices," 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2022

[2] Shalini Mukhopadhyay, Swarnava Dey, Avik Ghose, Pragya Singh, and Pallab Dasgupta. 2023. Generating Tiny Deep Neural Networks for ECG Classification on Micro-controllers, IEEE International Conference on Pervasive Computing and Communications (PerCom), 2023

[3] Arun M. George, Dighanchal Banerjee, Sounak Dey, Arijit Mukherjee and P. Balamurali, "A Reservoir-based Convolutional Spiking Neural Network for Gesture Recognition from DVS Input," International Joint Conference on Neural Networks (IJCNN), 2020

[4] Sayan Kahali, Sounak Dey, Chetan S. Kadway, Arijit Mukherjee, Arpan Pal and Manan Suri, "Low-Power Lossless Image Compression on Small Satellite Edge Using Spiking Neural Network," International Joint Conference on Neural Networks (IJCNN), 2023

Biosktech: Arpan Pal has more than 30 years of experience in the area of Intelligent Sensing, Signal Processing &AI, Edge Computing and Affective Computing. Currently, as Distinguished Chief Scientist and Research Area Head, Embedded Devices and Intelligent Systems, TCS Research, he is working in the areas of Connected Health, Smart Manufacturing, Smart Retail and Remote Sensing. He is on the editorial board of notable journals like ACM Transactions on Embedded Systems, Springer Nature Journal on Computer Science and is on the TPC of notable conferences like IEEE Sensors, ICASSP and EUSIPCO. He has filed 180+ patents (out of which 95+ granted in different geographies) and has published 160+ papers and book chapters in reputed conferences and journals. He has also written three complete books on IoT, Digital twins in Manufacturing and Application AI in Cardiac screening.

He is on the governing/review/advisory board of some of the Indian Government organizations like CSIR, MeitY, Educational Institutions like IIT, IIIT and Technology Innovation Hub. He is two times winner of Tata Group top Innovation award in Tata Innovista under Piloted technology category. Prior to joining Tata Consultancy Services (TCS), Arpan had worked for DRDO, India as Scientist for Missile Seeker Systems and in Rebeca Technologies as their Head of Real-time Systems. He is a B.Tech and M. Tech from IIT, Kharagpur, India and PhD. from Aalborg University, Denmark.

IES #6

Title: arTXTract – Extracting Text from Challenging Paper Documents in a FinST

Speaker: Oliver Giudice, Banca D'Italia, Italy

Abstract: Today we talk about Natural Language Processing (NLP) and Large Language Models: all useful tools that allow the extraction of value from large amounts of data with fascinating results. To date, however, before being able to use advanced NLP systems, it is necessary to have the data in a digitized form and without imperfections. This hypothesis is not often respected: in many institutions like mine, most of the information assets of the past are paper-based and therefore it is necessary to create increasingly advanced OCR tools in order to be able to extract the textual information components from documents.

The extraction is not simple: the documents have different resolutions, scanning methods, often have imperfections, images, tables, signatures, stamps. All of these problems makes it difficult if not impossible to extract information with OCR tools, whether they are free, open or commercial. In fact, not even the best commercial tools are able to solve all digitization problems and accurately extract all the desired information (i.e. tables should be treated in a specific way). Last but not least, even if the document is already digital it is not often easy to extract information due to the diversity of formats, encodings, etc.

In this talk, arTXTract will be presented: a prototype solution created in-house for handling documents from financial institutions. The pipeline, designed to successfully process a large amount of documents, will be presented, being able to process different kind of data and surpassing the results obtainable with commercial software.

A question therefore arises: is it possible to generalize the process? Or does each institution/company have to equip itself with its own text extraction systems to digitize its archives? At the end of the speech, we will try to give an answer to this problem which, unfortunately, is yet to be totally solved.

Biosktech: Oliver Giudice received his degree in Computer Engineering (summa cum laude) in 2011 at University of Catania and his Ph.D. in Maths and Computer Science in 2017 defending a thesis entitled “Digital Forensics Ballistics: Reconstructing the source of an evidence exploiting multimedia data”. From 2011 to 2014 he was involved in various research projects at University of Catania in collaboration with the Civil and Environmental Engineering department and the National Sanitary System. He was leader of the R&D team of University of Catania for project Farm.PRO (PO/FESR Misura 4.1.1.1) from 2012 to 2014. In 2014 he started his job as a researcher at the IT Department of Banca D’Italia dealing with text classification and crypto-currencies analysis. For various years since 2011 he collaborated with the IPLab (http://iplab.dmi.unict.it) working on Multimedia Forensics topics and being involved in various forensics cases as Digital Forensics Expert. Since 2016 he has been co-founder of “iCTLab s.r.l.”, spin-off of University of Catania, company that works in the field of Digital Forensics, Privacy and Security consulting and software development. His research interests include machine learning, computer vision, image coding, urban security, crypto-currencies and multimedia forensics.

IES #7

Title: Enhancing Content Experiences with Contextual Data

Speaker: Viswanathan (Vishy) Swaminathan, Adobe Research, Adobe, USA

Abstract: It took about 20 years for video over the Internet to be delightful for end users. This was built on a large body of research from both academia and the industry. Over the last few years, AI has provided generational transformations both in content understanding and in machine learning algorithms that learn and adapt using large amounts of contextual data. How do we stand on the shoulders of these giants (transformations) to make next-gen content experiences compelling? I will start with a glimpse of available data from user sessions for on-demand and live videos. I will elaborate on how this fine-grained contextual data can be combined with deep learning powered multimedia understanding to enhance content experiences. The first generation of our video research at Adobe focused on simple insights from content while the second generation focused on insights from video consumption data. Now, the explosion in compute and data enables the third generation of research that derives holistic insights simultaneously from both content and the behavioral data to close the feedback loop to improve content consumption experiences. With a few examples, I will show how to leverage these technological transformations to enhance end-user content experiences ranging from traditional video to immersive mixed-reality experiences. Some demos of past and current projects will be shown with a call to leverage contextual data to improve and personalize end-user media experiences.

Biosktech: Vishy (Viswanathan) Swaminathan is a Sr. Principal Scientist in Adobe Research leading the Enterprises, Platforms, Insights, and Content (EPIC) Research org, working at the intersection of insights from behavioral data and multimedia content. His areas of research include next generation video and immersive experiences, video streaming, and in general data-driven content and marketing technologies. His research work has substantially influenced various technologies in Adobe's video delivery, advertisement, and recommendations products including the guts of HTTP Dynamic Streaming which won the ‘Best Streaming Innovation of 2011′ Streaming Media Readers’ Choice Award. He has received several awards including for best papers, 2017 Distinguished alumnus from Utah State University ECE Department, and 3 ISO certificates of appreciation for contributions to MPEG Standards including editing the EMMY winning MPEG DASH Standard. Prior to joining Adobe, Vishy was a Senior Researcher at Sun Labs. He received his MS and Ph.D. in Electrical Engineering from Utah State University. He received his B.E degree in ECE from the College of Engineering, Guindy, Anna University, Chennai, India. Vishy has authored several papers, articles, RFCs, and book chapters, has about 100+ issued patents, and volunteers in organizing IEEE and ACM conferences.

DETAILS OF INDUSTRY WORKSHOPS

Meta & Google Workshop

Title: Alliance for Open Media Workshop, sponsored by Meta & Google

Abstract: The Alliance for Open Media was formed in 2015 and produced its first specification, AV1, in 2018 with the goal of being a royalty-free video coding standard. We will present an overall status update on the adoption and performance of AV1 as well as progress towards a potential future video coding standard, through sets of new coding tools contributed by AOMedia members. Advances in video quality metrics will also be presented, as well as related topics.

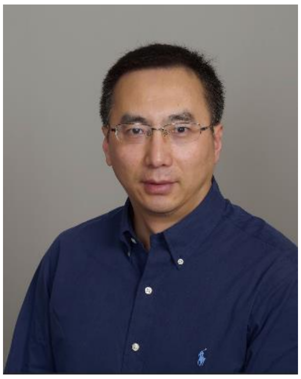
Organizers:

- Dr. Ioannis Katsavounidis, Research Scientist, Video Infrastructure, Meta
- Dr. Ryan Lei, Video Codec Specialist, Video Infrastructure, Meta
- Dr. Debargha Mukherjee, Principal Engineer, Google
- Dr. Balu Adsumilli, Head of Media Algorithms, YouTube/Google

Speakers:

- Dr. Ioannis Katsavounidis (Meta)
- Mr. Hassene Tmar (Meta)
- Dr. Ryan Lei (Meta)
- Mr. Anush Moorthy (Netflix)
- Dr. In Suk Chong (Google)
- Dr. Yilin Wang (YouTube/Google)
- Dr. Balu Adsumili (YouTube/Google)
- Dr. Debargha Mukherjee (Google)
- Dr. Yeping Su (Google)
- Mr. Xin Zhao (Google)
- Dr. Onur Guleryuz (Google)
- Mr. Joe Young (Google)

Dr. Ioannis Katsavounidis is part of the Video Infrastructure team, leading technical efforts in improving video quality and quality of experience across all video products at Meta. Before joining Meta, he spent 3.5 years at Netflix, contributing to the development and popularization of VMAF, Netflix's open-source video quality metric, as well as inventing the Dynamic Optimizer, a shot-based perceptual video quality optimization framework that brought significant bitrate savings across the whole video streaming spectrum. He was a professor for 8 years at the University of Thessaly's Electrical and Computer Engineering Department in Greece, teaching video compression, signal processing and information theory. He was one of the cofounders of Cidana, a mobile multimedia software company in Shanghai, China. He was the director of software for advanced video codecs at InterVideo, the makers of the popular SW DVD player, WinDVD, in the early 2000's and he has also worked for 4 years in high-energy experimental Physics in Italy. He is one of the co-chairs for the statistical analysis methods (SAM) and no-reference metrics (NORM) groups at the Video Quality Experts Group (VQEG). He is actively involved within the Alliance for Open Media (AOMedia) as co-chair of the software implementation working group (SWIG). He has over 150 publications, including 50 patents. His research interests lie in video coding, quality of experience, adaptive streaming, and energy efficient HW/SW multimedia processing.



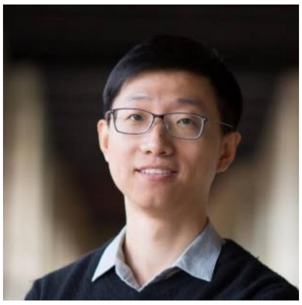
Dr. Ryan Lei is currently working as a video codec specialist and technical lead in the Video Infrastructure Media Algorithm team at Meta. His focus is on algorithms and architecture for cloud based video processing, transcoding, and delivery at large scale for various Meta products. Ryan Lei is also the co-chair of the Alliance for Open Media (AOM) testing subgroup and is actively contributing to the standardization of AV1 and AV2. Before joining Meta, Ryan worked at Intel as a principal engineer and codec architect. He worked on algorithm implementation and architecture definition for multiple generations of hardware based video codecs, such as AVC, VP9, HEVC and AV1. Before joining Intel, Ryan worked at ATI handhelp department, where he implemented embedded software for hardware encoder/decoder in mobile SoCs. Ryan received his Ph.D. in Computer Science from the University of Ottawa. His research interests include image/video processing, compression, adaptive streaming and parallel computing. He has (co-) authored over 50 publications, including 17 patents.

Dr. Debargha Mukherjee received his M.S./Ph.D. degrees in ECE from University of California Santa Barbara in 1999. Since 2010 he has been with Google LLC, where he is currently a Principal Engineer/Director leading next generation video codec research and development efforts. Prior to that he was with Hewlett Packard Laboratories, conducting research on video/image coding and processing. Debargha has made extensive research contributions in the area of image and video compression throughout his career, and was elected to IEEE Fellow for leadership in standard development for video-streaming industry. He has (co-)authored more than 120 papers on various signal processing topics, and holds more than 200 US patents, with many more pending. He currently serves as a Senior Area Editor of the IEEE Trans. on Image Processing, and as a member of the IEEE Visual Signal Processing and Communications Technical Committee (VSPC-TC).

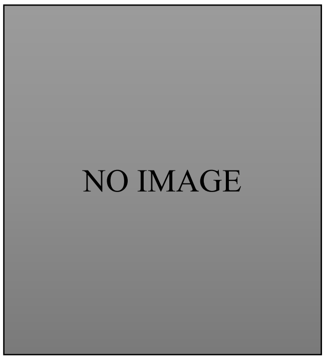
Dr. Balu Adsumilli is currently the Head of Media Algorithms group at YouTube/Google, leading transcoding infrastructure, audio/video quality, and media innovation at YouTube. Prior to this, he led the Advanced Technology group and the Camera Architecture group at GoPro, and before that, he was Sr. Staff Research Scientist at Citrix Online. He received his masters at the University of Wisconsin Madison, and his PhD at the University of California Santa Barbara. He has co-authored more than 120 papers and 100 granted patents with many more pending. He serves on the board of the Television Academy, on the board of NATAS Technical committee, on the board of Visual Effects Society, on the IEEE MMSP Technical Committee, and on ACM MHV Steering Committee. He is on TPCs and organizing committees for various conferences and workshops, and currently serves as Associate Editor for IEEE Transactions on Multimedia (T-MM). His fields of research include image/video processing, audio and video quality, video compression and transcoding, AR/VR, visual effects, video ML/AI models, and related areas.

Mr. Anush Moorthy currently leads the Video & Image Encoding team at Netflix, who's stellar engineers are responsible for the high-quality video content that one has come to expect of the service. He has been part of the Encoding Technologies team since 2016 where he contributes to video & image encoding and visual quality assessment at scale. He has previously worked at Qualcomm as a Senior Video Systems Engineer and at Texas Instruments Inc. as an

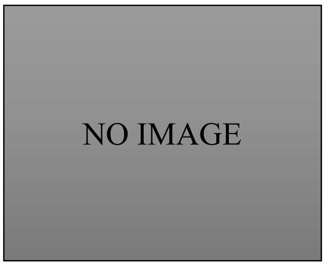
Advanced Imaging Engineer. His interests include image and video quality assessment, image and video compression, and computational vision.



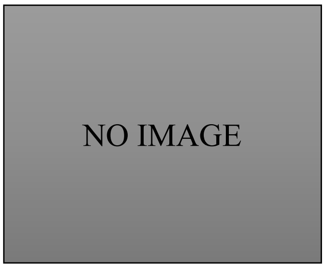
Dr. In Suk Chong holds a B.S. in Electrical Engineering from Seoul National University (1998) and earned his MS/Ph.D. in Electrical Engineering from the University of Southern California (USC) in 2004 and 2008, respectively. He worked at Qualcomm from 2008 to 2017 before joining Google as the Video Codec Lead, spearheading advancements in video compression technology.



Dr. Yilin Wang is a staff software engineer in the Media Algorithms team at YouTube/Google. He spent the last ten years on improving YouTube video processing and transcoding infrastructures, and building video quality metrics. Beside the video engineering work, he is also an active researcher in video quality related areas, and published papers in CVPR, ICCV, TIP, ICIP, etc. He received his PhD from the University of North Carolina at Chapel Hill in 2014, working on topics in computer vision and image processing.



Dr. Yeping Su, Google



Dr. Onur Guleryuz, Google

Mr. Joe Young graduated from the University of Washington with a B.S. degree in Computer Engineering and a M.S. in Electrical Engineering. He has extensive experience in the field of video compression, having developed both software and hardware-based encoders and transcoders at a variety of companies including Motorola Mobility and Google. Currently, Joe works at YouTube, where he is focused on video acceleration and coding efficiency, and contributing to the AV1 and AVM video standards.

Presentations:

Talk-01
Title: Introduction to the Alliance for Open Media
Speaker: Dr. Ioannis Katsavounidis (Meta)
Abstract: We will start by sharing the history of the Alliance Open Media (AOM), its current members, group structure and the standards it has already established (AV1 and AVIF).

Talk-02
Title: SIWG / SVT-AV1: 15-Month Milestones and Achievements
Speaker: Mr. Hassene Tmar (Meta)
Abstract: The SIWG group has been continuing to work on a product-level AV1 implementation since the release of SVT-AV1 1.0. This presentation will highlight the progress made in improving compression vs computational efficiency tradeoffs across various use cases, including VOD, Live, and RTC, as well as future plans for the group.

Talk-03
Title: AV1 Deployment at Meta
Speaker: Dr. Ryan Lei (Meta)
Abstract: AV1 was the first generation royalty-free coding standard developed by Alliance for Open Media, of which Meta is one of the founding members. Since AV1 was released in 2018, we have worked closely with the open source community to implement and optimize AV1 software decoder and encoder. Early in 2022, we believed AV1 was ready for delivery at scale for key VOD applications such as Facebook (FB) Reels and Instagram (IG) Reels. Since then, we have started delivering AV1 encoded FB/IG Reels videos to selected iPhone and Android devices. After roll out, we have observed great engagement win, playback quality improvement, and bitrate reduction with AV1. In this talk, we will share our journey on how we enabled AV1 end-to-end production and delivery. First, we will talk about AV1 production,

including encoding configuration and ABR algorithms. Second, since the main delivery challenge is on the decoder and client side, we will also talk about the learnings on integrating AV1 software decoder on both iOS and Android devices. Third, we will also talk about how we enabled Mixed Codec Manifest to expand AV1 delivery to low end Android phones. In the end, we will talk about how AV1 is leveraged within Meta for other use cases other than VOD.

Talk-04

Title: AV1 Deployment @ Netflix: Past, Present & Future

Speaker: Mr. Anush Moorthy (Netflix)

Abstract: TBA

Talk-05

Title: YouTube’s AV1 Activities

Speaker: Dr. In Suk Chong (Google)

Abstract: In this talk we are sharing how Google helps AV1 adoptions in the ecosystem. YouTube, Android, Chrome within Google really worked hard to expedite the adoptions of AV1 in the market. Furthermore, Google designed their own HW AV1 encoder/ decoder to exploit the gain that AV1 can provide in scale.

Talk-06

Title: UVQ in Video Compression

Speaker: Dr. Yilin Wang (YouTube/Google), Dr. Balu Adsumili (YouTube/Google)

Abstract: Universal Video Quality (UVQ) model is a deep learning based video quality metric proposed by Google, which has also been widely applied in production. In this talk, we will focus on the compression related development of UVQ, share insights from practical applications, and introduce improved UVQ models for compression.

Talk-07

Title: Towards a Next-Gen Video Codec

Speaker: Dr. Debargha Mukherjee (Google)

Abstract: TBA

Talk-08

Title: AOM Common Test Condition Design and Latest Result

Speaker: Dr. Yeping Su (Google), Dr. Ryan Lei (Meta)

Abstract: AV1 is an open, royalty-free video coding format designed by the Alliance for Open Media (AOMedia). Since it was finalized in 2018, AV1 has been supported by major content providers, such as YouTube, Meta and Netflix, and achieved great compression efficiency gain over previous generations of codecs. Since the middle of 2019, AOM member companies have started the research and exploration work for the next generation of the coding standard after AV1. The actual development work started from the beginning of 2021 in the codec working group, which is the main forum to discuss and review coding tool proposals from AOM member companies. Meanwhile, the testing sub-group also started the work to define the Common Test Conditions that are used to evaluate the compression efficiency gain and implementation complexity of the proposed coding tools.

In this talk, we will first present a high level overview of the Common Test Condition finalized by the testing sub group. We will focus on its design intent and present some details of a few unique test configurations that are close to production usage, but never supported in any previous coding standard development process. In the second part of the talk, We will present a high level summary of the latest compression efficiency result that has been achieved by the Alliance.

Talk-09

Title: Overview of Coding Tools Under Consideration in AVM

Speaker: Dr. Debargha Mukherjee, Mr. Xin Zhao, Dr. Onur Guleryuz, Mr. Joe Young (Google)

Abstract: TBA

DETAILS OF INDUSTRY SEMINARS

NEC Seminar (organized by NEC Corporation, Japan)

Seminar Title: Exploring NEC's Recognition AIs: Biometrics, Fairness, and Behavior Analysis

Seminar Abstract: Unveiling the latest advancements in AI recognition technology, our seminar will delve into real-time multi-object tracking for biometric solutions, fair face attribute recognition through Imbalance-Aware Adaptive Margin, annotation-free image recognition AIs in retail, and harnessing the power of retrieval for human behavior understanding in video analysis. Join us to explore groundbreaking research and the practical applications of these transformative technologies in shaping the digital society of the future.

Talk-01

Title: Toward Real-time End-to-End Multi-Object Tracking Model for Biometric Solution

Speaker: Dr. Hiroshi Fukui, NEC Corporation, Japan

Abstract: This session presents a new real-time multi-object tracking (MOT) model that is an important technology for computer vision, such as autonomous driving, video analysis, and biometrics systems. Recently, NEC has developed a technology called gateless access control system using biometric recognition, which is capable of authenticating 100 persons in one minute without any devices or gates by tracking everyone who passes through. The MOT process is a core component of this system, as it tracks each authenticator without missing. Although the deep-based MOT models have been proposed and resulted in improved accuracy, these models significantly slow down when including over 50 persons due to multi-step modules, such as deep- and rule-based methods. In contrast, our proposed MOT model, called TicrossNet, is a simple network design composed of a base detector and a cross-attention module only. TicrossNet, which runs all MOT processes by GPU in an end-to-end manner, achieves 31.0 FPS even if including as many as >100 persons per frame. As a result, the gateless system achieves a real-time authentication of a large number of pers.

Biosketch: Hiroshi Fukui is an assistant manager (senior researcher) of Biometrics Research Laboratories, NEC Corporation, Japan. He received B.E., M.E., and Ph.D. degrees from Chubu University in 2014, 2016, and 2019, respectively. His research interests include computer vision, machine learning, and video analysis. He is member of IEEE and IPSJ.

Talk-02

Title: Toward Fair Face Analysis Systems: Metric Learning Loss with Adaptive Margins for Fair Multi-label Face Attribute Recognition

Speaker: Dr. Masashi Usami, NEC Corporation, Japan

Abstract: NEC researchers are studying and developing face recognition systems using AI models and expanding their potential for social applications, such as gateless access control systems using biometric recognition. We are also investigating various areas of face image analysis, for example, healthcare research. With the development of AI model research and the growing potential of face image analysis, the fairness of AI systems is becoming increasingly important. NEC declares its corporate purpose to create social values of fairness to promote a more suitable world. It is imperative for us researchers to think deeply about the fairness of AI systems. This session mainly presents one of our studies on fair face attribute recognition. Datasets with imbalanced sample distributions can affect the fair discrimination ability of AI models. This imbalance problem becomes more complex in multi-label classification tasks due to the variety of imbalance levels. We proposed a novel training method to improve the fair classification ability in the multi-label classification tasks such as face attribute recognition.

Biosketch: Masashi Usami is a researcher of Biometrics Research Laboratories, NEC Corporation, Japan. Previously he specialized in the experimental particle physics, and received Ph.D. degree from Department of Physics, Graduate School of Science, The University of Tokyo. Currently, he is interested in the techniques and social applications of biometrics authentications, focusing on the fairness of face recognition and face attribute recognition.

Talk-03

Title: Towards annotation-free image recognition AIs

Speaker: Dr. Tomokazu Kaneko, NEC Corporation, Japan

Abstract: The manual annotation process is a critical bottleneck in implementing image recognition AIs. Training an object classification model requires an image dataset of each recognition target, and the annotation cost increases as the number of recognition target categories increases. Especially in retail stores, where hundreds of new products are coming in daily, updating the dataset requires continuous annotation costs. We propose an efficient product registration system for image recognition AIs in retail stores. The user only needs to shoot a video of the product being held and rotated in hand for 10-20 seconds. The proposed system focuses on the moving areas of the captured video to localize the target object's position and automatically crops the images to generate an image dataset. The proposed system approaches the problem of extending recognition targets not by methods such as few-shot learning but from the perspective of improving the efficiency of the registration process.

Biosketch: Tomokazu Kaneko, Ph.D. is an assistant manager (senior researcher) of Visual Intelligence Research Laboratories, NEC Corporation, Japan. His research covers object recognition, retail product detection, domain adaptation, and instant object registration system. He is also working on the research topic of object understanding based on the object-centric representation learning and world models.

Talk-04

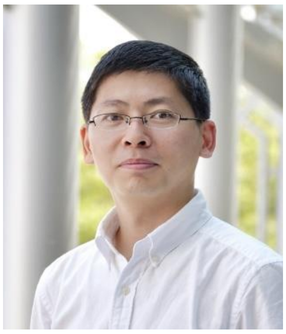
Title: Harnessing Domain Knowledge of Intra-Class Variations to Mitigate Label Scarcity Bias (in Satellite Imagery)

Speaker: Ms. Tsenjung Tai, NEC Corporation, Japan

Abstract: Domain adaptation utilizes labeled data from one domain (the 'source') to enhance the performance in another domain (the 'target') that is either scarcely labeled or unlabeled. However, when intra-class variability overshadows the distinctions between classes, severe misclassifications can arise due to the bias from label scarcity in the target domain. We introduce a feature conversion module that generates synthetic features from the few labeled target domain data by adapting inter-class knowledge and intra-class variations from the source domain. The synthetic features thus approximate a broader spectrum of the target domain's diversity. Our approach is assessed with satellite imagery classification tasks, where images from the same class can appear dramatically different depending on their capturing angles. The feature conversion module modifies labeled features as if they were extracted from images captured at various angles. Our classifier achieves enhanced accuracy training with only a few annotations in the target domain, particularly for images captured at angles that differ from the labeled training examples.

Biosketch: Tsenjung Tai earned her M.S. degree in Computer Science from the Hong Kong University of Science and Technology. At NEC's Visual Intelligence Research Laboratories, she specializes in domain adaptation for satellite imagery recognition and change detection, addressing challenges in limited data learning. In 2022, she was honored with the "Innovator under 35, Japan" award in recognition of her team's significant contribution to enabling rapid post-disaster responses.

Talk-05



Title: The Power of Retrieval for Video Analysis on Human Behavior Understanding

Speaker: Dr. Jianquan Liu, NEC Corporation, Japan

Abstract: In this talk, Dr. Liu will introduce an industrial level framework of utilizing the power of retrieval techniques for video analysis on human behavior sensing and understanding. This talk will mainly demonstrate a series of selected research achievements that contributed to both academia and industry in our framework. Our framework is composed of a series of cutting-edge technologies that sense the data generated in the real world, transform them into readable, visible, and modellable digital forms, and finally analyze these digital data to understand the human behavior. For example, such cutting-edge technologies include the human sensing by traditional cameras [MM'14, MM'16], 360 cameras [MM'19, WACV'20, ICIP'21], and microwave sensors [MM'20], the action recognition [MM'19, WACV'20, MM'20, ICIP'21], the object tracking [MIPR'19, BigMM'19], the human object interaction [CBMI'19], the scene recognition [MM'19], the behavioral pattern analysis [MM'16, ICMR'18 MIPR'19], the retrieval [MM'14, MM'16, MM'17, ICMR'18, CBMI'19, ICASSP'21] and the visualization [SIGGRAPH'16, ICMR'18], towards the fully understanding of human behavior. These works will be introduced in the way of a general overview with interactive technical demos and interesting insights, for human behavior sensing and understanding by adopting effective processing techniques and designing efficient algorithms. Finally, Dr. Liu will pick up and share some challenging issues and directions for the realization of digital society in the future.

Biosketch: Jianquan Liu is currently the Director/Head of Video Insights Discovery Research Group at the Visual Intelligence Research Laboratories of NEC Corporation, working on the topics of multimedia data processing. He is also an adjunct professor at Graduate School of Science and Engineering, Hosei University, Japan. Prior to NEC, he was a development engineer in Tencent Inc. from 2005 to 2006, and was a visiting researcher at the Chinese University of Hong Kong in 2010. His research interests include high-dimensional similarity search, multimedia databases, web data mining and information retrieval, cloud storage and computing, and social network analysis. He has published 70+ papers at major international/domestic conferences and journals, received 30+ international/domestic awards, and filed 70+ PCT patents. He also successfully transformed these technological contributions into commercial products in the industry. Currently, he is/was serving as the Industry Co-chair of IEEE ICIP 2023 and ACM MM 2023; the General Co-chair of IEEE MIPR 2021; the PC Co-chair of IEEE IRI 2022, ICME 2020, AIVR 2019, BigMM 2019, ISM 2018, ICSC 2018, ISM 2017, ICSC 2017, IRC 2017, and BigMM 2016; the Workshop Co-chair of IEEE AKIE 2018 and ICSC 2016; the Demo Co-chair of IEEE MIPR 2019 and MIPR 2018. He is a member of ACM, IEEE, IEICE, IPSJ, APSIPA and the Database Society of Japan (DBSJ), a member of expert committee for IEICE Mathematical Systems Science and its Applications (2017-), and IEICE Data Engineering (2015-2021), and an associate editor of IEEE TMM (2023-), ACM TOMM (2022-), EURASIP JIVP (2023-), IEEE MultiMedia Magazine (2019-2022), ITE Transaction on Media Technology and Applications (2021-), APSIPA Transactions on Signal and Information Processing (2022-), and the Journal of Information Processing (2017-2021). Dr. Liu received the M.E. and Ph.D. degrees from the University of Tsukuba, Japan.

NTT Seminar (organized by NTT Corporation, Japan)

Seminar Title: NTT's Media Processing AI and Its Industrial Applications

Seminar Abstract: This industry seminar introduces some of NTT's artificial intelligence technologies for image processing. Specifically, we will describe a multi-modal processing AI with all-in-one architecture that emulates human-like capabilities. We will also describe high-speed and efficient computing by the novel event-driven inference paradigm. Furthermore, we plan to explain other image media processing techniques in NTT R&D, like point cloud processing. In each presentation, we would like to introduce the motivation behind these technologies and their industrial applications. We are excited to share the details of the presentations soon. Stay tuned for more updates!

Talk-01

Title: MediaGnosis: the next-generation media processing artificial intelligence

Speaker: Dr. Ryo Masumura, NTT Corporation, Japan

Abstract: MediaGnosis provides the all-in-one cross-media processing module for visual, audio, and text media. One of the most notable distinctions of MediaGnosis has a human-like cross-media processing architecture. We will describe the motivation behind this architecture, some of the many novel algorithms for media processing in MediaGnosis, and the integrated method of each media processing. We will also describe our unique advertisement promotion efforts for making the research achievements and development widely known.

Biosketch: Ryo Masumura received B.E., M.E., and Ph.D. degrees in engineering from Tohoku University, Sendai, Japan, in 2009, 2011, 2016, respectively. Since joining Nippon Telegraph and Telephone Corporation (NTT) in 2011, he has been engaged in research on speech recognition, spoken language processing, and natural language processing. He received the Student Award and the Awaya Kiyoshi Science Promotion Award from the Acoustic Society of Japan (ASJ) in 2011 and 2013, respectively, the Sendai Section Student Awards The Best Paper Prize from the Institute of Electrical and Electronics Engineers (IEEE) in 2011, the Yamashita SIG Research Award and the SIG-NL Excellent paper award from the Information Processing Society of Japan (IPSJ) in 2014 and 2018, the Young Researcher Award and the Paper Award from the Association for Natural Language Processing (NLP) in 2015 and 2020, the ISS Young Researcher's Award in Speech Field and the ISS Excellent Paper Award from the Institute of Electronic, Information and Communication Engineers (IEICE) in 2015 and 2018. He is a member of the ASJ, the IPSJ, the NLP, the IEEE, and the International Speech Communication Association (ISCA).

Talk-02

Title: geoNebula: Elemental Technologies for Supporting the Integration of Real Space and Cyberspace

Speaker: Dr. Satoshi Suzuki, NTT Corporation, Japan

Abstract: To create a human-centered society in which everyone can lead a comfortable, vibrant, and high-quality life, we study a system that integrates real space and cyberspace. In this talk, point cloud processing technology, called geoNebula, is introduced. geoNebula analyzes data measured in real space, recognizes space and objects, and compresses the data into a compact form suitable for constructing cyberspace. We will describe our novel algorithms for point cloud processing, and the practical applications for constructing cyberspace that can precisely reproduce real space using geoNebula.

Biosketch: Satoshi Suzuki received B.E., M.E., and Ph.D. degrees from the University of Electro-Communications in 2015, 2017, and 2022, respectively. He joined Nippon Telegraph and Telephone (NTT) in 2017. He is currently a researcher at NTT Computer and Data Science Laboratories. His current research interests include neural networks, computer vision, surveillance systems, and machine learning. He received the IEEE CIS Japan Chapter Young Researcher Award in 2015. He is a member of the Information Processing Society of Japan (IPSJ).

Talk-03

Title: Conversational system that talks about the scenery seen from vehicles

Speaker: Dr. Hiroaki Sugiyama, NTT Corporation, Japan

Abstract: We are working on developing a passenger agent that can talk with people about the scenery seen from a moving vehicle. We believe that such casual dialogue with passengers enriches the driving. Such passenger agents should continuously understand input scenery images and talk about them. Recent advances in chatting dialogue systems based on huge-scale transformers promise to realize natural dialogue; however, most focus on text dialogues. While vision-based dialogue systems that aim to answer questions about the content in the given images are proposed, few studies tackle realizing casual dialogue systems that talk about the scenery. In this talk, we introduce our dialogue system that uses the changing scenery seen from a vehicle as a topic of conversation.

Biosketch: Hiroaki Sugiyama is a Senior Researcher, Interaction Research Group, Innovative Communication Laboratory, NTT Communication Science Laboratories. He received a B.E. and M.E. in information science and technology from the University of Tokyo in 2007 and 2009 and Ph.D. in engineering from Nara Institute of Science and Technology in 2016. He joined Nippon Telegraph and Telephone Corporation (NTT) in 2009. He has been engaged in research on chatting dialogue system for natural human interaction. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), Information Processing Society of Japan (IPSJ), Japanese Society for Artificial Intelligence (JSAI), and Association for Natural Language Processing.

Talk-04

Title: Distributed AI Video Analytics

Speaker: Ms. Monikka Roslianna Busto, NTT Corporation, Japan

Abstract: Deepack is a real-time video analytics framework developed for NTT groups' surveillance applications which allows analysis on multiple cameras at a lower cost by sharing GPU resources amongst a network of cameras. In a use case like smart city surveillance, thousands of cameras request large amounts of workloads that need to be filtered to only meaningful events to optimize resource consumption. In addition, widespread surveillance in public areas is a major concern for security. In this session, we discuss how distributed computing between the edge and the cloud make real-time analytics more feasible in terms of cost, efficiency, and privacy in Deepack's use cases. We also discuss how techniques called model cascade and model splitting offer dynamic decision making for distributing the workload to lower the computational cost while also preserving the privacy of the data being exchanged between the edge and cloud.

Biosketch: Monikka Roslianna Busto is a Researcher at NTT Software Innovation Center. She graduated from the Electrical and Electronics Engineering Institute, College of Engineering, the University of the Philippines in 2017, and received a master's degree from the Department of Information and Communications Engineering, Tokyo Institute of Technology in 2021. She joined Nippon Telegraph and Telephone Corporation in the same year. Her research interests include computer vision, collaborative intelligence for edge computing, remote sensing image analysis and multi-modal AI.

DETAILS OF INDUSTRY DEMONSTRATIONS

Demo #1: JPEG XS Low latency error robust low latency video transport

Presenter: Thomas Richter, Siegfried Fößel (Fraunhofer IIS)

Abstract: In this demo, we will show real-time video capture and recording with JPEG XS using forward error protection following SMPTE 2022-5. JPEG XS is a low-latency, low complexity video coding codec whose first and second edition have been standardized by ISO as ISO/IEC 21122, and whose third edition adding time differential coding is currently under standardization. In this demo, a mini-PC will compress a life signal from a camera in real time, and transport the JPEG XS codestream following IETF RFC 9134. While forward error correction schemes are currently not part of any JPEG XS related standard, an error protection mechanism following SMPTE 2022-5 will protect the stream in this demo from packet loss and burst errors. An error simulator, triggered by visitors of the demo, can be used to stress it such that visitors can observe the impact of errors with and without the additional error correction streams present. The decoder, also implemented in software on a mini-PC, will decode the signal and render it on a monitor in real time. If requested and considered useful by the program committee, a talk on the workings of JPEG XS, its standardization progress and SMPTE 2022-5 can be prepared.

Demo #6: Media over QUIC: Initial Testing, Findings and Results

Presenter: Zafer Gurel, Ozyegin University (Turkiye), Tugce Erkilic Civelek, Ozyegin University (Turkiye), Ali C. Begen, Ozyegin University (Turkiye) and Comcast (USA), Alex Giladi, Comcast (USA)

Abstract: Over-the-top (OTT) live sports has led sports broadcasting to a new level, where fans can stream their favorite games on connected devices. However, there are still challenges that need to be tackled. Nobody wants to hear a neighbor's cheers when a goal is scored before seeing it on the screen, making low-latency transport and playback indispensable. Synchronization among all the viewing devices and social media feeds is also essential. The existing HTTP ecosystem comprises solid foundational components such as distributed caches, efficient client applications and high-performant server software glued with HTTP. This formation allows efficient live media delivery at scale. However, the two popular approaches, DASH and HLS, are highly tuned for HTTP/1.1 and 2 running on top of TCP. The downside is the latency caused by the head-of-line (HoL) blocking experienced due to TCP's in-order and reliable delivery. The latest version of HTTP (HTTP/3) uses QUIC underneath instead of TCP. QUIC can carry different media types or parts in different streams. These streams can be multiplexed over a single connection avoiding the HoL blocking. The streams can also be prioritized (or even discarded) based on specific media properties (e.g., dependency structure and presentation timestamp) to trade off reliability with latency. DASH and HLS can readily run over HTTP/3, but they could only recoup the benefits if they could use its unique features. The IETF recently formed a new working group to develop a QUIC-based low-latency delivery solution for media ingest and distribution in browser and non-browser environments. Targeted use cases include live streaming, cloud gaming, remote desktop, videoconferencing and eSports. The work is still in its infancy, but we believe media over QUIC running in an HTTP/3 or WebTransport environment could potentially be a game-changer. This demo, among the first, presents the architectural design issues and preliminary results from the early prototypes.

Demo #7: Camera-based Gaze Tracker Driven Robotic and Assisted Living/Hospital Bed Use-cases

Presenter: Mithun B S,Tince Varghese, Aditya Choudhary, Rahul Dasharath Gavas, Ramesh Kumar Ramakrishnan, Arpan Pal (TCS Research, India)

Abstract: In recent years, hands free controlling of robots, hospital beds, senior assistance platforms using various modalities like handheld devices, physiological sensors and so on are gaining wide acceptance. Gaze tracking serves to be a better alternative owing to its non-obtrusiveness and freeing up of hands. Health care applications and human robot use cases can benefit largely using gaze. To meet this goal, we first developed an RGB camera-based gaze tracker that can be easily deployed in standard computer for daily usage. We deploy this gaze tracker to run two use-cases, viz., controlling robot’s movements and assisted living/hospital bed use cases. In the case of robotic use case, our primary goal is to control a robot’s movements using gaze tracking over handheld devices, to facilitate the usage of hands for other important tasks. In this regard, we provide a display screen to the users which shows the field of view of the robot. The user is required to gaze at the locations where he/she wants the robot to move and thus the locomotion of the robot is controlled accordingly. The system is tested exclusively in our lab, and it is found that the users can easily and effectively control the robot’s navigation in any direction using the proposed system. This can be used in hospital scenario, wherein the doctor, if physically not present near patient, can attend using the robot and assess patient’s health and have interactions with them, while his/her hands are free to make notes or to ascertain any other important actions and like in the case of telerobotic, where operator’s hands are free to manipulate the robot arms while moving the robot with the eye tracker. In the second use case, we intend to provide an assistive living platform for bed ridden patients or elderly people using an interface showing options for ordering food, snacks, calling doctor or nurse or even controlling the bed for adjusting the height or the angle of the head or the leg portions of the bed. We conducted an in-house pilot study to assess the ease and effectiveness of controlling this system over conventionally used input devices like mouse. It was seen that the time taken and the effectiveness of usage via gaze tracking as input modality was in-line with that of conventional input devices.

Demo #8: VVC in a large-scale streaming environment

Presenter: Kevin Rocard (Bitmovin), Jacob Arends (Bitmovin), Adam Wieckowski (Fraunhofer HHI), Benjamin Bross (Fraunhofer HHI)

Abstract: Versatile Video Coding (VVC) is the latest video coding standard released in mid-2020 as the successor to the High Efficiency Video Coding (HEVC) standard. VVC has been developed to provide up to 50% bit-rate savings compared to HEVC for the same perceived video quality. Unlike previous video coding standards, VVC already includes in its first version specific coding tools and systems functionalities for a wide range of applications. To video streaming, VVC offers several benefits, such as efficient streaming of demanding content, including UHD (8k+) and 360 video, lower distribution costs due to the lower bitrate and maintaining high visual quality even on slower networks. This demo shows how VVC performs in an actual end-to-end video streaming environment at low bitrates. In the past Bitmovin and Fraunhofer HHI already demonstrated VVC cloud encoding and playback in a browser. The videos have been encoded with the Bitmovin cloud transcoding solution using VVenC, an open VVC encoder from Fraunhofer HHI. This has been further evolved using a so-called smart chunking approach. After encoding, the assets are made available using Dynamic Adaptive Streaming over HTTP (DASH). On the player side, the open VVC decoder VVdeC has been integrated into an internal version of the Bitmovin Android demo app to evaluate VVC playback on a range of streams and devices. To test the VVC low bitrate performance, the same HD assets have been encoded in both VVC and HEVC at the same low bitrates below 500 kb/s. Upon playback on a Samsung Galaxy S8 Android tablet, the performance and image quality of the streams can be compared.

VVenC Software Repository on Github:
https://github.com/fraunhoferhhi/vvenc
VVdeC Software Repository on Github:
https://github.com/fraunhoferhhi/vvdec
Bitmovin’s Smart Chunking:
https://bitmovin.com/smart-chunking-encoding/
VVC – Its Benefits, Supported Devices, and How Bitmovin is Implementing it:
https://bitmovin.com/vvc-benefit-supported-devices-bitmovin-implementation/

Demo #9: A Real-time Chinese Food Auto Billing System based on Synthetic images

Presenter: Qiushi Guo, Yifan Chen, Jin Ma, Tengteng Zhang (China Merchants Bank)

Abstract: In recent years, the topic of food segmentation has gained significant attention in both academic and industrial circles. Various solutions have been proposed for the segmentation of Western food, demonstrating promising performance that aligns with the requirements of applications such as diet management and calorie estimation. Motivated by these accomplishments, we have undertaken the design of an automatic billing system for Chinese food prices based on instance segmentation methods. However, the segmentation of Chinese food poses a formidable challenge due to the extensive range of ingredients and cooking styles involved. It is infeasible to amass a sufficiently large image dataset that encompasses all potential variations of Chinese cuisine for training a segmentation model. To address this challenge, rather than attempting to detect individual dishes, we have reformulated the task by focusing on segmenting a curated selection of plates containing Chinese food. In this regard, we introduce a FoodSyn module, which employs image synthesis techniques by extracting food portions from the UECFoodPIX dataset and seamlessly integrating them into plate images. The resulting synthesized images are then utilized for training an encoder-decoder network to perform instance segmentation. Extensive experimentation has demonstrated the efficacy of our proposed approach in practical scenarios, achieving a mean Intersection over Union (mIoU) exceeding 95%, and the accuracy of final price estimation is over 99%. rate surpassing 20 frames per second (fps). We intend to release the source code once the paper detailing our research is accepted for publication.

Demo #10: Instant Object Registration System for Image Recognition of Retail Products

Presenter: Tomokazu Kaneko, Soma Shiraishi, Makoto Terao (NEC Corporation)

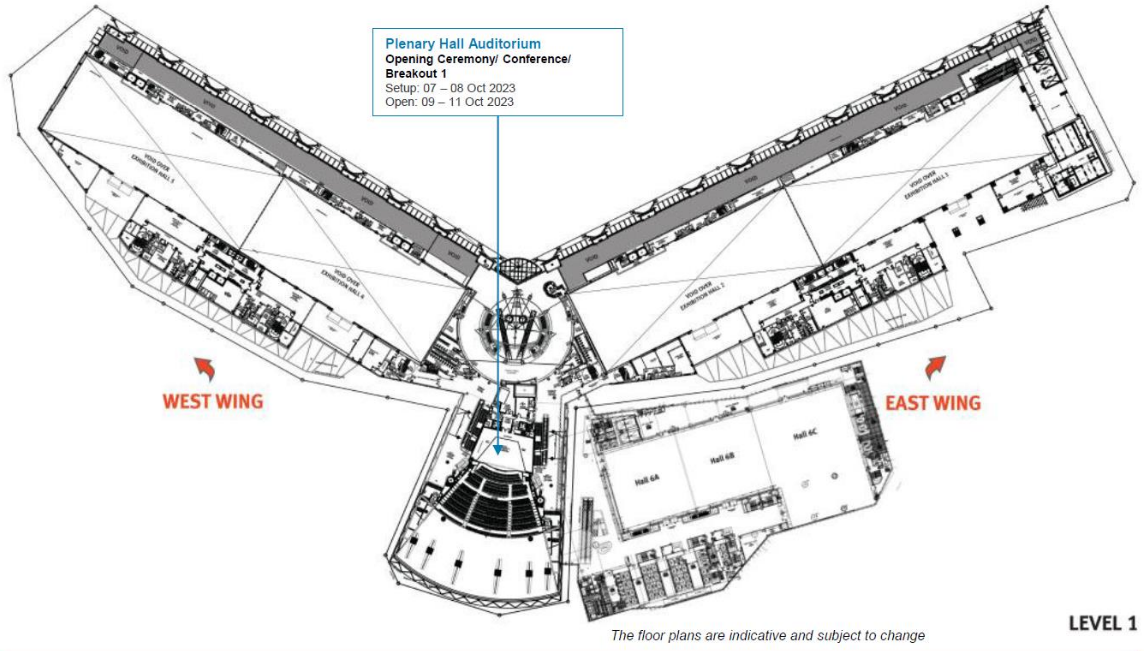
Abstract: We propose an efficient product registration system for image recognition AIs in retail stores. The system automatically generates a product image dataset for training AIs. The user only needs to shoot a video of the product being held and rotated in hand for 10-20 seconds. The proposed system approaches the problem of extending recognition products not by methods such as few-shot learning but from the perspective of improving the efficiency of the registration process. An image dataset of each product is required to train a classification model of retail products. However, hundreds of new products are introduced daily in the retail domain; therefore, an efficient dataset generation system is required to keep the dataset up to date. The conventional process of creating an image dataset requires tedious tasks, including taking pictures, manual annotation, and cleansing image sets. At first of the process, we take images of the products from various angles using a turntable. This process requires special equipment, such as a turntable, and the shooting environment must be set up so that other products do not appear in the background. Next, we manually annotate the location of the products in the captured images and cut out the product images. At the same time, we remove inappropriate images for training due to factors such as blurring or out-of-focus. These manual annotation processes are expensive and take about 30 minutes per product. The proposed system generates an image dataset from a video of the product moved by hand. The system estimates the product position in the video and cuts out images. At this localization step, the system focuses on the moving area in the video, taking advantage of the condition that the product is moved by hand. This method allows the system to specify only the product area to be registered, even in unknown background environments where other objects may appear. Furthermore, the system is equipped with a function to detect blur and occlusion of the product caused by hand movements. This function allows the automatic removal of inappropriate images for the dataset. The proposed system requires only an RGB camera and a PC or a mobile device; therefore, neither special equipment nor a particular environment is needed. With the proposed system, the registration process, which used to take 50 hours for 100 products, can be reduced to 30 minutes, allowing on-site registration. In the demonstration, we will present two versions of the proposed system, one is for PC, and the other is for smartphones. The PC version consists of a laptop PC and a camera to capture the video. The smartphone version requires a smartphone only. We also prepare some products and a small shelf displaying the products to show the registration process. Audiences will experience the process of creating a dataset by picking up the products and taking videos of them using the system.

FLOOR MAPS OF CONFERENCE ROOMS

IEEE International Conference on Image Processing (ICIP) 2023

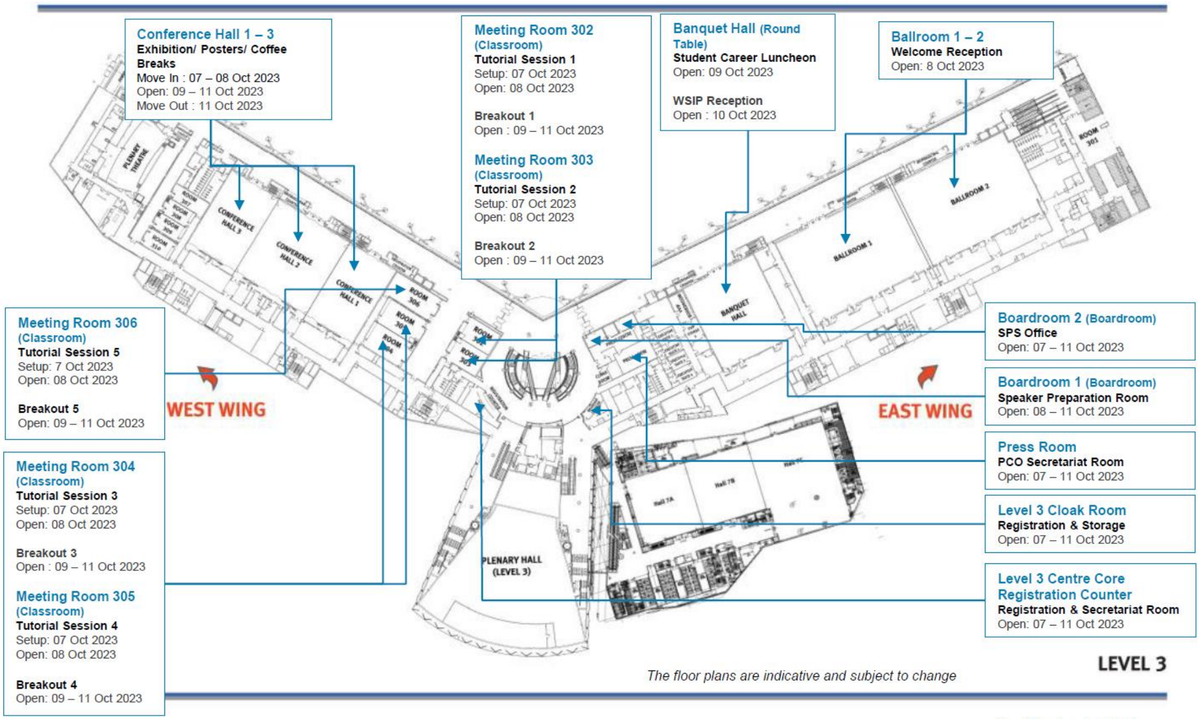
Saturday, 07 October – Wednesday, 11 October 2023

KUALA LUMPUR CONVENTION CENTRE



Plenary Hall Auditorium
Opening Ceremony/ Conference/ Breakout 1
Setup: 07 – 08 Oct 2023
Open: 09 – 11 Oct 2023

The floor plans are indicative and subject to change



Conference Hall 1 – 3
Exhibition/ Posters/ Coffee Breaks
Move In : 07 – 08 Oct 2023
Open: 09 – 11 Oct 2023
Move Out : 11 Oct 2023

Meeting Room 302 (Classroom)
Tutorial Session 1
Setup: 07 Oct 2023
Open: 08 Oct 2023

Breakout 1
Open : 09 – 11 Oct 2023

Meeting Room 303 (Classroom)
Tutorial Session 2
Setup: 07 Oct 2023
Open: 08 Oct 2023

Breakout 2
Open : 09 – 11 Oct 2023

Banquet Hall (Round Table)
Student Career Luncheon
Open: 09 Oct 2023

WSIP Reception
Open : 10 Oct 2023

Ballroom 1 – 2
Welcome Reception
Open: 8 Oct 2023

Meeting Room 306 (Classroom)
Tutorial Session 5
Setup: 7 Oct 2023
Open: 08 Oct 2023

Breakout 5
Open: 09 – 11 Oct 2023

Meeting Room 304 (Classroom)
Tutorial Session 3
Setup: 07 Oct 2023
Open: 08 Oct 2023

Breakout 3
Open : 09 – 11 Oct 2023

Meeting Room 305 (Classroom)
Tutorial Session 4
Setup: 07 Oct 2023
Open: 08 Oct 2023

Breakout 4
Open: 09 – 11 Oct 2023

Boardroom 2 (Boardroom)
SPS Office
Open: 07 – 11 Oct 2023

Boardroom 1 (Boardroom)
Speaker Preparation Room
Open: 08 – 11 Oct 2023

Press Room
PCO Secretariat Room
Open: 07 – 11 Oct 2023

Level 3 Cloak Room
Registration & Storage
Open: 07 – 11 Oct 2023

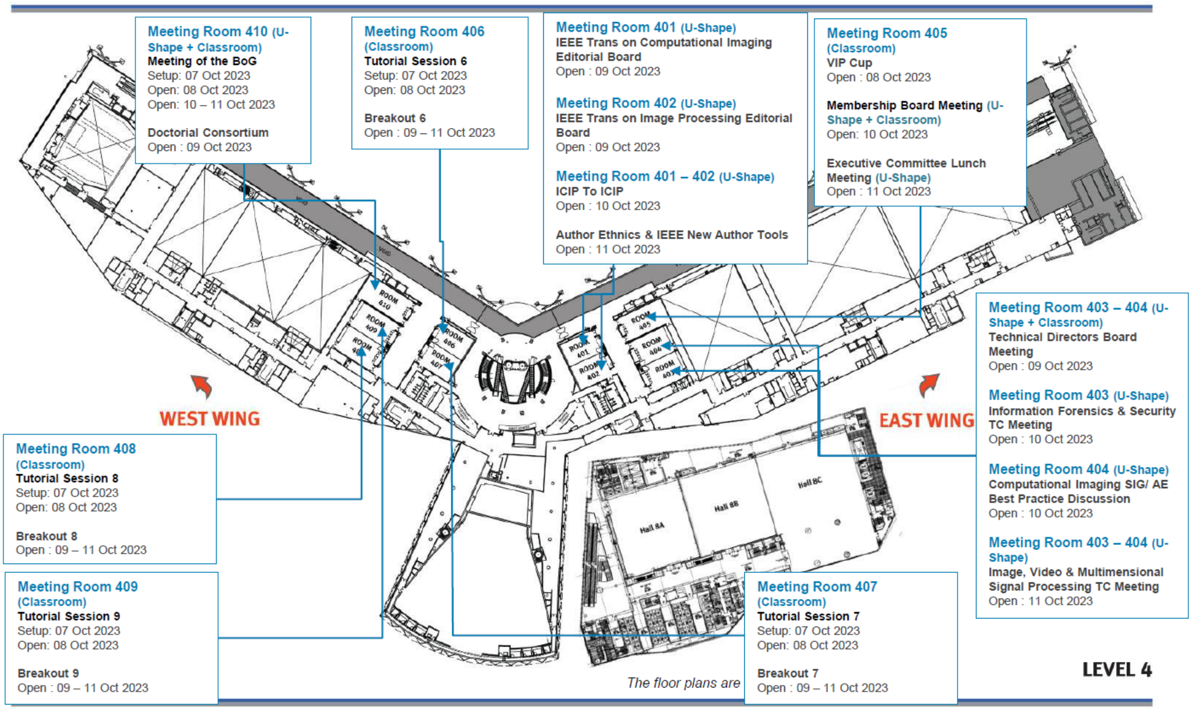
Level 3 Centre Core Registration Counter
Registration & Secretariat Room
Open: 07 – 11 Oct 2023

The floor plans are indicative and subject to change

IEEE International Conference on Image Processing (ICIP) 2023

Saturday, 07 October – Wednesday, 11 October 2023

KUALA LUMPUR CONVENTION CENTRE



Meeting Room 410 (U-Shape + Classroom)
Meeting of the BoG
Setup: 07 Oct 2023
Open: 08 Oct 2023
Open: 10 – 11 Oct 2023

Doctorial Consortium
Open : 09 Oct 2023

Meeting Room 406 (Classroom)
Tutorial Session 6
Setup: 07 Oct 2023
Open: 08 Oct 2023

Breakout 6
Open : 09 – 11 Oct 2023

Meeting Room 401 (U-Shape)
IEEE Trans on Computational Imaging Editorial Board
Open : 09 Oct 2023

Meeting Room 402 (U-Shape)
IEEE Trans on Image Processing Editorial Board
Open : 09 Oct 2023

Meeting Room 401 – 402 (U-Shape)
ICIP To ICIP
Open : 10 Oct 2023

Author Ethnics & IEEE New Author Tools
Open : 11 Oct 2023

Meeting Room 405 (Classroom)
VIP Cup
Open : 08 Oct 2023

Membership Board Meeting (U-Shape + Classroom)
Open: 10 Oct 2023

Executive Committee Lunch Meeting (U-Shape)
Open : 11 Oct 2023

Meeting Room 403 – 404 (U-Shape + Classroom)
Technical Directors Board Meeting
Open : 09 Oct 2023

Meeting Room 403 (U-Shape)
Information Forensics & Security TC Meeting
Open : 10 Oct 2023

Meeting Room 404 (U-Shape)
Computational Imaging SIG/ AE Best Practice Discussion
Open : 10 Oct 2023

Meeting Room 403 – 404 (U-Shape)
Image, Video & Multimensional Signal Processing TC Meeting
Open : 11 Oct 2023

Meeting Room 408 (Classroom)
Tutorial Session 8
Setup: 07 Oct 2023
Open: 08 Oct 2023

Breakout 8
Open : 09 – 11 Oct 2023

Meeting Room 409 (Classroom)
Tutorial Session 9
Setup: 07 Oct 2023
Open: 08 Oct 2023

Breakout 9
Open : 09 – 11 Oct 2023

Meeting Room 407 (Classroom)
Tutorial Session 7
Setup: 07 Oct 2023
Open: 08 Oct 2023

Breakout 7
Open : 09 – 11 Oct 2023

The floor plans are