

# **2023 IEEE International Symposium on Multimedia (ISM 2023)**

**Laguna Hills, California, USA  
11-13 December 2023**



**IEEE Catalog Number: CFP23197-POD  
ISBN: 979-8-3503-9577-8**

**Copyright © 2023 by the Institute of Electrical and Electronics Engineers, Inc.  
All Rights Reserved**

*Copyright and Reprint Permissions:* Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

***\*\*\* This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP23197-POD
ISBN (Print-On-Demand):	979-8-3503-9577-8
ISBN (Online):	979-8-3503-9576-1

**Additional Copies of This Publication Are Available From:**

Curran Associates, Inc  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: (845) 758-0400  
Fax: (845) 758-2633  
E-mail: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

CURRAN ASSOCIATES INC.  
**proceedings**  
.com

# 2023 IEEE International Symposium on Multimedia (ISM) **ISM 2023**

## Table of Contents

Message from the General Co-Chairs .....	xiii
Message from the Program Co-Chairs .....	xiv

### Best Paper Session (I)

Bitstream Organization for Parallel Entropy Coding on Neural Network-Based Video Codecs .....	1
<i>Amir Said (Qualcomm AI Research, USA), Hoang Le (Qualcomm AI Research, USA), and Farzad Farhadzadeh (Qualcomm AI Research, USA)</i>	
NN-VVC: Versatile Video Coding Boosted by Self-Supervisedly Learned Image Coding for Machines .....	10
<i>Jukka I. Ahonen (Nokia Technologies; Tampere University), Nam Le (Nokia Technologies; Tampere University), Honglei Zhang (Nokia Technologies), Antti Hallapuro (Nokia Technologies), Francesco Cricri (Nokia Technologies), Hamed Rezazadegan Tavakoli (Nokia Technologies), Miska M. Hannuksela (Nokia Technologies), and Esa Rahtu (Tampere University)</i>	
SmartCrop: AI-Based Cropping of Soccer Videos .....	20
<i>Sayed Mohammad Majidi Dorcheh (Oslo Metropolitan University, Norway; Forzasys, Norway), Mehdi Houshmand Sarkhoosh (Oslo Metropolitan University, Norway; Forzasys, Norway), Cise Midoglu (SimulaMet, Norway; Forzasys, Norway), Saeed S. Sabet (Forzasys, Norway), Tomas Kupka (Forzasys, Norway), Michael A. Riegler (SimulaMet, Norway), Dag Johansen (UIT The Arctic University of Norway, Norway), and Pål Halvorsen (SimulaMet, Norway; Oslo Metropolitan University, Norway; Forzasys, Norway)</i>	
360TripleView: 360-Degree Video View Management System Driven by Convergence Value of Viewing Preferences .....	28
<i>Qian Zhou (University of Illinois Urbana-Champaign), Mingyuan Wu (University of Illinois Urbana-Champaign), Yinjie Zhang (University of Illinois Urbana-Champaign), Michael Zink (University of Massachusetts Amherst), Ramesh Sitaraman (University of Massachusetts Amherst), and Klara Nahrstedt (University of Illinois Urbana-Champaign)</i>	

## Video Streaming

SAVG360: Saliency-Aware Viewport-Guidance-Enabled 360-Video Streaming System .....	36
<i>Yinjie Zhang (University of Illinois Urbana-Champaign, Urbana), Mingyuan Wu (University of Illinois Urbana-Champaign, Urbana), Beitong Tian (University of Illinois Urbana-Champaign, Urbana), Jiayi Li (University of Illinois Urbana-Champaign, Urbana), Bo Chen (University of Illinois Urbana-Champaign, Urbana), Qian Zhou (University of Illinois Urbana-Champaign, Urbana), and Klara Nahrstedt (University of Illinois Urbana-Champaign, Urbana)</i>	
Temporal Layer Injection for Fast Bitrate Ladder Creation in Video Live Streaming .....	44
<i>Hannes Mareen (Ghent University – imec, Belgium), Casper Haems (IDLab, Ghent University – imec), Tim Wauters (Ghent University – imec, Belgium), Filip De Turck (Ghent University – imec, Belgium), Peter Lambert (Ghent University – imec, Belgium), and Glenn Van Wallendael (Ghent University – imec, Belgium)</i>	
Real-Time Delivery of Visual Volumetric Video-Based Coding Data .....	52
<i>Lauri Ilola (Nokia Technologies, Finland), Sudarshan Bisht (Nokia Technologies, Finland), Ugurcan Budak (Nokia Technologies, Finland), Peter Fasogbon (Nokia Technologies, Finland), Jaakko Keränen (Nokia Technologies, Finland), and Lukasz Kondrad (Nokia Technologies, Finland)</i>	
CALC-VFS: Content-Adaptive Low-Complexity Video Frame Synthesis .....	57
<i>Nicola Giuliani (Technical University of Munich, Germany; Huawei Munich Research Center, Germany), Hongjie You (Technical University of Munich, Germany; Huawei Munich Research Center, Germany), A. Burakhan Koyuncu (Technical University of Munich, Germany; Huawei Munich Research Center, Germany), Atanas Boev (Huawei Munich Research Center, Germany), Elena Alshina (Huawei Munich Research Center, Germany), and Eckehard Steinbach (Technical University of Munich, Germany)</i>	

## Best Paper Session (II)

Multi-Scale Image Graph Representation: A Novel GNN Approach for Image Classification through Scale Importance Estimation .....	62
<i>João Pedro Oliveira Batisteli (Pontifícia Universidade Católica de Minas Gerais, Brazil), Silvio Jamil F. Guimarães (Pontifícia Universidade Católica de Minas Gerais, Brazil), and Zenilton K. G. Patrocínio (Pontifícia Universidade Católica de Minas Gerais, Brazil)</i>	
EHT-SR: An Entropy-Based Hybrid Approach for Faster Super-Resolution .....	69
<i>Abhilash Dharmavarapu (Cornell University, USA), Stefano Petrangeli (Adobe Research, USA), Jiashen Cao (Georgia Institute of Technology, USA), and Hyesoon Kim (Georgia Institute of Technology, USA)</i>	
Active Learning for Multi-Class Vehicle Categorization and Traffic Analysis in Complex Environments .....	79
<i>Gabriel Lugo (University of Alberta), Joey Quinlan (University of Alberta), Lingrui Zhou (University of Alberta), Md Nahid Sadik (University of Alberta), and Irene Cheng (University of Alberta)</i>	

Active Context Modeling for Efficient Image and Burst Compression .....	89
<i>Yang Li (University of North Carolina at Chapel Hill, USA), Gang Wu (Adobe Research, USA), Stefano Petrangeli (Adobe Research, USA), Haoliang Wang (Adobe Research, USA), Ryan Rossi (Adobe Research, USA), and Viswanathan Swaminathan (Adobe Research, USA)</i>	

## Object Detection

Angle Range and Identity Similarity Enhanced Gaze and Head Redirection Based on Synthetic Data .....	95
<i>Jiawei Qin (The University of Tokyo, Japan) and Xueting Wang (CyberAgent, Japan)</i>	
Illuminating the Bias in Pedestrian Detection .....	103
<i>Afnan Althoupey (Portland State University, USA), Li-Yun Wang (Portland State University, USA), Wu-chi Feng (Portland State University, USA), and Banafsheh Rekabdar (Portland State University, USA)</i>	
Validating Quantification Method for Object Visual Appeal to Motorists in Simulated Hazardous Driving Scene .....	108
<i>Takumi Inagawa (Meiji University, Japan), Tomoya Hayashi (Meiji University, Japan), and Yohei Nakada (Meiji University, Japan)</i>	
DeformableFormer for Classifying Endoscopic Ultrasound-Guided Fine-Needle Biopsy in Pancreatic Diseases .....	113
<i>Taiji Kurami (Meijo University), Takuya Ishikawa (Nagoya University), and Kazuhiro Hotta (Meijo University)</i>	

## Video Analysis

Video as Text: A New Paradigm for Flexible Video Analysis .....	115
<i>Li-Yun Wang (Portland State University, Oregon) and Wu-chi Feng (Portland State University, Oregon)</i>	
The Effect of Viewing Distances on 4K and 8K HDR Video Quality Perception .....	123
<i>Dominik Keller (Technische Universität Ilmenau, Germany), Rakesh Rao Ramachandra Rao (Technische Universität Ilmenau, Germany), Steve Göring (Technische Universität Ilmenau, Germany), and Alexander Raake (Technische Universität Ilmenau, Germany)</i>	
Towards Evaluation of Immersion, Visual Comfort and Exploration Behaviour for Non-Stereoscopic and Stereoscopic 360° Videos .....	131
<i>Stephan Fremerey (Technische Universität Ilmenau, Germany), Raja Faseeh Uz Zaman (Technische Universität Ilmenau, Germany), Touseef Ashraf (Technische Universität Ilmenau, Germany), Rakesh Rao Ramachandra Rao (Technische Universität Ilmenau, Germany), Steve Göring (Technische Universität Ilmenau, Germany), and Alexander Raake (Technische Universität Ilmenau, Germany)</i>	
Aerial 360-Degree Video Delivery for Immersive First Person View UAV Navigation .....	139
<i>Simran Singh (New Jersey Institute of Technology, USA) and Jacob Chakareski (New Jersey Institute of Technology, USA)</i>	

Progressive Coding for Neural Field Transmission .....	147
<i>Anustup Choudhury (Dolby Laboratories Inc., USA) and Guan-Ming Su (Dolby Laboratories Inc., USA)</i>	

## Image Analysis and Processing

A Framework for Multi-Plane Image Layer Merging .....	152
<i>Zachary Lazri (Dolby Laboratories, Inc., Sunnyvale, USA; University of Maryland, USA), Guan-Ming Su (Dolby Laboratories, Inc., Sunnyvale, USA), and Peng Yin (Dolby Laboratories, Inc., Sunnyvale, USA)</i>	
Mathematics as Interactive Multimedia .....	160
<i>Joshua Howell (Texas A&amp;M University, Texas), Angela Chan (Texas A&amp;M University, Texas), Glen Hordemann (Texas A&amp;M University, Texas), and Francis Quek (Texas A&amp;M University, Texas)</i>	
Self-Supervised Learning of Free-Hand Sketches with Bézier Curve Features .....	168
<i>Taner Gülez (Baskent University, Turkey) and Mustafa Sert (Baskent University, Turkey)</i>	
Deep Learning-Based Point Cloud Geometry Coding with Attention Models .....	172
<i>Mohammadreza Ghafari (Universidade de Lisboa, Portugal), André F. R. Guarda (Instituto de Telecomunicações, Portugal), Nuno M. M. Rodrigues (ESTG – Politécnico de Leiria, Instituto de Telecomunicações, Portugal), and Fernando Pereira (Universidade de Lisboa, Portugal)</i>	
Estimation of the Caloric Intake of Food Consumption using Convolutional Neural Network .....	177
<i>Jie Li (California State University at Channel Islands, USA) and Bahareh Abbasi (California State University at Channel Islands, USA)</i>	
Deep Layers Beware: Unraveling the Surprising Benefits of JPEG Compression for Image Classification Pre-Processing .....	182
<i>Guruprasad Nayak (Amazon AI) and Gerald Friedland (Amazon AI)</i>	

## Sound Event Detection

Weakly Labeled Sound Event Detection using Attention Mechanism with Teacher-Student Model .	186
<i>Yesim Akar (Baskent University, Turkey) and Mustafa Sert (Baskent University, Turkey)</i>	
The Impact of Parroting Mode on Cross-Lingual Speaker Recognition .....	193
<i>Wen-Hung Liao (National Chengchi University, Taiwan), Yen-Chun Ou (National Chengchi University, Taiwan), Po-Han Chen (National Chengchi University, Taiwan), and Yi-Chieh Wu (National Chengchi University, Taiwan)</i>	
Synthesis of Disparate Audio Species via Recurrent Neural Embedding .....	198
<i>Vania Miriam Ortiz Ramos (Sungkyunkwan University, Republic of Korea) and Sukhan Lee (Sungkyunkwan University, Republic of Korea)</i>	
War & Music: The Impact of the Ukrainian War on the Music Listening Behaviour in Eastern Europe .....	202
<i>Fernando Terroso-Sáenz (Polytechnic University of Cartagena, Spain), Andres Muñoz (University of Cadiz, Spain), and Philippe Roose (University of Pau, France)</i>	

Triplet Loss with Curriculum Learning for Audio-Visual Retrieval .....	206
<i>Donghuo Zeng (KDDI Research, Inc., Japan) and Kazushi Ikeda (KDDI Research, Inc., Japan)</i>	

## Image Enhancement and Generation

3DTextureNet: Neural 3D Texture Style Transfer .....	208
<i>Abhinav Upadhyay (Accenture Labs, India), Alpana Dubey (Accenture Labs, India), and Suma Mani Kuriakose (Accenture Labs, India)</i>	
Towards Imperceptible Adversarial Image Generation: Minimizing Perceptual Difference .....	216
<i>Sivaji Retta (AnimalEYEQ, India) and Ramarajulu Srinivasan (AnimalEYEQ, Singapore)</i>	
Content-Aware Progressive Image Compression and Syncing .....	221
<i>Junda Wu (University of California San Diego, CA), Haoliang Wang (Adobe Research, CA), Tong Yu (Adobe Research, CA), Gang Wu (Adobe Research, CA), Stefano Petrangeli (Adobe Research, CA), Handong Zhao (Adobe Research, CA), Sungchul Kim (Adobe Research, CA), and Viswanathan Swaminathan (Adobe Research, CA)</i>	
Towards Efficient Multi-View Representation Learning .....	225
<i>Kai Liu (Toronto Metropolitan University), Zheng Guo (Toronto Metropolitan University), Lei Gao (Toronto Metropolitan University), Naimul Mefraz Khan (Toronto Metropolitan University), and Ling Guan (Toronto Metropolitan University)</i>	
Adaptation of Bitstream-Based Video Quality Models for Image Quality Assessment .....	230
<i>Rakesh Rao Ramachandra Rao (Technische Universität Ilmenau, Germany), Steve Göring (Technische Universität Ilmenau, Germany), and Alexander Raake (Technische Universität Ilmenau, Germany)</i>	

## Papers from High Schoolers

Bridging Subjectivity and Objectivity in Evaluation of Machine-Generated Jazz Music: A Multimetric Approach .....	232
<i>Conrad Hsu (Carlmont High School, USA) and Ross Greer (University of California San Diego, USA)</i>	
Spectrogram-Based Deep Learning for Flute Audition Assessment and Intelligent Feedback .....	238
<i>Manu Agarwal (Herricks High School, USA) and Ross Greer (University of California San Diego, USA)</i>	

## Event Detection

Heart Rate Detection using an Event Camera .....	243
<i>Aniket Jagtap (Dublin City University, Ireland), RamaKrishna Venkatesh Saripalli (Dublin City University, Ireland), Joe Lemley (Xperi Corporation, Ireland), Waseem Shariff (University of Galway, Ireland), and Alan F. Smeaton (Dublin City University, Ireland)</i>	

Masked Feature Modelling for the Unsupervised Pre-Training of a Graph Attention Network Block for Bottom-up Video Event Recognition .....	247
<i>Dimitrios Daskalakis (CERTH, Greece), Nikolaos Gkalelis (CERTH, Greece), and Vasileios Mezaris (CERTH, Greece)</i>	
Interactive Scene Graph Analysis for Future Intelligent Teleconferencing Systems .....	251
<i>Mingyuan Wu (University of Illinois Urbana Champaign, IL), Yuhan Lu (University of Illinois Urbana Champaign, IL), Shiv Trivedi (University of Illinois Urbana Champaign, IL), Bo Chen (University of Illinois Urbana Champaign, IL), Qian Zhou (University of Illinois Urbana Champaign, IL), Lingdong Wang (University of Massachusetts Amherst, MA), Simran Singh (New Jersey Institute of Technology, NJ), Michael Zink (University of Massachusetts Amherst, MA), Ramesh Sitaraman (University of Massachusetts Amherst, MA), Jacob Chakareski (New Jersey Institute of Technology, NJ), and Klara Nahrstedt (University of Illinois Urbana Champaign, IL)</i>	
Efficient Automatic Annotation of Binary Masks for Enhanced Training of Computer Vision Models .....	256
<i>Dylan Seychell (University of Malta, Malta), Matthew Kenely (University of Malta, Malta), Matthias Bartolo (University of Malta, Malta), Carl James Debono (University of Malta, Malta), Mark Bugeja (University of Malta, Malta), and Matthew Sacco (SeyTravel Ltd, Malta)</i>	
Care3D: An Active 3D Object Detection Dataset of Real Robotic-Care Environments .....	260
<i>Michael G. Adam (Chair of Media Technology), Sebastian Eger (Chair of Media Technology), Martin Piccolrovazzi (Chair of Media Technology), Seongjien Bien (Technical University of Munich), Jon Skerlj (Technical University of Munich), Abdeldjallil Naceri (Technical University of Munich), Eckehard Steinbach (Chair of Media Technology), Sami Haddadin (Technical University of Munich), Maged Iskandar (German Aerospace Center (DLR), Germany), Joern Vogel (German Aerospace Center (DLR), Germany), Alexander Dietrich (German Aerospace Center (DLR), Germany), Alin Albu-Schaeffer (German Aerospace Center (DLR), Germany), and Wolfram Burgard (University of Technology Nuremberg, Germany)</i>	

## Applications

Maintaining Text Legibility Regarding Font Size Based on User Distance in Mobile Devices: Application Development and User Evaluation .....	262
<i>Andreas Mallas (University of Patras, Greece), Hara Papadatou (University of Patras, Greece), and Michalis Xenos (University of Patras, Greece)</i>	
Analysis of Hand Movement and Head Orientation in Hierarchical Menu Selection in Immersive AR .....	270
<i>Majid Pourmemar (Concordia University, Canada) and Charalambos Poullis (Concordia University, Canada)</i>	
Exploring Terrestrial Point Clouds with Google Street View for Discovery and Fine-Grained Catalog of Urban Objects .....	276
<i>Gabriel Lugo (University of Alberta, Canada), Rutvik Chauhan (University of Alberta, Canada), and Irene Cheng (University of Alberta, Canada)</i>	



Learning Individualized Automatic Content Magnification in Gaze-Based Interaction .....	282
<i>Florian Eggenkemper (HSW University of Applied Sciences, Germany), Lars Kölker (HSW University of Applied Sciences, Germany), Mike Valente (HSW University of Applied Sciences, Germany), Constantin A. Rothkopf (Technical University of Darmstadt, Germany), and Robert Mertens (HSW University of Applied Sciences, Germany; BHH University of Applied Sciences, Germany)</i>	
Semantic and Lexical Token Based Vectors Improve Precision of Recommendations for TV Programmes .....	287
<i>Taner Cagali (University College London, UK), Hadi Wazni (University College London, UK), Saba Nazir (University College London, UK), Mehrnoosh Sadrzadeh (University College London, UK), and Chris Newell (British Broadcasting Corporation, United Kingdom)</i>	

## **Multimedia Technologies for E-Learning (MTEL) Workshop**

An app-Based Spaced Repetition Learning Environment with User Generated Video Content .....	291
<i>Florian Schimanke (HSW University of Applied Sciences, Germany), Robert Mertens (HSW University of Applied Sciences, Germany), and Lars Kölker (HSW University of Applied Sciences, Germany)</i>	
On the Creation of Classifiers to Support Assessment of E-Portfolios .....	297
<i>Alexander Gantikow (University of Education Weingarten, Germany), Andreas Isking (University of Education Weingarten, Germany), Paul Libbrecht (University of Education Weingarten, Germany), Wolfgang Müller (University of Education Weingarten, Germany), and Sandra Rebholz (University of Education Weingarten, Germany; Electrical Engineering, Media and Computer Science, Ostbayerische Technische Hochschule Amberg-Weiden, Germany)</i>	
Mental Workload in Augmented Reality-Based Urban Planning Education .....	303
<i>Sanghamitra Das (Bauhaus-Universität Weimar, Germany), Mario Wolf (Bauhaus-Universität Weimar, Germany), Philippe Schmidt (Bauhaus-Universität Weimar, Germany), and Heinrich Söbke (Bauhaus-Universität Weimar, Germany)</i>	
Composing Music through Tile-Based Games .....	309
<i>Samuli Laato (Tampere University, Finland), Sampsa Rauti (University of Turku, Finland), Alexander Espeseth (Kristiania University College, Norway), Heinrich Söbke (Bauhaus-Universität Weimar, Germany), Juho Hamari (Tampere University, Finland), and Oguz Buruk (Tampere University, Finland)</i>	
Identification of Visual Objects in Lecture Videos with Color and Keypoints Analysis .....	315
<i>Dipayan Biswas (University of Houston, Houston), Shishir Shah (University of Houston, Houston), and Jaspal Subhlok (University of Houston, Houston)</i>	

# 6th International Workshop on New Frontiers in Computational Robotics + Semantic Multimedia Computing (Fall Edition)

GPU-Accelerated Lossless Image Compression with Massive Parallelization .....	321
<i>Yu Shen (Adobe Research, USA), Gang Wu (Adobe Research, USA), Viswanathan Swaminathan (Adobe Research, USA), Haoliang Wang (Adobe Research, USA), Stefano Petrangeli (Adobe Research, USA), and Tong Yu (Adobe Research, USA)</i>	
Complex Event Processing in Heterogeneous Domains .....	325
<i>Fabio Persia (University of L'Aquila) and Daniela D'Auria (Free University of Bozen-Bolzano)</i>	
<b>Author Index .....</b>	<b>331</b>