

2021 IEEE International Symposium on Multimedia (ISM 2021)

**Virtual Conference
29 November – 1 December 2021**



**IEEE Catalog Number: CFP21197-POD
ISBN: 978-1-6654-3735-6**

**Copyright © 2021 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:	CFP21197-POD
ISBN (Print-On-Demand):	978-1-6654-3735-6
ISBN (Online):	978-1-6654-3734-9

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
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

2021 IEEE International Symposium on Multimedia (ISM) ISM 2021

Table of Contents

Message from the ISM 2021 General Chairs	xii
Message from the ISM 2021 Program Chairs	xiii
Organizing Committee	xv
Program Committee	xvi

Session 1:

Image and Video Processing

360ViewPET: View Based Pose Estimation for Ultra-Sparse 360-Degree Cameras	1
<i>Qian Zhou (University of Illinois Urbana-Champaign, USA), Bo Chen (University of Illinois Urbana-Champaign, USA), Zhe Yang (University of Illinois Urbana-Champaign, USA), Hongpeng Guo (University of Illinois Urbana-Champaign, USA), and Klara Nahrstedt (University of Illinois Urbana-Champaign, USA)</i>	
Improving 360-Degree Video Field-of-View Prediction and Edge Caching	9
<i>Alihsan Samiei (University of Texas at Dallas, USA) and Ravi Prakash (University of Texas at Dallas, USA)</i>	
BOhance: Bayesian Optimization for Content Enhancement	17
<i>Trisha Mittal (University of Maryland, College Park, USA), Vishy Swaminathan (Adobe Research, USA), Somdeb Sarkhel (Adobe Research, USA), Ritwik Sinha (Adobe Research, USA), David Arbour (Adobe Research, USA), Saayan Mitra (Adobe Research, USA), and Dinesh Manocha (University of Maryland, College Park, USA)</i>	
A Web Service for Video Smart-Cropping	25
<i>Konstantinos Apostolidis (CERTH-ITI, Greece) and Vasileios Mezaris (CERTH-ITI, Greece)</i>	

Session 2:

Anomaly Detection

Anomaly Detection in Smart Home Environments Using Convolutional Neural Network	27
<i>Naci Mert Ercan (Baskent University, Turkey) and Mustafa Sert (Baskent University, Turkey)</i>	

Weakly Supervised Anomaly Detection for Streaming Data	31
<i>Wei Zhang (Adobe Inc., USA) and Chris Challis (Wells Fargo, USA)</i>	
Knowledge-Based Wavelet Filters Prominently Detect Spoofed Speech	35
<i>Alex M. G. de Almeida (São Paulo State University, Brazil) and Rodrigo Capobianco Guido (São Paulo State University, Brazil)</i>	
Damage Detection of the RC Building in TLS Point Clouds Using 3D Deep Neural Network PointNet++	39
<i>Wanpeng Shao (Kyushu Institute of Technology, Japan), Ken'ichi Kakizaki (Kyushu Institute of Technology, Japan), Shunsuke Araki (Kyushu Institute of Technology, Japan), and Tomohisa Mukai (National Research and Development Agency Japan, Japan)</i>	

Session 3:

Extended Reality

Head Rotation Model for Virtual Reality System Level Simulations	43
<i>Steve Blandino (National Institute of Standards and Technology, USA), Tanguy Ropitault (National Institute of Standards and Technology & Prometheus Computing LLC, USA), Raied Caromi (National Institute of Standards and Technology, USA), Jacob Chakareski (New Jersey Institute of Technology, USA), Mahmudur Khan (York College of Pennsylvania, USA), and Nada Golmie (National Institute of Standards and Technology, USA)</i>	
Space-Warp with Depth Propagation in XR Applications	50
<i>Yingen Xiong (Samsung Research America, USA) and Christopher Peri (Samsung Research America, USA)</i>	
RealityCheck: A Tool to Evaluate Spatial Inconsistency in Augmented Reality	58
<i>Carter Slocum (University of California-Riverside, USA), Xukan Ran (University of California-Riverside, USA), and Jiasi Chen (University of California-Riverside, USA)</i>	
SHECS: A Local Smart Hands-Free Elderly Care Support System on Smart AR Glasses with AI Technology	66
<i>Donghuo Zeng (KDDI Research, Inc., Japan), Jianming Wu (KDDI Research, Inc., Japan), Bo Yang (KDDI Research, Inc., Japan), Tomohiro Obara (KDDI Research, Inc., Japan), Akeri Okawa (KDDI Research, Inc., Japan), Nobuko Iino (KDDI Research, Inc., Japan), Gen Hattori (KDDI Research, Inc., Japan), Ryoichi Kawada (KDDI Research, Inc., Japan), and Yasuhiro Takishima (KDDI Research, Inc., Japan)</i>	
Stuffed Toy as an Appealing Tangible Interface for Children	75
<i>Zackary P. T. Sin (Hong Kong Polytechnic University, China), Peter H. F. Ng (Hong Kong Polytechnic University, China), and Hong Va Leong (Hong Kong Polytechnic University, China)</i>	

Session 4:

Matching and Recognition

Fast Startup Multicast Streaming on Operator IPTV Networks Using HESP	79
<i>Egon Okerman (THEO Technologies, Belgium) and Johan Vounckx (THEO Technologies, Belgium)</i>	
A Two-Stream Heterogeneous Network for Action Recognition Based on Skeleton and RGB Modalities	87
<i>Kai Liu (Zhengzhou University, China), Lei Gao (Ryerson University, Canada), Naimul Mefraz Khan (Ryerson University, Canada), Lin Qi (Zhengzhou University, China), and Ling Guan (Ryerson University, Canada)</i>	
Open-Source RTP Library for End-to-End Encrypted Real-Time Video Streaming Applications	92
<i>Joni Räsänen (Tampere University, Finland), Aaro Altonen (Tampere University, Finland), Alexandre Mercat (Tampere University, Finland), and Jarno Vanne (Tampere University, Finland)</i>	
The Study of Optimal Layout of Command Recognition Area in Implicit Eye-Controlled Interfaces	97
<i>Jibin Yin (Kunming University of Science and Technology, China), Chaoxi Lu (Kunming University of Science and Technology, China), and Haonan Qin (Kunming University of Science and Technology, China)</i>	

Session 5:

Classification and Filter

Content-Adaptive Convolutional Neural Network Post-Processing Filter	99
<i>Maria Santamaria (Tampere University, Finland), Yat-Hong Lam (Tampere University, Finland), Francesco Cricri (Nokia Technologies, Finland), Jani Lainema (Nokia Technologies, Finland), Ramin G. Youvalari (Nokia Technologies, Finland), Honglei Zhang (Nokia Technologies, Finland), Miska M. Hannuksela (Nokia Technologies, Finland), Esa Rahtu (Tampere University, Finland), and Moncef Gaubuj (Tampere University, Finland)</i>	
Comprehensive Saliency Fusion for Object Co-Segmentation	107
<i>Harshit Singh Chhabra (Indraprastha Institute of Information Technology Delhi, India) and Koteswar Rao Jerripothula (Indraprastha Institute of Information Technology Delhi, India)</i>	
Semi-Supervised Audio Classification with Partially Labeled Data	111
<i>Siddharth Gururani (Georgia Institute of Technology, USA) and Alexander Lerch (Georgia Institute of Technology, USA)</i>	

Session 6: Best Paper Session (I)

Dynamic Motion Matching: Context-Aware Character Animation with Subspaces Ensembling .	115
<i>Adan Häfliger (Cygames, Inc., Japan) and Shuichi Kurabayashi (Cygames, Inc., Japan)</i>	

Inverse Kinematics for Full-Body Self Representation in VR-Based Cognitive Rehabilitation	123
<i>Larissa Wagnerberger (Fraunhofer HHI, Germany), Detlef Runde (Fraunhofer HHI, Germany), Mustafa Tevfik Lafci (Fraunhofer HHI, Germany), David Przewozny (Fraunhofer HHI, Germany), Sebastian Bosse (Fraunhofer HHI, Germany), and Paul Chojecki (Fraunhofer HHI, Germany)</i>	
Ranking Micro-Influencers: A Novel Multi-Task Learning and Interpretable Framework	130
<i>Adam Elwood (lastminute.com, Switzerland), Alberto Gasparin (lastminute.com, Switzerland), and Alessandro Rozza (lastminute.com, Switzerland)</i>	
L3BOU: Low Latency, Low Bandwidth, Optimized Super-Resolution Backhaul for 360-Degree Video Streaming	138
<i>Ayush Sarkar (University of Illinois at Urbana-Champaign, USA), John Murray (University of Massachusetts Amherst, USA), Mallesh Dasari (Stony Brook University, USA), Michael Zink (University of Massachusetts Amherst, USA), and Klara Nahrstedt (University of Illinois Urbana-Champaign, USA)</i>	

Session 7: Best Paper Session (II)

A Hybrid NDN-IP Architecture for Live Video Streaming: A QoE Analysis	148
<i>Ishita Dasgupta (University of Massachusetts, Amherst, USA), Susmit Shannigrahi (Tennessee Tech University, USA), and Michael Zink (University of Massachusetts, Amherst, USA)</i>	
Adversarial Perturbation Suppression Using Adaptive Gaussian Smoothing and Color Reduction...	158
<i>Li-Yun Wang (Portland State University, USA)</i>	
Novel Datasets for Evaluating Song Popularity Prediction Tasks	166
<i>Michael Vötter (Universität Innsbruck, Austria), Maximilian Mayerl (Universität Innsbruck, Austria), Günther Specht (Universität Innsbruck, Austria), and Eva Zangerle (Universität Innsbruck, Austria)</i>	

Session 9:

Recommendation Systems in Multimedia

Open-Domain Trending Hashtag Recommendation for Videos	174
<i>Swapneel Mehta (New York University, USA), Somdeb Sarkhel (Adobe Research, USA), Xiang Chen (Adobe Research, USA), Saayan Mitra (Adobe Research, USA), Viswanathan Swaminathan (Adobe Research, USA), Ryan Rossi (Adobe Research, USA), Ali Aminian (Adobe Inc., USA), Han Guo (Adobe Inc., USA), and Kshitiz Garg (Adobe Inc., USA)</i>	
User Profiling for Tourist Trip Recommendations Using Social Sensing	182
<i>Vincenzo Emanuele Carusotto (University of Palermo, Italy), Giovanni Pilato (National Research Council of Italy, Italy), Fabio Persia (University of L'Aquila, Italy), and Mouzhi Ge (Deggendorf Institute of Technology, Germany)</i>	

Enhancing Personalised Recommendations with the Use of Multimodal Information	186
<i>Taner Cagali (University College London, UK), Mehrnoosh Sadrzadeh (University College London, UK), and Chris Newell (BBC Research and Development, UK)</i>	

Session 10:

Miscellaneous Applications (I)

Cross-Modality Wood Log Tracing	191
<i>Georg Wimmer (University of Salzburg, Austria), Rudolf Schraml (University of Salzburg, Austria), Lukas Lamminger (University of Salzburg, Austria), Alexander Petutschnigg (University of Applied Sciences Salzburg, Austria), and Andreas Uhl (University of Salzburg, Austria)</i>	
Combining Linked Open Data and Multimedia Knowledge Base for Digital Cultural Heritage Robotic Applications	196
<i>Kurosh Madani (LISSI Laboratory, University of Paris-Est Creteil, France), Antonio M. Rinaldi (University of Naples Federico II, Italy), and Cristiano Russo (University of Naples Federico II, Italy)</i>	
Social Interaction in Virtual Shopping	204
<i>Nada Nasser (The German University in Cairo, Egypt), Elhassan Makled (The German University in Cairo, Egypt), Nada Sharaf (The German University in Cairo, Egypt), and Slim Abdennadher (The German University in Cairo, Egypt)</i>	

Session 11:

Miscellaneous Applications (II)

Analyses and Benchmark of a Spontaneous Student Affect Database	206
<i>Bo Sun (Beijing Normal University at Zhuhai, China), Sixu Lu (Beijing Normal University, China), Yang Wen (Beijing Normal University, China), Jun He (Beijing Normal University at Zhuhai, China), and Lejun Yu (Beijing Normal University, China)</i>	
Automated Clipping of Soccer Events Using Machine Learning	210
<i>Joakim O. Valand (SimulaMet, Norway & University of Oslo, Norway), Haris Kadragic (SimulaMet, Norway & University of Oslo, Norway), Steven A. Hicks (SimulaMet, Norway & Oslo Metropolitan University, Norway), Vajira Thambawita (SimulaMet, Norway & Oslo Metropolitan University, Norway), Cise Midoglu (SimulaMet, Norway), Tomas Kupka (Forzasys, Norway), Dag Johansen (UIT The Arctic University of Norway, Norway), Michael A. Riegler (SimulaMet, Norway & UIT The Arctic University of Norway, Norway), and Pål Halvorsen (SimulaMet, Norway; Oslo Metropolitan University, Norway; & Forzasys, Norway)</i>	

Effect of Memory Soft Errors on Media Applications	215
<i>Pooja Sundar (Intel Corp., India), Suresh Vasu (Intel Corp., India), Nithin Venkatesh (Intel Corp., India), and Praveen Prasad (Intel Corp., USA)</i>	

Session 12:

Image and Video Coding

Enhancing Image Coding for Machines with Compressed Feature Residuals	217
<i>Joni Seppälä (Tampere University, Finland), Honglei Zhang (Nokia Technologies, Finland), Nam Le (Nokia Technologies, Finland), Ramin G. Youvalari (Nokia Technologies, Finland), Francesco Cricri (Nokia Technologies, Finland), Hamed Rezazadegan Tavakoli (Nokia Technologies, Finland), Emre Aksu (Nokia Technologies, Finland), Miska M. Hannuksela (Nokia Technologies, Finland), and Esa Rahtu (Tampere University, Finland)</i>	
Combining Global and Local Attention with Positional Encoding for Video Summarization	226
<i>Evlampios Apostolidis (CERTH-ITI, Greece & Queen Mary University of London, UK), Georgios Balaouras (CERTH-ITI, Greece), Vasileios Mezaris (CERTH-ITI, Greece), and Ioannis Patras (Queen Mary University of London, UK)</i>	
Learned Enhancement Filters for Image Coding for Machines	235
<i>Jukka I. Ahonen (Tampere University, Finland), Ramin G. Youvalari (Nokia Technologies, Finland), Nam Le (Nokia Technologies, Finland & Tampere University, Finland), Honglei Zhang (Nokia Technologies, Finland), Francesco Cricri (Nokia Technologies, Finland), Hamed Rezazadegan Tavakoli (Nokia Technologies, Finland), Miska M. Hannuksela (Nokia Technologies, Finland), and Esa Rahtu (Tampere University, Finland)</i>	
Adaptation and Attention for Neural Video Coding	240
<i>Nannan Zou (Nokia Technologies, Finland & Tampere Univesity, Finland), Honglei Zhang (Nokia Technologies, Finland), Francesco Cricri (Nokia Technologies, Finland), Ramin G. Youvalari (Nokia Technologies, Finland), Hamed R. Tavakoli (Nokia Technologies, Finland), Jani Lainema (Nokia Technologies, Finland), Emre Aksu (Nokia Technologies, Finland), Miska Hannuksela (Nokia Technologies, Finland), and Esa Rahtu (Tampere University, Finland)</i>	
A Sinusoidal Signal Reconstruction Method for the Inversion of the Mel-Spectrogram	245
<i>Anastasia Natsiou (Technological University of Dublin, Ireland) and Seán O'Leary (Technological University of Dublin, Ireland)</i>	

Session 13:

Neural Networks and Multimedia

Edge-Level Explanations for Graph Neural Networks by Extending Explainability Methods for Convolutional Neural Networks	249
<i>Tetsu Kasanishi (University of Tokyo, Japan), Xueting Wang (University of Tokyo, Japan), and Toshihiko Yamasaki (University of Tokyo, Japan)</i>	
Sequential Banner Design Optimization with Deep Reinforcement Learning	253
<i>Yusuke Kondo (University of Tokyo, Japan), Xueting Wang (University of Tokyo, Japan), Hiroyuki Seshime (Septeni Japan Co., Japan), and Toshihiko Yamasaki (University of Tokyo, Japan)</i>	
CCAP: Cooperative Context Aware Pruning for Neural Network Model Compression	257
<i>Li-Yun Wang (Portland Sate University, USA) and Zahid Akhtar (State University of New York Polytechnic Institute, USA)</i>	
An Analysis of Lightweight Convolutional Neural Networks for Parking Space Occupancy Detection	261
<i>Joshua D. Ellis (Missouri State University, USA), Anthony Harris (Missouri State University, USA), Naseem Saquer (Missouri State University, USA), and Razib Iqbal (Missouri State University, USA)</i>	

MTEL Workshop

Intelligent Voice Assistant to Facilitate Elementary School English Learning: A Case Study Using Amazon Echo Dot	269
<i>Yi-Chieh Wu (National Chengchi University, Taiwan) and Wen-Hung Liao (National Chengchi University, Taiwan)</i>	
The Impact of Spaced Repetition Learning on the Learning Success in Mobile Learning Games	275
<i>Florian Schimanke (HSW University of Applied Sciences, Germany)</i>	
Identifying Keyword Predictors in Lecture Video Screen Text	281
<i>Farah Naz Chowdhury (University of Houston, USA), Raga Shalini Koka (University of Houston, USA), Mohammad Rajiur Rahman (University of Houston, USA), Thamar Solorio (University of Houston, USA), and Jaspal Subhlok (University of Houston, USA)</i>	

Author Index	287
---------------------------	------------