

2024 International Symposium on Multimedia (ISM 2024)

**Tokyo, Japan
11-13 December 2024**



IEEE Catalog Number: CFP24197-POD
ISBN: 979-8-3315-1112-8

**Copyright © 2024 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:	CFP24197-POD
ISBN (Print-On-Demand):	979-8-3315-1112-8
ISBN (Online):	979-8-3315-1111-1

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
Web: www.proceedings.com

CURRAN ASSOCIATES INC.
proceedings
.com

2024 International Symposium on Multimedia (ISM) ISM 2024

Table of Contents

Message from the General Chairs	xiii
Message from the Program Chairs	xiv

ISM-1 - Image Processing (I)

S2MGen: A Synthetic Skin Mask Generator for Improving Segmentation	1
<i>Subhadra Gopalakrishnan (Dolby Laboratories Inc., USA), Trisha Mittal (Dolby Laboratories Inc., USA), Jaclyn Pytlarz (Dolby Laboratories Inc., USA), and Yuheng Zhao (University of South California, USA)</i>	
Generating and Evaluating Cursive Chinese Calligraphy by Semi-Classifying Style: A Case Study using a Diffusion Model	9
<i>Yi-Chieh Wu (National Chengchi University, Taiwan) and Yu-Jung Hsu (National Chengchi University, Taiwan)</i>	
StegoFusion-Net: Fusion of Convolutional Neural Networks for Spatial Image Steganalysis	17
<i>Yassine Belkhouche (Missouri State University, USA) and AlaaIdin Dwaik (Missouri State University, USA)</i>	
Disparity Correction Method of the Monocular Omnidirectional Stereo Camera	24
<i>Hisayoshi Kaneda (Hitachi, Ltd, Japan), Ryota Kawamata (Hitachi, Ltd, Japan), Kazuyoshi Yamazaki (Hitachi, Ltd, Japan), and Kazuya Shimizu (Hitachi-GE Nuclear Energy, Ltd, Japan)</i>	

ISM-2 - Audio Analysis

Unveiling the Potential of SSL-Generated Audio Embeddings for Cross-Lingual Speaker Recognition	26
<i>Wen-Hung Liao (National Chengchi University, Taiwan), Po-Han Chen (National Chengchi University, Taiwan), and Yi-Chieh Wu (National Chengchi University, Taiwan)</i>	
Two-Stage Instrument Timbre Transfer Method using RAVE	33
<i>Di Hu (Hosei University, Japan) and Katunobu Ito (Hosei University, Japan)</i>	
Speaker Pseudonymization for Japanese Speech using Duration Embeddings	41
<i>Aoi Ito (Hosei University, Japan) and Katunobu Ito (Hosei University, Japan)</i>	

ISM-3 - Best Paper Session

Modeling User Quality of Experience in Adaptive Point Cloud Video Streaming	49
<i>Duc V. Nguyen (Tohoku Institute of Technology, Japan), Nguyen Long Quang (Hanoi University of Science and Technology, Vietnam), Tran Thuy Hien (Hanoi University of Science and Technology, Vietnam), Nguyen Ngoc Huyen (Hanoi University of Science and Technology, Vietnam), Truong Thu Huong (Hanoi University of Science and Technology, Vietnam), and Pham Ngoc Nam (VinUniversity, Vietnam)</i>	
Appeal Prediction for AI Up-Scaled Images	55
<i>Steve Göring (Technische Universität Ilmenau, Germany), Rasmus Merten (Technische Universität Ilmenau, Germany), and Alexander Raake (Technische Universität Ilmenau, Germany)</i>	
Modelling Concurrent RTP Flows for End-to-End Predictions of QoS in Real Time Communications	63
<i>Tailai Song (Politecnico di Torino, Italy), Paolo Garza (Politecnico di Torino, Italy), Michela Meo (Politecnico di Torino, Italy), and Maurizio Matteo Munafò (Politecnico di Torino, Italy)</i>	
SoccerNet-Echoes: A Soccer Game Audio Commentary Dataset	71
<i>Sushant Gautam (SimulaMet, Norway; OsloMet, Norway), Mehdi Houshmand Sarkhoosh (OsloMet, Norway; Forzasys, Norway), Jan Held (University of Liège, Belgium), Cise Midoglu (SimulaMet, Norway; Forzasys, Norway), Anthony Cioppa (University of Liège, Belgium; KAUST, Saudi Arabia), Silvio Giancola (KAUST, Saudi Arabia), Vajira Thambawita (SimulaMet, Norway), Michael A. Riegler (SimulaMet, Norway), Pål Halvorsen (SimulaMet, Norway; OsloMet, Norway; Forzasys, Norway), and Mubarak Shah (University of Central Florida, USA)</i>	

ISM-4 - Object and Event

Ensuring Color Consistency in RGB-D Multi-Camera Setup	79
<i>Peter O. Fasogbon (Nokia Technologies, Finland)</i>	
Low Complexity Learning-Based Lossless Event-Based Compression	85
<i>Ahmadreza Sezavar (Instituto Superior Técnico; Instituto de Telecomunicações), Catarina Brites (Instituto Universitário de Lisboa (ISCTE-IUL); Instituto de Telecomunicações), and João Ascenso (Instituto Superior Técnico; Instituto de Telecomunicações)</i>	
PlayerTV: Advanced Player Tracking and Identification for Automatic Soccer Highlight Clips.....	93
<i>Håkon Solberg (University of Oslo, Norway), Mehdi H. Sarkhoosh (Forzasys, Norway), Sushant Gautam (OsloMet, Norway; SimulaMet, Norway), Saeed S. Sabet (Forzasys, Norway), Pål Halvorsen (OsloMet, Norway; SimulaMet, Norway), and Cise Midoglu (Forzasys, Norway; SimulaMet, Norway)</i>	

ISM-5 - Text and Data Analysis

Flexible and Faithful Data Insights Generation	98
<i>Wei Zhang (Adobe Inc., USA) and Victor Soares Bursztyn (Adobe Research, USA)</i>	

Holistic Visualization of Contextual Knowledge in Hotel Customer Reviews using Self-Attention	106
<i>Shuntaro Masuda (The University of Tokyo, Japan) and Toshihiko Yamasaki (The University of Tokyo, Japan)</i>	
Investigation of Feature Distribution and Network Weight Updates in the Machine Unlearning Process	110
<i>Wen-Hung Liao (National Chengchi University, Taiwan) and Yang-Jing Lin (National Chengchi University, Taiwan)</i>	
Platform for Endangered Language Education	114
<i>Greeshma Sree Parimi (University of Washington Bothell, USA), Gurkirat Singh Guliani (University of Washington Bothell, USA), and Min Chen (University of Washington Bothell, USA)</i>	

ISM-6 - Music Generation

Homophonic Music Composition using a GAN and LSTM Pipeline for Melody and Harmony Generation	116
<i>Clément Saint-Marc (Hosei University, Japan) and Katunobu Itou (Hosei University, Japan)</i>	
Instrumentality Classification Evaluation System for Natural Sounds	120
<i>Yuhuan Wang (Hosei University, Japan) and Katunobu Itou (Hosei University, Japan)</i>	
Generating Bass Phrases from Guitar Chord Backing with NMF	124
<i>Tomoo Kouzai (Nihon Univeristy, Japan), Junya Koguchi (Meiji University, Japan), and Tetsuro Kitahara (Nihon University, Japan)</i>	

ISM-7 - Video Analysis (I)

Watch Your Back! Dynamic Thumbnails for a 360-Degree Video Player to Enhance Viewing Experience on 2D Displays	126
<i>Jakub Kováč (Utrecht University, Netherlands) and Wolfgang Hürst (Utrecht University, Netherlands)</i>	
Influence of Display Devices and Field of View on Subjective Quality of Experience Evaluation of 8K 360° Videos	133
<i>Daichi Arai (NHK, Japan), Yuichi Kondo (NHK, Japan), Yasuko Sugito (NHK, Japan), and Yuichi Kusakabe (NHK, Japan)</i>	
VEMOCLAP: A Video Emotion Classification Web Application	137
<i>Serkan Sulun (INESC TEC, Portugal), Paula Viana (Polytechnic of Porto, Portugal), and Matthew E. P. Davies (Independent researcher)</i>	
A Power-Law Transformation Approach for Template-Based Cross-Component Prediction	141
<i>Zhikai Liu (Sun Yat-sen University, China), Kun Zhang (Sun Yat-sen University, China), Xin-Yi Cui (Sun Yat-sen University, China), Wei Sun (Sun Yat-sen University, China), and Fan Liang (Sun Yat-sen University, China)</i>	

Investigating the Impact of High Frame Rate on Video Quality: A SAMVIQ Approach	143
<i>Dominik Keller (Technische Universität Ilmenau, Germany), Paul Rudi Frank (Technische Universität Ilmenau, Germany), Steve Göring (Technische Universität Ilmenau, Germany), and Alexander Raake (Technische Universität Ilmenau, Germany)</i>	

ISM-8 - Video Streaming

A Server-Driven View-Aware Point Cloud Video Streaming Framework	145
<i>Tran Gia Minh (Hanoi University of Science and Technology, Vietnam), Truong Thu Huong (Hanoi University of Science and Technology, Vietnam), and Duc V. Nguyen (Tohoku Institute of Technology, Japan)</i>	
Evaluation of Strategies for Efficient Rate-Distortion NeRF Streaming	149
<i>Pedro Martin (Instituto de Telecomunicações - Instituto Superior Técnico, Portugal), António Rodrigues (Instituto de Telecomunicações - Instituto Superior Técnico, Portugal), João Ascenso (Instituto de Telecomunicações - Instituto Superior Técnico, Portugal), and Maria Paula Queluz (Instituto de Telecomunicações - Instituto Superior Técnico, Portugal)</i>	
Perceptual Quality Driven Point Cloud Compression for 6DoF 3D Point Cloud Streaming	154
<i>Yumeka Chujo (Waseda University, Japan), Yusuke Tagashira (Waseda University, Japan), Yukiko Harada (Waseda University, Japan), Kenji Kanai (Waseda University, Japan), and Jiro Katto (Waseda University, Japan)</i>	
On Multi-CDN Delivery Costs Optimization Problem	158
<i>Yuriy A. Reznik (Brightcove, Inc., USA) and Guillem Cabrera (Brightcove UK, Ltd., UK)</i>	

ISM-9 - Object and Event Detection (II)

Sliding Window Check: Repairing Object Identities	162
<i>Geerthan Srikantharajah (Toronto Metropolitan University, Canada) and Naimul Khan (Toronto Metropolitan University, Canada)</i>	
Data Augmentation with Diffusion Model for Hand Detection	170
<i>Genta Matsukawa (Japan Advanced Institute of Science and Technology, Japan) and Atsuo Yoshitaka (Japan Advanced Institute of Science and Technology, Japan)</i>	
AI Maintenance Techniques by Detecting Performance Degradation in Domain Shift using Model Ensembles	174
<i>Keita Yamane (Hitachi Ltd., Japan), Akira Kitayama (Hitachi Ltd., Japan), Keigo Hasegawa (Hitachi Kokusai Electric Inc., Japan), Yusuke Obonai (Hitachi Kokusai Electric Inc., Japan), and Hiroto Sasao (Hitachi Kokusai Electric Inc., Japan)</i>	

ISM-10 - Image Processing (II)

Cross-Modal 3D Model Retrieval	176
<i>Raphael Waltenspiül (University of Basel, Switzerland), Florian Spiess (University of Basel, Switzerland), and Heiko Schuldt (University of Basel, Switzerland)</i>	
Prevention of Unexpected Object Generation in Diffusion Model-Based Inpainting	181
<i>Takumi Komori (Kansai University, Japan) and Takahiro Hayashi (Kansai University, Japan)</i>	
LMM-Regularized CLIP Embeddings for Image Classification	185
<i>Maria Tzelepi (Centre of Research and Technology Hellas, Greece) and Vasileios Mezaris (Centre of Research and Technology Hellas, Greece)</i>	
Evaluation Framework for Novel View Synthesis	189
<i>Kolja Kieslich (Fraunhofer FOKUS, Germany), Louay Bassbouss (Fraunhofer FOKUS, Germany), Stephan Steglich (Fraunhofer FOKUS, Germany), and Stefan Arbanowski (Fraunhofer FOKUS, Germany)</i>	

ISM-11 - Video Transmission

A Simulation for the Evaluation of the Mean Opinion Score (MOS) for EVS-WB and AMR-WB	
Audio Codecs for 5G Mobile Networks	193
<i>Jussif Abularach Arnez (Sidia Institute of Science and Technology, Brazil), Cássio Antonio Tavares Alves (Sidia Institute of Science and Technology, Brazil), Wederson Medeiros Silva (Sidia Institute of Science and Technology, Brazil), Isaac Barros Gomes (Sidia Institute of Science and Technology, Brazil), Carla Lapa Nogueira (Sidia Institute of Science and Technology, Brazil), and Maria Gabriela Lima Damasceno (Sidia Institute of Science and Technology, Brazil)</i>	
FrameCorr: Adaptive, Autoencoder-Based Neural Compression for Video Reconstruction in Resource and Timing Constrained Network Settings	197
<i>John Li (University of Illinois at Urbana-Champaign), Deepak Nair (University of Illinois at Urbana-Champaign), Klara Nahrstedt (University of Illinois at Urbana-Champaign), Indranil Gupta (University of Illinois at Urbana-Champaign), and Shehab Sarar Ahmed (University of Illinois at Urbana-Champaign)</i>	
Ultra-Low-Latency 8K120p-Video-Transmission System Parallelizing SMPTE ST 2110	201
<i>Yasuhiro Mochida (Nippon Telegraph and Telephone Corporation, Japan), Takuro Yamaguchi (Nippon Telegraph and Telephone Corporation, Japan), Hirokazu Takahashi (Nippon Telegraph and Telephone Corporation, Japan), and Koichi Takasugi (Nippon Telegraph and Telephone Corporation, Japan)</i>	
Low-Latency Software-Based Uncompressed Video Transmission	203
<i>Takuro Yamaguchi (Nippon Telegraph and Telephone Corporation, Japan), Yasuhiro Mochida (Nippon Telegraph and Telephone Corporation, Japan), and Hirokazu Takahashi (Nippon Telegraph and Telephone Corporation, Japan)</i>	

ISM-12 - Video Analysis (II)

Visual Speech Recognition with Surrounding and Emotional Information	205
<i>Pengcheng Zeng (Japan Advanced Institute of Science and Technology) and Atsuo Yoshitaka (Japan Advanced Institute of Science and Technology)</i>	
Synchronized Object Sharing for Augmented Reality Virtual Conferencing	213
<i>John Murray (University of Massachusetts Amherst) and Michael Zink (University of Massachusetts Amherst)</i>	
Fusion-Based Human Pose Estimation using RGB and IR Images with Transformer-Based Decoding....	219
<i>Viviana Crescitelli (Hitachi Ltd., Japan) and Takashi Oshima (Hitachi Ltd., Japan)</i>	

ISM-13 - Augmented Reality (I)

Occlusion-Aware Real-Time Tiny Facial Alignment Model for Makeup Virtual Try-On	221
<i>Kin Ching Lydia Chau (ModiFace Inc., Canada), Zhi Yu (ModiFace Inc., Canada), and Ruowei Jiang (ModiFace Inc., Canada)</i>	
A Study on Mental Stress Test using Cybersickness Caused by Virtual Reality Contents	225
<i>Nan Bu (Kumamoto College) and Kakeru Nakano (Kumamoto College, Japan)</i>	

ISM-14 - Augmented Reality (II)

Exploring Augmented Table Setup and Lighting Customization in a Simulated Restaurant to Improve the User Experience	227
<i>Jana Motowilowa (TU Berlin, Germany), Maurizio Vergari (TU Berlin, Germany), Tanja Kojić (TU Berlin, Germany), Maximilian Warsinke (TU Berlin, Germany), Sebastian Möller (TU Berlin, Germany; DFKI, Germany), and Jan-Niklas Voigt-Antons (Hochschule Hamm-Lippstadt, Germany)</i>	
Human-in-the-Loop Knowledge Base Upkeep for Retrieval Augmented Generation Applications .	232
<i>Pedro Baptista de Castro (Hitachi, Ltd., Japan), Hiroko Sukeda (Hitachi, Ltd., Japan), and Soichi Takashige (Hitachi, Ltd., Japan)</i>	
LiveSkeleton: High-Quality Real-Time Human Tracking and Pose Estimation	234
<i>Hannes Fassold (JOANNEUM RESEARCH, Austria)</i>	

ISM: MTEL + SMC workshop

ISM-15 - Video Conferencing

A Technical Concept for Enhancing the Student Experience in Hybrid Lecture Scenarios	236
<i>Florian Schimanke (HSW University of Applied Sciences, Germany), Robert Mertens (HSW University of Applied Sciences, Germany), and Felix Prankel (Sennheiser electronic SE & Co. KG, Germany)</i>	

SpotiView: Partial Face Display Method for Smooth Communication While Protecting Privacy	242
<i>Ryota Kishimoto (Kobe University, Japan), Shuhei Tsuchida (Ochanomizu University, Japan), Tsutomu Terada (Kobe University, Japan), and Masahiko Tsukamoto (Kobe University, Japan)</i>	
Characterizing Students Behavior in Multi-User Multi-Computer Testing Environments	250
<i>Rajini Chittimalla (Missouri State University, USA), Sujung Choi (Missouri State University, USA), Madhu Sai Vineel Reka (Missouri State University, USA), and Yassine Belkhouche (Missouri State University, USA)</i>	
Evaluating Interactive Concept Maps Produced from E-Portfolios	255
<i>Alexander Gantikow (University of Education Weingarten, Germany), Andreas Isking (University of Education Weingarten, Germany), Wolfgang Müller (University of Education Weingarten, Germany), Paul Libbrecht (IU International University of Applied Science, Germany), and Sandra Rebholz (Ostbayerische Technische Hochschule Amberg-Weiden, Germany)</i>	
Gender Stereotypes in the Creation of Educational Cases with ChatGPT	261
<i>Gabriel Valerio-Ureña (Tecnológico de Monterrey, México), Giomara Sevilla-Campoverde (Tecnológico de Monterrey, México), Soledad Ortúzar (Universidad del Desarrollo, Chile), and Christian Lazcano (Universidad del Desarrollo, Chile)</i>	
Multi-View Gesture Recognition in Conflict Situations	267
<i>Karam Tomotaki-Dawoud (Fraunhofer HHI, Germany), Birgit Nierula (Fraunhofer HHI, Germany), Farelle Toumaleu Siewe (Fraunhofer HHI, Germany), Thomas Koch (Fraunhofer HHI, Germany), Daniel Johannes Meyer (Fraunhofer HHI, Germany), Andreas Bock (Akkon Hochschule, Germany), Marianne Heinze (Akkon Hochschule, Germany), Daniela Knuth (Akkon Hochschule, Germany), Denis Martin (Akkon Hochschule, Germany), Julia Schander (Akkon Hochschule, Germany), Anna Hilsmann (Fraunhofer HHI, Germany), Peter Eisert (Fraunhofer HHI, Germany; Humboldt University, Germany), and Sebastian Bosse (Fraunhofer HHI, Germany)</i>	
PanoramaViewer – A Framework for Educational Collaborative Virtual Field Trips	269
<i>Mario Wolf (Bauhaus-Universität Weimar, Germany), Sebastian Hartwig (Ulm University, Germany), Gregor Steinhöfel (Bauhaus-Universität Weimar, Germany), Heinrich Söbke (HSW University of Applied Sciences, Germany; Bauhaus-Universität Weimar, Germany), and Eckhard Kraft (Bauhaus-Universität Weimar, Germany)</i>	
Real-Time Multi-Modal Highlight Prediction for Simultaneous Viewing of Multiple Live Streams	275
<i>Yusuke Maeda (Kansai University, Japan) and Takahiro Hayashi (Kansai University, Japan)</i>	
Slide Analysis Method for Editing Lecture Materials Based on Hierarchical Structures of Subject Terminologies	279
<i>Itsuki Sano (Kwansei Gakuin University, Japan), Yuanyuan Wang (Yamaguchi University, Japan), Yukiko Kawai (Kyoto Sangyo University; Osaka University, Japan), and Kazutoshi Sumiya (Kwansei Gakuin University)</i>	

The «Huh?» Button: Improving Understanding in Educational Videos with Large Language Models	285
<i>Boris Ruf (AXA Group Operations, France) and Marcin Detyniecki (AXA Group Operations, France)</i>	
Author Index	291