# **2022 IEEE 5th International Conference on Multimedia Information Processing and Retrieval (MIPR 2022)**

**Virtual Conference** 2-4 August 2022



IEEE Catalog Number: CFP22K85-POD **ISBN:** 

978-1-6654-9549-3

# Copyright © 2022 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:	
ISBN (Print-On-Demand):	
ISBN (Online):	
ISSN:	

CFP22K85-POD 978-1-6654-9549-3 978-1-6654-9548-6 2770-4327

#### 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



## 2022 IEEE 5th International Conference on Multimedia Information Processing and Retrieval (MIPR) MIPR 2022

#### **Table of Contents**

Welcome Message from MIPR 2022 Organization Committee	xiv
Organizing Committee	xvi
Program Committee	<b>xviii</b>

#### **Session 1: Learning and Classification**

Feature Compression for Rate Constrained Object Detection on the Edge
CAPT: Contrastive Pre-training Based Semi-Supervised Open-Set Learning
A Robust Collaborative Learning Framework using Data Digests and Synonyms to Represent Absent Clients
AgileGCN: Accelerating Deep GCN with Residual Connections using Structured Pruning
Aerial Image Classification Through Thin Lensless Camera
SimpleView++: Neighborhood Views for Point Cloud Classification

### Session 2: Learning and Content Analysis

Learning Rotational Invariant Dictionary for Sparse Coding Based Key-Point Detection Phuoc-Thanh Hong (Ryerson University, Canada) and Ling Guan (Ryerson University, Canada)	35
Video Origin Camera Identification using Ensemble CNNs of Positional Patches Maryna Veksler (Florida International University), Ramazan Aygun (Kenessaw State University), Kemal Akkaya (Florida International University), and Sitharama Iyengar (Florida International University)	11
Interpretable Learning-Based Multi-modal Hashing Analysis for Multi-view Feature Representation Learning	17
Improving Angular Estimation using a Deep CNN Network in 6D Pose Estimation	53
Sparse Tensor-Based Point Cloud Attribute Compression5 Jianqiang Wang (Nanjing University, China) and Zhan Ma (Nanjing University, China )	59

### Session 3: Video Coding, Robotics

Machine-Learning Based high Efficiency Rate Control for AV1 Yi Chen (City University of Hong Kong, China), Yunhao Mao (City University of Hong Kong, China), Shiqi Wang (City University of Hong Kong, China), Xianguo Zhang (Shannon Lab, TEG, Tencent, China), and Sam Kwong (City University of Hong Kong, China)	65
Recursive Randomized Tree Coding of Speech Hoontaek Oh (University of California, Santa Barbara) and Jerry D. Gibson (University of California, Santa Barbara)	71
Information-Seeking in Localization and Mission Planning of Multi-agent Systems Kyriakos Lite (University of Klagenfurt, Austria) and Bernhard Rinner (University of Klagenfurt, Austria)	77
Fast VVC Intra Coding by Skipping Redundant Coding Block Structures and Unnecessary Directional Partition	84

<ul> <li>Evolution of AVIF Encoder: Speed and Memory Optimizations</li></ul>	90
Learning to Predict on Octree for Scalable Point Cloud Geometry Coding Yixiang Mao (New York University, USA), Yueyu Hu (New York University, USA), and Yao Wang (New York University, USA)	€
Session 4: New Applications	
Enabling Automatic Cinematography with Reinforcement Learning	03
Spatially Isotropic 3D Volumetric Reconstruction of Live Biological Cells with Multi-View Geometry	09
GARGI: Selecting Gaze-Aware Representative Group Image from a Live Photo	15

Man Ho (OPPO US Research Center, USA)	

York), Shashank Arora (University at Albany, State University of New York), and Pradeep K. Atrey (University at Albany, State University of

Shlomo Dubnov (UC San Diego), Gerard Assayag (IRCAM), and Vignesh

Jie Cai (OPPO US Research Center, USA), Zibo Meng (OPPO US Research Center, USA), Jiaming Ding (OPPO US Research Center, USA), and Chiu

#### **Session 5: Learning and Classification**

New York)

Gokul (UC San Diego)

RE2L: A Real-World Dataset for Outdoor Low-Light Image Enhancement Yuxin Lin (Nanyang Technological University, Singapore), Jian Li (Nanyang Technological University, Singapore), Lanqing Guo (Nanyang Technological University, Singapore), and Bihan Wen (Nanyang Technological University, Singapore)	133
MOAC: Multi-level Perception Optimizer Based on Dual Augmented Cost for Structure-from-Motion	139
Peixi Wu (Peking University Shenzhen Graduate School), Ge Li (Peking University Shenzhen Graduate School), and Thomas H. Li (Peking University Shenzhen Graduate School; Peking University; Center of Peking University)	
reking University)	

16
52
8
52

### Session 6: Learning and Content Analysis

Skipped-Connection Transformer for Image Captioning
CRNet: Collaborative Refinement Network for Self-Supervised Video Object Segmentation 172 Dexiang Hong (University of Chinese Academy of Sciences, China), Guorong Li (University of Chinese Academy of Sciences, China), Bineng Zhong (Guangxi Normal University, China), Zhenjun Han (University of Chinese Academy of Sciences, China), Li Su (University of Chinese Academy of Sciences, China), and Qinming Huang (University of Chinese Academy of Sciences, China)
Disparity-Guided Light Field Video Synthesis with Temporal Consistency
Lightning YOLOv4 for a Surface Defect Detection System for Sawn Lumber
Deep Unsupervised Blind Learning for Single Image Super Resolution

Action Semantic Alignment for Image Captioning	194
Da Huo (Nagoya University, Japan), Marc A. Kastner (Kyoto University,	
Japan), Takahiro Komamizu (Nagoya University, Japan), and Ichiro Ide	
(Nagoya University, Japan)	

### Session 7: Face Analysis

ExpressionHop: A Lightweight Human Facial Expression Classifier	18
Evaluating Image Quality Estimators for Face Matching	4
Detection-by-Simulation: Exposing DeepFake via Simulating Forgery using Face Reconstruction	0
Facial Expression Recognition in the Wild: Dataset Configurations	6
A Local-Global Metric Learning Method for Facial Expression Animation	:0
Open-Eye: An Open Platform to Study Human Performance on Identifying Al-Synthesized Faces . 224 Hui Guo (University at Albany, USA), Shu Hu (University at Buffalo, USA), Xin Wang (University at Buffalo, USA), Ming-Ching Chang (University at Albany, USA), and Siwei Lyu (University at Buffalo, USA)	

#### Session 8: Person, Pose, Gait, VR

Swin-Pose: Swin Transformer Based Human Pose Estimation Zinan Xiong (University of Massachusetts Lowell, USA), Chenxi Wang (University of Massachusetts Lowell, USA), Ying Li (Wuhan, China), Yan Luo (University of Massachusetts Lowell, USA), and Yu Cao (University of Massachusetts Lowell, USA)	228
Wasserstein Metric Attack on Person Re-identification Astha Verma (IIIT Delhi, India), A.V. Subramanyam (IIIT Delhi, India), and Rajiv Ratn Shah (IIIT Delhi, India)	234

Intuitive Gait Modeling using Mimetic-Words for Gait Description and Generation
Comparison of Virtual-Real Integration Efficiency Between Light Field and Conventional Near-Eye AR Displays
Contextual Active Learning for Person Re-Identification
Personalized Prediction of Indoor Comfort using Graph Convolutional Matrix Completion 258 Junyi Liu (University at Buffalo), Esha Naidu (University at Buffalo), Jialian Wu (University at Buffalo), Shira Gabriel (University at Buffalo), Edward Steinfeld (University at Buffalo), and Junsong Yuan (University at Buffalo)

### Session 9: New Applications

Image Based Food Energy Estimation with Depth Domain Adaptation
Bridging Script and Animation Utilizing a New Automatic Cinematography Model
Rate-Adaptive Streaming of 360-Degree Videos with Head-Motion-Aware Viewport Margins 274 Mehmet N. Akcay (Ozyegin University, Turkey), Burak Kara (Ozyegin University, Turkey), Ali C. Begen (Ozyegin University, Turkey), Saba Ahsan (Nokia, Finland), Igor D.D. Curcio (Nokia, Finland), and Emre Aksu (Nokia, Finland)
Ultra-Sparse 360-Degree Camera View Synthesis for Immersive Virtual Tourism
Multi-source Weak Supervision Fusion for Disaster Scene Recognition in Videos
A Review of Personalized Health Navigation for Drivers

#### Session 10: Security, Privacy

License Plate Privacy in Collaborative Visual Analysis of Traffic Scenes	0
Malware and Piracy Detection in Android Applications	6
A Systematic Review of Multimodal Approaches to Online Misinformation Detection	2
<ul> <li>FedTrust: Towards Building Secure Robust and Trustworthy Moderators for Federated Learning.</li> <li>318</li> <li>Chih-Fan Hsu (Inventec Corporation, Taiwan), Jing-Lun Huang (Inventec Corporation, Taiwan), Feng-Hao Liu (Florida Atlantic University, USA), Ming-Ching Chang (University at Albany, State University of New York,</li> </ul>	
USA), and Wei-Chao Chen (Inventec Corporation, Taiwan) An Overview of Recent Work in Multimedia Forensics	4
Automated Adaptive Playback for Encoder-Adjudicated Live Sports	0

#### **Session 11: Image Retrieval**

A Noise-Robust Locality Transformer for Fine-Grained Food Image Retrieval
Modeling Fake News Detection using BERT-CNN-BiLSTM Architecture
A Highly Optimized GPU Batched Elastic-net Solver (BENS) with Application to Real-Time Keypoint Detection for Image Retrieval
Personalized Fashion Sequential Recommendation with Visual Feature Based on Conditional Hierarchical VAE

### Session 12: Action, 3D LiDAR Processing

Adaptive Acquisition of Airborne LiDAR Point Cloud Based on Deep Reinforcement Learning 3 Chengxuan Huang (University of California, USA), Dalei Wu (University of Tennessee at Chattanooga, USA), and Yu Liang (University of Tennessee at Chattanooga, USA)	366
Multiscale Point Feature Object Localization for Hydrant Surveying using LiDAR Gabriel Lugo (University of Alberta, Canada), Amit Upreti (University of Alberta, Canada), and Irene Cheng (University of Alberta, Canada)	372
Generic Action Start Detection	379
Weakly Supervised Temporal Action Localization Through Contrastive Learning	383
Cross-Domain Knowledge Transfer for Skeleton-Based Action Recognition Based on Graph Convolutional Gradient Reversal Layer	387

BAPO: A Large-Scale Multimodal Corpus for Ball Possession Prediction in American Football

# Workshop - MAWA: Multimedia for Agriculture and Water Applications

Cloud Removal in Satellite Imagery using Adversarial Network and RGB-Optical Data Fusion .. 407 Sushil Ghildiyal (Indian Institute of Technology Ropar, India), Neeraj Goel (Indian Institute of Technology Ropar, India), and Mukesh Saini (Indian Institute of Technology Ropar, India)

#### Workshop - MCAUIS: Multimedia Computing for Automated Urban Intelligent Systems

Automatic Evaluation of Machine Generated Feedback For Text and Image Data Pratham Goyal (Indian Institute of Technology, India), Anjali Raj (Indian Institute of Technology, India), Puneet Kumar (Indian Institute of Technology, India), and Kishore Babu Nampalle (Indian Institute of Technology, India)	413
Fine-Tuning the Robust Temporal Feature Magnitude Model for Enhancing the Accuracy of Anomaly Detection Charan Charupalli (National Institute of Technology, Andhra Pradesh) and Karthick Seshadri (National Institute of Technology, Andhra Pradesh)	419
An Efficient Multi-functional Deep Learning Model for Effective Medical Image Classification using skin Lesion Database <i>Kishore Babu Nampalle (Indian Institute of Technology Roorkee, India)</i> <i>and Balasubramanian Raman (Indian Institute of Technology Roorkee, India)</i>	425