

2024 International Conference on Content-Based Multimedia Indexing (CBMI 2024)

**Reykjavik, Iceland
18-20 September 2024**



**IEEE Catalog Number: CFP2414C-POD
ISBN: 979-8-3503-7845-0**

**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:	CFP2414C-POD
ISBN (Print-On-Demand):	979-8-3503-7845-0
ISBN (Online):	979-8-3503-7844-3

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

Contents

Best Paper Candidates	1
Is CLIP the main roadblock for fine-grained open-world perception? . . .	1
<i>Lorenzo Bianchi, Fabio Carrara, Nicola Messina, Fabrizio Falchi</i>	
Evaluation of Deep Audio Representations for Semantic Sound Similarity	9
<i>Recep Oguz Araz, Dmitry Bogdanov, Pablo Alonso-Jiménez, Frederic Font</i>	
Elevating Video Retrieval Capabilities: A Cross-Modal Approach Utiliz- ing Text and Image Generative Models	16
<i>Kazuya Ueki, Yuma Suzuki, Haruki Sato, Takayuki Hori, Hiroki Takushima, Takumi Takada, Hayato Tanoue, Aiswariya Manoj Kumar, Hiroki Nishihara</i>	
Leveraging Query Expansion and Reformulation for Image Retrieval with Large Language and Vision-Language Models	23
<i>Sandrina Frunzä, Stevan Rudinac, Cees Diks</i>	
Research Papers	30
Predicting 3D Projectile Motion in Table Tennis Using Computer Vision and Physics-Informed Neural Network	30
<i>Zaïneb Chiha, Renaud Péteri, Laurent Mascarilla</i>	
A Comparison of Late-fusion Training Strategies for Quad-modal Joint Embeddings	37
<i>Domenic Furer, Abraham Bernstein, Luca Rossetto</i>	
A Hybrid AI system for fusion of object and context information: appli- cation to the rail line defect detection	44
<i>Alexey Zhukov, Jenny Benois-Pineau, Alain Rivero, Akka Zemmari, Mohamed Mosbah, Danilo Crispiani</i>	
Query Refinement for Non-Existent Items in Image Retrieval	51
<i>Naoto Naka, Shin'ichi Satoh</i>	
Leveraging Latent Diffusion Models for Training-Free In-Distribution Data Augmentation for Surface Defect Detection	58
<i>Federico Girella, Ziyue Liu, Franco Fummi, Francesco Setti, Marco Cristani, Luigi Capogrosso</i>	
PGNN-based Approach for Robust 3D Light Direction Estimation in Out- door Images	65
<i>Marcello Zanardelli, Mahyar Gohari, Riccardo Leonardi, Sergio Benini, Nicola Adami</i>	
IMSearch: An Interactive Multimedia Video-moment Search System . . .	72
<i>Duc-Tuan Luu, Duy-Ngoc Nguyen, Le-Khanh-Linh Bui, Vinh-Tiep Nguyen, Minh-Triet Tran</i>	
Coarse-to-Fine Pruning of Graph Convolutional Networks for Skeleton- based Recognition	79

Hichem Sahbi

SoccerRAG: Multimodal Soccer Information Retrieval via Natural Queries 86
Aleksander Theo Strand, Sushant Gautam, Cise Midoglu, Pål Halvorsen

MeshConv3D: Efficient convolution and pooling operators for triangular
 3D meshes 93
Germain Bregeon, Preda Marius, Radu Ispas, Zaharia Titus

WSESeg: Introducing a Dataset for the Segmentation of Winter Sports
 Equipment with a Baseline for Interactive Segmentation 99
Robin Schön, Daniel Kienzle, Rainer Lienhart

**AIMHDA: Advances in AI-Driven Medical and Health Data
 Analysis 106**

Combining Image and Region Uncertainty-based Active learning for Melanoma
 Segmentation 106
Nicolas Martin, Jean-Pierre Chevallet, Philippe Mulhem, Georges Quénot

SAM in the Pipeline: Transforming Axis-Aligned to Oriented Bounding
 Boxes for Superior Sperm Detection 113
*Pål Andreas Hoven Bentsen, Steven Hicks, Eric Jul, Pål Halvorsen, Vajira
 Thambawita*

A Multi-Instance Learning Approach for Improving Knee Osteoarthritis
 Diagnosis from MRI Data 120
*Mohamed Berrimi, Teoh Yun Xin, Aladine Chetouani, Lotfi Houam, Rachid
 Jennne*

**CB4AMAS: Content-Based Indexing for Audio and Music: From
 Analysis to Synthesis 126**

EMVD dataset: a dataset of extreme vocal distortion techniques used in
 heavy metal 126
*Modan Tailleur, Julien Pinquier, Laurent Millot, Corsin Vogel, Mathieu La-
 grange*

A quest through interconnected datasets: lessons from highly-cited ICASSP
 papers 131
Cynthia C. S. Liem, Doğa Taşçılar, Andrew M. Demetriou

Modeling Musical Knowledge with Quantum Bayesian Networks 139
Florian Krebs, Hermann Fürntratt, Roland Unterberger, Franz Graf

Invariant Audio Prints for Music Indexing and Alignment 145
Rémi Mignot, Geoffroy Peeters

Latent Space Exploration for Drum Sample Generation 152
Jake Drysdale, Jason Hockman

Towards training music taggers on synthetic data 159
Nadine Kroher, Steven Manangu, Aggelos Pikrakis

ExMA: Explainability in Multimedia Analysis	165
Exploring the Plausibility of Hate and Counter Speech Detectors with Explainable AI	165
<i>Adrian Jaques Böck, Djordje Slijepčević, Matthias Zeppelzauer</i>	
XAIface: A framework and toolkit for explainable face recognition	173
<i>Nelida Mirabet-Herranz, Martin Winter, Yuhang Lu, Naima Bousnina, Jonas Pfister, Chiara Galdi, Jean-Luc Dugelay, Werner Bailer, Touradj Ebrahimi, Paulo Lobato Correia, Fernando Pereira, Felix Schmautzer, Erich Schweighofer</i>	
TaxonoMap: An Interactive System for the Exploration and Explanation of Unsupervised Large-Scale News Classification	180
<i>Simon Ott, Daria Liakhovets, Mina Schütz, Medina Andresel, Moritz W. Rothmund-Burgwall, Armin Vogl, Heidi Scheichenbauer, Michael Suker, Alexander Schindler</i>	
Fine-grained rebalancing of datasets for correct demographic classification	184
<i>Andrea Bozzitelli, Pia Cavasinni di Benedetto, Maria De Marsico, Xing Di, Vishal M. Patel</i>	
IVR4B: Interactive Video Retrieval for Beginners	191
diveXplore at IVR4B 2024	191
<i>Mario Leopold, Klaus Schoeffmann</i>	
VISIONE 5.0: toward evaluation with novice users	198
<i>Giuseppe Amato, Paolo Bolettieri, Fabio Carrara, Fabrizio Falchi, Claudio Genaro, Nicola Messina, Lucia Vadicano, Claudio Vairo</i>	
Exquisitor: Studying the Interplay Between Conversational Search and Relevance Feedback	204
<i>Omar Shahbaz Khan, Ujjwal Sharma, Stevan Rudinac, Björn Þór Jónsson</i>	
VidBasys: A User-friendly Interactive Video Retrieval System for Novice Users in IVR4B	209
<i>Thao-Nhu Nguyen, Quang-Linh Tran, Hoang-Bao Le, Binh T. Nguyen, Liting Zhou, Gareth J. F. Jones, Cathal Gurrin</i>	
Expressive Multimedia Query Formulation for Novices in Virtual Reality with vitrivr-VR	214
<i>Florian Spiess, Heiko Schuldt</i>	
VERGE: Simplifying Video Search for Novice Users	218
<i>Nick Pantelidis, Maria Pegia, Damianos Galanopoulos, Konstantinos Apostolidis, Dimitris Georgalis, Klearchos Stavrothanasopoulos, Anastasia Moumtzidou, Konstantinos Gkoutakos, Ilias Gialampoukidis, Stefanos Vrochidis, Vasileios Mezaris, Ioannis Kompatsiaris</i>	
Improving the Flexibility of Video Events Retrieval through Dynamic Conditional Refinement with Multilingual Capabilities	224
<i>Thang-Long Nguyen-Ho, Tu Ninh-Van, Minh-Triet Tran, Graham Healy, Cathal Gurrin</i>	

MAS4DT: Multimedia Analysis and Wimulations for Digital Twins in the Construction Domain	228
A Framework for Vision-Based 3D Inspections for Maintenance Activities and Digital Twin Integration	228
<i>Panagiotis Vrachnos, Carlos Ramonell, Ilias Koulalis, Konstantinos Ioannidis, Irina Stipanovic, Stefanos Vrochidis</i>	
Enhanced Defect Detection in Airport Runway Infrastructure Using Image-Text Pairing	235
<i>Marios Krestenitis, Eftichia Badeka, Ilias Koulalis, Konstantinos Ioannidis, Stefanos Vrochidis</i>	
MIDRA: Multimodal Insights for Disaster Risk Management and Applications	242
Towards Advanced Wildfire Analysis: A Siamese Network-Based Change Detection Approach through Self-Supervised Learning	242
<i>Dimitrios Valsamis, Alexandros Oikonomidis, Chrysoula Chatzichristaki, Anastasia Moumtzidou, Ilias Gialampoukidis, Stefanos Vrochidis, Ioannis Kompatsiaris</i>	
Fire Detection for Emergency Responders using X	249
<i>Dimitrios Stefanopoulos, Aristeidis Bozas, Georgia Christodoulou, Maria Igarieva Maslioukova, Yiannis Kouloglou, Maria Pegia, Anastasia Moumtzidou, Ilias Gialampoukidis, Konstantinos Avgerinakis, Stefanos Vrochidis, Ioannis Kompatsiaris</i>	
MmIXR: Multimedia Indexing for XR	256
Descriptor Impact on Multimodal 3D Retrieval	256
<i>Maria-Eirini Pegia, Björn Þór Jónsson, Anastasia Moumtzidou, Sotiris Diplaris, Ilias Gialampoukidis, Stefanos Vrochidis, Ioannis Kompatsiaris</i>	
Learning Scene Semantics from Vehicle-centric Data for City-scale Digital Twins	263
<i>Hermann Fürntratt, Stefanie Onori-Wechtitsch, Werner Bailer, Isaac Agustí Ventura, Carles Sala Navarro, Aleksandar Jevtić, Jawad Haidar</i>	
Finding video shots for immersive journalism through text-to-video search	269
<i>Lyndon Nixon, Damianos Galanopoulos, Vasileios Mezaris</i>	
Data-efficient Domain Transfer for Instance Segmentations for AR Scenes	275
<i>Stefanie Onori-Wechtitsch, Hermann Fürntratt, Hannes Fassold, Werner Bailer</i>	
UHBer: Multimodal Data Analysis for Understanding of Human Behaviour, Emotions and their Reasons	282
Weakly Supervised Autism Severity Assessment in Long Videos	282
<i>Abid Ali, Mahmoud Ali, Camilla Barbini, Séverine Dubuisson, Jean-Marc Odobez, Francois Bremond, Susanne Thümmel</i>	

A Concept Design for a Positive Mood Supporting Application	289
<i>Aurora Saibene, Riccardo Giussani, Claudia Rabaioli, Nicolò Dozio, Francesca Gasparini, Francesco Ferrise</i>	
A survey on Graph Deep Representation Learning for Facial Expression Recognition	296
<i>Théo Gueuret, Akrem Sellami, Chaabane Djeraba</i>	
A behavior and emotion recognition framework for emotion-aware services in physical spaces	303
<i>Sari Järvinen, Johanna Kallio, Johannes Peltola, Satu-Marja Mäkelä</i>	
From Controlled to Chaotic: Disparities in Laboratory vs Real-World Stress Detection	309
<i>Simao Ferreira, Fatima Rodrigues, Johanna Kallio, Filipe Coelho, Vesa Kyllonen, Nuno Rocha, Matilde A. Rodrigues, Elena Vildjiounaite</i>	
Predicting Multiple Reading Tasks Using Eye Movement Measures	316
<i>Onanong Kongmeesub, Cathal Gurrin, Prapaporn Rattanatamrong</i>	
HRV stress detection based on neural networks approach	323
<i>Salome Quoy, Dan Istrate, Mouna Bencheikroun, Vincent Zalc</i>	
Motion consistency constraint map for facial expression spotting	328
<i>Ouala Benjemaa, Amel Aissaoui, Benjamin Allaert, Ioan Marius Bilasco</i>	
Demonstrations	335
Music Scope Pad: Video Selecting Application by Natural Movement in VR Space	335
<i>Masatoshi Hamanaka</i>	
Video Shot Discovery through Text2Video Embeddings in a News Analytics Dashboard	339
<i>Lyndon Nixon, Damianos Galanopoulos, Vasileios Mezaris, Alexander Hubmann-Haidvogel, Daniel Fischl, Arno Scharl</i>	
Demo: Creating Player-Specific Soccer Highlight Clips with PlayerTV . .	344
<i>Håkon Maric Solberg, Mehdi Houshmand Sarkhoosh, Sushant Gautam, Saeed Shafiee Sabet, Pål Halvorsen, Cise Midoglu</i>	
Toolchain for Comprehensive Audio/Video Analysis Using Deep Learning Based Multimodal Approach: Use Case of Riot or Violent Context Detection	349
<i>Lam Pham, Tin Nguyen, Phat Lam, Hieu Tang, Alexander Schindler</i>	
Real-time Musical Collaboration with a Probabilistic Model	353
<i>Karl Johann Johannsson, Victor Shepardson, Enrique Hurtado, Thor Magnusson, Hannes Vilhjálmsson</i>	
Lowering Barriers to Entry for Fully-Integrated Custom Payloads on a DJI Matrice	357
<i>Joshua David Springer, Gylfi Þór Guðmundsson, Marcel Kyas</i>	

Demo: Soccer Information Retrieval via Natural Queries using SoccerRAG	362
<i>Aleksander Theo Strand, Sushant Gautam, Cise Midoglu, Pål Halvorsen</i>	
Enabling Domain Experts to Train Efficient Few-Shot Incremental Land-	
mark Recognition	367
<i>Helmut Neuschmied, Werner Bailer</i>	