# The Fifth IEEE International Image Processing, Applications and Systems Conference (IPAS'22)

**Organized by:**

**The National Engineering School of Sfax, Sfax University, Tunisia**
**Genoa University, Italy**
**INRIA, Sophia Antipolis, France**

**Hosted by:**

**Genoa University, Italy**
**5-7 December 2022**

***The Fifth IEEE IPAS is technically sponsored by:***

**IEEE Tunisia Section**
**IEEE Italy Section**
**Region 08 - Europe, Middle East, Africa**
**IEEE Italy Section Geoscience and Remote Sensing Society Chapter**
**IEEE Tunisia Section Geoscience and Remote Sensing Society Chapter**
**IEEE Italy Section Computer Chapter**
**IEEE Italy Section Engineering in Medicine and Biology Chapter**
**IEEE Tunisia Section Engineering in Medicine and Biology Chapter**
**IPASS'22 is financially sponsored by:**
**IEEE Tunisia Section**
**Association Boussole de la Recherche Scientifique - ABORS Tunisia**

# Preface

IPAS 2022 will take place in Genova, Italy during December 5-7, 2022. For more convenience and safety to some authors, the conference is managed in hybrid mode. Most of the presentations are on in-person mode. Nevertheless, some presentations are on-line and they are managed in the same days of the conference so that authors attending IPAS'22 in Genova can follow them and can interact with authors presenting on-line.
Image Processing Application in modern Systems makes them smart, interactive integrating features going sometimes beyond human intelligence. The international Image Processing Applications and Systems conference aims at grouping from all over the world challenging researchers, innovators, academicians, and practitioners in image processing theory and tools, for exchanging their research achievements and discussing their main issues. The conference will also offer an opportunity to make image processing researchers and industrial parts collaborate together for innovation.

The conference is devoted to image processing, computer vision algorithms and applications. It includes high quality original papers in all areas of computer vision applications.

The $5^{th}$ IEEE International Conference of Image Processing Applications and Systems has 11 regular sessions composed of about 60 participations including about 65 regular papers from 32 countries ( Austria, Arab Emirates, Belgium, Bangladesh, Brazil, Canada, China, Cyprus, Croatia, Denmark, Egypt, Japan, Korea of the south, France, Germany, Greece, India, Italy, Iran, Ireland, Macedonia, Mauritius, New-Zealand, Nigeria, Pakistan, Portugal, Romania, Spain, Tunisia, Turkey, United Kingdom, United States of America). The review process has been double blinded. On behalf the program committee of IPAS'22, we are grateful to all the authors for their fine contributions and for the high quality of their papers.

The $5^{th}$ IEEE International Conference of Image Processing Applications and Systems has five invited keynote speakers on different topics of image processing applications. On behalf the organizing committee of IPAS'22, we would like to thank all the speakers for their support and their valuable contributions to the success of the conference.

*General Co-Chairs*
*Fabio Solari*
*François Brémond*
*Dorra Sellami*

# Contents

# Committees

## General Chairs

Fabio Solari ( Genoa University, Italy)
François Brémond (Inria Sophia Antipolis - France)
Dorra Sellami ( University of Sfax, Ecole Nationale d'Ingénieurs de Sfax, Tunisia)

## Technical Program Chairs

Andrea Trucco (Genoa, Italy)
Sébastien Ambellouis (Lille, France)
Riadh Abdelfattah (Tunis, Tunisia)
Ali Wali (Sfax, Tunisia)
Alima Damak (FSS, Tunisia)
Giovanni Maria Farinella (Catania, Italy)
Jacques Boonaert (IMT-Lille-Douai, France)
Cyril Meurie (IFSTTAR, France)
Nicolas Dobigeon (Toulouse University, France)
Petra Perner ( Ibai, Germany)
Nicolas Gillis (Mons University, Belgium)
Cecilia Zanni Merk( Rouen University, France)
Stella Mark Swecker (Strasbourg University, France)
Chao-Cheng Wu (Taipei, Taiwan)
Pau-Choo Chung (Tainan, Taiwan)

## Plenary, tutorial and special session Chairs

Sergio Cerutti (Politecnico di Milano, Italy)
Anthony Fleury (IMT-Lille-Douai, France)
Cédric Wemmert (Strasbourg, France)
Hassan Rabah (Lorraine, France)
Abdesslam Benzinou (ENIB, France)
Mehrez Abdellaoui (ISATK, Tunisia)
Hela Boulehmi Chtara (ISET, Bizerte)

## Publication Chairs

Habib Kamoun (FSS, Tunisia)
Marco Domenico Santambrogio ( Italy)
Amir Gargouri (ENET-COM, Tunisia)

## Exhibits and Industry Liaison Chairs

Mohamed Louati (Logidas, Tunisia)
Anis Benabdennebi (GFI, Tunisia)

## Local Arrangement Chairs

Marianna Pizzo (Genoa, Italy)
Razeen Hussain (Genoa, Italy)
Jihen Frikha ( ENIS, Tunisia)
Majd Belaaj ( ENIS, Tunisia)

## Organizing Chairs

Andrea Trucco ( Genoa, Italy)
Mouna Zouari ( CRNS, Tunisia)
Norhene Gargouri ( CRNS, Tunisia)
Mondher Frikha (ENET-COM, Tunisia)

## Plenary, tutorial and special session committee members

Nawres Khlifa (Tunisia)
Mounir Sayadi ( ENSIT, Tunisia)
Aicha Bouzid (ENIT, Tunisia)
Ali Douik (ENIM, Tunisia)
Mohamed Atri ( ENIM, Tunisia)
Mohamed Ali Mahjoub (ENISO, Tunisia)
Zied Lachiri ( ENIT, Tunisia)
Hassene Seddik ( ENSIT, Tunisia)
Monia Turki (ENIT, Tunisia)
Kamel Hamrouni (ESPRIT, Tunisia)
Sabeur Abid (ENSIT, Tunisia)

## Registration Chairs

Mohamed Krid (ISGI, Tunisia)
Hejer Elloumi (PRIMATEC, Tunisia) Wiem Abbes (ENIS, Tunisia)

## Publicity Chairs

Manuela Chessa (Genoa, Italy)
Randa Boukhris (ISIMS, Tunisia)
Kamel Abderrahim ( Gabes, Tunisia)

## Organizing Committee members

Andrea Trucco (University of Genoa, Italy)
Fabio Solari (University of Genoa, Italy)
Mouna Zouari (ENIS, Tunisia)
Manel Loumi (ENIS, Tunisia)
Jihène Frikha (ENIS, Tunisia)
Wiem Abbes (ENIS,Tunisia)
Hejer Elloumi (ENIS,Tunisia)

# Technical Program Committee Members

Daniele Caviglia (ITL)
Lorenzo Natale (ITL)
Manuela Chessa (ITL)

Fiora Pirri (ITL)
Gobron Stéphane (SWZ)
Francisco Barranco(ES)
Hela Boulemi Chtara (TN)
Tièche François(CH)
Antonio Rodríguez-Sanchez(AUS)
Alaa Balti (TN)
Nicolas Thome (FR)
Jean-Marc Odobez (SWZ)
Ludovic Macaire (FR)
Pascal Vasseur (FR)
Serge Miguet (FR)
James Ferryman (UK)
Andreas Koschan (USA)
Alexander Loui (CA)
Andrea Cavallaro (UK)
Anoop Cherian (AUS)
Bogdan Raducanu (ES)
Simone Calderara (ITL)
Carlo Regazzonui (ITL)
Li Cheng (CHN)
Christian Micheloni (ITL)
Curtis Padgett ( USA)
Shiloh Dockstader (USA)
Riad Hammoud (UK)
Hazim Ekenel (SWZ)
Federico Tombari (DE)

Mahesh Gellaboina (India)
Guillaume Ginolhac (FR)
Pierre-Antoine Thouvenin (UK)
Emilie Chouzenoux (FR)
Cédric Demonceaux (FR)

Chiara Bartolozzi(ITL)
Maurizio Valle(ITL)
Giovanni Maria Farinella(ITL)
Dimitri Ognibene(UK)
Norbert Kruger(DK)
Eris Chinellato(UK)
Hanene Sehli (TN)
Carlos Fernando (FR)
Javier Diaz(ES)
Sabeur Abid (TN)
Dominique gruyer (FR)
Sergio Velastin (ES)
Vincent Fremont (FR)
Olivier Losson (FR)
Remy Boutteau (FR)
Juan Carlos Niebles ( USA)
François Fleuret (FR)
Aydin Alatan (UK)
Andrea Prati (ITL)
Andreas Hutter (DE)
Atsushi Shimada (JPN)
Boxin Shi (CHN)
Caifeng Shan (UK)
Chan Fang (CHN)
Anustup Choudhury (USA)
Cong Phuoc Huynh (AUS)
Dimitrios Tzovaras ( GR)
Donatello Conte (USA)
Ehsan ZareBorzeshi ( KOR)
FahimehRezazadegan (KOR)
Frank Pagel (DE)
Yoann Altmann (UK)
Marcelo Pereyra (UK)
Mauro Dalla-Mura (FR)
Emmanuel Soubies (SWZ)

| | |
|---|---|
| Paul Honeine (FR) | Olivier Lezoray (FR) |
| Adrian Basarab (FR) | Jérôme Bobin (FR) |
| Masoud faraki(AUS) | Massimo Piccardi(SYD) |
| Kamel Hamrouni (TN) | Kamel Abderrahim (TN) |
| Mohamed Akil (FR) | Taher Battikh(TN) |
| Amel Benazza (TN) | Mohamed Hédi Bedoui(TN) |
| Faouzi Benzarti (TN) | Dhibi(TN) |
| Afef Elloumi(TN) | Mourad Zaeid(TN) |
| Kais Ouni (TN) | Issam Jabri (TN) |
| Mohamed Atri (TN) | Anis Kricha (TN) |
| Sakly Anis (TN) | Kouraichi Maher (TN) |
| Mourad Moussa (TN) | Olfa Jemai (TN) |
| Salwa Said (TN) | Sofiane Hachicha (TN) |
| Marwa Hmida (TN) | Ines Rahmeny(TN) |

# Keynote I by Prof. François Brémond
# Session chair: Prof. Fabio Solari (Monday 5 Dec 2022 at 10:10-10:50)

[5 Dec 2022
10:10-10:50 ]

## Action Recognition for People Monitoring

5 Dec 2022
10:10-10:50

François Bremond
STARS – INRIA – Sophia Antipolis
Francois.Bremond@inria.fr

**Summary**

In this talk, we will discuss how Video Analytics can be applied to human monitoring using as input a video stream. Existing work has either focused on simple activities in real-life scenarios, or on the recognition of more complex (in terms of visual variabilities) activities in hand-clipped videos with well-defined temporal boundaries. We still lack methods that can retrieve multiple instances of complex human activity in a continuous video (untrimmed) flow of data in real-world settings. Therefore, we will first review few existing activity recognition/detection algorithms. Then, we will present several novel techniques for the recognition of ADLs (Activities of Daily Living) from 2D video cameras. We will illustrate the proposed activity monitoring approaches through several home care application datasets: Toyota SmartHome, NTU-RGB+D, Charades and Northwestern UCLA. We will end the talk by presenting some results on home care applications.

**François Bremond biography**

François Brémond is a Research Director at Inria Sophia Antipolis-Méditerranée, where he created the STARS team in 2012. He has pioneered the combination of

Artificial Intelligence, Machine Learning and Computer Vision for Video Understanding since 1993, both at Sophia-Antipolis and at USC (University of Southern California), LA. In 1997 he obtained his PhD degree in video understanding and pursued this work at USC on the interpretation of videos taken from UAV (Unmanned Airborne Vehicle). In 2000, recruited as a researcher at Inria, he modeled human behavior for Scene Understanding: perception, multi-sensor fusion, spatio-temporal reasoning and activity recognition. He is a co-founder of Keeneo, Ekinnox and Neosensys, three companies in intelligent video monitoring and business intelligence. He also co-founded the CoBTek team from Nice University in January 2012 with Prof. P. Robert from Nice Hospital on the study of behavioral disorders for older adults suffering from dementia. He is author or co-author of more than 250 scientific papers published in international journals or conferences in video understanding. He has (co)- supervised 20 PhD theses. More information is available at: http://www-sop.inria.fr/members/Francois.Bremond/

---

# Regular Session 1: Remote sensing and smart sensors for earth preserval
# Session chair: Prof. Manuela Chessa
# (Monday 5 Dec 2022 at 10:50-12:20 )

## Glacier-surface velocities in Gangotri from Landsat8 satellite imagery

5 Dec 2022
10:50-12:20
session 1

Reem Tawfiq Klaib[1], Hajer Alabdoulli [2] Mritunjay K Singh [3]

[1]: reemtk24@gmail.com, [2]: hajeralabdouli@hotmail.com, [3]: mritunjaysingh@uaeu.ac.ae

National Space Science and Technology Center, United Arab Emirates University

**Summary**

A glacier's mass balance and dynamics are regulated by changes in ice velocity. As a result, estimating glacier flow velocity is a crucial part of temporal glacier monitoring of its health, response to climate change parameters, and its effect on increasing the sea level rise. In this study, we estimated the Gangotri glacier surface velocities from 2014 to 2021. We used remote sensing-based techniques to estimate the Gangotri surface velocity since it provides such measurements regularly for a vast geographical area. Sub-pixel correlation of Landsat 8 imagery was used by using the COSI- Corr (co-registration of optically sensed images and correlation) tool to determine surface velocities over the Gangotri glacier. Our derived velocities values appear to match the ground truth velocities values comparatively well. Gangotri surface velocities vary over the various regions of the glacier, and from year to year. Our study indicated that the middle region of the ablation zone and the accumulation zone had higher velocities across all the years, while the boundary regions of the glacier show lower speeds. The average velocities range varies from 13 m/year in the accumulation zone to 22m/year. For the ablation zone, the average velocities range

from 11m/year in the ablation zone to 18m/year. The whole average surface velocity showed a high decrement of 26 % from 2014 to 2021. The general surface velocities in Gangotri vary from 8 m/y to 61 m/y± 1.9 from 2014-to 2021.

---

5 Dec 2022
10:50-12:20
Session 1

# Innovative tools for investigation on flame dynamics by means of fast imaging

Antonio Ferrante[1], Giuseppe Molfetta[2]

[1]:antonio.ferrante@ccaspa.it, [2]:giuseppe.molfetta@ccaspa.it

Centro Combustione Ambiente Spa
Gioia del Colle, Italy

**Summary**

Any type of flame, due to its turbulent nature, is subject to instability phenomena that are often unwanted in technical devices, such as gas turbines but also in aeronautical and aerospace engines. These phenomena generate fluctuations in heat release and pressure fluctuations inside the combustion chambers; these pressure fluctuations, in extreme cases, can irreparably damage the combustion chambers or other parts of the engines. In order to investigate the underlying mechanisms that drive flame-induced oscillation phenomena found in modern gas turbines, a methodology able to spatially localize such oscillations has been developed. This methodology is based on the use of fast cameras, computer vision and image post processing techniques and provides a two-dimensional map of the flame points where flame dynamics phenomena are formed, if present. One of the strengths of this methodology is its simplicity in the execution of measurements which allows it to be applied not only in the laboratory but also on full-scale test rigs and, with appropriate modifications, on real gas turbines. As a demonstration, we show the results obtained by applying this method to a full scale gas turbine burner operated at atmospheric pressure in a test rig.. Images sequences have been recorded at 1000 fps and each set of sequences is made up of 500 frames (Figure 1b); each frame consists of 388 x 288 pixels. Through an optimized algorithm 111,744 FFTs have been calculated from the sequence of 500 frames. Under certain conditions of flame stoichiometry and inflow speed, the flame oscillates simultaneously in several modes, as shown in Figure 1c. Moreover these oscillating modes persist over time and they can be considered coherent structures.

---

## Domestic Solid Waste Classification

5 Dec 2022
10:50-12:20
Session 1

Surajsingh Dookhee
sookdook11@gmail.com

University of Technology, Mauritius

### Summary

The overwhelming amount of household solid waste generated daily is alarming, and this contributes to the rise in pollution and drastic climate change. In such a context, automated waste classification at the initial stage of disposal can be an effective solution to separate recyclable items. Convolutional Neural Networks based on deep learning are used for waste classification, but however, research works are often limited to TrashNet dataset consisting of 2527 images and 6 categories of waste. This dataset lacks other important varieties of household waste such as battery, biological, and clothing items to reflect real-life environmental problems. In this paper, a large garbage dataset consisting of 15,515 images and 12 categories of common household items was used to compare the performance of DenseNet121, DenseNet169, EfficientNetB0, InceptionV3, MobileNetV2, ResNet50, VGG16, VGG19, and Xception CNN models. Data augmentation was applied to solve the problem of class imbalance, and findings of my first research showed that the Xception model compiled with Adam optimiser outperformed all the other models with an accuracy of 88.77% and an F1-score of 0.89. The performance of the model was improved to 89.57% with an F1-score of 0.90 when compiled with Nadam optimiser. However, further experimentation without data augmentation showed that the model did not generalise well despite reaching an accuracy of 93.42% and an F1-score of 0.93. This demonstrates the feasibility of the proposed model for real-life environmental problems.

---

## Comparative studies on similarity distances for remote sensing image classification

5 Dec 2022
10:50-12:20
Session 1

Omid Ghozatlou[1], Mihai Datcu[2]
[1]: oqozat@gmail.com, omid.ghozat@gmail.com

[1,2]: Research Center for Spatial Information, Germany,
[1]: University Politechnica of Bucharest, Romania

### Summary

Scene classification is one of the most important tasks in the remote sensing field. In general, remotely sensed data comprises targets of different natures with many detailed classes. Therefore, the classification of patches in a satellite scene is a challenging issue. To address the problem, the preferred alternative is to transform to polar coordinates and analyze angular distances. Prior works have so far considered angular distances between points while ignoring that the target class is not a point,

but a distribution. In this paper, we take advantage of this critical fact by using a point-to-probability distribution measure rather than an L norm. In this paper, two similarity measures (Euclidean and Mahalanobis) in two different feature spaces are experimentally investigated through some remote sensing datasets.

---

5 Dec 2022
10:50-12:20
Session 1

## Experimental Investigation of Non-contact 3D Sensors for Marine-growth Cleaning Operations

Christian Mai[1], Jesper Liniger [2], Anna Lyhne Jensen [3], Henrik Sørensen[4], Simon Pedersen[5]

[1]: chrimai@energy.aau.dk, [2]: jel@energy.aau.dk, [3]: alj@energy.aau.dk, [4]: hs@energy.aau.dk, [5]: spe@energy.aau.dk

Aalborg University, Denmark

**Summary**

Marine growth on submerged structures causes additional mechanical loads from drag and mass increases. In order to ensure structural integrity, regular inspection and cleaning procedures are carried out on the submerged structures, most commonly using remote-operated vehicles (ROVs). Often, the measurement methodology in these inspections is spot-checks using simple mechanical gauges, which yield a rough estimate of marine growth thickness. Expanding on this method, in order to optimize these inspection and cleaning procedures, modern methods for 3D surface measurement can be applied to increase inspection quality and ensure that superfluous cleaning is not carried out. This work investigates three state-of-the-art sensor technologies: a Time-of-Flight depth camera based on modulated visible blue laser illumination, a commercial stereo-vision solution based on visible-light sensors, and high-frequency imaging sonar. The sensors' performance has been compared in a laboratory environment to assess their suitability for use as a measurement device for marine-growth measurement in terms of accuracy, resolution, and noise/artifacts. It is concluded that the measurement fidelity of all evaluated sensors shows promise for the application, pending future evaluation in a real-world test.

---

5 Dec 2022
10:50-12:20
Session 1

## Likelihood ratio map for direct exoplanet detection

Hazan Daglayan[1], Simon Vary [2], Faustine Cantalloube [3], P.-A. Absil [4], Olivier Absil [5],

hazan.daglayan@uclouvain.be[1], simon.vary@uclouvain.be [2], faustine.cantalloube@lam.fr [3], pa.absil@uclouvain.be [4], olivier.absil@uliege.be [5]

[1,2,4]: UCLouvain, Belgium,
[3]: Aix Marseille University, France,
[5]: University of Liège, Belgium

**Summary**

Direct imaging of exoplanets is a challenging task due to the small angular distance and high contrast relative to their host star, and the presence of quasi-static noise. We propose a new statistical method for direct imaging of exoplanets based on a likelihood ratio detection map, which assumes that the noise after the background subtraction step obeys a Laplacian distribution. We compare the method with two detection approaches based on signal-to-noise ratio (SNR) map after performing the background subtraction by the widely used Annular Principal Component Analysis (AnnPCA). The experimental results on the Beta Pictoris data set show the method outperforms SNR maps in terms of achieving the highest true positive rate (TPR) at zero false positive rate (FPR).

---

# Keynote II by Prof. Nicolas Gillis Session chair: Prof. François Brémond (Monday 5 Dec 2022 at 12:20-13:00 )

## Nonnegative Matrix Factorization: Sparsity and Linear-Quadratic Model for Hyperspectral Image Unmixing

7 Dec 2022
14:00-14:40 pm

Nicolas Gillis

University de Mons, Belgium

**Summary**

Given a nonnegative matrix X and a factorization rank r, nonnegative matrix factorization (NMF) approximates the matrix X as the product of a nonnegative matrix W with r columns and a nonnegative matrix H with r rows such that WH approximates X as well as possible. NMF has become a standard linear dimensionality reduction technique. Although it has been extensively studied in the last 20 years, many questions remain open. In this talk, we address two such questions. The first one is about the uniqueness of NMF decompositions, also known as the identifiability, which is crucial in many applications. We provide a new model and algorithm based on sparsity assumptions that guarantee the uniqueness of the NMF decomposition. The second problem is the generalization of NMF to non-linear models. We consider the linear-quadratic NMF (LQ-NMF) model that adds as basis elements the component-wise product of the columns of W, that is, W(:,j).*W(:,k) for all j,k where .* is the component-wise product. We show that LQ-NMF can be solved in polynomial time, even in the presence of noise, under the separability assumption which requires the presence of the columns of W as columns of X. We illustrate these new results on the blind unmixing of hyperspectral images.

**Nicolas Gillis biography**

Nicolas Gillis is Professor with the Department of Mathematics and Operational Research, University of Mons, Belgium. He is the recipient of the Householder award 2014 and ERC starting grant in 2016. He is senior area editor for IEEEE Transactions on Signal Processing, and associate editor for the SIAM Journals on Matrix Analysis and Applications and on Mathematics of Data Science.

---

# Keynote III by Prof. Dimitri Ognibene
# Session chair: Prof. Dorra Sellami (Monday 5 Dec 2022 at 14:00-14:40 )

### Adaptive Vision for Human Robot Collaboration

5 Dec 2022
14:00-14:40

Dimitri Ognibene

University of Essex, Kings College London and Universita' Milano-Bicocca, Italy

**Summary**

Unstructured social environments, e.g. building sites, release an overwhelming amount of information yet behaviorally relevant variables may be not directly accessible. Currently proposed solutions for specific tasks, e.g. autonomous cars, usually employ over redundant, expensive, and computationally demanding sensory systems that attempt to cover the wide set of sensing conditions which the system may have to deal with. Adaptive control of the sensors and of the perception process input is a key solution found by nature to cope with such problems, as shown by the foveal anatomy of the eye and its high mobility and control accuracy. The design principles of systems that adaptively find and select relevant information are important for both Robotics and Cognitive Neuroscience. At the same time, collaborative robotics has recently progressed to human-robot interaction in real manufacturing. Measuring and modeling task specific gaze behaviours is mandatory to support smooth human robot interaction. Indeed, anticipatory control for human-in-the-loop architectures, which can enable robots to proactively collaborate with humans, heavily relies on observed gaze and actions patterns of their human partners. The talk will describe several systems employing adaptive vision to support robot behavior and their collaboration with humans.

**Dimitri Ognibene biography**

Dimitri Ognibene is Associate Professor of Human Technology Interaction at University of Milano-Bicocca, Italy. His main interest lies in understanding how social agents with limited sensory and computational resources adapt to complex and uncertain environments, how this can induce suboptimal behavior such as addiction or antisocial behaviors, and how this understanding may be applied to real life problems. To this end he develops both neural and Bayesian models and applies them both in physical, e.g. robots, and virtual, e.g. social media, settings. Before joining Milano Bicocca University, he was at the University of Essex as Lecturer in Computer Science and Artificial Intelligence from October 2017 having moved from University Pompeu Fabra (Barcelona, Spain) where he was a Marie Curie Actions COFUND fellow. Previously he developed algorithms for active vision in industrial robotic tasks as a Research Associate (RA) at Centre for Robotics Research, Kings' College London; He developed Bayesian methods and robotic models for attention in social and dynamic environments as an RA at the Personal Robotics Laboratory in Imperial College London. He studied the interaction between active vision and autonomous learning in neuro-robotic models as an RA at the Institute of Cognitive Science and Technologies of the Italian Research Council (ISTC CNR). He also collaborated with the Wellcome Trust Centre for Neuroimaging (UCL) to study how to model exploration in the active inference modelling paradigm. He has been Visiting Researcher at Bounded Resource Reasoning Laboratory in UMass and at University of Reykjavik (Iceland) exploring the symmetries between active sensor control and active computation or metareasoning. He obtained his PhD in Robotics in 2009 from University of Genoa with a thesis titled "Ecological Adaptive Perception from a Neuro-Robotic perspective: theory, architecture and experiments" and graduated in Information Engineering at the University of Palermo in 2004. He is handling editor of Cognitive Processing, review editor for Paladyn - The journal of Behavioral Robotics, Frontiers Bionics and Biomimetics, and Frontiers Computational Intelligence in Robotics, guest associate editor for Frontiers in Neurorobotics and Frontiers in Cognitive Neuroscience. He has been chair of the robotics area of several conferences and workshops.

---

# Regular Session 2: Video Processing and Security Session chair: Dr Razeen Hussain (Monday 5 Dec 2022 at 14:40-16:10)

## DWT Collusion Resistant Video Watermarking Using Tardos Family Codes

5 Dec 2022
14:40-16:10
Session 2

Abdul Rehman[1], Gaëtan Le Guelvouit[2], Jean Dion[3], Frédéric Guilloud[4], Matthieu Arzel[5]

[1]:abdul.rehman@b-com.com, [2]:gaetan.leguelvouit@b-com.com, [3]:jean.dion@b-com.com, [4]:frederic.guilloud@imt-atlantique.fr, [5]:matthieu.arzel@imt-atlantique.fr

[1,2,3]:IRT b-com Cesson-Sevigné, France
[4,5]:IMT Atlantique, Lab-STICC Brest, France

**Summary**

A fingerprinting process is an efficient means of protecting multimedia content and preventing illegal distribution. The goal is to find individuals who were engaged in the production and illicit distribution of a multimedia product. We investigated discrete wavelet transform (DWT) based blind video watermarking strategy tied with probabilistic fingerprinting codes to avoid collusion among higher-resolution videos. We used FFmpeg to run a variety of collusion attacks (e.g., averaging, darkening, and lighten) on high resolution video and compared the most often suggested code generator and decoders in the literature to find at least one colluder within the necessary code length. The Laarhoven codes generator and nearest neighbor search (NNS) decoder outperforms all other suggested generators. and decoders in the literature in terms of computational time, colluder detection and resources.

---

5 Dec 2022
14:40-16:10
Session 2

## Real or Fake? A Novel Method for Detecting Tempered Images

Ching-Yu Kao[1], Hongjia Wan[2], Karla Markert[3], Konstantin Böttinger[4]
[1]: ching-yu.kao@aisec.fraunhofer.de, [2]: hongjiawan@yahoo.com,
[3]: karla.markert@aisec.fraunhofer.de, [4]: konstantin.boettinger@aisec.fraunhofer.de

Fruanhofer AISEC, Germany

**Summary**

Tempering images has become a technology that almost everyone can complete, including fake news, fake evidence presented in court, or forged documents. The main reason is that these editing tools, such as Photoshop, are simple to use, which is an urgent issue we need to solve. Hence, automatic tools helping to find manipulated images apart are critical for fighting misinformation campaigns. Here we propose and evaluate a novel neural network-based method. It can detect whether images have been artificially modified (classification), and further indicate the forged parts (segmentation). Our proposed method has better performance than most baseline methods. Last but not least, our method is not only effective in JPEG format but can also be used in other formats.

---

5 Dec 2022
14:40-16:10
Session 2

## Hybrid Watermarking Algorithm to Protect and Authenticate KhalifaSat Imagery using DWT-SVD and SHA3 Hash Function

Alavikunhu Panthakkan[1]
[1]: apanthakkan@ud.ac.ae

University of Dubai, United Arab Emirates

**Summary**

Using DWT-SVD and SHA3 Hash function, this research aims to develop an ownership protection and image authentication technique that embeds the watermark information and hash authentication key in a hybrid domain. The experiment was conducted with multispectral images from the KhalifaSat. The Performance of the proposed method is evaluated using wavelet domain signal to noise ratio (WSNR), structural similarity index measurement (SSIM) and peak signal to noise ratio (PSNR). To analyse the efficacy of the recovered watermark, two metrics are used: Normalized Correlation (NC) and Image Quality Index (IQI). The method presented is robust against many intended and unintended attacks. Without sacrificing transparency, our proposed watermarking approach meets the objectives of imperceptibility and robustness. It accurately detects the manipulated locations on the satellite image and is sensitive to even small changes.

---

## Vision Transformer for Automatic Student Engagement Estimation

5 Dec 2022
14:40-16:10
Session 2

Sandeep Mandia[1], Kuldeep Singh[2], Rajendra Mitharwal[3]
[1]: 2018rec9129@mnit.ac.in, [2]: kuldeep.ece@mnit.ac.in, [3]: rajendra.ece@mnit.ac.in

Malaviya National Institute of Technology, Jaipur, India

### Summary

Availability of the internet and quality of content attracted more learners to online platforms stimulated by COVID-19. Learners of different cognitive capabilities join the learning process. However, it is challenging for the instructor to identify the level of comprehension of the individual learner, specifically when they waver in responding to feedback. The learner's facial expressions relate to content comprehension and engagement. This paper presents use of the vision transformer (ViT) for automatic estimation of student engagement with 32 and 64 input image patch sizes and various loss functions to handle class imbalance. The ViT was evaluated on Dataset for Affective States in E-Environments (DAiSEE); it outperformed frame level baseline result by approximately 8% and the other two video level benchmarks by 8.78% and 2.78% achieving an overall accuracy of 55.18%. In addition, ViT with focal loss was also able to produce well distribution among classes except for one minority class.

---

## AI based Automatic Vehicle Detection from Unmanned Aerial Vehicles (UAV) using hybrid YOLO Model

5 Dec 2022
14:40-16:10
Session 2

Alavikunhu Panthakkan[1]
[1]: apanthakkan@ud.ac.ae

University of Dubai, United Arab Emirates

### Summary

Unmanned aerial vehicle (UAV) detection of moving vehicles is becoming into a significant study area in traffic control, surveillance, and military applications. The challenge arises in keeping minimal computational complexity allowing the system to be real-time as well. Applications of vehicle detection from UAVs include traffic parameter estimation, violation detection, number plate reading, and parking lot monitoring. Tracking vehicle direction and speed include as an important aspect of surveillance. The one stage detection model, YOLOv5 is used in this research work to develop a deep neural model-based vehicle detection system on highways from UAVs. In our system, several improvised strategies are put forth that are appropriate for small vehicle recognition under an aerial view angle which can accomplish real-time detection and high accuracy by incorporating an optimal pooling approach

and dense topology method. Tilting the orientation of aerial photographs can improve the system's effectiveness. By comparing consecutive frames and identifying changes in vehicle location, the proposed model can also determine the speed and direction of the vehicles coming from the UAV. Metrics like hit rate, accuracy, and precision values are used to assess the performance of the proposed hybrid model, and performance is compared to that of other state-of-the-art algorithms.

---

5 Dec 2022
14:40-16:10
Session 2

## A Deep Reinforcement Learning based Recommender System Using Explicit Neural Collaborative Filtering

Maryam Bukhari[1], Muazzam Maqsood[2], Farhan Aadil[3]

[1]: maryambukhari09@gmail.com, [2]: muazzam.maqsood@ciit-attock.edu.pk,
[3]: Farhan.aadil@ciit-attock.edu.pk

COMSATS University Islamabad, Pakistan

**Summary**

Currently, the information explosion is continuously increasing over the internet, and there is a desperate need for such systems in which the most relevant information is chosen to be shown to different users based on their current interests. In the existing studies, several traditional and machine learning approaches are suggested to design the recommendation models. However, these approaches use a static view and hence dynamic sequential-decision-making feature of recommendation is overlooked. From this aspect, we proposed a deep reinforcement-learning-based recommender agent that can adjust itself depending on the most recent feedback in terms of ratings. However, in deep reinforcement learning models, state formation is critical. As a result, in this research study, states are formed by item embedding's of recently clicked items of users, which are learned explicitly through user and item interaction with neural collaborative filtering (NCF). Since implicit learning with NCF simply gives noisy indications regarding user preferences because it regards the least rated item of the user as a positive interaction. As a result, we view explicit ratings as the key label of each interaction of users with an item during learning. The explicitly learned item embedding's then employed as the state in the reinforcement learning model. Furthermore, the proposed model is validated on the MovieLensIM dataset in terms of precision and mean average precision (MAP) respectively.

---

# XAI enhancing enhancing cyber defence against adversarial attacks in industrial applications

5 Dec 2022
14:40-16:10
Session 2
Live stream

Georgios Makridis[1], Spyros Theodoropoulos[2], Dmimitrios Dardanis[3], Ioannis Makridis[4], Maria-Margarita Separdani[5], Georgios Fatouros[6], Dimosthenis Kyriazis[7], Panagiotis Koulouris[8]

[1]:gmakridis@unipi.gr, [2]:sptheod@unipi.gr,
[3]:ddardanis@unipi.gr, [4]:imakridi@unipi.gr, [5]:ritasepa8@gmail.com, [6]:gfatouros@unipi.gr,
[7]:dimos@unipi.gr, [8]:pa.koulouris@gmail.com

[1,2,3,4,6,7,8]: University of Piraeus, Greece
[5]: University of Crete, Greece

**Summary**

In recent years there is a surge of interest in the interpretability and explainability of AI systems, which is largely motivated by the need for ensuring the transparency and accountability of Artificial Intelligence (AI) operations, as well as by the need to minimize the cost and consequences of poor decisions. Another challenge that needs to be mentioned is the Cyber security attacks against AI infrastructures in manufacturing environments. This study examines eXplainable AI (XAI)-enhanced approaches against adversarial attacks for optimizing Cyber defense methods in manufacturing image classification tasks. The examined XAI methods were applied to an image classification task providing some insightful results regarding the utility of Local Interpretable Model-agnostic Explanations (LIME), Saliency maps, and the Gradient-weighted Class Activation Mapping (Grad-Cam) as methods to fortify a dataset against gradient evasion attacks. To this end, we "attacked" the XAI-enhanced Images and used them as input to the classifier to measure their robustness of it. Given the analyzed dataset, our research indicates that LIME-masked images are more robust to adversarial attacks. We additionally propose an Encoder-Decoder schema that timely predicts (decodes) the masked images, setting the proposed approach sufficient for a real-life problem.

# Regular Session 3: 3D image reconstruction and analysis for smart applications
# Session chair: Dr Hela Boulehmi (Monday 5 Dec 2022 at 16:20-17:50)

## Experimental Validation of Photogrammetry based 3D Reconstruction Software

5 Dec 2022
16:20-17:50
session 3

Hussain Razeen[1], Pizzo Marianna [2], Ballestin Giorgio[3], Chessa, Manuela [4], Solari Fabio [5]

[1]: razeen.hussain@edu.unige.it, [2]: marianna.pizzo@edu.unige.it, [3]: giorgio.ballestin@gmail.com, [4]: manuela.chessa@unige.it, [5]: fabio.solari@unige.it

University of Genova, Italy

**Summary**

3D reconstruction is of interest to several fields. However, obtaining the 3D model is usually a time-consuming task that involves manual measurements and reproduction of the object using CAD software, which is not always feasible (e.g. for organic shapes). The necessity of quickly obtaining a dimensionally accurate 3D model of an object has led to the development of several reconstruction techniques, either vision based (with photogrammetry), using laser scanners, or a combination of the two. The contribution of this study is in the analysis of the performances of currently available 3D reconstruction frameworks with the aim of providing a guideline to novice users who may be unfamiliar with 3D reconstruction technologies. We evaluate various software packages on a synthetic dataset representing objects of various shapes and sizes. For comparison, we consider various metrics such as mean errors in the reconstructed cloud point and meshes and reconstruction time. Our results indicate that Colmap produces the best reconstruction.

---

5 Dec 2022
16:20-17:50
session 3

# A Novel Automated Classification and Segmentation for COVID-19 using 3D CT

Shiyi Wang [1], Yang Guang [2]

[1]: s.wang22@imperial.ac.uk, [2]: g.yang@imperial.ac.uk

Imperial College London, United Kingdom

**Summary**

Medical image classification and segmentation based on deep learning (DL) are emergency research topics for diagnosing variant viruses of the current COVID-19 situation. In COVID-19 computed tomography (CT) images of the lungs, ground glass turbidity is the most common finding that requires specialist diagnosis. Based on this situation, some researchers propose the relevant DL models which can replace professional diagnostic specialists in clinics when lacking expertise. However, although DL methods have a stunning performance in medical image processing, the limited datasets can be a challenge in developing the accuracy of diagnosis at the human level. In addition, deep learning algorithms face the challenge of classifying and segmenting medical images in three or even multiple dimensions and maintaining high accuracy rates. Consequently, with a guaranteed high level of accuracy, our model can classify the patients' CT images into three types: Normal, Pneumonia and COVID. Subsequently, two datasets are used for segmentation, one of the datasets even has only a limited amount of data (20 cases). By feeding with different datasets, the COVID image segmentation of the infected area will be carried out according to classification results. Our model achieves 94.52% accuracy in the classification of lung lesions by 3 types: COVID, Pneumonia and Normal. For 2 labels (ground truth, lung lesions) segmentation, the model gets 99.57% of accuracy, 0.2191 of train loss and 0.78±0.03 of MeanDice±Std, while the 4 labels (ground truth, left lung, right lung, lung lesions) segmentation achieves 98.89% of accuracy, 0.1132 of train loss and 0.83±0.13 of MeanDice±Std. For future medical use, embedding the model into the medical facilities might be an efficient way of assisting or substituting doctors with diagnoses, therefore, a broader range of the problem of variant viruses in the COVID-19 situation may also be successfully solved.

---

5 Dec 2022
16:20-17:50
session 3

# Improving 3D Point Cloud Reconstruction with Dynamic Tree-Structured Capsules

Antonio J Rodriguez-Sanchez[1], Engelhardt Chris[2], Mittelberger Jakob [3], Peer David[4], Stabinger Sebastian[5]

[1]: antonio.rodriguez-sanchez@uibk.ac.at, [2]: Chris.engelhardt@student.ac.at, [3]: jakob.mittelberger@student.uibk.ac.at, [4]: david.peer@outlook.com, [5]: sebastian@stabinger.name

KIOS Research and Innovation Center of Excellence University of Cyprus, Cyprus

**Summary**

When applying convolutional neural networks to 3D point cloud reconstruction, these do not seem to be able to learn meaningful 2D manifold embeddings, suffer a lack of explainability and are vulnerable to adversarial attacks. Except for the latter, these shortcomings can be overcome with capsule networks. In this work we introduce an auto-encoder based on dynamic tree-structured capsule networks for sparse 3D point clouds with SDA-routing. Our approach preserves the spatial arrangements of the input data and increases the adversarial robustness without introducing additional computational overhead. Our experimental evaluation shows that our architecture outperforms the current state-of-the-art capsule and CNN-based networks.

---

## Continual Learning in an Industrial Scenario: Equipment Classification on Edge

5 Dec 2022
16:20-17:50
session 3

Morgado Ana C.[1], Carvalho Rafaela[2], Andrade Catarina B[3],Barbosa Telmo[4], Santos Gonçalo[5], Vasconcelos Maria Joao M[6]

[1]: ana.morgado@fraunhofer.pt, [2]: rafaela.carvalho@fraunhofer.pt,
[3]: catarina.andrade@fraunhofer.pt, [4]: telmo.barbosa@fraunhofer.pt,
[5]: goncalo.santos@glartek.com, [6]: maria.vasconcelos@fraunhofer.pt

Fraunhofer Portugal AICOS, Portugal

**Summary**

The ability to incrementally learn to categorize objects is a key feature for a personalized system in real-world applications. The major constraint for such scenario relies on the catastrophic forgetting problem, which negatively impacts the performance of the models on previously learned representations. In this work, we developed an equipment classification model to be deployed on edge devices by applying regularization and memory-based class-incremental strategies, such that it can detect new classes while preserving its ability to detect previously known classes, mitigating the forgetting phenomenon. The strategies were tested on three datasets: CIFAR100 to validate the implementation, Stanford Dogs to ensure the reliability of the results as it is a more representative dataset, and SINATRA, which is the work's industrial dataset for equipment recognition. Experimental results on these datasets show that the Experience Replay strategy performed better. For the SINATRA dataset, average accuracy values of 95.57% and of 100% were achieved for Aguas e Energias do Porto and Plastaze subsets, respectively. The outcomes of this work proved that by retaining only a limited number of exemplars from old classes, it is possible to update a pre-existing system to classify new devices in a shorter period and avoid catastrophic forgetting.

---

5 Dec 2022
16:20-17:50
session 3

# Surface Crack Detection using Deep Convolutional Neural Network in Concrete Structures

Alireza Rahai[1], Mohammad Rahai[2], Mostafa Iraniparast[3], Mehdi Ghatee[4]
[1]:rahai@aut.ac.ir, [2]:mohammad.rahai@alum.sharif.edu
, [3]:m.iraniparast@aut.ac.ir, [4]: ghatee@aut.ac.ir

[2]: Sharif University of Technology, Iran
[1,3,4]: Amirkabir University of Technology, Iran

**Summary**

Regular safety inspections of concrete and steel structures during their serviceability are essential since they directly affect the reliability and structural health. Early detection of cracks helps prevent further damage. Traditional methods involve the detection of cracks by human visual inspection. However, it is difficult to visually find cracks and other defects for extremely large structures because of time and cost constraints. Therefore, the development of smart inspection systems has been given utmost importance. We provide a deep convolutional neural network (DCNN) with transfer learning (TF) technique for crack detection. To reduce false detection rates, the images used to train in the TF technique come from three separate datasets. Moreover, the designed CNN is trained on 3200 images of $256 \times 256$ pixel resolutions. Different deep learning networks are considered and the experiments on test images show that the accuracy of the damage detection is more than 99%. Results illustrate the viability of the suggested approach for crack observation and classification.

# Regular Session 4: Tracking for activity understanding
# Session chairs: Prof. François Brémond and Sébastien Ambellouis
# (Tuesday 6 Dec 2022 at 09:30-11:00)

## Optical tracking system based on COTS components

6 Dec 2022
09:30-11:00
Session 4

Débora NP Oliveira[1], Marcos R Morais[2], Antonio MN Lima[3]
[1]: debora.oliveira@ee.ufcg.edu.br, [2]: morais@dee.ufcg.edu.br,
[3]: amnlima@dee.ufcg.edu.br

Universidade Federal de Campina Grande (UFCG), Brazil

**Summary**

This paper deals with the design of an indoor optical tracking system based on commercially available off-the-shelf products. In the proposed system, four V2 NoIR Raspberry cameras are connected to various Raspberry Pi boards (Model 3B, 3B+, and 4) as capture stations. In this work, algorithms for clock synchronization, rapid contour extraction, and intrinsic camera calibration are discussed. The size, layout, illumination, and safety of an arena are also addressed, as well as construction issues like non-uniform lighting or noisy reflections. The system's accuracy is sub-centimeter at a frame rate of 100Hz, which is comparable to the performance of the proprietary and commercially available optical tracking systems. These results demonstrate that the proposed solution is feasible and show the correctness of the suggested methodology.

---

6 Dec 2022
09:30-11:00
Session 4

## DUNE: Deep UNcertainty Estimation for tracked visual features

Katia Sousa Lillo[1], Andrea De Maio[2], Simon Lacroix[3], Amaury Negre[4], Michele Rombaut[5], Nicolas Marchand[6], Nicolas Vercier[7]

[1]: katia.lillo@grenoble-inp.fr, [2]: andrea.de-maio@laas.fr, [3]: simon.lacroix@laas.fr, [4]: amaury.negre@gipsa-lab.grenoble-inp.fr, [5]: michele.rombaut@gipsa-lab.grenoble-inp.fr, [6]: nicolas.marchand@gipsa-lab.grenoble-inp.fr, [7]: nicolas.vercier@fr.thalesgroup.com

LAAS-CNRS and Thales AVS, France

**Summary**

Uncertainty estimation of visual feature is essential for vision-based systems, such as visual navigation. We show that errors inherent to visual tracking, in particular using KLT tracker, can be learned using a probabilistic loss function to estimate the covariance matrix on each tracked feature position. The proposed system is trained and evaluated on synthetic data, as well as on real data, highlighting good results in comparison to the state of the art. The benefits of the tracking uncertainty estimates are illustrated for visual motion estimation.

---

6 Dec 2022
09:30-11:00
Session 4
Live stream

## A Benchmark Database for Animal Re-Identification and Tracking

Ludmila I Kuncheva[1], Frank Williams[2], Samuel Hennessay[3], Juan Rodriguez[4]

[1]: mas00a@bangor.ac.uk, [2]: eeub05@bangor.ac.uk, [3]: s.hennessay@bangor.ac.uk, [4]: jjrodriguez@ubu.es

[1,2,3]:Bangor University, United Kingdom
[4]: Universitat de Burgos, Spain

**Summary**

While there are multiple sources of annotated images and videos for human and vehicle re-identification, databases for individual animal recognition are still in demand. We present a database containing five annotated video clips. The database can be used for testing novel methods for animal reidentification, object detection and tracking. The main challenge of the database is that multiple animals are present in the same video frame, leading to problems with occlusion and noisy, cluttered bounding boxes. To set-up a benchmark on individual animal recognition, we trained and tested 26 classification methods for the five videos and three feature representations. We also report results with state-of-the-art deep learning methods for object detection (MMDet) and tracking (Uni-Track).

---

## Analysis of Real-Time Hostile Activitiy Detection from Spatiotemporal Features Using Time Distributed Deep CNNs, RNNs and Attention-Based Mechanisms

6 Dec 2022
09:30-11:00
Session 4
Live stream

Labib Ahmed Siddique[1], Rabita Junhai[2], Md Tanzim Reza[3], Tanvir Rahman[4]
[1]: labib.ahmed.siddique@g.bracu.ac.bd, [2]: rabita.junhai@g.bracu.ac.bd, [3]: tanzim.reza@bracu.ac.bd,[4]: rtanvir@udel.edu

University of Delaware, Bangladesh

**Summary**

Real-time video surveillance, through CCTV camera systems has become essential for ensuring public safety which is a priority today. Although CCTV cameras help a lot in increasing security, these systems require constant human interaction and monitoring. To eradicate this issue, intelligent surveillance systems can be built using deep learning video classification techniques that can help us automate surveillance systems to detect violence as it happens. In this research, we explore deep learning video classification techniques to detect violence as they are happening. Traditional image classification techniques fall short when it comes to classifying videos as they attempt to classify each frame separately for which the predictions start to flicker. Therefore, many researchers are coming up with video classification techniques that consider spatiotemporal features while classifying. However, deploying these deep learning models with methods such as skeleton points obtained through pose estimation and optical flow obtained through depth sensors, are not always practical in an IoT environment. Although these techniques ensure a higher accuracy score, they are computationally heavier. Keeping these constraints in mind, we experimented with various video classification and action recognition techniques such as ConvLSTM, LRCN (with both custom CNN layers and VGG-16 as feature extractor) CNNTransformer and C3D. We achieved a test accuracy of 80% on ConvLSTM, 83.33% on CNN-BiLSTM, 70% on VGG16-BiLstm, 76.76% on CNN-Transformer and 80% on C3D.

---

## 6D pose estimation for precision assembly

6 Dec 2022
09:30-11:00
Session 4
Live stream

Ola Skeik[1], Mustafa Suphi Erden[2], Xianwen Kong[3]
[1]: eng.3ola1988@hotmail.com, [2]: m.s.erden@hw.ac.uk, [3]: X.Kong@hw.ac.uk

Coventry University and Heriot-Watt University, United Kingdom

**Summary**

The assembly of 3D products with complex geometry and material, such as in a concentrator photovoltaics solar panel unit, are typically conducted manually. This results in low efficiency, precision and throughput. This study is motivated by an actual industrial need and targeted towards automation of the currently manual

assembly process. By replacing the manual assembly with robotic assembly systems, the efficiency and throughput could be improved. Prior to assembly, it is essential to estimate the pose of the objects to be assembled with high precision. The choice of the machine vision is important and plays a critical role in the overall accuracy of such a complex task. Therefore, this work focuses on the 6D pose estimation for precision assembly utilizing a 3D vision sensor. The sensor we use is a 3D structured light scanner which can generate high quality point cloud data in addition to 2D images. A 6D pose estimation method is developed for an actual industrial solar-cell object, which is one of the four objects of an assembly unit of concentrator photovoltaics solar panel. The proposed approach is a hybrid approach where a mask R-CNN network is trained on our custom dataset and the trained model is utilized such that the predicted 2D bounding boxes are used for point cloud segmentation. Then, the iterative closest point algorithm is used to estimate the object's pose by matching the CAD model to the segmented object in point cloud.

---

## Computer Vision-Based Bengali Sign Language To Text Generation

6 Dec 2022
09:30-11:00
Session 4
Live stream

Zarin Anan Aunshu[1], Tonjih Tazalli[2], Sumaya Sadbeen Liya[3], Magfirah Hossain[4], Zareen Mehjabeen[5], Md. Sabbir Ahmed[6], Muhammad Iqbal Hossain[7]

[1]: zarin.anan.aunshu@g.bracu.ac.bd, [2]: tonjih.tazalli@g.bracu.ac.bd, [3]: sumaya.sadbeen.liya@g.bracu.ac.bd, [4]: mst.magfirah.hossain@g.bracu.ac.bd, [5]: zareen.mehjabeen@g.bracu.ac.bd, [6]: msahmeds1997@gmail.com, [7]: iqbal.hossain@bracu.ac.bd

Brac University, Bangladesh

### Summary

In the whole world, around 7% of people have hearing and speech impairment problems. They use sign language as their communication method. People from various countries use a variety of sign languages. As an example, there are ASL, CSL, JSL, etc. Even in our country, there are lots of people born with hearing and speech impairment problems. So, our primary focus is to work for those people by converting Bangla sign language into text. There are already various projects on Bangla sign language done by other people. However, they focused more on the separate alphabets and numerical numbers. That is why we want to concentrate on Bangla word signs since these people prefer to communicate using words or phrases rather than alphabets. There is not any proper database for Bangla word sign language, so we are making a database for Bangla word sign language for our work. In recognition of sign language (SLR), there usually are two types of scenarios: isolated SLR, which takes words by word and completes recognize action, and the other one is continuous SLR, which completes action by translating the whole sentence at once. We are working on isolated SLR. We introduce a method where we are going to use PyTorch and YOLOv5 for a video classification model to convert Bangla sign language into the text from the video where each video has only one sign language

word. Here,we have achieved an accuracy rate of 76.29% on the training dataset and 51.44% on the testing dataset. We are working to build a system that will make it easier for hearing and speech-disabled people to interact with the general public.

---

# Keynote IV by Prof. Jean-Mark Odobez
# Session chair: Prof. Fabio Solari
# (Tuesday 6 Dec 2022 at 11:10-11:50)

## Towards gaze analysis in the wild

6 Dec 2022
11:10-11:50

Jean Mark-Odobez

Idiap Research Institue, Switzerland

**Summary**

As a display of attention and interest, gaze is a fundamental cue in understanding people's activities, behaviors, and state of mind, and plays an important role in many applications and research fields. In psychology and sociology, gaze information helps to infer inner states of people or their intention, and to better understand the interaction between individuals. In particular, gaze plays a major role in human communication, like for showing attention to the speaker or indicating who is addressed, which makes the automatic extraction of gaze highly relevant for the design of intuitive human computer or robot interfaces, or for medical diagnosis like for children with Autism Spectrum Disorders (ASD).

Gaze (estimating the 3D line of sight) and Visual Focus of Attention (VFOA) estimation, however, are challenging, even for humans. It often requires not only analysing the person's face and eyes, but also the scene content including the 3D scene structure and the person's situation (What is in the field of view of the person? How many people are around? Who is talking? manipulating objects? interacting or observing others?) to detect obstructions in the line of sight or apply attention priors that humans typically have when observing others. In this presentation, we will present three methods that address these challenges: first, a method that leverages standard activity-related priors about gaze to perform online calibration; secondly, an approach for VFOA inference which casts the scene in the 3D field of view of a person, enabling the use of audio-visual information as well as dealing with an arbitrary number of targets, and providing better cross-scene generalization;

and third, moving towards gaze estimation in the wild, an approach for the gaze-following task explicitly leveraging derived multimodal cues like depth and pose. Finally, we will shortly describe a gaze-interactive scene demonstration developed for the 'musée de la main'.

**Jean-Mark Odobez biography**

He is leading the Perception and Activity Understanding group at the Idiap Research Institue. His main research interests are on human activities analysis from multi-modal data. This entails the investigation of fundamental tasks like the detection and tracking of people, the estimation of their pose or the detection of non-verbal behaviors, and the temporal interpretation of this information in forms of gestures, activities, behavior or social relationships. These tasks are addressed through the design of principled algorithms extending models from computer vision, multi-modal signal processing, and machine learning, in particular probabilistic graphical models and deep learning techniques, surveillance, traffic and human behavior analysis.

---

# Panel discussion I (Tuesday 6 Dec 2022 at 11:50-12:20) Panelists: François Bremond, Jean-Mark Odobez, Nicolas Gillis, Sébastien Ambellouis, and Dimitri Ognigene

**Machine Learning Transformers and current issues**

5 Dec 2022
11:50-12:20

Panelists: François Brémond, Jean-Mark Odobez, Nicolas Gillis, Sébastien Ambellouis, and Dimitri Ognigene

---

# Keynote V by Prof. Riadh Abdelfattah
# Session chair: Prof. Nicolas Gillis (Tuesday 6 Dec 2022 at 12:20-13:00)

### Contribution of Interferometric SAR to water resource management

6 Dec 2022
12:20-13:00

Riadh Abdelfattah
SUP'COM, University of Carthage, Tunisia
riadh.abdelfattah@supcom.tn

**Summary**

Water resources are essential for the development of life on Earth. The majority of water is stored in the oceans (97.19%), the rest - fresh water - is distributed between frozen surfaces and glaciers (2.20%), groundwater (0.60%) and surface water which includes streams, lakes and soil moisture (about 0.01%). While water needs are generally associated with access to drinking water for the population, it is also essential for many industrial and agri-food sectors. The use of water has thus continuously intensified and diversified since the beginning of the 20th century, leading to an increased volumes of water used. Agriculture, which uses water for the irrigation of cultures, represents 3/4 of the current demand in the world: Mapping the distribution and persistence of surface water as well as groundwater in a timely fashion has broad value for tracking dynamic events like flooding, and for monitoring the effects of climate and human activities on natural resource values and biodiversity. Synthetic Aperture Radar (SAR) sensors are adequate for monitoring surface water extent. However, this is not the case with groundwater. In this talk, the potential of using interferometric SAR (InSAR) for detecting intensive exploitation of aquifers (groundwater) will be addressed. We show that this could be possible using SAR interferometry for mapping landslides caused by intensive pumping. Moreover, we present the added value of the InSAR coherence in addition to optical data for accurately mapping surface water extent. This talk will first provide an overview on the state of the art in SAR, Interferometric SAR and

their environmental applications. Then the InSAR coherence for monitoring change detection will be detailed. The second part of this talk describes an analysis on the monitoring of the evolution of water surface related to the Lebna dam in the North of Tunisia based on optical and SAR (synthetic aperture radar) data from Sentinel 2 and 1. The InSAR (Interferometric SAR) coherence maps over the Lebna watershed, Tunisia ( 210 km2) were generated using the Sentinel-1 data acquired from 2017 to 2022. Then, these maps simultaneously with the NDWI (Normalized Difference Water Index) extracted from the Sentinel-2 data were employed to assess the separability of different wetland types and their trends over time. Furthermore, the landslide maps generated using the Differential Interferometric SAR processing chain P-SBAS (Parallel Small BAseline Subset) will be analyzed with respect to intensive exploitation of aquifers.

**Riadh Abdelfattah biography**

Dr. Riadh Abdelfattah is Professor at the Higher School of Engineering in Communications (SUP'COM) at the University of Carthage In Tunisia. He was the President, and a Vice- President of the University of Carthage (2017-2020), in charge of research activities, technologic development and environmental partnership. He is also Associate Researcher at the Department ITI (Image Traitement de l'Information) at IMT-Atlantique, the "Institut de Télécom", Brest, France. He was member of the scientific council of AUF (Agence Universitaire de la Francophonie) and member of the Expert Regional Committee (2016- 2020). He received the engineer degree from the Telecommunication Engineering School of Tunis, Tunisia in 1995, the Master Degree (DEA) and the the Ph.D degree in Electrical Engineering from the " Ecole Nationale Ingénieurs de Tunis", in 1995 and 2000 respectively, and "le Diplôme de l'Habilitation Universitaire" from the Higher School of Communications (SUP'COM) at the University of Carthage in Tunisia (2008). Between 2000 and 2002 he was a postdoctoral researcher at the " Ecole Nationale des Télécommunications", Paris, France consecutively at the department TSI and then at the department COMELEC. He is a senior member of the IEEE and he served as a member of the Executive Committee of the IEEE Tunisia Section (2013-2015). He has authored and co-authored more than 80 journal papers, conference papers and book chapters. His main research interests include interferometric radar imagining, multiemporal and multiscale image analysis, desertification, flooding and soil salinity mapping from remote sensed data, applied AI for water resource management and SAR-nanosatellite development.

---

# Regular Session 5: Image analysis for medical and environmental applications
# Session chair: Dr. Mouna Zouari (Tuesday 6 Dec 2022 at 14:00-15:30)

## A Combined Acute and Chronic Risk Assessment Rolling Window for Type 1 Diabetes

6 Dec 2022
14:00-15:30
Session 5

Faizan Munawar[1], Etain Kiely[2], John Donovan[3], Konrad Mulrennan[4]
[1]:faizan.munawar@research.atu.ie, [2]:etain.kiely@atu.ie, [3]:john.donovan@atu.ie, [4]:konrad.mulrennan@atu.ie

[1,2,4]:The Health and Biomedical (HEAL) Research Centre
[1,2,4]:Atlantic Technological University Sligo, Ireland
[3]:School of Science and Computing, Atlantic Technological University Galway, Galway, Ireland

**Summary**

Monitoring the control of persons with type 1 diabetes based on their history of blood glucose levels is essential for self-management. Persons with diabetes must keep their blood glucose levels in a very narrow glycaemic region (70-180 mg/dl) to avoid hypoglycaemia and hyperglycaemia. An extended period of time in the hypoglycaemic or hyperglycaemic region can lead to short-term and long-term complications, respectively. Many measures have been proposed for the management of diabetes, such as the Glucose Management Indicator (GMI) and the Average Daily Risk Range (ADRR). A major drawback of these measures is that they only address acute (ADRR) or chronic (GMI) complications and provide no information on the trend. This paper proposes a rolling window to calculate ADRR and GMI. Calculating ADRR and GMI using a rolling window results in new data, which provides information on the efficacy of self-management of an individual and their risk trend. Use of a rolling window for the risk analysis provides novel information about

glycaemic variability and can be used for improved personal diabetes management plans. Furthermore, ADRR and GMI are combined to propose four new risk levels, which represent the lowest to the highest probable risk of complications. The analysis was performed on 12 subjects from the OhioT1DM data set. The results presented include a detailed examination and summary of all risks to the subjects and information about their ADRR and GMI trend.

---

## Evaluating Attention in Convolutional Neural Networks for Blended Images

6 Dec 2022
14:00-15:30
Session 5

Andrea Portscher[1], Sebastian Stabinger[2], Antonio J Rodriguez-Sanchez [3]

[1]: andrea.portscher@gmail.com, [2]: sebastian@stabinger.name,
[3]: antonio.rodriguez-sanchez@uibk.ac.at

University of Innsbruck, Austria

### Summary

In neuroscientific experiments, blended images are used to examine how attention mechanisms in the human brain work. They are particularly suited for this research area, as a subject needs to focus on particular features in an image to be able to classify superimposed objects. As Convolutional Neural Networks (CNNs) take some inspiration from the mammalian visual system – such as the hierarchical structure where different levels of abstraction are processed on different network layers – we examine how CNNs perform on this task. More specifically, we evaluate the performance of four popular CNN architectures (ResNet18, ResNet50, CORnet-Z, and Inception V3) on the classification of objects in blended images. Since humans can rather easily solve this task by applying object-based attention, we also augment all architectures with a multi-headed self-attention mechanism to examine its effect on performance. Lastly, we analyse if there is a correlation between the similarity of a network architecture's structure to the human visual system and its ability to correctly classify objects in blended images. Our findings showed that adding a self-attention mechanism reliably increases the similarity to the V4 area of the human ventral stream, an area where attention has a large influence on the processing of visual stimuli.

---

## Complex Network for Complex Problems: A comparative study of CNN and Complex-valued CNN

6 Dec 2022
14:00-15:30
Session 5

Soumick Chatterjee[1], Pavan Tummala[2], Oliver Speck [3], Andreas Nurnberger [4]
[1]: soumick.chatterjee@ovgu.de, [2]: pavan.tummala@ovgu.de, [3]: oliver.speck@ovgu.de, [4]: andreas.nuernberger@ovgu.de

Otto von Guericke University Magdeburg, Germany

**Summary**

Neural networks, especially convolutional neural networks (CNN), are one of the most common tools these days used in computer vision. Most of these networks work with real-valued data using real-valued features. Complex-valued convolutional neural networks (CV-CNN) can preserve the algebraic structure of complex-valued input data and have the potential to learn more complex relationships between the input and the ground-truth. Although some comparisons of CNNs and CV-CNNs for different tasks have been performed in the past, a large-scale investigation comparing different models operating on different tasks has not been conducted. Furthermore, because complex features contain both real and imaginary components, CV-CNNs have double the number of trainable parameters as real-valued CNNs in terms of the actual number of trainable parameters. Whether or not the improvements in performance with CV-CNN observed in the past have been because of the complex features or just because of having double the number of trainable parameters has not yet been explored. This paper presents a comparative study of CNN, CNNx2 (CNN with double the number of trainable parameters as the CNN), and CV-CNN. The experiments were performed using seven models for two different tasks - brain tumour classification and segmentation in brain MRIs. The results have revealed that the CV-CNN models outperformed the CNN and CNNx2 models.

---

## A CNN Architecture for Detection and Segmentation of Colorectal Polyps from CCE Images

6 Dec 2022
14:00-15:30
Session 5
Live stream

Ashkan Tashk[1], Esmaeil S. Nadimi[2], Jurgen Herp[3], K.E. Sahin [4]
[1]: tashk-ashkan@hotmail.com, [2]: esi@mmmi.sdu.dk, [3]: herp@mmmi.sdu.dk, [4]: kasah19@student.sdu.dk

Southern Denmark University, Denmark

**Summary**

Colon capsule endoscopy (CCE) as a novel biomedical image modality provides a higher perspective of the potential gastrointestinal lesions like polyps within the small and large intestines than the conventional colonoscopy. As the quality of images acquired via CCE imagery is low, so the artificial intelligence methods are proposed to help detect and localize polyps within an acceptable level of efficiency

and performance. In this paper, a new deep neural network architecture known as AID-U-Net is proposed. AID-U-Net consists of two distinct types of paths: a) Two main contracting/expansive paths, and b) Two sub-contracting/expansive paths. The playing role of the main paths is to localize polyps as the target objectives in high resolution and multi-scale manner, while the two sub paths are responsible for preserving and conveying the information of low resolution and low-scale target objects. Furthermore, the proposed network architecture provides simplicity so that the model can be deployed for real time processing. AID-U-Net with an implementation of a VGG19 backbone shows better performance to detect polyps in CCE images in comparison with the other state-of-the-art U-Net models like conventional U-Net, U-Net++, and U-Net3+ with different pre-trained backbones like ImageNet, VGG19, ResNeXt50, Resnet50, InceptionV3 and InceptionResNetV2.

---

## Real-time Powered Wheelchair Assistive Navigation System Based on Intelligent Semantic Segmentation for Visually Impaired Users

6 Dec 2022
14:00-15:30
Session 5
Live stream

Elhassan Mohamed [1] , Konstantinos Sirlantzis [2], Gareth Howells[3]
[1]: enrm4@kent.ac.uk,[2]: k.sirlantzis@kent.ac.uk, [3]: w.g.j.howells@kent.ac.uk

University of Kent, United Kingdom

**Summary**

People with movement disabilities may find powered wheelchair driving a challenging task due to their comorbidities. Certain visually impaired persons with mobility disabilities are not prescribed a powered wheelchair because of their sight condition. However, powered wheelchairs are essential to the majority of these disabled users for commuting and social interaction. It is vital for their independence and well being. In this paper, we propose to use a semantic segmentation (SS) system based on deep learning algorithms to provide environmental cues and information to visually impaired wheelchair users to aid with the navigation process. The system classifies the objects of the indoor environment and presents the annotated output on a display customised to the user's condition. The user can select a target object, for which the system can display the estimated distance from the current position of the wheelchair. The system runs in real-time, using a depth camera installed on the wheelchair, and it displays the scene in front of the wheelchair with every pixel annotated with distinguishable colour to represent the different components of the environment along with the distance to the target object. Our system has been designed, implemented and deployed on a real powered wheelchair for practical evaluation. The proposed system helped the users to estimate more accurately the distance to the target objects with a relative error of 19.8% and 18.4% for the conditions of a) semi-neglect and b) short-sightedness, respectively, compared to errors of 47.8% and 5.6% without the SS system. In our experiments, healthy participants were put in simulated conditions representing the above visual impairments using

instruments commonly used in medical research for this purpose. Finally, our system helps to visualise, on the display, hidden areas of the environment and blind spots that visually impaired users would not be able to see without it.

---

# Regular Session 6: Scene Analysis for activity recognition
# Session chairs: Prof. Ali Douik and Prof. François Brémond (Tuesday 6 Dec 2022 at 15:40-17:10)

## A novel license plate detection based Time-To-Collision calculation for forward collision warning using Azure Kinect

6 Dec 2022
15:40-17:10
Session 6

Zhouyan Qiu[1], Joaquín Martínez-Sánchez[2], Pedro Arias-Sánchez[3]
[1]: zhouyan.qiu@uvigo.es, [2]: joaquin.martinez@uvigo.es, [3]: parias@uvigo.es

University of Vigo, Spain

**Summary**

Forward Collision Warning (FCW) system constantly measures the relative position of the vehicle ahead and then predicts collisions. This paper proposes a new cost-effective and computationally efficient FCW method that uses a time-of-flight (ToF) camera to measure relevant distances to the front vehicle based on license plate detection. First, a Yolo V7 model is used to detect license plates to identify vehicles in front of the ego vehicle. Second, the distance between the front vehicle and the ego vehicle is determined by analyzing the captured depth map by the time-of-flight camera. In addition, the relative speed of the vehicle can be calculated by the direct distance change between the license plate and the camera between two consecutive frames. With a processing speed of 25-30 frames per second, the proposed FCW system is capable of determining relative distances and speeds within 26 meters in the real-time.

---

6 Dec 2022
15:40-17:10
Session 6

## Combination of Object Tracking and Object Detection for Animal Recognition

Frank Williams[1], Ludmila I Kuncheva[2], Juan Rodriguez[3], Samuel Hennessay[4]

[1]: eeub05@bangor.ac.uk, [2]: mas00a@bangor.ac.uk, [3]: jjrodriguez@ubu.es,
[4]: s.hennessay@bangor.ac.uk

[1,2,4]: Bangor University, United Kingdom
[3]: Universidad de Burgos, Spain

**Summary**

While methods for object detection and tracking are well-developed for the purposes of human and vehicle identification, animal identification and re-identification from images and video is lagging behind. There is no clarity as to which object detection methods will work well on animal data. Here we compare two state-of-the art methods which output bounding boxes: the MMDetector and the UniTrack video tracker. Both methods were chosen for their high ranking on benchmark data sets. Using a bespoke pre-annotated database of five videos, we calculated the Average Precision (AP) of the outputs from the two methods. We propose a combination method to fuse the outputs of MMDetection and UniTrack and demonstrate that the proposed method is capable of outperforming both.

---

6 Dec 2022
15:40-17:10
Session 6

## Cell tracking for live-cell microscopy using an activity-priortized assignment strategy

Karina Ruzaeva[1], Jan-Christopher Cohrs[2], Keitaro Kasahara[3], Dietrich Kohlheyer[4], Katharina Noeh[5], Benjamin Berkels[6]

[1]: ruzaeva@aices.rwth-aachen.de, [2]: cohrs@aices.rwth-aachen.de,
[3]: k.kasahara@fz-juelich.de, [4]: d.kohlheyer@fz-juelich.de, [5]: k.noeh@fz-juelich.de,
[6]: berkels@aices.rwth-aachen.de

[1,2,6]: North Rhine-Westphalia Technical University of Aachen, Germany
[3,4,5]: Forschungs Zentrum Jülich, Germany

**Summary**

Cell tracking is an essential tool in live-cell imaging to determine single-cell features, such as division patterns or elongation rates. Unlike in common multiple object tracking, in microbial live-cell experiments cells are growing, moving, and dividing over time, to form cell colonies that are densely packed in mono-layer structures. With increasing cell numbers, following the precise cell-cell associations correctly over many generations becomes more and more challenging, due to the massively increasing number of possible associations. To tackle this challenge, we propose a fast parameter-free cell tracking approach, which consists of activity-prioritized nearest

neighbor assignment of growing (expanding) cells and a combinatorial solver that assigns splitting mother cells to their daughters. As input for the tracking, Omnipose is utilized for instance segmentation. Unlike conventional nearest-neighbor-based tracking approaches, the assignment steps of our proposed method are based on a Gaussian activity-based metric, predicting the cell-specific migration probability, thereby limiting the number of erroneous assignments. In addition to being a building block for cell tracking, the proposed activity map is a standalone tracking-free metric for indicating cell activity. Finally, we perform a quantitative analysis of the tracking accuracy for different frame rates, to inform life scientists about a suitable (in terms of tracking performance) choice of the frame rate for their cultivation experiments, when cell tracks are the desired key outcome.

---

## Union Embedding and Backbone-Attention boost Zero-Shot Learning Model (UBZSL)

6 Dec 2022
15:40-17:10
Session 6
Live Stream

ziyu Li
Liziyu0104@gmail.com

North Rhine-Westphalia Technical University of Aachen, Germany

**Summary**

Zero-Shot Learning (ZSL) aims to identify categories that are never seen during training. There are many ZSL methods available, and the number is steadily increasing. Even then, there are still some issues to be resolved, such as class embedding and image functions. Human-annotated attributes have been involved in recent work on class embedding. However, this type of attribute does not adequately represent the semantic and visual aspects of each class, and these annotating attributes are time-consuming. Furthermore, ZSL methods for extracting image features rely on the development of pre-trained image representations or finetuned models, focusing on learning appropriate functions between image representations and attributes. To reduce the dependency on manual annotation and improve the classification effectiveness, we believe that ZSL would benefit from using Contrastive Language Image Pretraining (CLIP) or combined with manual annotation. For this purpose, we propose an improved ZSL model named UBZSL. It uses CLIP combined with manual annotation as a class embedding method and uses an attention map for feature extraction. Experiments show that the performance of our ZSL model on the CUB dataset is greatly improved compared to the current model. Index Terms—Zero-shot Learning, Image Classification, Backbone-Attention, CLIP.

---

6 Dec 2022
15:40-17:10
Session 6
Live stream

# Image Processing and Control of Tracking Intelligent Vehicle Based on Grayscale Camera

Jian Zhang[1], Yufan Liu[2], Ao Li[3], Jinshan Zeng[4], Hongtu Xie[5]
[1]: zhangj765@mail2.sysu.edu.cn, [2]:528045193@qq.com, [3]:1311603179@qq.com, [4]:2836592406@qq.com, [5]:xiehongtu@mail.sysu.edu.cn

[1,5]: School of Electronic and Communication Engineering Shenzhen Campus of Sun Yat-sen University Shenzhen, China
[2,3,4]: School of Information Engineering NanChang Hangkong University, Nanchang, China

**Summary**

In order to realize the rapid and stable recognition and automatic tracking of various complex roads by the intelligent vehicles, this paper proposes image processing and cascade Proportion Integration Differentiation (PID) steering and speed control algorithms based on CMOS grayscale cameras in the context of the national college student intelligent vehicle competition. First, the grayscale image of the track is acquired by the grayscale camera. Then, the Otsu method is used to binarize the image, and the information of black boundary guide line is extracted. In order to improve the speed of the race, various track elements in the image are identified and classified, and the deviation between the actual centerline position and the ideal centerline position of the intelligent vehicle is calculated. Third, the discrete incremental cascade PID control algorithm is used to calculate the pulse width modulation (PWM) signal corresponding to the deviation. And the PWM signal is acted on the steering motor through the driving circuit, driving the intelligent vehicle to always drive along the middle road, so as to achieve the purpose of automatic tracking guidance. Experiments prove that the intelligent vehicle of this design can identify complex roads quickly and in a stable way, accurately complete automatic tracking, and obtain higher speed performance.

---

6 Dec 2022
15:40-17:10
Session 6
Live stream

# A Light Weight Approach for Real-time Background Subtraction in Camera Surveillance Systems

Ege Ince[1], Sevdenur Kutuk[2], Rayan Abri[3], Sara Abri[4], Salih Cetin[5]
[1]:ege.ince@mavinci.com.tr, [2]:sevdenur.kutuk@mavinci.com.tr, [3]:rayan.abri@mavinci.com.tr, [4]:sara.abri@mavinci.com.tr, [5]:salih.cetin@mavinci.com.tr

Mavinci Informatics Inc. Ankara, Turkey

**Summary**

Real time processing in the context of image processing for topics like motion detection and suspicious object detection requires processing the background more times. In this field, background subtraction solutions can overcome the limitations

caused by real time issues. Different methods of background subtraction have been investigated for this goal. Although more background subtraction methods provide the required efficiency, they do not make produce a real-time solution in a camera surveillance environment. In this paper, we propose a model for background subtraction using four different traditional algorithms; ViBe, Mixture of Gaussian V2 (MOG2), Two Points, and Pixel Based Adaptive Segmenter (PBAS). The presented model is a lightweight real time architecture for surveillance cameras. In this model, the dynamic programming logic is used during preprocessing of the frames. The CDnet 2014 data set is used to assess the model's accuracy, and the findings show that it is more accurate than the traditional methods whose combinations are suggested in the paper in terms of Frames per second (fps), F1 score, and Intersection over union (IoU) values by 61.31, 0.552, and 0.430 correspondingly.

---

# Regular Session 7: Image Processing and Medical Applications
# Session chairs: Prof. Mehrez Abdellaoui and Dr Shadi Khan (Tuesday 6 Dec 2022 at 17:10-18:10)

## Hyperspectral Brain Tissue Classification using a Fast and Compact 3D CNN Approach

6 Dec 2022
17:10-18:10
Session 7

Hamail Ayaz [1], David Tormey [2], Muhammad Ahmad[3], Ian McLoughlin [4], Saritha unnikrishnan [5]

[1]:s00215992@mail.itsligo.ie, [2]:Tormey.David@itsligo.ie, [3]:mahmad00@gmail.com, [4]:ian.mcloughlin@gmit.ie, [5]:unnikrishnan.saritha@itsligo.ie

[1,2,5]:Atlantic Technological University, Ireland,
[4]:National University of Computer and Emerging Sciences, Pakistan,
[3]:Galway-Mayo Institute of Technology, Ireland

**Summary**

Glioblastoma (GB) is a malignant brain tumor and requires surgical resection. Although complete resection of GB improves prognosis, supratotal resection may cause neurological abnormalities. Therefore, intraoperative tissue classification techniques are needed to delineate infected tumor regions to remove reoccurrences. To delineate the affected regions, surgeons mostly rely on traditional magnetic resonance imaging (MRI) which often lacks accuracy and precision due to the brain-shift phenomenon. Hyperspectral Imaging (HSI) is a non-invasive advanced optical technique and has the potential to classify tissue cells accurately. However, HSI tumor classification is challenging due to overlapping regions, high interclass similarity, and homogeneous information. Additionally, HSI models using 2D Convolutional Neural Network (CNN) models work with spectral information eliminating spatial features, and 3D

followed by 2D hybrid model lacks abstract spatial information. Therefore, in this study, we have used a minimal layer 3D CNN model to classify the GB tumor region from normal tissues using an intraoperative Vivo-HSI dataset. The HSI data have normal tissue (NT), tumor tissue (TT), hypervascularized tissue or blood vessels (BV), and background (BG) tissue cells. The proposed 3D CNN model consists of only two 3D layers using limited training samples (20%), which are further divided into 50% for training and 50% for validation and blind tested (80%) on the rest of the data. This study outperformed then state-of-the-art hybrid architecture by achieving an overall accuracy of 99.99%.

---

## Cluster Analysis: Unsupervised Classification for Identifying Benign and Malignant Tumors on Whole Slide Image of Prostate Cancer

6 Dec 2022
17:10-18:10
Session 7

Subrata Bhattacharjee[1], Yeong-Byn Hwang [2], Rashadul Islam Sumon [3], Hafizur Islam [4], Dong-Woo Hyeon [5], Damin Moon[6], Kouayep Sonia Carole [7], Hee-Cheol Kim [8], Heung-Kook Choi [9]

[1]:subrata-bhattacharjee@outlook.com, [2]:hyb1345679@gmail.com , [3]:sumon39.cst@gmail.com , [4]:hafizurdiucsc@gmail.com, [5]:dwhyeon@jlkgroup.com , [6]:dmmoon@jlkgroup.com, [7]:carolesonia39@gmail.com, [8]:heeki@inje.ac.kr , [9]:cschk@inje.ac.kr

[1,2,3,4,7,8,9]: Inje University, Korea of the south
[5,6]: JLK Artificial Intelligence R and D Center, Korea of the south

**Summary**

Recently, many fields have widely used cluster analysis: psychology, biology, statistics, pattern recognition, information retrieval, machine learning, and data mining. Diagnosis of histopathological images of prostate cancer is one of the routine tasks for pathologists and it is challenging for pathologists to analyze the formation of glands and tumors based on the Gleason grading system. In this study, unsupervised classification has been performed for differentiating malignant (cancerous) from benign (non-cancerous) tumors. Therefore, the unsupervised-based computer-aided diagnosis (CAD) technique would be of great benefit in easing the workloads of pathologists. This technique is used to find meaningful clustering objects (i.e., individuals, entities, patterns, or cases) and identify useful patterns. Radiomic-based features were extracted for cluster analysis using the gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), and gray-level size zone matrix (GLSZM) techniques. Multi-clustering techniques used for the unsupervised classification are K-means clustering, K-medoids clustering, Agglomerative Hierarchical (AH) clustering, Gaussian mixture model (GMM) clustering, and Spectral clustering. The quality of the clustering algorithms was determined using Purity, Silhouettes, Adjusted Rand, Fowlkes Mallows, and Calinski Harabasz (CH) scores. However, the best-performing algorithm (i.e., K-means) has been applied to predict

and annotate the cancerous regions in the whole slide image (WSI) to compare with the pathologist annotation.

---

## Liver Segmentation in Time-resolved C-arm CT Volumes Reconstructed from Dynamic Perfusion Scans using Time Separation Technique

6 Dec 2022
17:10-18:10
Session 7

Soumick Chatterjee[1], Hana Haseljić[2], Robert Frysch [3], Vojtěch Kulvait [4], Vladimir Semshchikov [5], Bennet Hensen [6], Frank Wacker [7], Inga Brüsch [8], Thomas Werncke [9], Oliver Speck [10], Andreas Nurnberger [11] , Georg Rose [12]

[1]:soumick.chatterjee@ovgu.de, [2]:hana.haseljic@ovgu.de, [3]:robert.frysch@ovgu.de,
[4]:vojtech.kulvait@hereon.de, [5]:vladimir.semshchikov@ovgu.de,
[6]:hensen.bennet@mh-hannover.de, [7]:wacker.frank@mh-hannover.de,
[8]:Bruesch.Inga@mh-hannover.de, [9]:Werncke.Thomas@mh-hannover.de,
[10]:oliver.speck@ovgu.de,[11]:andreas.nuernberger@ovgu.de, [12]:georg.rose@ovgu.de

[1,3,5,9,10,11,12]: Otto von Guericke University Magdeburg, Germany
[2]:University of Sarajevo, Faculty of Electrical Engineering, Bosnia
[4]:Helmholtz-Zentrum hereon, Germany
[6,7,8]:Hannover Medical School, Germany,

**Summary** Perfusion imaging is a valuable tool for diagnosing and treatment planning for liver tumours. The time separation technique (TST) has been successfully used for modelling C-arm cone-beam computed tomography (CBCT) perfusion data. The reconstruction can be accompanied by the segmentation of the liver - for better visualisation and for generating comprehensive perfusion maps. Recently introduced Turbolift learning has been seen to perform well while working with TST reconstructions, but has not been explored for the time-resolved volumes (TRV) estimated out of TST reconstructions. The segmentation of the TRVs can be useful for tracking the movement of the liver over time. This research explores this possibility and shows the robustness of Turbolift learning that it can work efficiently even with the TRVs, resulting in a Dice score of 0.864±0.004.

---

6 Dec 2022
17:10-18:10
Session 7

# A Tool for Thermal Image Annotation and Automatic Temperature Extraction around Orthopedic Pin Sites

Sowmya Annadatha [1], Ming Shen [2], Marie Fridberg [3], Søren Kold [4], Ole Rahbek [5]
[1]:sann@es.aau.dk, [2]:mish@es.aau.dk, [3]:m.fridberg@rn.dk,
[4]:sovk@rn.dk, [5]:o.rahbek@rn.dk

Aalborg University, Denmark

**Summary**

Existing annotation tools are mainly designed for visible images to support supervised learning problems for machine learning. A few tools exist for extracting temperature information from thermal images. However, they are time and manpower consuming, require different stages of data management, and are not automated. This paper focuses on addressing the limitation of existing tools in handling big thermal datasets for annotation, temperature distribution extraction in the Region of Interest (ROI) of Orthopedic surgical wounds and provides flexibility for a researcher to integrate thermal image analysis into wound care machine learning models. We present an easy to use research tool for one click annotation of Orthopedic pin sites for extraction of thermal information, which is a preliminary step of research to estimate the reliability of thermography for home based surveillance of post-operative infection. The proposed tool maps annotations from visible registered image onto thermal and radiometric images. Mapping these annotations from visible registered images avoids manual bias in annotating thermal images. Integrating the functionality of an annotation tool by processing thermal images to acquire single-click manual annotations and extracting temperature distributions in the ROI with those acquired annotations is the novelty of the proposed work and is also crucial for research on deep learning-based investigation on surgical wound infections.

# Regular Session 8: Image and Video Processing for Smart City and Smart environments Session chair: Prof. Nawres Khalifa and Dr Norhen Gargouri (Wednesday 7 Dec 2022 at 9:30-11:00)

## DONEX: Real-time occupancy grid based dynamic echo classification for 3D point cloud

7 Dec 2022
9:30-11:00
Session 8

Niklas Stralau [1], Chengxuan Fu [2]

[1]:Niklas.Stralau@de.bosch.com , [2]:chengxuan.fu@de.bosch.com

Department of LiDAR System Engineering Robert Bosch GmbH Schwieberdingen, Germany

**Summary**

For driving assistance and autonomous driving systems, it is important to differentiate between dynamic objects such as moving vehicles and static objects such as guard rails. Among all the sensor modalities, RADAR and FMCW LiDAR can provide information regarding the motion state of the raw measurement data. On the other hand, perception pipelines using measurement data from ToF LiDAR typically can only differentiate between dynamic and static states on the object level. In this work, a new algorithm called DONEX was developed to classify the motion state of 3D LiDAR point cloud echoes using an occupancy grid approach. Through algorithmic improvements, e.g. 2D grid approach, it was possible to reduce the runtime. Scenarios, in which the measuring sensor is located in a moving vehicle, were also considered.

---

7 Dec 2022
9:30-11:00
Session 8

## An image processing based classifier to support safe dropping for delivery-by-drone

Assem Alsawy Abdel Hak [1], Dan Moss [2], Alan Hicks [3], susan mckeever [4]
[1]:assem.abdelhak@tudublin.ie, [2]: dan.moss@manna.aero,
[3]:alan@manna.aero, [4]:susan.mckeever@tudublin.ie

[1,4]:Technological University Dublin, Ireland
[2,3]:Manna Drone Delivery, Ireland, [1]:Ahram Canadian University (ACU), Egypt

**Summary**

Autonomous delivery-by-drone of packages is an active area of research and commercial development. However, the assessment of safe dropping/ delivery zones has received limited attention. Ensuring that the dropping zone is a safe area for dropping, and continues to stay safe during the dropping process is key to safe delivery. This paper proposes a simple and fast classifier to assess the safety of a designated dropping zone before and during the dropping operation, using a single onboard camera. This classifier is, as far as we can tell, the first to address the problem of safety assessment at the point of delivery-by-drone. Experimental results on recorded drone videos show that the proposed classifier provides both average precision and average recall of 97% in our test scenarios.

---

7 Dec 2022
9:30-11:00
Session 8

## Image Interpolation with Edges Preserving and Implementation on the Real ADAS Platform

Božidar Kelava[1], Mario Vranješ[2], Denis Vranješ [3], Vojtěch Kulvait [4], Željko Lukač
[1]:bozidar.kelava@tttech-auto.com, [2]:mario.vranjes@etfos.hr, [3]:denis.vranjes@ferit.hr,
[4]:zeljko.lukac@uns.ac.rs

[1]:TTTechAuto CEE Osijek, Croatia, [2,3]:Faculty of Electrical Engineering,
Computer Science and Information Technology Osijek, Croatia
[4]:University of Novi Sad, Faculty of Technical Sciences, Novi Sad, Serbia

**Summary**

To save transmission, processing and memory resources in Advanced Diver Assistance Systems (ADAS), it is often necessary to reduce the image resolution. Sometimes it is necessary to increase it after the transmission. Both resolution changes involve an image interpolation process. This paper describes the implementation of three well-known interpolation methods, nearest neighbour interpolation (NN), bilinear interpolation (BL) and bicubic interpolation (BC), on a real automotive AMV ALPHA platform, using multiple processors on the same System on Chip (SoC). Implementation was done using C programming language and Vision Software Development Kit (VSDK). Specific attention is given to the optimal distribution of tasks to a certain processor. The results have shown that, on the real automotive

AMV ALPHA platform, BL interpolation achieves the best trade-off between the quality of interpolated image for the usage in automotive image-processing based algorithms and execution time, especially for the algorithms where the lower frame rate is acceptable (like surround-view, park assist, etc.).

---

## Face Mask Detection using VGG-16 and YOLOv3 Architectures

7 Dec 2022
9:30-11:00
Session 8

Oladapo Ibitoye[1], Oluwafunso Oluwole Osaloni[2]
[1]:ibitoyeo@abuad.edu.ng,[2]:wole-osaloni@yahoo.com

Department of Electrical/Electronics and Computer Engineering
Afe Babalola University Ado Ekiti, Nigeria

**Summary**

Object detection approaches have received a lot more attention in recent years, especially in the areas of face mask detection, classification, and masked face recognition. Facial biometrics is now the most secure alternative for authentication and access control due to the contact-based nature of other biometric techniques and the potential for a pandemic. According to experts, proper preparedness must be undertaken in case there is ever another respiratory-related pandemic. The issue of face mask detection and categorization is one of the areas where it is worthwhile to look for and acquire absolute technological advances. The principles of fair-skinned people served as the foundation for the development of the current face mask detection and identification technologies. The goal of this project was to develop a convolutional neural network with VGG-16 and YOLOv3 architectures that would enable existing systems to perform magnificently in real-time on dark-skinned faces. A performance evaluation of the system revealed improved outcomes.

---

## AI assisted pothole detection and depth estimation-A data driven approach

7 Dec 2022
9:30-11:00
Session 8

Eshta Ranyal[1], Ayan Sadhu[2], Kamal Jain[3]
[1]:eshtaranyal@ce.iitr.ac.in, [2]:asadhu@uwo.ca,
[3]:kjainfce@iitr.ac.in

[1,3]:Indian Institute of Technology Roorkee, India
[2]:Western University of Canada, Canda

**Summary**

AI-assisted engineering solutions integrated with commercial RGB sensors and computationally intensive Graphical Processing Units (GPUs) promise a low-cost solution, to prevent deterioration of premature pavement disintegration. Potholes a

common pavement distress are a severe threat to road safety and demand time and cost-effective state-of-the-art technologies for road inspection and condition monitoring. An intelligent pavement pothole detection system is proposed in this study by modifying the single stage CNN architecture-RetinaNet to detect potholes and perform metrological studies using 3D vision. The photogrammetric technique of structure from motion based on image frames extracted from pavement video recordings is used to model the 3D point cloud structure of potholes to assess the severity of the detected potholes as a function of its depth and is integrated with the CNN based pothole detection system. High F1 scores on benchmark dataset with a high value of 0.98, validate the model's performance. A mean error below 5% is obtained on the measured depths thus promising an intelligent and practical solution to be implemented as part of a potential pavement health assessment system for future practice.

---

7 Dec 2022
9:30-11:00
Session 8

## Vehicle Color Identification Framework using Pixel-level Color Estimation from Segmentation Masks of Car Parts

Klearchos Stavrothanasopoulos[1], Konstantinos Gkountakos: [2], Kostas Ioannidis [3], Theodora Tsikrika [4], Stefanos Vrochidis [5], Yiannis Kompatsiaris [6]

[1]:klearchos-stav@iti.gr,[2]:gountakos@iti.gr,
[3]:kioannid@iti.gr, [4]:theodora.tsikrika@iti.gr,
[5]:stefanos@iti.gr, [6]:ikom@iti.gr

[1,2,3,4,5,6]:Information Technologies Institute, CERTH, Greece
[2]:Centre for Research and Technology Hellas, Greece

**Summary**

Color comprises one of the most significant and dominant cues for various applications. As one of the most noticeable and stable attributes of vehicles, color can constitute a valuable key component in several practices of intelligent surveillance systems. In this paper, we propose a deep-learning-based framework that combines semantic segmentation masks with pixels clustering for automatic vehicle color recognition. Different from conventional methods, which usually consider only the features of the vehicle's front side, the proposed algorithm is able for view-independent color identification, which is more effective for the surveillance tasks. To the best of our knowledge, this is the first work that employs semantic segmentation masks along with color clustering for the extraction of the vehicle's color representative parts and the recognition of the dominant color, respectively. To evaluate the performance of the proposed method, we introduce a challenging multi-view dataset of 500 car-related RGB images extending the publicly available DSMLR Car Parts dataset for vehicle parts segmentation. The experiments demonstrate that the proposed approach achieves excellent performance and accurate results reaching an accuracy of 93.06% in the multi-view scenario. To facilitate further research, the evaluation dataset and the pre-trained models will be released at https://github.com/klearchos-stav/vehicle-color-recognition.

---

# RailSet: A Unique Dataset for Railway Anomaly Detection

7 Dec 2022
9:30-11:00
Session 8
Live stream

Arij Zouaoui[1], Konstantinos Ankur Mahtani [2], Mohamed-Amine Hadded [3], Sebastien Ambellouis [4], Jacques boonaert [5], Hazem Wannous [6]

[1]:arijzouaoui@gmail.com, [2]:ankur.mahtani@railenium.eu,
[3]:mohamedhadded27@gmail.com, [4]:sebastien.ambellouis@univ-eiffel.fr,
[5]: jacques.boonaert@imt-lille-douai.fr , [6]:hazem.wannous@imt-nord-europe.fr

[1,2,3]:IRT Railenium, France,
[4]:Gustave Eiffel University, France,
[5]:IMT Lille Douai, France
[6]:IMT Nord Europe, CRIStAL UMR 9189, France

**Summary**

Understanding the driving environment is one of the key factors in achieving an autonomous vehicle. In particular, the detection of anomalies in the traffic lane is a high priority scenario, as it directly involves vehicle's safety. Recent state of the art image processing techniques for anomaly detection are all based on deep learning of neural networks. These algorithms require a considerable amount of annotated data for training and test purposes. While many datasets exist in the field of autonomous road vehicles, such datasets are extremely rare in the railway domain. In this work, we present a new innovative dataset relevant for railway anomaly detection called RailSet. It consists of 6600 high-quality manually annotated images containing normal situations and 1100 images of railway defects such as hole anomaly and rails discontinuity. Due to the lack of anomaly samples in public images and difficulties to create anomalies in the railway environment, we generate artificially images of abnormal scenes, using a deep learning algorithm named StyleMapGAN. This dataset is created as a contribution to the development of autonomous trains able to perceive tracks damage in front of the train. The dataset is available at this link.

---

# Panel discussion II (Wednesday 7 Dec 2022 at 11:10-11:40) Panelists: François Brémond, Fabio Solari, Nicolas Gillis, Dorra Sellami, and Shadi Khan

### The Gap between industrials and researchers, international collaborations, and Covid-19 impact.

5 Dec 2022
11:10-11:40

Panelists: François Brémond, Fabio Solari, Nicolas Gillis, Dorra Sellami, and Shadi Khan

**Summary**

---

# Regular Session 9: Image processing and biological applications
# Session chair: Prof. Dorra Sellami (Wednesday 7 Dec 2022 at 11:40-13:10)

## Bacterial Blight and Cotton Leaf Curl Virus Detection Using Inception V4 Based

7 Dec 2022
11:40-13:10
session 9

Shadi Khan Baloch[1], Sohail Anwar[2]

[1]:Shadi.baloch@faculty.muet.edu.pk, [2]:sohailanwar.es16@gmail.com

Dept. of Mechatronics Eng. Mehran University of Engineering and Technology Jamshoro, Pakistan

**Summary**

Agriculture sector is an important pillar of the country's economy and cotton crop is considered as one of the prominent agriculture resource. It is widely cultivated in India, China, Pakistan, USA, Brazil and in other countries of the world. Worldwide, cotton crop production is affected by numerous diseases such as cotton leaf curl virus (CLCV/CLCuV), bacterial blight and ball rot. Image processing techniques together with machine learning algorithms are successfully employed in numerous fields and it is also used for crop disease detection. In this study, we describe a deep learning-based method for classifying disease of cotton crop, including bacterial blight and cotton leaf curl virus (CLCV). The dataset of cotton leaf showing diseases symptoms are collected from various locations in Sindh Pakistan. We employ the Inception v4 architecture as a convolutional neural network to identify diseased plant leaves in particular bacterial blight and CLCV. The accuracy of the designed model is 98.26% which is more as compared to the existing systems.

7 Dec 2022
11:40-13:10
session 9
Live stream

# Accurate Medicinal Plant Identification in Natural Environments by Embedding Mutual Information in a Convolution Neural Network Model

Shahmiri Lida[1], Wong Patrick[2], Dooley Laurence[3]

[1]: ls26755@open.ac.uk, [2]: patrick.wong@open.ac.uk, [3]: laurence.dooley@open.ac.uk

The Open University, United Kingdom

**Summary**

Medicinal plants are a primary source of disease treatment in many countries. As most are edible however, consumption of the wrong herbal plants can have serious consequences and even lead to death. Automatic accurate recognition of plant species to help users who do not have specialist knowledge of herbal plants is thus a desirable aim. Several automatic medicinal plant identification systems have been proposed, though most are significantly constrained either in the small number of species or in requiring manual image segmentation of plant leaves. This means they are captured on a plain background rather than being readily identified in their natural surroundings, which often involve complex and noisy backgrounds. While deep learning (DL) based methods have made considerable strides in recent times, their potential has not always been maximised because they are trained with samples which are not always fully representative of the intra-class and interclass differences between the plant species concerned. This paper addresses this challenge by incorporating mutual information into a Convolutional Neural Network (CNN) model to select samples for the training, validation, and testing sets based on a similarity measure. A critical comparative evaluation of this new CNN medicinal plant classification model incorporating a mutual information guided training (MIGT) algorithm for sample selection, corroborates the superior classification performance achieved for the VNPlant-200 dataset, with an average accuracy of more than 97%, while the precision and recall values are also consistently above 97%. This is significantly better than existing CNN classification methods for this dataset as it crucially means false positive rates are substantially lower thus affording improved identification reliability.

## Bioacoustic augmentation of Orcas using TransGAN

7 Dec 2022
11:40-13:10
session 9

Yella Nishant[1], Eppakayala Manisai[2], Bhargavi Divyasri[3], Pasha Tauqir[4]
[1]:nishanty1219@gmail.com, [2]:manisaieppakayala1@gmail.com,
[3]:divyasri2018@mrec.ac.in, [4]:tauqirp003@gmail.com

Malla Reddy Engineering College, India

**Summary**

The Southern Resident Killer Whale (Orcinus Orca) is an apex predator in the oceans. Currently, these are listed as endangered species and have slowly declined in number over the past two decades. There is a lack of availability of data on audio vocalizations of killer whales, which in itself creates a demanding task to acquire labelled audio sets. The vocalizations of orcas are usually categorized into two groups namely, whistles and pulsed calls. There is a significant amount of scarcity on audio sets of these two types of vocalizations. Hence this creates a challenge to address the lack of availability of data on these vocalizations. Methods of data augmentations have proven over the years to be very effective in generating synthetically created data for the use of labelled training of a given feed-forward neural network. The Transformer based Generative Adversarial neural network (Trans-GAN) has performed phenomenally well on tasks pertaining to visual perception. In this paper, we would like to demonstrate the use of trans-GAN on audio datasets, which would be used to perform bioacoustics augmentation of the killer whale audio vocalizations obtained from existing open-source libraries to generate a synthetically substantial amount of audio data on the killer whale vocalizations for tasks pertaining to audio perception. To validate the Trans-GAN generated audio to the original killer Whale vocalization sample, we have implemented a time-sequence-based algorithm called Dynamic Time Wrapping (DTW), which compares the similarity index between these two audio samples.

---

## A microfluidic system, utilising image processing methods, for the detection of blood coagulation and erythrocyte aggregation

7 Dec 2022
11:40-13:10
session 9

Louka Marinos[1], Passos Andreas[2], Inglezakis Antonis[3], Loizou Constantinos[4], Kaliviotis Efstathios[5]
[1]:ma.louka@edu.cut.ac.cy, [2]:andreas.passos@cut.ac.cy,
[3]:a.inglezakis@embiodiagnostics.eu, [4]:k.loizou@embiodiagnostics.eu,
[5]:e.kaliviotis@cut.ac.cy

[1,2]:Cyprus University of Technology, Cyrpus,
[3,4]:EMBIO Diagnostics, Cyrpus

**Summary**
Hemostasis is a defence mechanism that prevents blood losses in cases of vessel injuries, and other related disorders. In many cases, patients need to frequently

monitor their blood coagulation tendency in order to regulate their medication. In addition, red blood cell aggregation (RBCA) is related to blood inflammation, and it appears elevated in many pathological conditions. Blood coagulation and RBCA can be studied by analysing the dynamic changes of light transmittance though a clotting/aggregating sample, and indeed various works in the literature exploit this approach. In this work, blood coagulation and RBCA are examined by utilising single drops of blood in an inexpensive camera-based microfluidic system, designed for low computational and production cost. Results are compared with a microscopy-camera system, with both setups utilizing the same custom made microchannel. Three image processing algorithms are developed to analyze the averaged light intensity, and the local structural chracteristics of blood, through a binarization and region classification method, using logical operations. The results illustrate the repeatability of the technique and the donor-to-donor variation within the proposed approach. Based on the image processing analysis, the developed coagulation and aggregation indices show great potential of utilisation in an inexpensive and robust point of care device.

---

7 Dec 2022
11:40-13:10
session 9

## A Novel Resource-Constrained Insect Monitoring System based on Machine Vision with Edge AI

Kargar Barzi Amin[1], P. Wilk Mariusz[2], Zorbas Dimitrios[3], T. Gaffney Michael[4], O'Flynn Brendan[5]

[1]:amin.kargar@tyndall.ie,[2]:mariusz.wilk@tyndall.ie, [3]:dimzorbas@ieee.org, [4]:Michael.gaffney@teagasc.ie, [5]:brendan.oflynn@tyndall.ie

[1,2,5]:Tyndall National Institute, Ireland,
[3]:Department of Computer Science Nazarbayev University, Ireland,
[4]:Horticulture Development Department Teagasc Ashtown Food Research Centre, Ireland

### Summary

Effective insect pest monitoring is a vital component of Integrated Pest Management (IPM) strategies. It helps to support crop productivity while minimising the need for plant protection products. In recent years, many researchers have considered the integration of intelligence into such systems in the context of the Smart Agriculture research agenda. This paper describes the development of a smart pest monitoring system, developed in accordance with specific requirements associated with the agricultural sector. The proposed system is a low-cost smart insect trap, for use in orchards, that detects specific insect species that are detrimental to fruit quality. The system helps to identify the invasive insect, Brown Marmorated Stink Bug or Halyomorpha halys (HH) using a Microcontroller Unit-based edge device comprising of an Internet of Things enabled, resource-constrained image acquisition and processing system. It is used to execute a lightweight image analysis algorithm and a Convolutional Neural Network (CNN) model for insect detection and classification, respectively. The prototype device is currently deployed in an orchard

in Italy. The preliminary experimental results show over 70 percent of accuracy in HH classification, demonstrating the proposed system feasibility and effectiveness in monitoring this invasive insect species.

---

## Human Dendritic Cells Classification based on Possibility Theory

7 Dec 2022
11:40-13:10
session 9

Mouna Zouari Mehdi [1], Abdesslam Benzinou [2], Jihen Frikha Elleuch [3], Kamal Nasreddine [4], Ammari DhiaEddine[5], Dorra Sellami[6]

[1]:mouna.zouari.enis@gmail.com,
[2]:benzinou@enib.fr,[3]:jihen_frikha4@yahoo.fr,[4]:nasreddine@enib.fr,
[5]:Ammaridhiaedin@gmail.com, [5]:dorra.sellami@enis.tn

[1]:Digital Research Center of Sfax, P.B275, Sakiet Ezzit, 3021 Sfax, Tunisia,
[2,4]:Brest National Engineering school of Brest, France,
[3,5,6]:Control and Energy Management Laboratory, National Engineering school of Sfax, University of Sfax, Tunisa.

**Summary** Dendritic cells can be seen as a mirror of our immune system. Based on their in vitro analysis, biological experts are now able to study the impact of food contaminants on the human immune system. Accordingly, a visual characterization of dendritic cell morphology can provide an indirect estimation of the toxicity. In this paper, we propose an automatic classification of dendritic cells that could serve as a second non-subjective opinion for pathologists. The proposed approach is built on pre-processing steps for segmentation and cell detection in microscopic images. Then, a set of features such as shape descriptors are extracted for cell characterization. At this step, three cell classes are distinctively identified by experts. Nevertheless, a high ambiguity is revealed between cell classes. Possibility theory can offer a realistic framework for making reliable decisions under high ambiguity. It exploits a human natural concept of the implicit use of probability distribution for deciding on the possibility of some assertions in some contexts where a cognitive conflict is observed while interfering existing related postulates, leading to high ambiguity. Based on the consistency concept of Dubois and Prade, a transformation of the probability into a possibility distribution is undertaken. Under possibility paradigm, a further feature selection in the possibility space using the Shapely index. Compared to state-of-the art methods the proposed approach yielded on a real dataset of nearly 630 samples an improvement in terms of the mean precision rate, the Recall rate, and the F1-measure.

---

# Keynote VI by Prof. Fabio Solari Session chair: Prof. Riadh Abdelfattah (Wednesday 7 Dec 2022 at 14:00-14:40)

## Computational Models for Ecological perception and interaction in Virtual and Augmented Reality

5 Dec 2022
12:20-13:00

Fabio Solari

University of Genoa, Italy

**Summary**

A bio-inspired computational model of visual perception for action tasks is proposed to provide clues to better design virtual and augmented reality (VR and AR) systems. The proposed neural model is based on space-variant mapping, disparity, and optic flow computation by implementing paradigms of the dorsal visual processing stream. The cortical representation of the visual information is directly exploited to infer features related to the real world without devising ad-hoc computer vision algorithms. Besides artificial vision applications, the proposed model can mimic and describe human behavioral data of both motion and depth perception. By leveraging previous outcomes, we can employ the modeled perception to improve the experience in VR and AR environments: as a case study, to implement a foveated depth-of-field blur that mitigates cybersickness.

**Fabio Solari biography**

Fabio Solari is Associate Professor of Computer Science at the Department of Informatics, Bioengineering, Robotics and Systems Engineering of the University of Genoa. His research activity concerns the study of visual perception with the aim to design novel bio-inspired artificial vision systems and to develop natural human-computer interaction techniques in virtual and augmented reality. In particular,

his research interests are related to: (i) neural algorithms for motion and depth estimation, space-variant visual processing and scene interpretation; (ii) perceptual assessment of virtual/augmented reality systems and the development of systems that allow a natural experience and ecological human-computer interactions in mixed reality environments. He teaches "Augmented Reality" for the master of Computer Science, "Software Technologies for HCI" for the master of Bioengineering, and "Computer Vision" for the European Master on Advanced Robotics. He is Member of the Board of the Doctoral Course in Bioengineering and Robotics at University of Genoa, and Member of the Faculty Joint Commission for teaching and the right to study. He has participated to eight European projects: Interreg Alcotra CLIP "E-Santé/Silver Economy", PROSOL "Jeune" and PROSOL "Senior"; FP7-ICT, EYESHOTS and SEARISE; FP6-IST-FET, DRIVSCO; FP6-NEST, MC-COOP; FP5-IST-FET, ECOVISION. He is a reviewer for Italian PRIN and FIRB projects, and EU Marie Curie fellowships and ERC. He has a pending International Patent Application (WO2013088390) on augmented reality, and two Italian Patent Applications on virtual (No. 0001423036) and augmented (No. 0001409382) reality. More information is available at http://sites.google.com/site/fabiosolarifs/

---

# Regular Session 10: Low level Image Processing for image synthesis and fusion
# Session chair: Prof. Abdesslam Benzinou
# (Wednesday 7 Dec 2022 at 14:40-16:10)

### A review of photorealistic image stylization techniques

7 Dec
14:40-16:10
Live stream
Session 10

Hassaan A Qazi
hassaanali2006@hotmail.com

University of Waterloo, Canada

**Summary**

Rendering photorealistic images from the image stylization technique is still considered as a challenging task. In this paper, we compare three recent state-of-the-art approaches. All three algorithms are mainly driven by Convolution Neural Network (CNN) technique. A brief discussion of the selected approaches is followed by some comparisons and results. Both Structural Similarity Index (SSIM) and Learned Perceptual Image Patch Similarity (LPIPS) metrics are used to generate new findings of the methodologies. Finally, subjective analysis is also presented to gauge the efficacy of the algorithms in discussion.

---

7 Dec
14:40-16:10
Live stream
Session 10

## Unrolling Alternating Direction Method of Multipliers for Visible and Infrared Image Fusion

Bakan Altuğ[1], Erer Işın[2]
[1]:bakan20@itu.edu.tr, [2]:ierer@itu.edu.tr

[1,2]:Department of Electronics and Communication Engineering Istanbul Technical University, Turkey
[1]:Microelectronics, Guidance and Electro-Optics Business Sector ASELSAN Ankara, Turkey

**Summary**

In this paper a new infrared and visible image fusion (IVIF) method which combines the advantages of optimization and deep learning based methods is proposed. This model takes the iterative solution used by the alternating direction method of the multiplier (ADMM) optimization method, and uses algorithm unrolling to obtain a high performance and efficient algorithm. Compared with traditional optimization methods, this model generates fusion with 99.6% improvement in terms of image fusion time, and compared with deep learning based algorithms, this model generates detailed fusion images with 99.1% improvement in terms of training time. Compared with the other state-of-the-art unrolling based methods, this model performs 26.7% better on average in terms of Average Gradient (AG), Cross Entropy (CE), Mutual Information (MI), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Loss (SSIM) metrics with a minimal testing time cost.

---

7 Dec
14:40-16:10
Live stream
Session 10

## A fast method for impulse noise reduction in digital color images using anomaly median filtering

Gantenapalli Srinivasa Rao[1], Choppala Praveen Babu[2], Gullipalli Vandana[3], Meka James Stephen[4], Teal Paul D[5]
[1]:srinivasa.gantenapalli@gmail.com, [2]:praveen@wistm.edu.in, [3]:vandanaece@wistm.edu.in, [4]:james@wistm.edu.in, [5]:paul.teal@vuw.ac.nz

[1]:Andhra University, India,
[2,3,4]:Wellfare Institute of Science Technology & Management, Andhra University, India
[5]:Victoria University of Wellington, Victoria, New Zeeland

**Summary**

The traditional vector median filtering and its variants used to reduce impulse noise in digital color images operate by processing over all the pixels in the image sequentially. This renders these filtering methods computationally expensive. This paper presents a fast method for reducing impulse noise in digital color images. The key idea here is to slice each row of the image as a univariate data vector, identify

impulse noise using anomaly detection schemes and then apply median filtering over these to restore the original image. This idea ensures fast filtering as rows of pixels are processed in batches. Using simulations, we show that the proposed method scales super efficiently with respect to accuracy and time.

---

## Segmentation of Shipping Bags in RGB-D Images

7 Dec
14:40-16:10
Live stream
Session 10

Vasileva Elena[1], Ivanovski Zoran A[2]

[1]:el.vasileva.7@gmail.com, [2]:mars@feit.ukim.edu.mk

[1]:Faculty of Electrical Engineering and Information Technologies Ss. Cyril, Macedonia
[2]:Methodius University Skopje, North Macedonia, Macedonia

**Summary**

This paper presents a convolutional neural network for segmenting partially transparent shipping bags in RGB-D images of cluttered scenes containing different packaging items in unstructured configurations, optimized for training with a limited number of samples with high variability. The analysis of the results with regards to the input type, network architecture, and lighting conditions proves that including low-resolution depth information improves the segmentation of objects with similar color and objects in previously unseen lighting conditions, and the high resolution color photographs greatly improve the segmentation of details, thus motivating the proposed multi-input architecture with feature fusion to fully utilize the benefits of high-resolution photographs and low-resolution depth information. The proposed architecture performs successful segmentation of shipping bags in a cluttered environment among packages and items of different color and materials with irregular shape and provides accuracy improvement over well-known semantic segmentation architectures while significantly reducing the required processing time, making it suitable for real-time application.

---

## Visual Data Enciphering via DNA Encoding, S-Box, and Tent Mapping

7 Dec
14:40-16:10
Live stream
Session 10

Gabr Mohamed K[1], Younis Hana[2], Ibrahim Marwa[3], Alajmy Sara[4], Alexan Wassim[5]

[1]:mohamed.kamel-gabr@guc.edu.eg, [2]:hana.younis@student.guc.edu.eg, [3]:marwa.mohamedibrahim@student.guc.edu.eg, [4]:sara.alajmy@student.guc.edu.eg, [5]:wassim.joseph@guc.edu.eg

German University in Cairo, Egypt

**Summary**

The ever-evolving nature of the Internet and wireless communications, as well as the production of huge amounts of multimedia every day has created a dire need for their security. In this paper, an image encryption technique that is based on 3 stages is proposed. The first stage makes use of DNA encoding. The second stage proposed and utilizes a novel S-box that is based on the Mersenne Twister and a linear descent algorithm. The third stage employs the Tent chaotic map. The computed performance evaluation metrics exhibit a high level of achieved security.

---

## C-ESRGAN: Synthesis of super-resolution images by image classification

7 Dec
14:40-16:10
Live stream
Session 10

Liu, Jingan[1], N.P., Chandrasiri[2]

[1]:em22028@ns.kogakuin.ac.jp, [2]:chandrasiri@kogakuin.ac.jp

[1]:Intelligent Media Laboratory, Kogakuin University, Japan,
[2]:Kogakuin University, Japan

**Summary**

With the development of deep learning, super-resolution image synthesis techniques for enhancing low-resolution images have advanced remarkably. However, mainstream algorithms focus on improving the quality of the entire image on average and this may result in blurring. In this paper, we propose three key components for synthesizing super-resolution images that can reflect the fine details of an image. We synthesize super-resolution images by image classification. First, the neural network weights learned using the images in the same image category were utilized in synthesizing super-resolution images. For this purpose, image classification was performed using a transfer-trained ResNet. Second, SENet was applied to the generators in our proposed method to obtain detailed information about the images. Finally, the feature extraction network was changed from VGG to ResNet in order to get more important features. As a result, we achieved better image evaluation values (PSNR, NIQE) for the super-resolution images of dogs and cats compared to the previous studies. Furthermore, the images were generated more naturally on the benchmark dataset.

---

# Regular Session 11: Image Understanding and Biological Applications
# Session chairs: Prof. Hassan Rabah
# (Wednesday 7 Dec 2022 at 16:20-17:20)

## Fractional Vegetation Cover Estimation using Hough Lines and Linear Iterative Clustering

7 Dec 2022
16H20-17H20
Session 11
Live stream

Venkata Siva Kumar Margapuri [1], Trevor Rife [2], Chaney Courtney [3], Kai Zhao [4], Mitchell Neilsen [5], Brandon Schlautman [6]

[1]:marven@ksu.edu, [2]:trife@ksu.edu,
[3]:chaneylc@ksu.edu, [4]:kaizhao@ksu.edu
[5]:neilsen@ksu.edu, [6]:schlautman@landinstitute.org

[1,4]:Department of Computing Sciences Villanova University Villanova, Philadelphia, United States of America
[2,3]:Department of Plant and Environmental Sciences Clemson University Clemson, SC, United States of America
[4]:The Land Institute Salina, KS, United States of America
[5,6]:Department of Computer Science Kansas State University Manhattan, KS, United States of America

**Summary**

A common requirement of plant breeding programs across the country is companion planting – growing different species of plants in close proximity so they can mutually benefit each other. However, the determination of companion plants requires meticulous monitoring of plant growth. The technique of ocular monitoring is often laborious and error prone. The availability of image processing techniques can be

used to address the challenge of plant growth monitoring and provide robust solutions that assist plant scientists to identify companion plants. This paper presents a new image processing algorithm to determine the amount of vegetation cover present in a given area, called fractional vegetation cover. The proposed technique draws inspiration from the trusted Daubenmire method for vegetation cover estimation and expands upon it. Briefly, the idea is to estimate vegetation cover from images containing multiple rows of plant species growing in close proximity separated by a multi-segment PVC frame of known size. The proposed algorithm applies a Hough Transform and Simple Linear Iterative Clustering (SLIC) to estimate the amount of vegetation cover within each segment of the PVC frame. When applied as a longitudinal study, this analysis provides crucial insights into plant growth. As a means of comparison, the proposed algorithm is compared with SamplePoint and Canopeo, two trusted applications used for vegetation cover estimation. The comparison shows a 99% similarity with both SamplePoint and Canopeo demonstrating the accuracy and feasibility of the algorithm for fractional vegetation cover estimation.

---

7 Dec 2022
16H20-17H20
Session 11
Live stream

## Drought Stress Segmentation on Drone captured Maize using Ensemble U-Net framework

Tejasri Nampally[1], G Ujwal Sai[2], Rajalakshmi pachamuthu [3], Balaji Banothu [4], Uday B Desai [5]

[1]:ai19resch11002@iith.ac.in, [2]:ai20mtech01001@iith.ac.in,
[3]:Raji@iith.ac.in, [4]:balajinaikbanoth789@gmail.com
[5]:ubdesai@iith.ac.in

[1,2,3,5]:Indian Institute of Technology Hyderabad, India,
[4]:Professor Jayashankar Telangana State Agricultural University (PJTSAU), Hyderabad, India

**Summary**

Water is essential for any crop production. Lack of sufficient supply of water supply causes abiotic stress in crops. Accurate identification of the crops affected by drought is required for achieving sustainable agricultural yield. The image data plays a crucial role in studying the crop's response. Recent developments in aerial-based imaging methods allow us to capture RGB maize data by integrating an RGB camera with the drone. In this work, we propose a pipeline to collect data rapidly, pre-process the data and apply deep learning based models to segment drought affected/stressed and unaffected/healthy RGB maize crop grown in controlled water conditions. We develop an ensemble-based framework based on U-Net and U-Net++ architectures for the drought stress segmentation task. The ensemble framework is based on the stacking approach by averaging the predictions of fine-tuned U-Net and U-Net++ models to generate the output mask. The experimental results showed that the ensemble framework performed better than individual U-Net and U-Net++ models on the test set with a mean IoU of 0.71 and a dice coefficient of 0.74.

---

## Animal Video Retrieval System using Image Recognition and Relationships Between Concepts of Animal Families and Species

7 Dec 2022
16H20-17H20
Session 11
Live stream

Chinatsu Watanabe [1], Chandrasiri N.P. [2], Mayu Kaneko [3]
[1]:em22029@g.kogakuin.jp, [2]:chandrasiri@kogakuin.ac.jp [3]:mayu.kaneko@mizuho-ir.co.jp

[1,3]: Intelligent Media Processing Laboratory, Kogakuin University Tokyo, Japan
[2]:Mizuho Research & Technologies, Ltd., Tokyo, Japan

**Summary**

In recent years, video streaming services have become increasingly popular. In general, the search function in a video sharing service site evaluates the relevance of a search query to the title, tags, description, and so on given by the creator of the video. Then, the search results with the highest relevance are displayed. Therefore, if a title is given to a video that does not match its content, there is a possibility that a video with low relevance will be found. In this research, (1) we built a new system that retrieves animal videos that are relevant to its content using image recognition. (2) By describing the relationships between the concepts of animal families and species and incorporating them into the retrieval system, it is possible to retrieve animal videos by their family names. Adding retrieval by animal family name enabled us to find species that have not been learned. In this research, (3) we confirmed the usefulness of our video retrieval system using trained neural networks, GoogLeNet and ResNet50, as animal species classifiers.

---

## White Flies and Black Aphids Detection in Field Vegetable Crops using Deep Learning

7 Dec 2022
16H20-17H20
Session 11
Live stream

Nikolaos Giakoumoglou[1], Eleftheria Maria Pechlivani: [2], Nikolaos Katsoulas [3], Dimitrios Tzovaras[4]
[1]: ngiakoumoglou@iti.gr, [2]:riapechl@iti.gr,
[3]:nkatsoul@uth.gr, [4]:dimitrios.tzovaras@iti.gr

[1,2,4]:Centre for Research and Technology Hellas, Greece
[3]:University of Thessaly, Greece

**Summary**

Digital image processing for the early detection of plant pests as insects in vegetable crops is essential for plant's yield and quality. In recent years, deep learning has made strides in the digital image processing, opening up new possibilities for pest monitoring. In this paper, state-of-the-art deep learning models are presented to detect common insect pests in vegetable cultivation named whiteflies and black aphids. Due to the absence of data sources addressing the aforementioned insect pests, adhesive traps for catching the target insects were used for the creation of

an annotated image dataset. In total 225 images were collected, and 5904 insect instances were labelled by expert agronomists. This dataset faces many challenges such as the tiny size of objects, occlusions and resemblance. Object detection models were used like YOLOv3, YOLOv5, Faster R-CNN, Mask R-CNN, and RetinaNet as baseline algorithms for benchmark experiments. For achieving accurate results, data augmentation was used. This study has addressed these challenges by applying deep learning models which are able to deal with tiny object detection ascribed to very small insect size. The experiment results exhibit a mean Average Precision (mAP) of 75%. Dataset is available for download at https://zenodo.org/record/7139220.

---