The Fourth IEEE International Image Processing, Applications and Systems Conference (IPAS'20)

The National Engineering School of Sfax, Sfax University, Tunisia
Genoa University, Italy
INRIA, Sophia Antipolis, France
was planned to be hosted by Genoa University, Italy
The conference is scheduled online
9-11 December 2020

The Fourth IEEE IPAS is sponsored by:

IEEE Italy Section
IEEE Italy Section Computer Chapter
IEEE Italy Section Engineering in Medicine and Biology Chapter
IEEE Tunisia Section Signal Processing Chapter
IEEE Italy Section Oceanic Engineering Chapter
Association Boussole de la Recherche Scientifique- ABORS Tunisia

Preface

IPAS 2020 will take place in Genova, Italy during December 9-11, 2020. Due to covid-pandemic problems that may be encountered, the safety of our authors is our priority, the conference is managed 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 fourth IEEE International Conference of Image Processing Applications and Systems has seven sessions composed of 35 regular papers from 22 countries (Algeria, Bengladesh, Canada, Cyprus, Japan, Kazakhstan, Korea, France, Indonesia, Italy, Lybia, Morocco, Philippine, Portugal, Russia, South Africa, Spain, Sri Lanka, Tunisia, Turkey, United Kingdom, United States of America). The review process has been double blinded. On behalf the program committee of IPAS'20, we are grateful to all the authors for their fine contributions and for the high quality of their papers.

The fourth IEEE International Conference of Image Processing Applications and Systems has seven invited keynote speakers on different topics of image processing applications. On behalf the organizing committee of IPAS'20, we would like to thank all the speakers for their support and their valuable contributions.

General Co-Chairs
Dorra Sellami
François Bremond
Fabio Solari

Contents

Committees

General Chairs

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

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

Anthony Fleury (IMT-Lille-Douai, France)
Cédric Wemmert (Strasbourg, France)
Carlos Fernando Crispim Junior (Lyon, France)
Hassan Rabah (Nancy, France)
Abdesslam Benzinou (Brest, France)
Abdellatif Mtibaa (ENIM, Tunisia)

Exhibits and Industry Liaison Chairs

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

Local Arrangement Chairs

Antitza Dantcheva (Sophia Antipolis, France)
Abhijit Das (Sophia Antipolis, France)
Chiara Bassano (University of Genoa, Italy)
Giorgio Ballestin (University of Genoa, Italy)
Adel Khlifi (Emirates)
Mohamed Krid (ENIS, Tunisia)

Publicity Chairs

Manuela Chessa (University of Genoa, Italy)
Randa Boukhris (ISIMS, Tunisia)
Norhene Gargouri (CRNS, Tunisia)
Mondher Frikha (ENET-COM, Tunisia)
Mohamed Atri (ENIM, Tunisia)
Zied Lachiri (ENIT, Tunisia)
Ratnesh Kumar (Nvidia, USA)
Aicha Bouzid (ENIT, Tunisia)
Hassene Seddik (ENSIT, Tunisia)
Slim Yacoob (SupCom, Tunisia)

Plenary, tutorial and special session committee members

Sergio Cerutti (Politecnico di Milano, Italy)
Nawres Khlifa (Tunisia)
Mounir Sayadi (ENSIT, Tunisia)
Aicha Bouzid (ENIT, Tunisia)
Ali Douik (ENIM, Tunisia)
Monia Turki (ENIT, Tunisia)
Sabeur Abid (ENSIT, Tunisia)

Registration Chairs

Mohamed Krid (ISGI, Tunisia)
Julien Moreau (IFSTTAR, France)

Publication Chairs

Hejer Elloumi (Sfax, Tunisia) Marco Domenico Santambrogio (Politecnico di Milano, Italy)
Kamel Abderrahim (Gabes, Tunisia)
Kamel Hamrouni (Tunis, 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)
Rihab Idoudi (TN)
Tarek Ouni (TN)
Hela Boulemi Chtara (TN)
Tièche François(CH)
Antonio Rodríguez-Sanchez(AUS)
Alaa Balti (TN)
Nicolas Thome (FR)
Jean-Marc Odobez (SWZ)
Chetouani (FR)
Veronique Cherfaoui (FR)
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)
Mohamedali Mahjoub(TN)
Boughariou Jihene(TN)
Hanene Sehli (TN)
Carlos Fernando (FR)
Javier Diaz(ES)
Sabeur Abid (TN)
Dominique gruyer (FR)
Sergio Velastin (ES)
Catherine Achard (FR)
Philippe Bonnifait (FR)
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)	Samia Fraj Hachmi(TN)
Nedra Ben Romdhane (TN)	Issam Jabri (TN)
Ikbel Azaiez (TN)	Emna BenSlimene (TN)
Nadia Sraieb (TN)	Hassan Makki (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 Giovanni Maria Farinella (Live Stream1)

Future Predictions in Egocentric Vision

9 Dec
9:00-9:50 am
Live stream1

Giovanni Maria Farinella
University of Catania, Italy

Summary

The ability to predict the future is fundamental for humans to explore the world, learn, perceive, navigate, and act. It is hence expected that future AI systems shall be able to reproduce such abilities, reasoning about the world in terms of events and stimuli which live in the near or distant future. Predictive abilities are also fundamental for wearable systems to understand the user's short- and long-term goals, offer appropriate guidance based on the user's objectives, and improve user's safety anticipating future actions. In this talk, we will present recent research on future predictions from egocentric video which has been carried out at the Image Processing Laboratory (IPLAB) at the University of Catania, Italy. We will first introduce the main motivations behind research on egocentric perception, then discuss approaches to predict future interacted objects and actions from egocentric video. The talk will also focus on the relevant datasets to support the study of future prediction tasks from egocentric video, such as the EPIC-KITCHENS series of datasets and challenges, as well as our newly introduced MECCANO dataset for studying human-object interaction recognition and future prediction in industrial-like scenarios.

Giovanni Maria Farinella biography

Giovanni Maria Farinella is an Associate Professor at the Department of Mathematics and Computer Science, University of Catania, Italy. His research interests lie in the fields of Computer Vision and Machine Learning with focus on First Person (Egocentric) Vision. He is author of more than 120 papers in international book chapters, journals and conference proceedings, and co-inventor of 6 patents involving industrial partners. Dr. Farinella serves as a reviewer and on the programme committee board of major international journals and conferences (CVPR, ICCV,

ECCV, BMVC). He has been Area Chair for ICCV 2017/19, CVPR 2020/21, Video Proceedings Chair for ECCV 2012 and ACM MM 2013, Guest Editor for Special Issues on Computer Vision and Image Understanding, Pattern Recognition Letters, and IEEE Journal of Biomedical and Health Informatics. He is currently Associate Editor of the international journals IEEE Transactions Pattern Analysis and Machine Intelligence (2019-) Pattern Recognition (2017-) and IET Computer Vision (2015-). Dr. Farinella founded (in 2006) and currently directs the International Computer Vision Summer School (ICVSS). He was awarded the PAMI Mark Everingham Prize in October 2017.

Session 1: Image understanding and activity recognition (Live Stream1)

AI-Enabled High-Level Layer for Posture Recognition Using The Azure Kinect in Unity3D

9 Dec
10:30-12:10 am
Live stream1

Hamza Alaoui[1], Mohamed Tarik Moutacalli [2] Mehdi Adda [3]

Hamza.Alaoui@uqar.ca[1], MohamedTarik_Moutacalli@uqar.ca[2] Mehdi_Adda@uqar.ca[3]

University of Quebec in Rimouski Lèvis, Canada

Summary

Posture recognition is one of the challenging tasks in computer vision. It lays on top of the pose estimation of the different body joints, and can be used in many applications. In the medical field, it can serve to assist patients in rehabilitation. In games it can be an elegant form of computer human interaction. Different Artificial Intelligence techniques were used over the years to precisely output the joint positions of the body from a single or stream of images. One of the great solutions that tacked well the pose estimation challenge is the Kinect camera, however further process is required to create and detect body postures. This article presents a customizable high-level layer that allows its users to easily create and manage body postures in unity3d projects allowing them to focus more on the other aspects of their project. The layer offers two detection methods, both scored more than 95% accuracy in each of the tested postures.

Automotive Interior Sensing - Towards a Synergetic Approach between Anomaly Detection and Action Recognition Strategies

9 Dec
10:30-12:10 am
Live stream1

Pedro Augusto[1], Jaime S. Cardoso [2] Joaquim Fonseca [3]
up201503495@fe.up.pt[1], jaime.cardoso@inesctec.pt[2] Joaquim.Fonseca2@pt.bosch.com [3]

University of Porto, Portugal

Summary

With the appearance of Shared Autonomous Vehicles there will no longer be a driver responsible for maintaining the car interior and well-being of passengers. To counter this, it is imperative to have a system that is able to detect any abnormal behaviors, more specifically, violence between passengers. Traditional action recognition algorithms build models around known interactions but activities can be so diverse, that having a dataset that incorporates most use cases is unattainable. While action recognition models are normally trained on all the defined activities and directly output a score that classifies the likelihood of violence, video anomaly detection algorithms present themselves as an alternative approach to build a good discriminative model since usually only non-violent examples are needed. This work focuses on anomaly detection and action recognition algorithms trained, validated and tested on a subset of human behavior video sequences from Bosch's internal datasets. The anomaly detection network architecture defines how to properly reconstruct normal frame sequences so that during testing, each sequence can be classified as normal or abnormal based on its reconstruction error. With these errors, regularity scores are inferred showing the predicted regularity of each frame. The resulting framework is a viable addition to traditional action recognition algorithms since it can work as a tool for detecting unknown actions, strange/violent behaviors and aid in understanding the meaning of such human interactions.

Audiovisual Classification of Group Emotion Valence Using Activity Recognition Networks

9 Dec
10:30-12:10 am
Live stream1

João Ribeiro Pinto[1], Tiago Gonçalves[2],Carolina Pinto [3], Luís Sanhudo[4], Joaquim Fonseca[5], Filipe Gonçalves[6], Pedro Carvalho[7], Jaime S. Cardoso [8]

[1]: joao.t.pinto@inesctec.pt, [2]: tiago.f.goncalves@inesctec.pt, [3]: up201506006@fe.up.pt, [4]: luis.sanhudo@inesctec.pt, [5]: joaquim.fonseca2@pt.bosch.com, [6]: filipe.goncalves@pt.bosch.com, [7]:pedro.m.carvalho@inesctec.pt, [8]:jaime.cardoso@inesctec.pt

[1], [3], [4], [7] , [8]: University of Porto, Portugal, [1], [2], [4], [7], [8]:INESC TEC,[5], [6]: Bosch Car Multimedia

Summary

Despite recent efforts, accuracy in group emotion recognition is still generally low. One of the reasons for these underwhelming performance levels is the scarcity of available labeled data which, like the literature approaches, is mainly focused on still images. In this work, we address this problem by adapting an inflated ResNet-50 pretrained for a similar task, activity recognition, where large labeled video datasets are available. Audio information is processed using a Bidirectional Long Short-Term Memory (Bi-LSTM) network receiving extracted features. A multimodal approach fuses audio and video information at the score level using a support vector machine classifier. Evaluation with data from the EmotiW 2020 AV Group-Level Emotion sub-challenge shows a final test accuracy of %65.74 for the multimodal approach, approximately %18 higher than the official baseline. The results show that using activity recognition pretraining offers performance advantages for group-emotion recognition and that audio is essential to improve the accuracy and robustness of video-based recognition.

Towards real-time activity recognition

9 Dec
10:30-12:10 am
Live stream1

Shane Reid[1], Philip Vance [2], Sonya Coleman [3],Dermot Kerr[4], Siobhan ONeill[5]
Reid-S22@ulster.ac.uk[1], p.vance@ulster.ac.uk[2], sa.coleman@ulster.ac.uk [3],
d.kerr@ulster.ac.uk[4], sm.oneill@ulster.ac.uk [5]

Ulster University, United Kingdom

Summary

Activity recognition relates to the automatic visual detection and interpretation of human behaviour and is emerging as an active domain of computer vision. It has important applications such as identifying individuals who are at risk of suicide in public locations such as bridges or railway stations. These individuals are known to exhibit easily observable activities and behaviours such as pacing, looking up and down the railway tracks, and leaving objects on the platform. In order to detect these behaviours, an approach to individual person activity recognition is needed which can run in real time and monitor multiple individuals in parallel. We present a method for human activity recognition using skeletal keypoints and investigate how using varying sample rates and sequence lengths impacts accuracy. The results show that for any given sequence length, optimising the sample rate can result in an overall increase in classification accuracy and improvement in run-time. Results demonstrate that finding the optimal time period over which to sample frames is more important than simply decreasing the number of frames sampled. Further, we show that keypoint based activity recognition approaches outperform other state of the art approaches. Finally, we show that this approach is fast enough for real time activity recognition when up to 14 people are present in the image whilst maintaining a high degree of accuracy.

---

9 Dec
10:30-12:10 am
Live stream1

Emotion Recognition on large video dataset based on Convolutional Feature Extractor and Recurrent Neural Network

Muhammad Fahim[1], Denis Rangulov [2]
m.fahim@innopolis.ru[1],d.rangulov@innopolis.ru [2]

Innopolis University, Russia

Summary

For many years, the emotion recognition task has remained one of the most interesting and important problems in the field of human-computer interaction. In this study, we consider the emotion recognition task as a classification as well as a regression task by processing encoded emotions in different datasets using deep learning models. Our model combines convolutional neural network (CNN) with recurrent neural network (RNN) to predict dimensional emotions on video data. At the first step, CNN extracts feature vectors from video frames. In the second step, we fed these feature vectors to train RNN for exploiting the temporal dynamics of video. Furthermore, we analyzed how each neural network contributes to the system's overall performance. The experiments are performed on publicly available datasets including the largest modern Aff-Wild2 database. It contains over sixty hours of video data. We discovered the problem of overfitting of the model on an unbalanced dataset with an illustrative example using confusion matrices. The problem is solved by downsampling technique to balance the dataset. By significantly decreasing training data, we balance the dataset, thereby, the overall performance of the model is improved. Hence, the study qualitatively describes the abilities of deep learning models exploring enough amount of data to predict facial emotions. Our proposed method is implemented using Tensorflow Keras. The code is publicly available in repository.

Keynote II by Dimitri Ognibene (Live stream1)

Adaptive Vision for Human Robot Collaboration

9 Dec
14:00-14:50 pm
Live stream1

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.

Session 2: Low level image processing and Image Retrieval (Live Stream1)

Statistical and spatial information association for clusters number and mean values estimation in image segmentation

9 Dec
15:00-17:00 pm
Live stream1

Aicha Baya Goumeidane[1], Nafaa Nacereddine [2]
a.goumeidane@crti.dz,[1], n.nacereddine@crti.dz[2]

Research Center in Industrial, Algeria

Summary

A straightforward way to assess the number of clusters as well as their mean values for clustering-based gray scale image segmentation initialization purpose, is presented in this paper. It is based on an association of statistical and spatial information of the image so that, the final classes in the segmented image correspond to (meaningful) modes in an histogram built from the image characteristics. The experiments testify the efficiency of this method in case of synthetic and real images.

Scale Proposals for Region dependent Super Resolution for Object Detection

9 Dec
15:00-17:00 pm
Live stream1

Kazutoshi Akita[1], Muhammad Hari [2], Norimichi Ukita [3]
[1]: sd19401@toyota-ti.ac.jp, [2]: muhammad.haris@bukalapak.com, [3]: ukita@toyota-ti.ac.jp

Toyota Technological Institute of Nagoya, Japan

Summary

This paper presents a method for estimating objectscale proposals applied to super resolution (SR) for scaleoptimized object detection. With the region-dependent scale proposals, we achieve scale-independent object detection. This object detection scheme consists of three functions; regiondependent scale proposals, SR, and object detection. While SR and object detection have been fused in deep end-to-end

networks in previous works, region-dependent scale proposals are not provided or are performed independently of SR and object detection processes. The proposed region-dependent scaleproposal network is designed to explicitly estimate appropriate SR scales depending on the image region in accordance with scene contexts. Qualitative and quantitative experimental results show that our method can provide appropriate SR'scales for improving detection accuracy. Our proposed method gains 2.7 points in AP with Centernet used as the base detector. Index Terms Super resolution, Object Detection, Object scales, Region-dependent scale proposals.

Analysis of Speckle Tracking Methods: Correlation and RF Interpolation

9 Dec
15:00-17:00 pm
Live stream1

Brandon Rebholz[1], Mohamed Almekkawy [2]

[1]: $ORCID$: $0000-0003-1631-5499$, [2]: mka9@psu.edu

Penn State University, USA

Summary

Speckle tracking are well known methods in ultrasound imaging to estimate the motion of each pixel from consecutive frames. Due to the low lateral resolution of ultrasound images, many methods of speckle tracking rely on interpolation to generate subsample accurate displacement. This interpolation is commonly performed on the correlation map generated by the sample level cross correlation calculation. This study applies interpolation to the correlation map, a standard speckle tracking practice, and compares this to data that instead interpolates the radio frequency (RF) data before correlation, which is shown to be beneficial. Estimation accuracy and the shape of the correlation map are compared for both interpolation of the correlation map and interpolation prior to correlation. These methods are tested on data sets generated by Field II to simulate an elastography phantom, with subsample lateral displacement. The RF interpolation method is more accurate than correlation interpolation with cubic spline fitting, showing that interpolation at the correlation level is not perfectly analogous to correlating an interpolated RF image.

Improving CNN-based colorization of BW photographs

9 Dec
15:00-17:00 pm
Live stream1

Sanae Boutarfass[1], Bernard Besserer [2]
[1]: sanae.boutarfass1@univ-lr.fr, [2]: bernard.besserer@univ-lr.fr

University of La Rochelle, France

Summary

Colorization is the process of converting a black and white image - lot of different shades of gray - into a realistic colour image. Our contribution highlights the fact that the CNNs used for colorization (and specifically the data set used for their training) are not adequate to colourize legacy BW pictures. In fact the data sets are exclusively composed of colour images, and turning these colour images into greyscale pushes the CNN to learn the contribution of each colour to the resulting luminance. Anyway, CNN are quite good in classification task and can easily recognize skies, trees and foliage, faces and persons. The colorization works well for these elements, but it fails even on memorable objects such as flags and well-known monuments albeit their representation is present in the data sets. We make improvement by using a pretrained network and add hints in the colorization process. Global hints (a colour palette given to the CNN conjointly to the image) still leave a high degree of automation, and these hints could be provided once for a collection of images related to the same theme, which is very suitable for movie colorization. Adding manual hints by scribbling colours onto the BW image leads to even better results but needs user interaction.

Neural Networks for Cross-Section Segmentation in Raw Images of Log Ends

9 Dec
15:00-17:00 pm
Live stream1

Rémi Decelle[1], Ehsaneddin Jalilian [2]
[1]: remi.decelle@loria.fr, [2]: ejalilian@cs.sbg.ac.at

University of Loraine, France and University of Salzburg, Austria

Summary

In this paper, wood cross-section (CS) segmentation of RGB images is treated. CS segmentation has already been studied for computed tomography images, but few study focuses on RGB images. CS segmentation in rough log ends is an important feature for the both assessment of wood quality and wood traceability. Indeed, it allows to extract other features like pith, eccentricity (distance between the pith and the geometric centre) or annual tree rings which are related to mechanical strength. In image processing, neural networks have been widely used to solve the problem of objects segmentation. In this paper, we propose to compare different state-of-the-art neural networks for CS segmentation task. In particular, we consider U-Net, Mask R-CNN, RefineNet and SegNet. We create an imageset which has been split into

6 subsets . Considered neural networks have been trained on each subset in order to compare their performance on different type of images. Results show different behaviors between neural networks. On the one hand, overall U-Net learns better on small dataset than the others. On the other hand, RefineNet learns well on huge dataset. While SegNet is less efficient and Mask R-CNN does not provide a detailed segmentation. This offers a preliminary result on neural network performances for CS segmentation.

---

9 Dec
15:00-17:00 pm
Live stream1

News2Image: Automated System of Image Recommendation to News Articles

Zhanibek Darimbekov[1], Aslan Ubingazhibov [2], Zarina Serikbulatova [2], and Fatih M Demirci [2]

[1]: zhanibek.darimbekov@nu.edu.kz, [2]: aslan.ubingazhibov@nu.edu.kz, [3] : zarina.serikbulatova@nu.edu.kz,[4] : muhammed.demirci@nu.edu.kz

Nazarbayev University, Kazakhstan

Summary

This paper presents an image recommendation system called (News2Image), which takes as input a news text and outputs a small set of images selected from the database such that the selected images present contexts similar to the keywords of the input text. The proposed framework starts by preprocessing and summarizing the news for the selection of its keywords, which are then converted to vectors using word2vec. The framework associates each database image with a set of sentences describing its content using image captioning. After preprocessing the sentences and translating them into the word vectors as performed for input news texts, the word subspaces are compared using the Mutual Subspace Method (MSM). Experimental evaluation of the proposed framework including a comparison with the previous approach demonstrates its overall effectiveness.

Keynote III by Prof. François Bremond(Live stream2)

Video Analytics for People Monitoring

10 Dec
9:00-9:50 am
Live stream2

François Bremond
Research Center INRIA, France

Summary

In this talk, we will discuss how Video Analytics can be applied to human monitoring in general within a video camera network. Specifically, we will present an efficient technique for people detection and several techniques for tracking people in multi-camera settings (People Re-Identification). In particular, we will present several categories of algorithms to recognize human identity based on CNN, with more or less supervision performed during training. We will then present several techniques for the recognition and detection of human activities from 2D video cameras. With the emergence of deep learning and large-scale datasets from internet sources, substantial improvements have been made in video understanding. For instance, state-of-the-art 3D convolutional networks like I3D pre-trained on huge datasets like Kinetics have successfully boosted the recognition of actions from internet videos, but challenges remain to address Activities of Daily Living (ADL), such as – (i) fine-grained actions with short and subtle motion like pouring grain and pouring water, (ii) actions with similar visual patterns differing in motion patterns like rubbing hands and clapping, and finally (iii) long complex actions like cooking. In order to address these challenges, we will discuss 1) multi-modal fusion strategy, 2) pose driven attention mechanism and 3) a Temporal Model to represent long complex actions which is crucial for ADL. We will illustrate the proposed activity monitoring approaches through several home care application datasets: CAD-120, Toyota SmartHome, NTU-RGB+D, Charades and Northwestern UCLA.

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/

Session 3: Virtual reality and vision systems(Live stream2)

Virtual to Real Unsupervised Domain Adaptation for Image-Based Localization in Cultural Sites

10 Dec
10:00-11:20 pm
Live stream2

Santi Andrea Orlando[1], Antonino Furnari [2], Giovanni Maria Farinella[3]
santi.orlando@unict.it,[1], furnari@dmi.unict.it[2] , gfarinella@dmi.unict.it[3]

DMI, University of Catania DWORD-XENIA progetti s.r.l. Catania, Italy

Summary

The ability to localize the visitors of a cultural site from egocentric images can allow applications to understand where people go and what they pay attention to in the site. Current pipelines to tackle the problem require the collection and labeling of large amounts of images, which is challenging, especially in large-scale indoor environments. On the contrary, virtual images of a cultural site can be generated and automatically labeled using dedicated tools with minimum effort. In this paper, we investigate whether unsupervised domain adaptation techniques can be used to train localization models on labeled virtual data and unlabeled real data, and deploy them to work with real images. To perform this study, we propose a new dataset of both real and virtual images acquired in a cultural site which are labeled for room-based localization as well as for 3 DOF camera pose estimation. We hence compare two approaches to unsupervised domain adaptation: mid-level representations and image-to-image translation. Our analysis shows that both approaches can be used to reduce the domain gap arising from the different data sources and that the proposed dataset is a challenging benchmark for unsupervised domain adaptation for image-based localization.

---

22 Nov
2:15pm
R104

Modelling Foveated Depth-of-field Blur for Improving Depth Perception in Virtual Reality

Razeen Hussain[1], Manuela Chessa [2], Fabio Solari[3]
razeen.hussain@edu.unige.it[1], manuela.chessa@unige.it[2], fabio.solari@unige.it [3]

University of Genoa Genoa, Italy

Summary

This paper presents a technique to incorporate spatial blur effects in virtual reality devices. The considered spatial blur is based on foveation and defocus blur: concepts inspired by the human visual system. The proposed technique can be applied to any head-mounted display as a post-processing step. Our foveated depth-of-field method removes intensity leakage artifacts in the transitory regions and works in real-time. We verify the usefulness of our technique by conducting a pilot study on depth perception in virtual environments. In the conducted user study, systems integrated with our blur effect provided a better estimation of object depth in the peripheral regions.

---

22 Nov
2:15pm
R104

Imitation-Based Active Camera Control with Deep Convolutional Neural Network

PChristos Kyrkou[1]
kyrkou.christos@ucy.ac.cy

KIOS Research and Innovation Center of Excellence University of Cyprus

Summary

The increasing need for automated visual monitoring and control for applications such as smart camera surveillance, traffic monitoring, and intelligent environments, necessitates the improvement of methods for visual active monitoring. Traditionally, the active monitoring task has been handled through a pipeline of modules such as detection, filtering, and control. In this paper we frame active visual monitoring as an imitation learning problem to be solved in a supervised manner using deep learning, to go directly from visual information to camera movement in order to provide a satisfactory solution by combining computer vision and control. A deep convolutional neural network is trained end-to-end as the camera controller that learns the entire processing pipeline needed to control a camera to follow multiple targets and also estimate their density from a single image. Experimental results indicate that the proposed solution is robust to varying conditions and is able to achieve better monitoring performance both in terms of number of targets monitored as well as in monitoring time than traditional approaches, while reaching up to 25 FPS. Thus making it a practical and affordable solution for multitarget active monitoring in surveillance and smart-environment applications.

Rigid Registration of Monomodal and Multimodal Images for Wood Pieces Analysis

22 Nov
2:15pm
R104

Dahbi Radouan[1], Bombardier Vincent[2], Brie David[3], and Masson Eric [4]
radouan.dahbi@univ-lorraine.fr[1], vincent.bombardier@univ-lorraine.fr[2],
david.brie@univ-lorraine.fr[3], and eric.masson@cribois.net[4]

UCRAN-CNRS UMR 7039, Université de Lorraine Vandoeuvre-les-Nancy and
CRITT Bois Epinal, France

Summary

This article shows a comparison of rigid image registration methods of monomodal and multimodal images. These methods are applied on different images of oak wood pieces. The work presented in this article is a part of a complete vision system which aims to analyze the visual and physicochemical

aspect of oak wood piece surface. In this context, a multi-sensor acquisition using a multimodal imagery platform is performed. The acquired images are not superimposable due to the rigid deformations resulting from the use of different sensor scanners and the presence of imperfections during the image acquisition process. This leads to consider image registration as a preprocessing step. The efficiency of the registration method depends on the deformation itself and the type of image to be registered. That is why we propose a comparison of different built-in and extended MATLAB registration methods, based on cross-correlation, phase correlation, mutual information and alignment of geometric primitives. The registration evaluation of the different methods is done by using quantitative measure of image alignment, visual inspection and computational time. Finally, the choice of the most adapted image registration methods, based on the existing differences between the acquired images, in terms of type of deformations and image modality, is discussed.

Session 4: Video Compression and processing (Live stream2)

A Feature Oriented Framework and Enhanced Assessments for Imaging Compression

10 Dec
11:30-12:50am
Live stream 2

Jesse Redford[1], and Xingjie Li [2]
tredford@uncc.edu[1], xli47@uncc.edu[2]

Department of Mechanical Engineering University of North Carolina Charlotte
Charlotte, NC, USA

Summary

This paper offers a new feature-oriented compression algorithm for flexible reduction of data redundancy commonly found in images and videos streams. Using a combination of image segmentation and face detection techniques as a preprocessing step, we derive a compression framework to adaptively treat 'feature' and 'ground' while balancing the total compression and quality of 'feature' regions. We demonstrate the utility of a feature compliant compression algorithm (FC-SVD), a revised peak signal-to-noise ratio to control artificial distortion. The goal of this investigation is to provide new contributions to image and video processing research via multi-scale resolution and the block-based adaptive singular value decomposition.

Video Quality Model for Space-Time Resolution Adaptation

10 Dec
11:30-12:50am
Live stream 2

Dae Yeol Lee[1], Hyunsuk Ko [2], Jongho Kim[3], Alan C. Bovik[4]
daelee711@utexas.edu[1], hyunsuk@hanyang.ac.kr[2] pooney@etri.re.kr [3], and
bovik@ece.utexas.edu[4]

Austin, TX, USA and Ansan/Daejeon, Korea

Summary

Delivering voluminous amounts of video data through limited bandwidth channels is a challenge affecting billions of viewers. Accordingly, it is becoming more important to understand the perceptual effects that arise from various dimension reduction methodologies. Towards this direction, we propose a new video quality model

that predicts the perceptual quality of videos undergoing varying levels of spatio-temporal subsampling and compression. The new model is established upon the natural statistics principle of videos, which leverage the fact that pristine videos obey statistical regularities that are disturbed by distortions. We found that there exist space-time paths between video frames that best preserve the statistical regularity inherent in the spatial structure of the video frames. The distribution features extracted from frame differences displaced in the direction of these paths correlate more highly with human subjective quality opinions than those from non-displaced frame differences. Given that non-displaced frame differences are widely utilized in video quality models, the improved efficiency of spatially and/or temporally displaced (possibly by more than one frame) frame differences, is an important finding that may significantly elevate the success of studies on temporal features and video quality.

---

10 Dec
11:30-12:50am
Live stream 2

Real-time Object Detection using Deep Learning for helping People with Visual Impairments

Matteo Terreran[1], Andrea G. Tramontano [2],Jacobus C. Lock[3], Stefano Ghidoni[4],Nicola Bellotto[5]

matteo.terreran@dei.unipd.it[1],andreagaetano.tramontano@dei.unipd.it[2],jlock@lincoln.ac.uk[3],stefano.ghidon

Dept. of Information Engineering University of Padova Padova, Italy and School of Computer Science University of Lincoln Lincoln, United Kingdom

Summary

Object detection plays a crucial role in the development of Electronic Travel Aids (ETAs), capable to guide a person with visual impairments towards a target object in an unknown indoor environment. In such a scenario, the object detector runs on a mobile device (e.g. smartphone) and needs to be fast, accurate, and, most importantly, lightweight. Nowadays, Deep Neural Networks (DNN) have become the state-of-the-art solution for object detection tasks, with many works improving speed and accuracy by proposing new architectures or extending existing ones. A common strategy is to use deeper networks to get higher performance, but that leads to a higher computational cost which makes it impractical to integrate them on mobile devices with limited computational power. In this work we compare different object detectors to find a suitable candidate to be implemented on ETAs, focusing on lightweight models capable of working in real-time on mobile devices with a good accuracy. In particular, we select two models: SSD Lite with Mobilenet V2 and TinyDSOD. Both models have been tested on the popular OpenImage dataset and a new dataset, named L-CAS Office dataset, collected to further test models' performance and robustness in a real scenario inspired by the actual perception challenges of a user with visual impairments.

Real time System Implementation for Stereo 3D Mapping and Visual Odometry

10 Dec
11:30-12:50pm
Live stream 2

Yashwant Temburu[1], Mandar Datar[2], Simranjeet Singh, Vaibhav Malviya[3], Sachin Patkar[4]

temburuyk@gmail.com[1] ,mandardatar@ee.iitb.ac.in[2],simranjeet@ee.iitb.ac.in [3], vaibhav.m@iitb.ac.in[4],patkar@iitb.ac.in [5]

Indian Institute of Technology Bombay Mumbai, India

Summary

Depth Estimation and localisation is a critical and computation heavy task. Due to its computation complexity, it is hard to achieve the high rate performance on the normal CPU. We here have as purpose compact, low power architecture for real time stereo depth estimation and stereo visual odometry that can be easily used on the UAV (unmanned aerial vehicle) and other autonomous navigation vehicles. A novel implementation of Stereo odometry based on careful feature selection and tracking [2] on GPU is described. It accelerates core computations like feature tracking, RANSAC based non linear solver using the GPU. 3D Mapping using stereo disparity estimation based on More Global Matching (MGM) a variant of SGM is implemented on FPGA. A Pipeline Architecture is introduced to increase throughput of the 3D map by leveraging multiple ARM cores. The programs are tested on Jetson Nano (an embedded GPU) and Ultra96 (ARM-FPGA Soc) which have less form factor and consume less power. Update rate of 20 is achieved for 6 degrees of freedom pose on Jetson Nano (60 on i7 core with Nvidia 1050ti) and an update rate of 16 fps is achieved for 3D point cloud on Ultra96 making the system desirable for above mentioned applications.

Session 5: Image processing and smart city applications (Live stream2)

Robust Road Region Extraction in Video Under Various Illumination and Weather Conditions

10 Dec
4:00-5:40pm
Live stream 2

Hadi Ghahremannezhadi[1], Hang Shi[2], and Chengjun Liu[3]
[1]: hg255@njit.edu, [2]: hs328@njit.ed, [3]: cliu@njit.edu

Institute of Technology Newark, New Jersey , United State of America

Summary

Robust road region extraction plays a crucial role in many computer vision applications, such as automated driving and traffic video analytics. Various weather and illumination conditions like snow, fog, dawn, daytime, and nighttime often pose serious challenges to automated road region detection. This paper presents a new real-time road recognition method that is able to accurately extract the road region in traffic videos under adverse weather and illumination conditions. Specifically, the novel global foreground modeling (GFM) method is first applied to subtract the ever-changing background in the traffic video frames and robustly detect the moving vehicles which are assumed to drive in the road region. The initial road samples are then obtained from the subtracted background model in the location of the moving vehicles. The integrated features extracted from both the grayscale and the RGB and HSV color spaces are further applied to construct a probability map based on the standardized Euclidean distance between the feature vectors. Finally, the robust road mask is derived by integrating the initially estimated road region and the regions located by the flood-fill videos demonstrate the feasibility of the proposed method for automated road recognition in real-time.

---

10 Dec
4:00-5:40pm
Live stream 2

Traffic Sign Recognition Based On Scaled Convolutional Neural Network For Advanced 1st Driver Assistance System

Riadh Ayachi[1], Mouna Afif[2], Yahia Said [3], and Abdessalem Ben Abdelali[4]
riadh.ayachi@fsm.rnu.tn[1], Mouna.afif@outlook.fr [2], yahia.said@fsm.rnu.tn [3],
Abdessalem.BenAbdelali@enim.rnu.tn [4]

University of Monastir, Tunisia

Summary

Advanced driver assistance system (ADAS) is one of the most important systems for human assistance. It assists the drivers to control the vehicle by providing essential information about the environment objects. In this paper, we propose a traffic signs recognition application for ADAS. The proposed application is based on the deep learning technique. In particular, we used the convolutional neural networks (CNN) to process the data provided by the system cameras. The proposed CNN was scaled in a way to get a light model size without decreasing the accuracy. The proposed CNN is suitable for embedded implementation while keeping high performance and real-time processing. The evaluation of the proposed CNN on the European dataset results in 99.32% accuracy and 250 FPS of inference speed when implemented on an Nvidia GTX960 GPU. The achieved results proved the efficiency of the scaling technique. It is a very good technique to get a small model size and high performance.

---

10 Dec
4:00-5:40pm
Live stream 2

Boosting car plate recognition systems performances with agile re-training

Giorgio Cantarini[1], Nicoletta Noceti[2], and Francesca Odone[3]
[1]:cantarini.giorgio261@gmail.com, [2]: nicoletta.noceti@unige.it, [3]:
Francesca.Odone@unige.it

University of Genova, Italy

Summary

In this work, we report an experimental study on an Automatic Licence Plate Recognition system developed and commercialized by a partner company, with the main goals of critically analysing the original system and of devising effective but minimally invasive design changes. From a scientific point of view, ours is an attempt of reducing the gap between the different experimental approaches in academia and industry. The system is organized in layers, with an initial car plate proposal step followed by a OCR step. To cope with the drawbacks of the pre-existing system, we inserted an intermediate CNN binary classi?cation step to discriminate between plates and non plates independently from the OCR module. Our solution incorporates new data available from working installations, in a closed refinement loop. We

evaluate the modi?ed system on 8 different installations. With respect to the original performances, we obtained significant improvements with an impact on both false positive (9,8%) and false negatives (5%).

Detection of Microconidia in Microscopy Images of Fusarium oxysporum f. sp. cubense Using Image Processing Techniques and Neural Networks

10 Dec
4:00-5:40pm
Live stream 2

Erinn Giannice T. Abigan [1] , Luis Gabriel A. Cajucom [2], Josh Daniel L. Ong[3], Patricia Angela R. Abu and Ma. Regina Justina E. Estuar[4]

erinn.abigan@obf.ateneo.edu[1],luis.cajucom@obf.ateneo.edu[2], josh.ong@obf.ateneo.edu[3], pabu@ateneo.edu,restuar@ateneo.edu[4]

Ateneo de Manila University Quezon City, Philippines

Summary

Fusarium oxysporum f. sp. cubense (Foc) is a soilborne fungus and the causative agent of the deadly Fusarium wilt disease in banana plants. Left alone, the fungus is able to survive for years and infect multiple plants through the soil. External symptoms only manifest in late stages of infection, with the destruction of all plants within a 7.5 meter radius of the diseased through burning being the only way to eradicate the fungus. Foc Tropical Race 4 (TR4) is capable of infecting the widely used Cavendish cultivars, threatening global banana production. It is imperative then that Foc be detected as soon as possible. To achieve this, the study endeavors to detect microconidia, a reproductive structure of the Foc species, in microscopy images of stained soil specimen under three microscopy con?gurations using image processing techniques and convolutional neural networks (CNNs). The networks were built using the ResNet-50 architecture, and results were validated via Gradient-weighted class activation mapping (Grad-CAM). The network classifying ?uorescent images achieved the highest accuracy with 95.24%, followed by bright field images with 94.94%, all (bright field, dark field, and fluorescent) images with 93.75%, and lastly, dark ?eld images with 92.86%. Grad-CAM results indicate the networks are able to identify Foc structures and correctly distinguish clean from Foc-infected images. This study contributes towards the early detection of Foc, and is a step toward mitigating the threat it presents.

---

10 Dec
4:00-5:40pm
Live stream 2

Pedestrian Detection and Classification for Autonomous Train

Ankur Mahtan[1], Wael Ben-Messaoud[2], Abdelmalik Taleb Ahmed[3], Smail Niar [4], and Clement Strauss[5]

ankur.mahtani@railenium.eu[1],wael.benmessaoud@railenium.eu[2], abdelmalik.taleb-ahmed@uphf.fr[3],Smail.Niar@uphf.fr[4], clement.strauss@railenium.eu[5]

FCS Railenium, FCS Railenium, IEMN, LAMIH, France

Summary

In this paper, we present a combined approach for human localization and classification in Autonomous Train application. Our contribution is threefold. (a) The creation of a new dataset for workers wearing orange vests in a railway environment context. (b) A deep learning supervised YOLO object detector for persons detection combined with a linear SVM (Support Vector Machine) classifier for persons classification into workers wearing orange vests or travelers. (c) A realtime vision-based technique for the environment monitoring in a driverless train application. Experimental results evaluate the parameters of our two stages detection approach and show that our algorithm is robust in detecting and classifying railway workers for a real-time implementation on an embedded system. Our implementation on an embedded system allows a detection with a correct classification rate of 98.5% of accuracy and a classification time of 1 ms per frame.

---

14 Image-Preprocessing and Segmentation Techniques for Vehicle-Plate Recognition Mohamed A I Alkalai (University of Benghazi)*; Ahmed lawgali (University of Benghazi) Mohamed A I Alkalai mohamed.alkalai@uob.edu.ly Lybia

10 Dec
4:00-5:40pm
Live stream 2

Image-Preprocessing and Segmentation Techniques for Vehicle-Plate Recognition

Mohamed A I Alkalai[1], Ahmed lawgali[2]

mohamed.alkalai@uob.edu.ly[1],Ahmed.lawgali@uob.edu.ly[2]

University of Benghazi

Summary

Recently, the number of vehicles on the roads has been rapidly growing which needs frequent expanding of concerned authorities budget to cover the new demands such as hire more road traffic officers as well as buy their special equipment. This often overburdens the government with unwelcome financial commitments. The mainstream solution for these kind of issues is to mechanise some of the current officers duties. In this paper, we proposed a vehicle plate recognition approach for monitoring traffic violation purposes. This is accomplished by firstly enhancing the quality

of vehicle plate images using a proposed image pre-processing method. Then, applying a proposed connected-component segmentation method to extract plate numbers boundaries from each image content. Finally, with the help of open-source classification libraries, the segmented plate's numbers are recognised. The experiments show promising results compared with those that are obtained by classifying vehicle plate numbers using Otsu, one of the most common thresholding method.

Keynote IV by Prof. Jean-Mark Odobez (Live stream2)

Using less data for training: investigating weak labels and unsupervised training for the robust sound localization and gaze estimation

10 Dec
14:00-14:50 am
Live stream2

Jean-Mark Odobez

Idiap Research Institue, Switzerland

Summary

Supervised training is a comfortable option for training deep learning networks. However, it usually comes at the cost of requiring large labeled training datasets, a resource that might be difficult to produce in practice. For instance, in gaze estimation, creating such data usually requires both the video of the camera sensing the person of interest, and information in the 3D scene (calibrated with the sensing camera) about what the sensed person might be looking at. This means that internet videos or captured videos can not be annotated post-acquisition by humans, limiting the ability to build large scale datasets with hundreds of persons. Regarding audio localization based on microphone-array, the situation is slightly different: the need for training data arises each time a new device configuration (number of microphones, location, microphone type) is used. In this talk, I will present our work on these two tasks (sound source and human voice localization, gaze estimation), with an emphasis on neural network architectures and methods relying on synthetic data, weak labels, and unsupervised approaches to address the learning problem, including adaptation to users.

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 multimodal 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, multimodal signal processing, and machine learning, in particular probabilistic graphical models and deep learning techniques, surveillance, traffic and human behavior analysis.

Keynote V by Prof. Nicolas Gillis (Live stream2)

Some recent results on nonnegative matrix factorizations with application in hyperspectral imaging

10 Dec
15:00-15:50 pm
Live stream2

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 X?WH. NMF has become a standard linear dimensionality reduction technique in data mining and machine learning. 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 received the Master's and Ph.D. degrees in applied mathematics from the University catholique de Louvain, Louvain-la-Neuve, Belgium, in 2007 and 2011, respectively. He is currently an Associate Professor with the Department of Mathematics and Operational Research, Faculty polytechnique, University de Mons, Mons, Belgium. His research interests include optimization, numerical linear algebra, machine learning, signal processing, and data mining. Dr. Gillis received the Householder award in 2014, and an ERC starting grant in 2015. He is currently serving

as an Associate Editor of the IEEE Transactions on Signal Processing and of the SIAM Journal on Matrix Analysis and Applications.

Keynote VI by Prof. Nicolas Dobigeon (Live stream3)

Fusion-based change detection for remote sensing images of different resolutions and modalities

10 Dec
9:00-9:50 am
Live stream3

Nicolas Dobigeon
University of Toulouse, France

Summary

Change detection is one of the most challenging issues when analyzing remotely sensed images. It consists in detecting alterations occurred in a given scene from images acquired at different time instants. Archetypal scenarios for change detection generally compare two images acquired through the same kind of sensor, i.e., with the same modality and the same spatial/spectral resolutions. This talk will address the problem of detecting changes between images of different resolutions or modalities. We will show that this challenging task can be tackled from a fusion perspective.

Nicolas Dobigeon biography

Since 2008, Nicolas Dobigeon has been with Toulouse INP (INP-ENSEEIHT, University of Toulouse) where he is currently a Professor. He conducts his research within the Signal and Communications (SC) group of IRIT and is an associate member of the Apprentissage Optimisation Complexité (AOC) project-team of CIMI. He currently holds an AI Research Chair at the Artificial and Natural Intelligence Toulouse Institute (ANITI) and he is a Junior Member of the Institut Universitaire de France (IUF, 2017-2022). His recent research activities have been focused on statistical signal and image processing, with a particular interest in Bayesian inverse problems and applications to remote sensing, biomedical imaging and microscopy.

Session 6: Image processing and Medical Applications (Live stream3)

Convolutional Neural Network Based Desktop Applications to Classify Dermatological Diseases

11 Dec
10:00-11:40 am
Live stream3

Evgin Göçeri[1]
evgin@akdeniz.edu.tr:[1]

Biomedical Engineering Department Akdeniz University Antalya, Turkey

Summary

Convolutional Neural Networks (CNNs) have the potential to assist medical doctors in diagnosis and treatment stage. This paper has been prepared to help dermatologists by presenting (i) fundamental information on deep learning and CNNs, and (ii) applications of CNNs for skin diseases classification. Also, this work shows that although CNN based methods have a strong potential for automated diagnosis, further researches and new techniques are still required in image processing and pattern recognition area to provide diagnosis of dermatological diseases with higher performances. In this work, these two groups of applications of CNNs in dermatology have been handled: (i) disease classification from medical images (e.g., dermoscopy and pathological images); (ii) disease classification from digital photographs. Therefore, important contributions of this work are two-fold: First, main concepts of deep learning and CNNs are presented, which will be helpful for dermatologists to understand and follow up CNN based computerized methods. Second, the state-of-the-art applications developed for lesion classifications from medical images and colored photographs are presented. Also, disadvantages or limitations of these applications are explained. In addition, this paper indicates shortage of desktop applications developed for other dermatological diseases except skin cancer.

---

10 Dec
10:00-11:40 am
Live stream3

On the Exploitation of Temporal Redundancy to Improve Polyp Detection in Colonoscopy

Giovanna Pappalardo[1], Dario Allegra [2], Filippo Stanco [3], Giovanni Maria Farinella[4]

[1]: giovanna.pappalardo1@unict.it, [2]: allegra@unict.it, [3]: fstanco@dmi.unict.it, [4]:gfarinella@dmi.unict.it

Department of Mathematics and Computer Science University of Catania Catania, Italy

Summary

Colonoscopy is currently the most effective screening method to find precancerous colon polyps and plan their removal. Computer-aided polyp detection can reduce polyp miss detection rates and help doctors find the most critical regions to pay attention to. The challenge in detecting polyps is due to the polyp's morphology and size, and these fall into false negative. Indeed, polyps may exhibit high variability in shapes (e.g., depressed, flat, pedunculated, etc...). Moreover, the water injected from the endoscope results in artifacts which impede the detection, and the lubricating mucus causes light artifacts due its glossiness. To address this problem, we propose a mask-based attention mechanism to ensure that the employed detector focuses on particular regions of the image in order to reduce misdetection rate. Our contribution takes advantage of information on polyp's position over time within a video sequence. We provide such information through a binary mask which points out the lastknown polyp's position. The proposed approach is validated on a dataset that has been labeled by colonoscopy experts. It contains about 200 videos and more than 500 different polyps with high variability in size and textures. Experimental results show that the proposed attention mechanism recover a smaller number of false negatives and achieves an F1-score of %80.21.

---

10 Dec
10:00-11:40 am
Live stream3

Binary View Classification of Echocardiograms of the Heart Using Transfer Learning

Nontokozo Mpofu[1], Michael Sears [2]

[1]: nontokozo.mpofu@spu.ac.za, [2]: michael.sears@wits.ac.za

School of Computer Science University of Witwatersrand Johannesburg, South Africa

Summary

Echocardiogram metadata annotation is an entry level to the interpretation and analysis of echocardiographs for clinical usage. Automation of this task has been the subject of research over the past few years. In this paper we evaluate how

far some state-of-the-art convolutional neural networks (CNNs) trained with non-radiological images can go at extracting features from cardiac structures in apical two-chamber (2CH) and four-chamber (4CH) echocardiograms for automatic binary view echocardiogram classification on a publicly available ultrasound dataset.

EczemaNet: A Deep CNN-based Eczema Diseases Classification

10 Dec
10:00-11:40 am
Live stream3

Masum Shah Junayed[1], Abu Noman Md Sakib [2], Nipa Anjum[3], Md Baharul Islam[4], Afsana Ahsan Jeny [5]

[1]: masumshahjunayed@gmail.com, [2]: abunomanmd.sakib@gmail.com, [3] : nipaanjum1998@gmail.com, [4] : mdbaharul.islam@eng.bau.edu.tr, [5] : ahsan15-5278@diu.edu.bd

Bahcesehir University,Bangladesh; niversity of Khulna, Turkey; ahcesehir University, Malta

Summary

Eczema is the most common among all types of skin diseases. A solution for this disease is very crucial for patients to have better treatment. Eczema is usually detected manually by doctors or dermatologists. It is tough to distinguish between different types of Eczema because of the similarities in symptoms. In recent years, several attempts have been taken to automate the detection of skin diseases with much accuracy. Many methods such as Image Processing Techniques, Machine Learning algorithms are getting used to execute segmentation and classification of skin diseases. It is found that among all those skin disease detection systems, particularly detection work on eczema disease is rare. There is also insufficiency in eczema disease dataset. In this paper, we propose a novel deep CNN-based approach for classifying five different classes of Eczema with our collected dataset. Data augmentation is used to transform images for better performance. Regularization techniques such as batch normalization and dropout helped to reduce over?tting. Our proposed model achieved an accuracy of 96.2%, which exceeded the performance of the state of the arts.

---

10 Dec
10:00-11:40 am
Live stream3

Image Augmentation for Deep Learning Based Lesion Classification from Skin Images

Evgin Goceri[1]

[1]: evgin@akdeniz.edu.tr

Biomedical Engineering Department Akdeniz University Antalya, Turkey

Summary

Skin lesion classification based on deep learning models, which are data-hungry, is a challenging issue because of the shortage of annotated images and unbalanced classes in image sets. The lack of sufficient number of labeled data or class unbalancing in image sets lead to overfitting problems affecting robustness and generalization ability of network models. Image augmentation is an efficient approach to deal with this issue using existing images more efficiently. In addition to image augmentations, various solutions have been developed in the literature to solve the overfitting problem and to obtain well generalizing network models. However, there is not a clear way how the most appropriate solution should be selected. Therefore, in this paper,those alternative solutions and image augmentations applied recently in deep learning based skin lesion classifications are presented.

Keynote VII by Prof. Fabio Solari (Live stream3)

Natural perception in virtual and augmented reality: a computational model

11 Dec
11:50-12:40 am
Live stream3

Fabio Solari

University of Genoa, Italy

Summary

The current virtual and augmented reality (VR and AR) technologies provide new experiences for users, though they might cause discomfort and unnatural perception during the interaction in VR and AR environments. Here, a bio-inspired computational neural model of visual perception for action tasks is considered in order to provide a tool to better design VR and AR systems. In particular, the log-polar mapping, disparity and optic flow computation is presented. Then, such a computational model is exploited to mimic (thus to describe) human behavioral data. By leveraging previous outcomes, we employ the modeled perception to improve the experience in VR and AR environments by implementing a foveated depth-of-field blur.

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 interests are related to computational models of motion and depth estimation, space-variant visual processing and scene interpretation. Such models are able to replicate relevant aspects of human experimental data. This can help to improve virtual and augmented reality systems in order to provide a natural perception and human-computer interaction. He is principal investigator of three international projects: Interreg Alcotra CLIP E-Santé/Silver Economy", PROSOL Jeune" and PROSOL "Senior". He has participated in five European projects: FP7-ICT, EYESHOTS and SEARISE; FP6-IST-FET, DRIVSCO; FP6-NEST, MCCOOP; FP5-IST-FET, ECOVISION. 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 www.dibris.unige.it/en/solari-fabio .

Session 7: Image processing for biological applications and biometry (Live stream3)

Detection of Microconidia in Microscopy Images of Fusarium oxysporum f. sp. cubense Using Image Processing Techniques and Neural Networks

10 Dec
10:00-11:40 am
Live stream3

Erinn Giannice T. Abigan[1], Luis Gabriel A. Cajucom [2], Josh Daniel L. Ong[3], Patricia Angela R. Abu [4], Ma. Regina Justina E. Estuar [5]

[1]:erinn.abigan@obf.ateneo.edu , [2]: luis.cajucom@obf.ateneo.edu , [3]: josh.ong@obf.ateneo.edu, [4]: pabu@ateneo.edu, [5]: restuar@ateneo.edu

University of Quezon City, Philippines

Summary

Fusarium oxysporum f. sp. cubense (Foc) is a soilborne fungus and the causative agent of the deadly Fusarium wilt disease in banana plants. Left alone, the fungus is able to survive for years and infect multiple plants through the soil. External symptoms only manifest in late stages of infection, with the destruction of all plants within a 7.5 meter radius of the diseased through burning being the only way to eradicate the fungus. Foc Tropical Race 4 (TR4) is capable of infecting the widely used Cavendish cultivars, threatening global banana production. It is imperative then that Foc be detected as soon as possible. To achieve this, the study endeavors to detect microconidia, a reproductive structure of the Foc species, in microscopy images of stained soil specimen under three microscopy con?gurations using image processing techniques and convolutional neural networks (CNNs). The networks were built u sing the ResNet-50 architecture, and results were validated via Gradient-weighted class activation mapping (Grad-CAM). The network classifying ?uorescent images achieved the highest accuracy with 95.24followed by bright ?eld images with 94.94%, all (bright ?eld, dark ?eld, and ?uorescent) images with 93.75%, and lastly, dark ?eld images with 92.86%. Grad-CAM results indicate the networks are able to identify Foc structures and correctly distinguish clean from Foc-infected images. This study contributes towards the early detection of Foc, and is a step toward mitigating the threat it presents.

10 Dec
10:00-11:40 am
Live stream3

Coconut Disease Prediction System Using Image Processing and Deep Learning Techniques

Dhapitha Nesarajan[1], Lokini Kunalan [2], Mithun Logeswaran [3], Sanvitha Kasthuriarachchi[4], Dilani Lunugalage[5]

[1]: dhapitha1996@gmail.com, [2]: lokini2196@gmail.com , [3]: adomithun@gmail.com , [4]: sanvitha.k@my.sliit.lk , [5]: dilani.l@my.sliit.lk

University of Malabe, Sri Lanka

Summary

Coconut production is the most important and one of the main sources of income in the Sri Lankan economy. The recent time it has been observed that most of the coconut trees are affected by the diseases which gradually reduces the strength and production of coconut. Most of the tree leaves are affected by pest diseases and nutrient deficiency. Our main intensive is to enhance the livelihood of coconut leaves and identify the diseases at the early stage so that farmers get more benefits from coconut production. This paper proposes the detection of pest attack and nutrient deficiency in the coconut leaves and analysis of the diseases. Coconut leaves monitorization has been taken place after the use of pesticides and fertilizer with the help of the finest machine learning and image processing techniques. Rather than human experts, automatic recognition will be beneficial and the fastest approach to identify the diseases in the coconut leaves very efficiently. Thus, in this project, we developed an android mobile application to identify the pests by their food behaviors, pest diseases and the nutrition deficiencies in the coconut trees. As an initial step, all datasets for image processing technology met pre-processing steps such as converting RGB to greyscale, filtering, resizing, horizontal flip and vertical flip. After completing the above steps, the classification was performed by analyzing several algorithms in the literature review. SVM and CNN were chosen as the best and appropriate classifier with 93.54% and 93.72% of accuracy respectively. The outcome of this project will help the farmers to increase the coconut production and undoubtedly will make a revolution in the agriculture sector.

---

10 Dec
10:00-11:40 am
Live stream3

Spatial characteristics of pigeon tracks depending on distribution of visual elements of urban and natural terrain

Margarita Zaleshina[1], Alexander Zaleshin[2]

[1]: zaleshina@gmail.com, [2]: terbiosorg@gmail.com

Moscow Institute of Physics and Technology Moscow, Russia

Summary

Visual characteristics of terrain affect the properties of pigeon trajectories in medium-distance flights. Pigeon flight often provides a solution to the task of searching for

food (foraging), returning home (homing), or exploring territory (surveying). In this work, we considered the flights of single pigeons and pigeon flocks, calculated flight characteristics such as direction, altitude and its deviations, and analyzed reactions to the boundaries between different areas. Based on remote sensing datasets, we identified visual characteristics of terrain, such as the density of surface fill and its distribution over the study terrain, boundaries of single objects, and boundaries between homogeneous areas. Applying spatial analysis, we compared the characteristics of pigeon GPS tracks and features of object distributions on terrain over which birds fly. Our analysis revealed which flight parameters are stable and which, on the contrary, are very sensitive to visually perceived terrain characteristics. We found that the properties of flight over an urbanized area often differ from the properties of flight over a natural landscape. Spatial data- pigeon GPS track records and open-access remote sensing datasets-were processed using the geographical information system QGIS. Our results show that adaptive visual perception can help solve navigation tasks when pigeons fly over mixed terrain. Knowledge of the characteristic features of bird flights can be used both for a better understanding of the spatial behavior of living creatures (humans and animals) and for optimization of artificial intelligence algorithms.

Are Adaptive Face Recognition Systems still Necessary? Experiments on the APE Dataset

10 Dec
10:00-11:40 am
Live stream3

Giulia Orru[1], Marco Micheletto [2], Julian Fierrez[3], Gian Luca Marcialis[4]
[1]: giulia.orru@unica.it, [2]: marco.micheletto@unica.it, [3] : julian.fierrez@uam.es, [4] : marcialis@unica.it

University of Cagliari, DIEE Cagliari, Italy

Summary

In the last five years, deep learning methods, in particular CNN, have attracted considerable attention in the field of face-based recognition, achieving impressive results. Despite this progress, it is not yet clear precisely to what extent deep features are able to follow all the intra-class variations that the face can present over time. In this paper we investigate the performance the performance improvement of face recognition systems by adopting self updating strategies of the face templates. For that purpose, we evaluate the performance of a wellknown deep-learning face representation, namely, FaceNet, on a dataset that we generated explicitly conceived to embed intraclass variations of users on a large time span of captures: the APhotoEveryday (APE) dataset. Moreover, we compare these deep features with handcrafted features extracted using the BSIF algorithm. In both cases, we evaluate various template update strategies, in order to detect the most useful for such kind of features. Experimental results show the effectiveness of "optimized" self-update methods with respect to systems without update or random selection of templates.

---

10 Dec
10:00-11:40 am
Live stream3

Handwritten Recognition: A survey

May Mowaffaq AL-Taee[1], Sonia Ben Hassen Neji: [2],Mondher Frikha: [3]
[1]: may.m.altaie@gmail.com

University of Sfax, Tunisia

Summary

Handwritten recognition has received considerable attention in the domain of pattern recognition, image processing, over the last few decades. As a consequence of this research effort, several algorithms were developed using different techniques. Particularly, Deep Learning has shown a remarkable capability to handle handwritten recognition in very recent years. The well-known Deep learning techniques are the Convolutional Neuronal Networks (CNNs) and Recurrent Neuronal Networks (RNNs). This paper provides a survey of the most recent handwritten recognition systems. Thus, we present the most significant algorithms for handwritten character/word/text recognition by explaining the different approaches used in the recognition process and we compare them in terms of accuracy.
