

# **16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2021)**

Volume 5: VISAPP

Online  
8 – 10 February 2021

## **Editors:**

**Giovanni Maria Farinella  
Petia Radeva**

**Jose Braz  
Kadi Bouatouch**

ISBN: 978-1-7138-4026-8

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2021) by SCITEPRESS – Science and Technology Publications, Lda.  
All rights reserved.

Printed with permission by Curran Associates, Inc. (2021)

For permission requests, please contact SCITEPRESS – Science and Technology Publications, Lda.  
at the address below.

SCITEPRESS – Science and Technology Publications, Lda.  
Avenida de S. Francisco Xavier, Lote 7 Cv. C,  
2900-616 Setúbal, Portugal

Phone: +351 265 520 185  
Fax: +351 265520 186

[info@scitepress.org](mailto:info@scitepress.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# CONTENTS

---

## INVITED SPEAKERS

### KEYNOTE SPEAKERS

3D Indoor Scene Understanding with Scene Graphs and Self-supervision <i>Federico Tombari</i>	5
Visualization in the Real World: Confluence of Visualization and Augmented Reality <i>Dieter Schmalstieg</i>	7
Keynote Lecture <i>Nathalie Henry Riche</i>	9

## IMAGE AND VIDEO UNDERSTANDING

### FULL PAPERS

Clothing Parsing using Extended U-Net <i>Gabriela Vozáriková, Richard Staňa and Gabriel Semanišin</i>	15
Lightweight SSD: Real-time Lightweight Single Shot Detector for Mobile Devices <i>Shi Guo, Yang Liu, Yong Ni and Wei Ni</i>	25
Reduction in Communication via Image Selection for Homomorphic Encryption-based Privacy-protected Person Re-identification <i>Shogo Fukuda, Masashi Nishiyama and Yoshio Iwai</i>	36
Automatically Generating Websites from Hand-drawn Mockups <i>João Silva Ferreira, André Restivo and Hugo Sereno Ferreira</i>	48
On the Transferability of Winning Tickets in Non-natural Image Datasets <i>Matthia Sabatelli, Mike Kestemont and Pierre Geurts</i>	59
Point Cloud Upsampling and Normal Estimation using Deep Learning for Robust Surface Reconstruction <i>Rajat Sharma, Tobias Schwandt, Christian Kunert, Steffen Urban and Wolfgang Broll</i>	70
Building Synthetic Simulated Environments for Configuring and Training Multi-camera Systems for Surveillance Applications <i>Nerea Aranjuelo, Jorge García, Luis Unzueta, Sara García, Unai Elordi and Oihana Otaegui</i>	80
CUPR: Contrastive Unsupervised Learning for Person Re-identification <i>Khadija Khaldi and Shishir K. Shah</i>	92
Latent Video Transformer <i>Ruslan Rakhimov, Denis Volkhonskiy, Alexey Artemov, Denis Zorin and Evgeny Burnaev</i>	101
Feature Map Upscaling to Improve Scale Invariance in Convolutional Neural Networks <i>Dinesh Kumar and Dharmendra Sharma</i>	113
Image Synthesis and Data Augmentation for Safe Object Detection in Aircraft Auto-landing System <i>Najda Vidimlic, Alexandra Levin, Mohammad Loni and Masoud Daneshtalab</i>	123

A Human Ear Reconstruction Autoencoder <i>Hao Sun, Nick Pears and Hang Dai</i>	136
Few-shot Linguistic Grounding of Visual Attributes and Relations using Gaussian Kernels <i>Daniel Koudouna and Kasim Terzić</i>	146
Detecting Object Defects with Fusioning Convolutional Siamese Neural Networks <i>Amr M. Nagy and László Czúni</i>	157
Towards Real-time Object Recognition and Pose Estimation in Point Clouds <i>Marlon Marcon, Olga Regina Pereira Bellon and Luciano Silva</i>	164
A Neural Network with Adversarial Loss for Light Field Synthesis from a Single Image <i>Simon Evain and Christine Guillemot</i>	175
Domain Adaptation for Traffic Density Estimation <i>Luca Ciampi, Carlos Santiago, Joao Paulo Costeira, Claudio Gennaro and Giuseppe Amato</i>	185
Contextualise, Attend, Modulate and Tell: Visual Storytelling <i>Zainy M. Malakan, Nayyer Aafaq, Ghulam Mubashar Hassan and Ajmal Mian</i>	196
Fairer Evaluation of Zero Shot Action Recognition in Videos <i>Kaiqiang Huang, Sarah Jane Delany and Susan Mckeever</i>	206
Global Point Cloud Descriptor for Place Recognition in Indoor Environments <i>Jacek Komorowski, Grzegorz Kurzejamski, Monika Wysoczańska and Tomasz Trzcinski</i>	216
Towards Combined Open Set Recognition and Out-of-Distribution Detection for Fine-grained Classification <i>Alexander Gillert and Uwe Freiherr von Lukas</i>	225
Analysis of Recent Re-Identification Architectures for Tracking-by-Detection Paradigm in Multi-Object Tracking <i>Haruya Ishikawa, Masaki Hayashi, Trong Huy Phan, Kazuma Yamamoto, Makoto Masuda and Yoshimitsu Aoki</i>	234
<b>SHORT PAPERS</b>	
Facial Exposure Quality Estimation for Aesthetic Evaluation <i>Mathias Gudiksen, Sebastian Falk, Lasse Nymark Hansen, Frederik Brønnum Jensen and Andreas Møgelmoose</i>	247
Learning Unsupervised Cross-domain Image-to-Image Translation using a Shared Discriminator <i>Rajiv Kumar, Rishabh Dabral and G. Sivakumar</i>	256
Approaching the Semantic Segmentation in Medical Problems: A Solution for Pneumothorax Detection <i>Călin Timbus, Vlad Miclea and Camelia Lemnaru</i>	265
Crowd Behavior Analysis based on Convolutional Neural Network: Social Distancing Control COVID-19 <i>Fatma Bouhlel, Hazar Mliki and Mohamed Hammami</i>	273
Three-step Alignment Approach for Fitting a Normalized Mask of a Person Rotating in A-Pose or T-Pose Essential for 3D Reconstruction based on 2D Images and CGI Derived Reference Target Pose <i>Gerald Adam Zwettler, Christoph Praschl, David Baumgartner, Tobias Zucali, Dora Turk, Martin Hanreich and Andreas Schuler</i>	281

Weakly-supervised Human-object Interaction Detection <i>Masaki Sugimoto, Ryosuke Furuta and Yukinobu Taniguchi</i>	293
Online Point Cloud Object Recognition System using Local Descriptors for Real-time Applications <i>Yacine Yaddaden, Sylvie Daniel and Denis Laurendeau</i>	301
ChartSight: An Automated Scheme for Assisting Visually Impaired in Understanding Scientific Charts <i>Mandhatya Singh and Puneet Goyal</i>	309
Unsupervised Domain Adaptation from Synthetic to Real Images for Anchorless Object Detection <i>Tobias Scheck, Ana Perez Grassi and Gangolf Hirtz</i>	319
No Need for a Lab: Towards Multi-sensory Fusion for Ambient Assisted Living in Real-world Living Homes <i>Alessandro Masullo, Toby Perrett, Dima Damen, Tilo Burghardt and Majid Mirmehdi</i>	328
Tropical Skin Disease Classification using Connected Attribute Filters <i>Fred N. Kiwanuka, Omar Eltaher Abuelmaatti, Anang Hudaya Muhamad Amin and Brian J. Mukwaya</i>	338
Are Image Patches Beneficial for Initializing Convolutional Neural Network Models? <i>Daniel Lehmann and Marc Ebner</i>	346
Improving Car Model Classification through Vehicle Keypoint Localization <i>Alessandro Simoni, Andrea D'Eusanio, Stefano Pini, Guido Borghi and Roberto Vezzani</i>	354
Long-term Behaviour Recognition in Videos with Actor-focused Region Attention <i>Luca Ballan, Ombretta Strafforello and Klamer Schutte</i>	362
A Multi-level Rank Correlation Measure for Image Retrieval <i>Nikolas Gomes de Sá, Lucas Pascotti Valem and Daniel Carlos Guimarães Pedronette</i>	370
Learning Joint Twist Rotation for 3D Human Pose Estimation from a Single Image <i>Chihiro Nakatsuka, Jianfeng Xu and Kazuyuki Tasaka</i>	379
Multimodal Neural Network for Sentiment Analysis in Embedded Systems <i>Quentin Portes, José Mendès Carvalho, Julien Pinquier and Frédéric Lerasle</i>	387
A Snapshot-based Approach for Self-supervised Feature Learning and Weakly-supervised Classification on Point Cloud Data <i>Xingye Li and Zhigang Zhu</i>	399
Early Bird: Loop Closures from Opposing Viewpoints for Perceptually-aliased Indoor Environments <i>Satyajit Tourani, Dhagash Desai, Udit Singh Parihar, Sourav Garg, Ravi Kiran Sarvadevabhatla, Michael Milford and K. Madhava Krishna</i>	409
3D Object Classification via Part Graphs <i>Florian Teich, Timo Lüddecke and Florentin Wörgötter</i>	417
Graph Convolutional Networks Skeleton-based Action Recognition for Continuous Data Stream: A Sliding Window Approach <i>Mickael Delamare, Cyril Laville, Adnane Cabani and Houcine Chafouk</i>	427
Multi-Stage Dynamic Batching and On-Demand I-Vector Clustering for Cost-effective Video Surveillance <i>David Montero, Luis Unzueta, Jon Goenetxea, Nerea Aranjuelo, Estibaliz Loyo, Oihana Otaegui and Marcos Nieto</i>	436

Non-Maximum Suppression for Unknown Class Objects using Image Similarity <i>Yoshiaki Homma, Toshiki Kikuchi and Yuko Ozasa</i>	444
Road Lane Detection and Classification in Urban and Suburban Areas based on CNNs <i>Nima Khairdoost, Steven S. Beauchemin and Michael A. Bauer</i>	450
Detection of Distraction-related Actions on DMD: An Image and a Video-based Approach Comparison <i>Paola Natalia Cañas, Juan Diego Ortega, Marcos Nieto and Oihana Otaegui</i>	458
Lightweight Filtering of Noisy Web Data: Augmenting Fine-grained Datasets with Selected Internet Images <i>Julia Böhlke, Dimitri Korsch, Paul Bodesheim and Joachim Denzler</i>	466
Line2depth: Indoor Depth Estimation from Line Drawings <i>Pavlov Sergey, Kanamori Yoshihiro and Endo Yuki</i>	478
Automatic Annotation and Segmentation of Sign Language Videos: Base-level Features and Lexical Signs Classification <i>Hussein Chaaban, Michèle Gouiffès and Annelies Braffort</i>	484
Collaborative Learning of Generative Adversarial Networks <i>Takuya Tsukahara, Tsubasa Hirakawa, Takayoshi Yamashita and Hironobu Fujiyoshi</i>	492
LAMV: Learning to Predict Where Spectators Look in Live Music Performances <i>Arturo Fuentes, F. Javier Sánchez, Thomas Voncina and Jorge Bernal</i>	500
Occluded Iris Recognition using SURF Features <i>Anca Ignat and Ioan Păvăloi</i>	508
Convolutional Neural Networks with Fixed Weights <i>Tyler C. Folsom</i>	516
Extracting Accurate Long-term Behavior Changes from a Large Pig Dataset <i>Luca Bergamini, Stefano Pini, Alessandro Simoni, Roberto Vezzani, Simone Calderara, Rick B. D'Eath and Robert B. Fisher</i>	524
Multi-Task Architecture with Attention for Imaging Atmospheric Cherenkov Telescope Data Analysis <i>Mikaël Jacquemont, Thomas Vuillaume, Alexandre Benoit, Gilles Maurin and Patrick Lambert</i>	534
GAPF: Curve Text Detection based on Generative Adversarial Networks and Pixel Fluctuations <i>Jun Yang, Zhaogong Zhang and Xuexia Wang</i>	545
Automated Infant Monitoring based on R-CNN and HMM <i>Cheng Li, A. Pourtaherian, L. van Onzenoort and P. H. N. de With</i>	553
On Power Jaccard Losses for Semantic Segmentation <i>David Duque-Arias, Santiago Velasco-Forero, Jean-Emmanuel Deschaud, François Goulette, Andres Serna, Etienne Decencière and Beatriz Marcotegui</i>	561
3D Object Detection with Normal-map on Point Clouds <i>Jishu Miao, Tsubasa Hirakawa, Takayoshi Yamashita and Hironobu Fujiyoshi</i>	569
Feature Sharing Cooperative Network for Semantic Segmentation <i>Ryota Ikedo and Kazuhiro Hotta</i>	577

Analysing Adversarial Examples for Deep Learning <i>Jason Jung, Naveed Akhtar and Ghulam Mubashar Hassan</i>	585
U-Net based Zero-hour Defect Inspection of Electronic Components and Semiconductors <i>Florian Kälber, Okan Köpüklü, Nicolas Lehment and Gerhard Rigoll</i>	593
SCAN: Sequence-character Aware Network for Text Recognition <i>Heba Hassan, Marwan Torki and Mohamed E. Hussein</i>	602
Supervised versus Self-supervised Assistant for Surveillance of Harbor Fronts <i>Jinsong Liu, Mark P. Philipsen and Thomas B. Moeslund</i>	610
Segment My Object: A Pipeline to Extract Segmented Objects in Images based on Labels or Bounding Boxes <i>Robin Deléarde, Camille Kurtz, Philippe Dejean and Laurent Wendling</i>	618
Embedding Human Knowledge into Deep Neural Network via Attention Map <i>Masahiro Mitsuhashi, Hiroshi Fukui, Yusuke Sakashita, Takanori Ogata, Tsubasa Hirakawa, Takayoshi Yamashita and Hironobu Fujiyoshi</i>	626
Temporal Bilinear Encoding Network of Audio-visual Features at Low Sampling Rates <i>Feiyan Hu, Eva Mohedano, Noel O'Connor and Kevin Mcguinness</i>	637
S*ReLU: Learning Piecewise Linear Activation Functions via Particle Swarm Optimization <i>Mina Basirat and Peter M. Roth</i>	645
Player Identification in Different Sports <i>Ahmed Nady and Elsayed E. Hemayed</i>	653
Performance Benchmarking of YOLO Architectures for Vehicle License Plate Detection from Real-time Videos Captured by a Mobile Robot <i>Amir Ismail, Maroua Mehri, Anis Sahbani and Najoua Essoukri Ben Amara</i>	661
Deep Emotion Recognition through Upper Body Movements and Facial Expression <i>Chaudhary Muhammad Aqdas Ilyas, Rita Nunes, Kamal Nasrollahi, Matthias Rehm and Thomas B. Moeslund</i>	669
Efficient Multi-task based Facial Landmark and Gesture Detection in Monocular Images <i>Jon Goenexea, Luis Unzueta, Unai Elordi, Oihana Otaegui and Fadi Dornaika</i>	680
Audience Shot Detection for Automatic Analysis of Soccer Sports Videos <i>Kazimierz Choroś</i>	688

## MOTION, TRACKING AND STEREO VISION

### FULL PAPERS

- A Lightweight Real-time Stereo Depth Estimation Network with Dynamic Upsampling Modules 701  
*Yong Deng, Jimin Xiao and Steven Zhiying Zhou*
- BirdSLAM: Monocular Multibody SLAM in Bird's-eye View 711  
*Swapnil Daga, Gokul B. Nair, Anirudha Ramesh, Rahul Sajjani, Junaid Ahmed Ansari and K. Madhava Krishna*
- Independently Moving Object Trajectories from Sequential Hierarchical Ransac 722  
*Mikael Persson and Per-Erik Forssén*
- A Benchmark for 3D Reconstruction from Aerial Imagery in an Urban Environment 732  
*Susana Ruano and Aljosa Smolic*
- Normalized Convolution Upsampling for Refined Optical Flow Estimation 742  
*Abdelrahman Eldesokey and Michael Felsberg*
- Investigating 3D Convolutional Layers as Feature Extractors for Anomaly Detection Systems Applied to Surveillance Videos 753  
*Tiago S. Nazare, Rodrigo F. de Mello and Moacir A. Ponti*
- Real-time Monocular 6DoF Tracking of Textureless Objects using Photometrically-enhanced Edges 763  
*Lucas Valença, Luca Silva, Thiago Chaves, Arlindo Gomes, Lucas Figueiredo, Lucio Cossio, Sebastien Tandel, João Paulo Lima, Francisco Simões and Veronica Teichrieb*

### SHORT PAPERS

- Real-time and Online Segmentation Multi-target Tracking with Track Revival Re-identification 777  
*Martin Ahrnbom, Mikael Nilsson and Håkan Ardö*
- Object based Hybrid Video Compression 785  
*Rhoda Gbadeyan and Chris Joslin*
- Modeling a priori Unknown Environments: Place Recognition with Optical Flow Fingerprints 793  
*Zachary Mueller and Sotirios Diamantas*
- Quantifying Wind Turbine Blade Surface Roughness using Sandpaper Grit Sizes: An Initial Exploration 801  
*Ivan Nikolov and Claus Madsen*
- A Study on the Influence of Omnidirectional Distortion on CNN-based Stereo Vision 809  
*Julian Bruno Seuffert, Ana Cecilia Perez Grassi, Tobias Scheck and Gangolf Hirtz*
- Procam Calibration from a Single Pose of a Planar Target 817  
*Ghani O. Lawal and Michael Greenspan*
- Object Hypotheses as Points for Efficient Multi-Object Tracking 828  
*Shuhei Tarashima*



## MOBILE AND EGOCENTRIC VISION FOR HUMANS AND ROBOTS

### FULL PAPERS

- A Surface and Appearance-based Next Best View System for Active Object Recognition 841  
*Pourya Hoseini, Shuvo Kumar Paul, Mircea Nicolescu and Monica Nicolescu*
- Multi-view Planarity Constraints for Skyline Estimation from UAV Images in City Scale Urban Environments 852  
*Ayyappa Swamy Thatavarthy, Tanu Sharma, Harshit Sankhla, Mukul Khanna and K. Madhava Krishna*
- Boosting Self-localization with Graph Convolutional Neural Networks 861  
*Takeda Koji and Tanaka Kanji*
- Practical Auto-calibration for Spatial Scene-understanding from Crowdsourced Dashcamera Videos 869  
*Hemang Chawla, Matti Jukola, Shabbir Marzban, Elahe Arani and Bahram Zonooz*
- Dual CNN-based Face Tracking Algorithm for an Automated Infant Monitoring System 881  
*Cheng Li, Genyu Song, A. Pourtaherian and P. H. N. de With*

### SHORT PAPERS

- Sewer Defect Classification using Synthetic Point Clouds 891  
*Joakim Bruslund Haurum, Moaaz M. J. Allahham, Mathias S. Lynge, Kasper Schøn Henriksen, Ivan A. Nikolov and Thomas B. Moeslund*
- Learning to Correct Reconstructions from Multiple Views 901  
*Ștefan Săftescu and Paul Newman*
- Integration of Multiple RGB-D Data of a Deformed Clothing Item into Its Canonical Shape 910  
*Yasuyo Kita, Ichiro Matsuda and Nobuyuki Kita*
- Fiducial Points-supported Object Pose Tracking on RGB Images via Particle Filtering with Heuristic Optimization 919  
*Mateusz Majcher and Bogdan Kwolek*
- Visual-based Global Localization from Ceiling Images using Convolutional Neural Networks 927  
*Philip Scales, Mykhailo Rimel and Olivier Aycard*
- Unsupervised Gaze Prediction in Egocentric Videos by Energy-based Surprise Modeling 935  
*Sathyannarayanan N. Aakur and Arunkumar Bagavathi*
- Detecting Anomalies from Human Activities by an Autonomous Mobile Robot based on “Fast and Slow” Thinking 943  
*Muhammad Fikko Fadrimiratno, Yusuke Hatae, Tetsu Matsukawa and Einoshin Suzuki*
- Unsupervised Domain Adaptation for 6DOF Indoor Localization 954  
*Daniele Di Mauro, Antonino Furnari, Giovanni Signorello and Giovanni Maria Farinella*
- AUTHOR INDEX 963