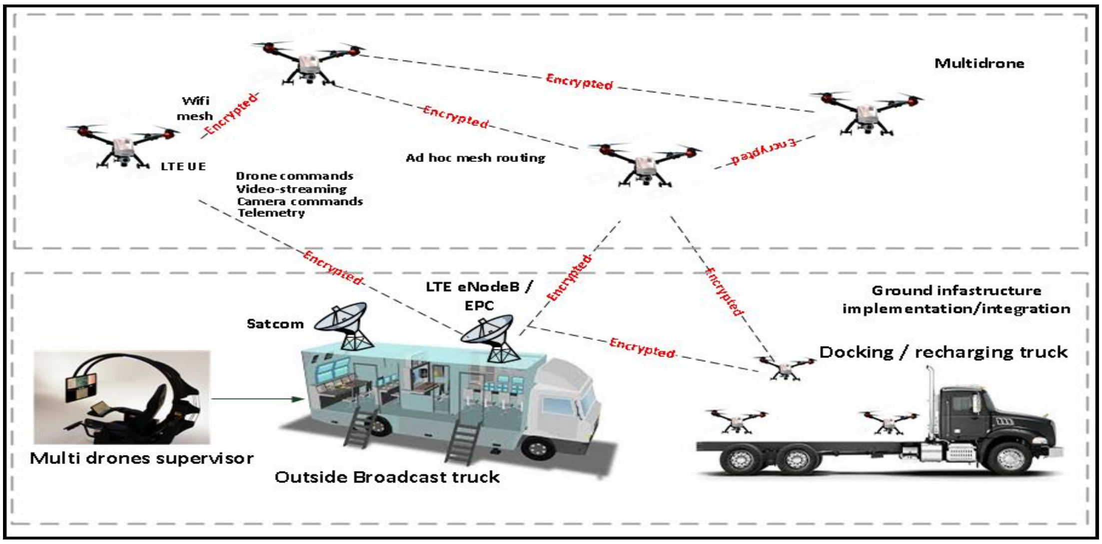
AVSS2018 Tutorial 27/11/2018
Deep learning and multiple drone vision

Tutorial Summary

Computer vision pays pivotal role both for drone cinematographic shooting and for drone safety. The aim of drone cinematography is to develop innovative intelligent single- and multiple-drone platforms for media production. Such systems should be able to cover outdoor events (e.g. sports) that are typically distributed over large expanses, ranging, for example, from a stadium to an entire city. In most cases, drone shooting has a target, e.g. cyclists, or boats in case of sports events. Deep learning is currently the principal approach in various computer vision tasks, notably object (shooting target, crowd, landing site) detection. The drone or drone team, to be managed by the production director and his/her production crew, shall have: a) increased multiple drone decisional autonomy for tracking/following the target and allowing event coverage in the time span of around one hour in an outdoor environment and b) improved multiple drone robustness and safety mechanisms (e.g., communication robustness/safety, embedded flight regulation compliance, enhanced crowd avoidance and emergency landing mechanisms), enabling it to carry out its mission against errors or crew inaction and to handle emergencies. Such robustness is particularly important, as the drones will operate close to crowds and/or may face environmental hazards (e.g., wind). Therefore, it must be contextually aware and adaptive, towards maximizing shooting creativity and productivity, while minimizing production costs. Drone vision and machine learning play a very important role towards this end, covering the following topics: a) drone localization, b) drone visual analysis for target, obstacle, crowd, point of interest detection, c) 2D/3D target tracking, d) privacy protection technologies in drones (e.g. face de-identification).

The tutorial will offer an overview of all the above plus other related topics, stressing the algorithmic aspects, such as:

a) drone imaging b) drone/target localization and world mapping c) target detection and tracking, d) privacy protection in drones.



Lecturer
Ioannis Pitas (pitas@aiia.csd.auth.gr), Aristotle University of Thessaloniki, Greece

Prof. Ioannis Pitas (IEEE fellow, IEEE Distinguished Lecturer, EURASIP fellow) received the Diploma and PhD degree in Electrical Engineering, both from the Aristotle University of Thessaloniki, Greece. Since 1994, he has been a Professor at the Department of Informatics of the same University. He served as a Visiting Professor at several Universities.

His current interests are in the areas of image/video processing, intelligent digital media, machine learning, human centered interfaces, affective computing, computer vision, 3D imaging and biomedical imaging. He has published over 860 papers, contributed in 44 books in his areas of interest and edited or (co-)authored another 11 books. He has also been member of the program committee of many scientific conferences and workshops. In the past he served as Associate Editor or co-Editor of 9 international journals and General or Technical Chair of 4 international conferences. He participated in 69 R&D projects, primarily funded by the European Union and is/was principal investigator/researcher in 41 such projects. He has 27600+ citations (Google Scholar) to his work and h-index 80+ (Google Scholar).

Prof. Pitas leads the big European R&D project MULTIDRONE: https://multidrone.eu/

Tutorial outline

The tutorial will consist of 4 talks, as detailed below:

Drone vision and cinematography: an overview

Abstract: This part of the tutorial will provide the general context for this new and emerging topic, presenting the aims of drone vision for cinematography and media production, the challenges (especially from an image/video analysis and computer vision point of view), the

important issues to be tackled, the limitations imposed by drone hardware, regulations and safety considerations etc.

1. Introduction to multiple drone imaging

This lecture will provide the general context for this new and emerging topic, presenting the aims of drone vision for cinematography and media production, the challenges (especially from an image/video analysis and computer vision point of view), the important issues to be tackled, the limitations imposed by drone hardware, regulations and safety considerations etc.
Use case scenarios for sports broadcasting, the challenges to be faced and the adopted methodology will be discussed first, followed by scenario-specific, media production and system platform requirements. The multiple drone platform will be also detailed, beginning with platform hardware overview, issues and requirements and proceeding by discussing safety and privacy protection issues. Finally, platform integration will be the closing topic of the lecture, elaborating on drone mission planning, object detection and tracking, UAV-based cinematography, target pose estimation, privacy protection, ethical and regulatory issues, potential landing site detection, crowd detection, semantic map annotation and simulations.

2. Mapping and localization

Abstract: The lecture includes the essential knowledge about how we obtain/get 2D and/or 3D maps that robots/drones need, taking measurements that allow them to perceive their environment with appropriate sensors. Semantic mapping includes how to add semantic annotations to the maps such as POIs, roads and landing sites. The section Localization is exploited to find the 3D drone or target location based on sensors using specifically Simultaneous mapping and localization (SLAM). Finally, the Fusion in drone localization section describes the improving accuracy on localization and mapping in Multidrone to exploit the synergies between different sensors.

3. Deep learning for target detection

Abstract: Target detection using deep neural networks, Detection as search and classification task, Detection as classification and regression task, Modern architectures for target detection, RCNN, Faster RCNN, YOLO, SSD, lightweight architectures, Data augmentation, Deployment, Evaluation and benchmarking.

Utilizing Machine Learning techniques in the video shooting process can assist the capturing of opportunistic shots, e.g., by detecting objects of importance and subsequently tracking them. Hence, Visual Object Detection can greatly aid the task of video shooting using drones. Drones with Graphics Processing Units (GPUs) in particular can be aided by Deep Learning techniques, as GPUs routinely speed up common operations such as matrix multiplications. Recently, Convolutional Neural Networks (CNNs) have been used for the task of object detection with great results. However, using such models on drones for real-time face detection is prohibited by the hardware constraints that drones impose. We examine various architectures and settings to facilitate the use of CNN-based object detectors on a drone with limited computational capabilities.

4. 2D Target tracking and 3D target localization

Abstract: Target tracking is a crucial component of many computer vision systems. Many approaches regarding face/object detection and tracking in videos have been proposed. In this lecture, video tracking methods using correlation filters or convolutional neural networks are presented, focusing on video trackers that are capable of achieving real time performance for long-term tracking on a UAV platform.

Intended audience

The target audience is expected to be researchers, engineers and computer scientists working in the areas of computer vision, machine learning, image and video processing and analysis that would like to enter in the new and exciting field of drone visual information analysis and processing for cinematography, media production or other related applications (e.g. surveillance).

Material to be distributed to attendees.

The presentations slides (in PPT or PDF) will be handed to the participants, in electronic form. A list of related publications will be also distributed.

List of related publications:

1. Multidrone Project (MULTIple DRONE platform for media production), funded by the EU (2017-19), within the scope of the H2020 framework. URL: https://multidrone.eu/
2. I. Mademlis, V. Mygdalis, N. Nikolaidis, I. Pitas, Challenges in Autonomous UAV Cinematography: An Overview, IEEE ICME2018, San Diego.
3. P. Nousi, E. Patsiouras, A. Tefas, I. Pitas, 'Convolutional Neural Networks for Visual Information Analysis with Limited Computing Resources', IEEE ICIP2018, Athens, Greece.
4. I. Karakostas, I. Mademlis, N. Nikolaidis, I. Pitas, 'UAV CINEMATOGRAPHY CONSTRAINTS IMPOSED BY VISUAL TARGET TRACKING', IEEE ICIP2018, Athens, Greece.
5. I. Mademlis, V. Mygdalis, C. Raptopoulou, N. Nikolaidis, N. Heise, T. Koch, J Grünfeld, T. Wagner, A. Messina, F. Negro, S Metta, I. Pitas, “Overview of Drone Cinematography for Sports Filming”, CVMP 2017, London, UK, December 2017
6. O. Zachariadis, V. Mygdalis, I. Mademlis, N. Nikolaidis, I. Pitas, “2D visual tracking for sports UAV cinematography applications”, 5th IEEE Global Conference on Signal and Information Processing (GLOBALSIP), Montreal, Canada, November 14-16, 2017
7. P. Chriskos, O.Zoidi, A.Tefas and I.Pitas, "De-identifying facial images using singular value decomposition and projections", Multimedia Tools and Applications, 2016.
8. "Cooperative Unmanned Aerial Systems for Fire Detection, Monitoring and Extinguishing", L. Merino, J.R. Martinez-de Dios, A. Ollero. In "Handbook of Unmanned Aerial Vehicles", ISBN 978-90-481-9706-4, Springer 2015.
9. M Tzelepi, A Tefas “Human crowd detection for drone flight safety using convolutional neural networks”, 25th European Signal Processing Conference (EUSIPCO) 2017 743-747
10. A. Torres-Gonzalez, J. Capitan, R. Cunha, A. Ollero and I. Mademlis, “A MULTIDRONE Approach for Autonomous Cinematography Planning”, Third Iberian Robotics Conference (ROBOT), 2017