ICCV2017 Tutorial
Drone vision for cinematography: An overview

Organizer, Lecturer: Prof. Ioannis Pitas
Aristotle University of Thessaloniki
pitas@aiia.csd.auth.gr

Contributors: Aristotle University of Thessaloniki, Alerion, Deutche Welle, RAI, Thales, Thales Services, University of Bristol, University of Seville, University of Lisbon (IST)
Multidrone case study: sports AV shooting
Rowing boat race
Giro d’Italia

<<Accident Detected>>
Multi-angled shooting of a football match
Drone vision objectives and challenges

• A) Improved multiple drone decisional autonomy, robustness and safety.

• B) Innovative, safe and fast multiple drone active perception and AV shooting.
Challenges in boat race shooting

A) Drone decisional autonomy, robustness and safety:
   • Crowd, obstacle detection and avoidance
   • Emergency landing site detection.

B) Multiple drone active perception and AV shooting:
   • Target tracking and following
A1) Improved multiple drone decisional autonomy, robustness and safety.

- Adaptive/cooperative/dynamic (online) multiple drone (re)planning:
  - cover a large-scale -over space and time- event.
- Improved, easy and transparent interaction with the production director and his/her crew.
- Improved decisional/cognitive autonomy and robustness, e.g. by: to be obtained by:
  - proper dynamic docking/recharge/emergency landing planning;
  - intelligent autonomous emergency handling;
A2) Improved multiple drone decisional autonomy, robustness and safety.

Improved safety during multiple drone mission execution by e.g.:
- a) embedded flight regulation compliance;
- b) enhanced vision-based crowd avoidance;
- c) a-priori emergency landing planning and autonomous multiple drone emergency landing re-planning.

• Robust video streaming, communication and synchronization.
B1) Innovative, safe and fast multiple drone active perception and AV shooting.

• Fast multiple drone semantic world modelling during pre-production, for e.g.:
  a) identification of shooting flight paths/formations
  b) establishing flight/shooting constraints.

• Fast innovative multiple drone vision- and GPS-/RFID-based target (e.g., boat, cycler, football player) tracking and shooting techniques:
  • to track people, crowds, or objects.
B2) Innovative, safe and fast multiple drone active perception and AV shooting.

- **Multiple drone AV shooting intelligence:**
  - Novel path/formation/camera control techniques.

- **Improved multiple drone human-centered visual information analysis both for individual persons and for crowds for e.g.**:
  - Better multiple drone AV shooting
  - Detecting unexpected events (e.g., audience/crowd reaction to goals/accidents).
Other issues-challenges

• Development of a mission-specific safe multiple drone platform having enhanced autonomy.
• Security and privacy issues.
• Overcoming barriers/obstacles due to regulations, public acceptance and other factors.
• Boosting public awareness and dialog.
Methodology

• End user requirements.
• HW/SW system specs, design, implementation, integration.
• Strong interplay between:
  a) mission (AV shooting) planning, mission control/execution.;
  b) active perception
• Pre-production:
  • semantic world mapping
  • mission planning.
• Production:
  • multiple drone flight/formation control
  • active perception (multiple drone and target localization tracking),
  • cinematographic AV shooting.
  • safety/emergency monitoring/sensing
  • emergency handling at the production phase.
Media production requirements

• Distinction between pre-production (planning) and production (live) requirements:
  • Pre-production requirements include e.g., maps management functionalities, definition of events of interest and shooting actions, visual target specification by examples.
  • Production requirements include e.g., management of unexpected events, emergency control by pilots, online replanning capabilities.
• Other general requirements are about video quality, video formats and codecs, camera control and camera control latency:
  • E.g., we want camera command latency < 1s.
Media production requirements - shooting

• Important part of media production requirements are about shooting capabilities of the system …

- No mutual shooting
- Defined flight patterns
- Leave looking room for subjects
- Central framing shot selection
- Rule of thirds shot selection
- 180 degree rule shot transition selection
- Reverse angles shot transition selection
- Avoid same camera perspectives

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)
Media production requirements – basic multi-drone shooting
System Platform requirements

• These requirements are from the point of view of media production, therefore they are mainly about:
  • Drone physical parameters (weight, maximum speed etc.)
  • Autonomy in flight and in perception
  • Storage, communication, control
  • Logistics
  • Interfaces to studio
  • Director ans flight supervisor dashboards.
Personnel and Roles

**Director.** Person in charge of the media production. Specify the shots to be taken by the drone team. He will interface with the system through the Dashboard.

**Supervisor Operator.** Person in charge of the security of the system. Throughout the Supervisor module, this person will validate plans as safe, and he will give a green light to the Director.

**Drone Pilots(?).** For security reasons, each drone will have a human pilot in charge to take over in case of emergency.

**Cameramen (?).** There will be a cameraman in charge of each camera on board the drones to take manual control if required by Director.
Drone vision for cinematography: HW issues

1. Drone platform:
   1. Flight machine
   2. AV and visual perception payload

2. Ground station platform

3. Drone-ground station communications

4. Human centered interfaces:
   1. Director, (photographers?)
   2. Flight supervisor, (pilots?).
Overview of UAV Parts

- **Drone core**
  - Flight Control Unit with main sensors, RTK GPS, Thales LTE & WiFi module

- **Batteries**
  - At least 2 batteries on the drone

- **Drone platform**
  - Frame, arms, landing gears, propulsion systems, ESCs

- **Audio-Visual Payload**
  - Audio-Visual camera and its motorised lens, 3 axis gimbal (for stabilisation), high quality images storage

- **Flight Payload**
  - Navigational camera, LIDAR, onboard computer, safety system with possibly parachute system
Communication infrastructure

- Drone 2 Drone Communication
- Drone 2 Ground communication
- Live broadcasting
Integrated software and hardware target

Reliable communication

QoS manager
ROS pub/sub
OpenWRT
WiFi driver
LTE/4G driver

Virtualisation infrastructure

Virtual Machine #1
Virtual Machine #2
Virtual Machine #3

Linux KVM/KQEMU hypervisor

On-board computing (multi-cores CPU Intel/ARM + GPU)

Mini PCIe
LTE/4G
WiFi

Serial/USB, Ethernet, ...

Ethernet, PCIe, USB, ...

Proposal

MultiDrone

Cooperative, intelligent, autonomous planning

Auto Pilot Software & Controller (ArduPilot)

Actuator and sensor interface with sensor fusion
Energy Management
Geofence & Safety controller & Emergency landing re-planning

ROS pub/sub bus

Attitude Control
Trajectory Control

Linux Companion computer

Activation perception

Sense and avoid
SCAM
Privacy (burning)

Media production

No way it field of view
Person & body parts recognition

Track and follow

NUTIX

Swappable execution location

Ubuntu

Incident reporting

Navigation devices

Preferably IP camera

AV shooting devices
This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)
Drone vision for cinematography: Functionalities (1)

1. (Multiple) drone mission planning.
2. (Multiple) drone mission control.
3. Active perception.
4. AV shooting.
High-level pre-production/production mission planning

- **High-level planner** to assign different behaviours/tasks to the multidrone team according to director and environmental requirements.

- The multidrone planner needs to be **scalable** with multiple actors, since on-line re-planning could be needed as events happen or execution is performed.
Mission Planning Vocabulary

• MULTIDRONE **Shooting Mission**: list of actions

• Types of actions:
  
  • **Shooting Actions**: drone + camera  
    e.g. Lateral Tracking, Fly-Over, Orbit, …

  • **Navigation Actions**: drone action only, does not involve shooting  
    e.g. Take-off, Land, Go-to-waypoint, …

• Shooting Actions are *event-triggered*:
  
  • A start event is associated to each Shooting Action, which will trigger the action when it occurs.  
    E.g. target reaches a milestone, start of race, …
Shooting Action Parameters

- **Shot type:**
  - Lateral shot, Orbital shot, etc.

- **Zoom type:**
  - Long shot, Medium shot, Close-up, etc.

- **Start position** for the drone and the camera look-at position

- **Triggering event**

- **Duration**

- **Target ID.**
Mission Planning/Control
On ground modules

• **Mission Controller:**
  • Interacts with **High-level Planner** to produce a mission plan.
  • Monitors mission execution.
  • Asks for replanning if needed.

• **Event Manager:**
  • Receives, manages, and generates events.
  • Sends events to drones to start and stop action execution.
Mission Execution
On drone modules

Onboard Scheduler:
- Receives list of actions
- Receives events to trigger action execution
- Activates the Action Executer
- Sends drone status to ground

Action Executer:
- Translates Shooting Actions into desired drone+camera configurations
- Interacts with other modules to produces commands for autopilot, camera and gimbal

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)
Drone vision for cinematography: Functionalities (2)

1. Perception and localization. SLAM:
   1. Semantic 3D world mapping
   2. Drone localization.

2. Visual and perception data analysis for AV Shooting:
   1. 2D target (athlete, boat, cycle) detection and tracking
   2. 3D target localization and following
   3. Drone cinematography
   4. Target pose estimation.
UAV Simultaneous Localization and Mapping

On-board the UAV
- Intermittent GPS
- Small-baseline stereo
- Other sensors e.g. IMU
- Rough path
- Fusion and estimation
- Control Guidance
- Navigational 3D Map
- Onboard Visual SLAM
- 2D Map (if available)
- Recorded or live imagery

Tactical input
Field user
UAV in urban canyon

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)
Object detection

- State-of-the-art object detectors are based on very deep and multiple-channel CNNs.
- Current neural detectors are very accurate.
- Multiple layers of many convolutional filters are applied to the input image, forming a very deep architecture of successive convolutions and optionally some fully connected components.

- Trained on large-scale datasets, such as
  - VOC2007 with 10k images depicting ~24k objects belonging to 20 classes
  - VOC2012 with ~11k images depicting ~27k objects belonging to the same 20 classes as VOC2007
  - COCO with 328k images, about 2.5 million objects belonging to 91 classes.
UAV Object Tracking

- 2D visual tracking will be employed for target following.
- Satisfactory performance in UAV sports footage is required.
- Target tracking should be performed in real-time i.e. > 25 FPS.
- On drone implementation might be required as well, thus low computational complexity is preferred.
- Parallel or parallelizable methods (e.g., with CUDA implementations) should be preferred as well.
- Assuming 2D target tracking methods operate faster than combining target detection and recognition methods, long-term UAV tracking is also preferred.
UAV Object Tracking

• 14 top performing 2D trackers [VOT 2016] were implemented in MATLAB using the UAV123 dataset interface.
• Performance was evaluated in 26 UAV videos obtained from UAV123 and YouTube, including long term videos as well.
• 3-fold evaluation:
  • Precision plot (the ratio of successful frames whose tracker output is within the given threshold (x-axis of the plot, in pixels) from the ground-truth, measured by the center distance between bounding boxes)
  • Mean time before success rates falls below y%, y=10,..,100
  • Operation speed
• Evaluation platform: Ubuntu 16.04, 8GB ram,i7
UAV Object Tracking

- ASMS provides a good compromise between accuracy and fps rate
Optimal multi-sensor multi-drone 3D target localization, tracking & following

• Problem: maximize a merit metric resulting from multi-drone object tracking

• Assumptions:
  • M drones
  • All cameras can be oriented
  • Drones motion is assumed known

• Objective: maximize a merit metric
  • resulting from multi-drone
UAV Shot Type Taxonomy

• There is no prior, comprehensive work on identifying aerial shot types suitable for UAV-based cinematography

• In this context, shot type refers to a combination of camera/UAV motion (with regard to the target) and target composition/framing

• 23 motion types have been identified, each one compatible with a number of framing types

• Not all of these motion types include shooting a specific target/subject (e.g., establishing shots)
UAV Shot Type Identification

- Example: CHASE
Mathematical UAV Shot Type Modelling

• The identified UAV motion types have been modeled mathematically, using two coordinate systems:
  • A global World Coordinate System (WCS)
  • A Target Coordinate System (TCS), having its origin on the current target location at each time instance.

• This analysis can prove useful for several tasks.
  • Example: analytically determine maximum allowable camera focal length when the UAV orbits a moving target, so that 2D visual tracking does not fail

• 3 additional shot types, specifically for multiple drone shooting, have been specified:
  • 2-UAV Chase, 3-UAV Orbit, Dancing Drones.
Target Pose Estimation

• **Computer Vision Approach**
  • Relies on detecting a set of *predefined points* (e.g., facial landmarks) and then using a method for solving the respective *Perspective-n-Point* (PnP) *problem*, i.e., estimation of the camera position with respect to the object.

• **Limitations:**
  • The 3-D coordinates for the landmark points must be known, i.e., a 3-D model of the object is needed
  • The landmarks points must be precisely tracked, i.e., the texture of the object must allow for setting enough discriminative landmarks
Target Pose Estimation

• **Machine Learning Approach**
  • A neural network receives the object and directly *regress* its pose
  • Only a set of pose-annotated object pictures are needed
    • There is no need to manually develop 3-D models
    • The model are more robust to variations of the object for which we want to estimate its pose
    • The pose estimation can run entirely on GPU and (possibly) incorporated into a unified detection+pose estimation neural network
  • Very few pre-trained models are available
    • Models must be trained for the objects of interest (faces, bicycles, boats, etc.)
Target Pose Estimation

- **Machine Learning Approach**
  - We integrated a pre-trained yaw estimation model of facial pose (DeepGaze library) into the SSD-300 object detector (trained to detect human faces)
  - Varying illumination conditions seem to affect the estimation.
Drone vision for cinematography: Functionalities (3)

1. Visual and perception data analysis for safety and security:
   1. Obstacle detection.
   2. Event detection.
   3. Privacy protection.
   4. Emergency landing site detection.
   5. Crowd detection.
Privacy Protection

• An issue of ethics and security
• Post-production stage
• Approaches
  • Face de-detection (Face detector obfuscation)
    • Naïve approach
    • SVD-DID
  • Face de-identification (face recognizer obfuscation)
    • Gaussian blur
    • Hypersphere projection
Privacy Protection: acceptable facial image quality?

Original Image

Gaussian blur with std. deviation of 5

Hypersphere projection with radius of 8
Privacy protection: Trade-off between de-identification performance and facial image quality
Face recognition/de-identification/privacy protection
Potential Landing Site Detection

INPUT (3D projection)

INPUT (2D projection)

OUTPUT: Safe landing areas in Blue color

INPUT/OUTPUT of the novel algorithm

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 731667 (MULTIDRONE)
Potential Landing Site Detection

• **Algorithm for identification of potential landing areas using digital elevation models (DEM):**
  • Utilizes:
    • Digital Terrain Models (DTM): just terrain, no manmade structure or vegetation
    • Digital Surface Models (DSM): terrain+buildings+vegetation
  • Detects buildings & vegetation by subtracting the DTM from DSM and applying a threshold
  • Uses Sobel operator and a threshold to extract the gradient (slope) of the surface in order to classify the DTM pixels in flat or non-flat areas
  • Uses mean shift segmentation to identify connected regions and classify them in landing zones or not based on thresholding
    • Small areas are discarded.
  • **Final result:** sufficiently large map areas with no buildings/vegetation and small slope
Crowd Detection

- A *Fully Convolutional Neural Network* can be trained for Crowd Detection
- The result is a heatmap
Semantic 3D Map Annotation

• Annotations derived through drone video analysis are projected on the 3D map.
• Starting point: projective texture mapping (graphics) – project a textured image onto a scene as if by a slide projector
• Assumes that we know the camera extrinsic and intrinsic parameters
Semantic 3D Map Annotation
ICCV2017 Tutorial lectures

• SLAM (Prof. Montiel, U Zaragossa)
• Tracking fusion (Prof. Martniez-de Dios, U Seville, MULTIDRONE)
• Drone flight and formation control (Dr. Cunha, U Lisbon, MULTIDRONE)
• NN visual analysis (Prof. Tefas, U Thessaloniki, MULTIDRONE)
• Drone cinematography (Prof. Nikolaidis, U Thessaloniki, MULTIDRONE)
• Privacy protection (Prof. Pitas, U Thessaloniki, MULTIDRONE).
Multidrone Consortium

1. Aristotle University of Thessaloniki, Greece (Coordinator)
2. Thales Communications & Security SAS
   a. Thales Services
3. University Of Bristol
4. The University of Seville
5. Deutsche Welle (DW)
6. RAI Radiotelevisione Italiana (RAI)
7. Alerion
8. Instituto Superior Técnico, Portugal
Get involved: Cooperation with other groups and projects

• Have a look at www.multidrone.eu
• Student forum on drone technologies.
• Creation of a European (and not only) SIG on drone technologies.
• Competitions on drone cinematography.
• Organization of an open call workshop in 2019.
• Special sessions and special issues.
• Open to any new idea and cooperation options!
  • Send message to pitas@aiia.csd.auth.gr
Thank you very much for your attention
Enjoy the rest of the tutorial!

Contact: Prof. I. Pitas
pitas@aiia.csd.auth.gr
www.multidrone.eu