ULTINOUS

# INTELLIGENT VIDEO ANALYSIS TECHNOLOGIES

## TECHNICAL WHITE PAPER

Ultinous has developed a series of computer vision technologies utilizing the latest deep learning research results. These technologies provide state-of-the-art accuracy along with high speed processing to make them capable for processing HD video streams in real-time.

The main goal of these technologies is to transform the huge amount of raw video data to compact summaries and alerts. Extracting useful information from video data can be applied in many areas from security to home or retail. With the technology of Ultinous potentially we can seat an intelligent observer in front of every camera feed and transform the video data into valuable information in real-time.

Running these algorithms require high-end GPU infrastructure. Ultinous provides a cloud based and an appliance based packaging option of these technologies to best meet client requirements of deployment. Our optimized server typically contains more Nvidia GPU-s (Nvidia Titan X) and is capable to process dozens of real-time video streams parallel.

---

# TECHNOLOGIES

The computer vision technologies presented below can be applied to images as well, however they are all optimized to efficiently process video data in real-time so the main input of these algorithms are video streams typically coming from IP cameras.

## BACKGROUND EXTRACTION

The algorithms can be best used if the cameras are steady, as typical in most surveillance scenarios. For steady cameras the algorithms maintain a background model and analyse only the “difference” from the background model. This method significantly decreases compute resources.

## FACE/HEAD DETECTION

This technology can detect faces and/or heads in images or video streams. In face mode it only detects faces with a wide tolerance in pose:
+-40 degrees in yaw
+-30 degrees in pitch
+-30 degrees in roll

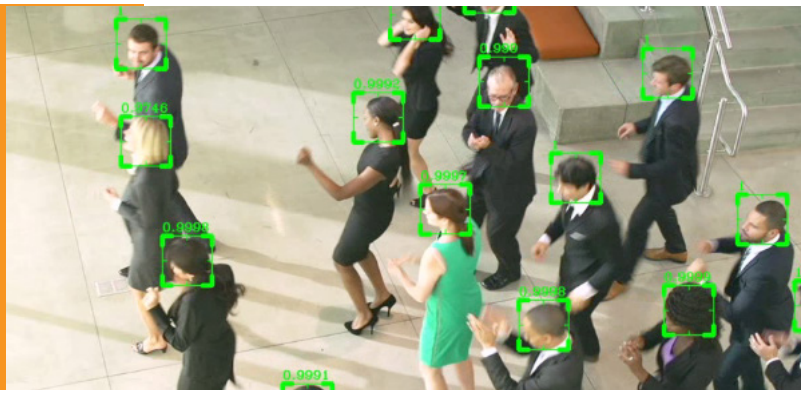
In  head mode it can detect human heads in any pose from any angle if most of the head is visible. Typical surveillance camera positions (as shown in the picture) are optimal for counting people (heads do cover each other). Minimum detectable face/head size is 30x30 pixels. The user can specify region of interest and face/head size range.

OUTPUT:

- 2D bounding box for each detected face/head

EXAMPLE USES:

- Measure waiting line length (retail, customer service)
- Measure resource utilization (eg.: promotion game console)
- Measure crowd size
- Input for face or facial expression recognition



## PASSAGE COUNTING

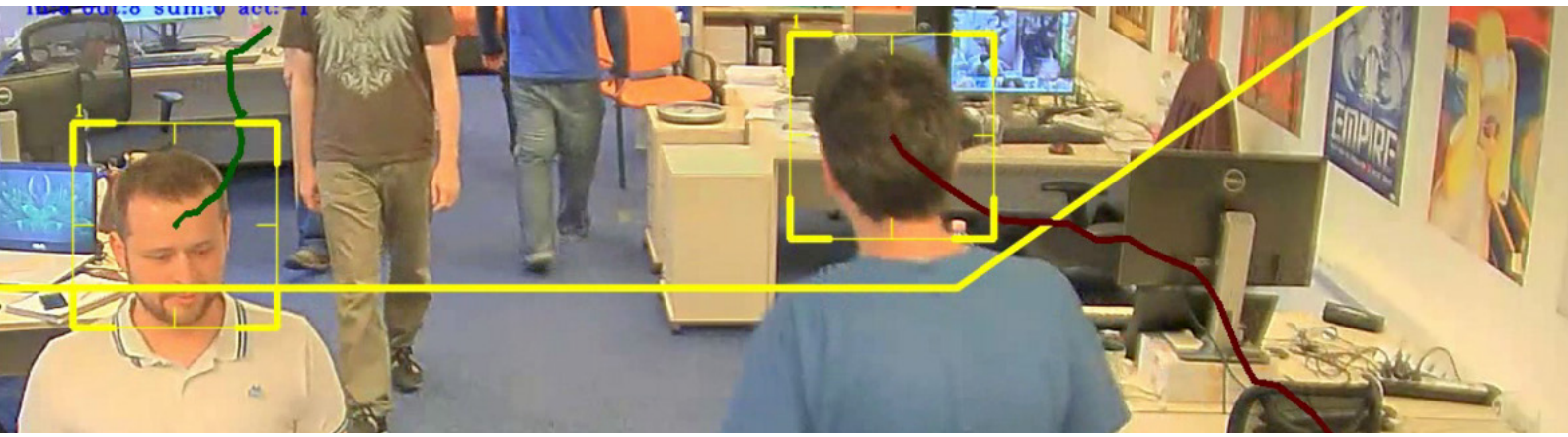
This technology is capable of counting people crossing an imaginary line (eg.: enter or leave a shop) in real-time. People can be counted from head size as small as 30x30 pixels. Heads of people must be visible to be able to count. Line crossing events (enter, leave an area) are generated real-time and able to trigger alerts. Multiple simultaneous crossing people are counted accurately making the system capable of achieving above 99% people counting accuracy.

OUTPUT:

- Time of entry and exit events in millisecond precision

EXAMPLE USES:

- retail visitor counting



## DEMOGRAPHICS DETECTION

This feature provides information on a person's gender and age. Face detection requirements are the same as for face recognition (see below).

### EXAMPLE USES:

Demographic distribution measurement for retail
Better target advertisements

## FACE RECOGNITION

Our face recognition technology uses state-of-the-art deep learning methods and matches human level face recognition performance. Face recognition works on face crops greater than 100x100 pixels, however, for optimal performance face size should be at least 150x150.
The recognition algorithm tolerates wide range of poses
+-30 degrees in yaw
+-20 degrees in pitch
+-20 degrees in roll

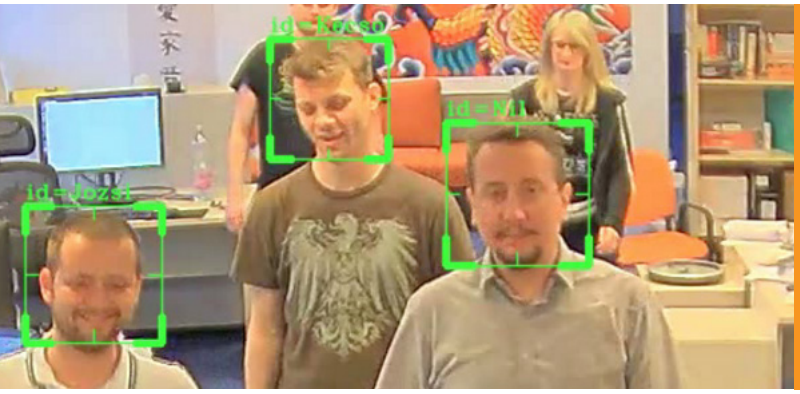
The technology is ideal for recognizing people in a non-cooperative mode (people do not have to stop and look at the camera). In video processing mode the recognition uses all the frames that belong to the same person's track to further improve recognition accuracy.

### OUTPUT:

2D bounding boxes (or track ID), ID of registered persons with millisecond timestamp of observation
Feature vector of the faces. This can be used to measure face similarity.

### EXAMPLE USES:

- Entrance system
- Measure similarity between persons
- Reidentify visitors, measure visit time distribution



# REAL-TIME 3D TRACKING

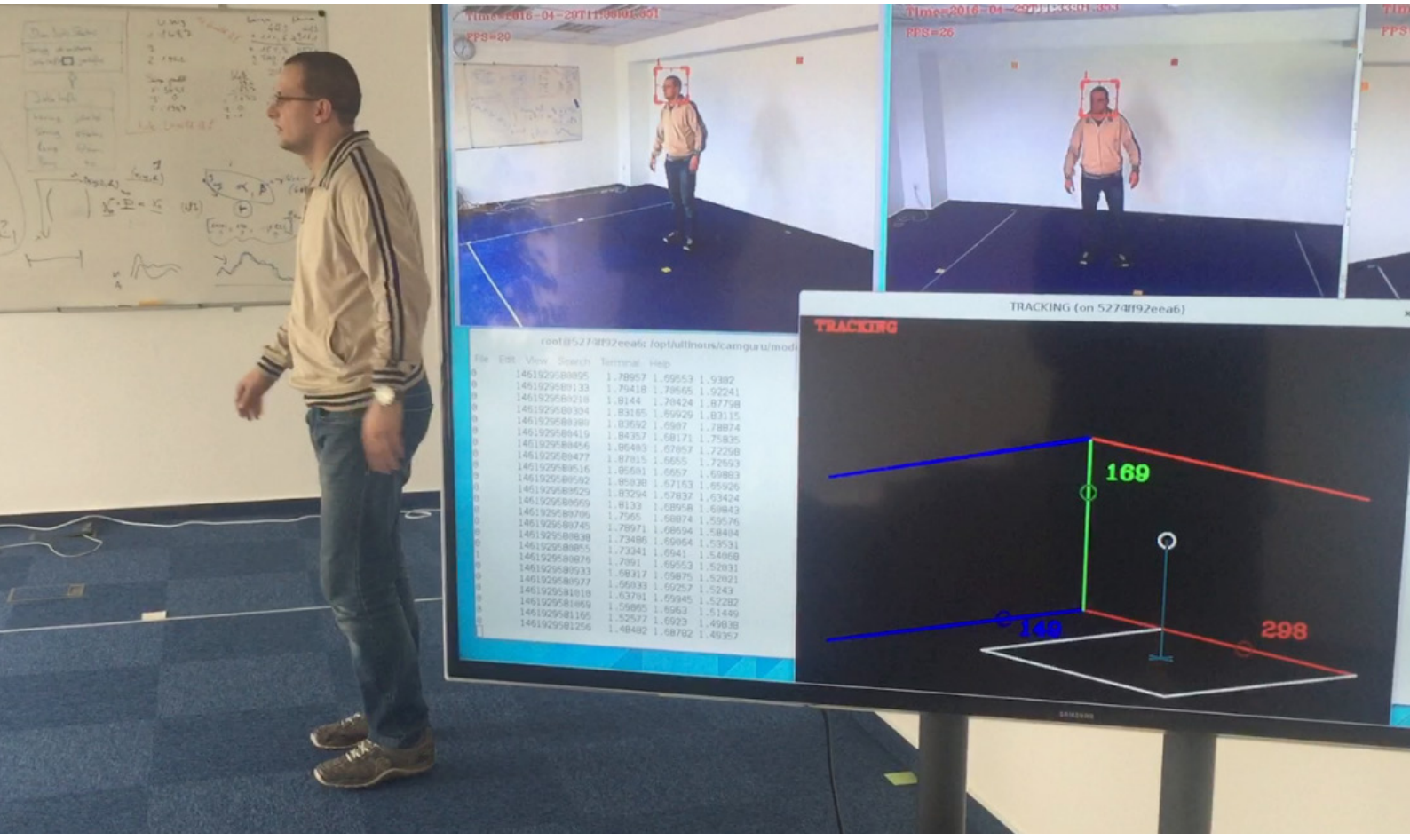
Our real-time tracking technology is capable of tracking persons in 3D using two or three camera streams. 3D tracking error can be under 4 cm while maintaining 30 FPS. (Higher frame rate can be achieved with high FPS cameras and additional GPU resources.)

## OUTPUT:

- 3D position stream (30 FPS) of persons

## EXAMPLE USES:

- sport statistics
- control theater lights




ULTINOUS