Traffic Detector with FW 8.10

Technical Note

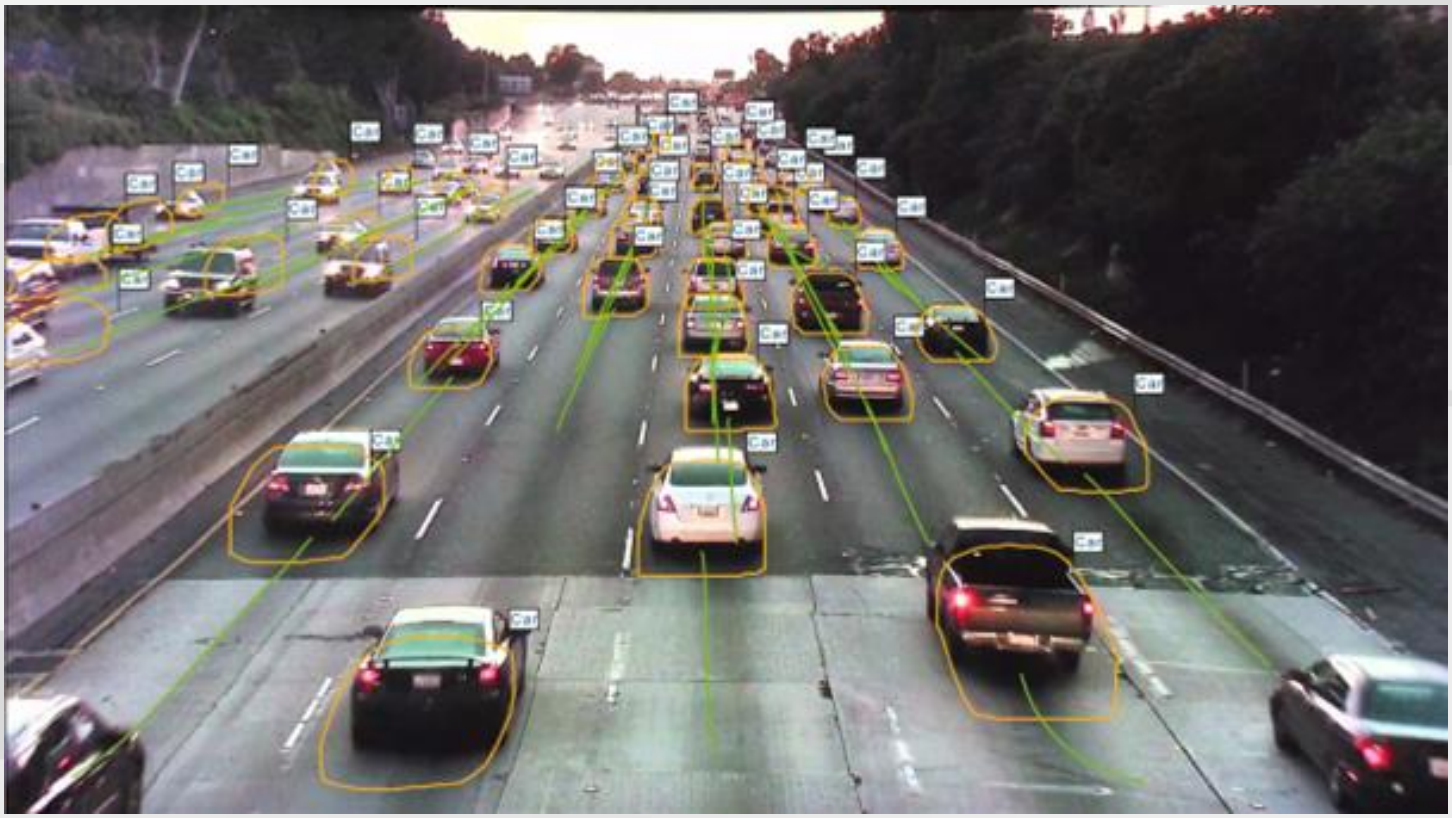


Table of contents

1 Introduction 3

1.1 Applications 3

1.2 Limitations 3

1.3 Supported cameras 3

2 Intelligent Video Analytics, Camera Trainer and Traffic Detector 5

3 Technical Background 7

3.1 Machine learning: Finding threshold between target object and world 7

3.2 Will machine learning, especially deep learning, replace current video analytics? 8

4 Configuration 9

4.1 Activating Traffic Detector 9

4.2 Configuring alarm or counting tasks 9

1 Introduction

Traffic detector has been developed to cope with all traffic densities up to high density like congestions and queues in front of traffic lights. It also detects non-moving vehicles and is thus applicable for smart parking. In addition to vehicles, it also detects pedestrians.

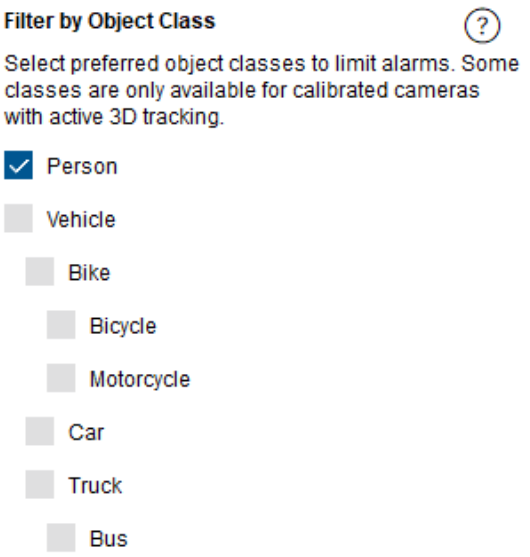
1.1 Applications

- Tunnels & Highways:
  - Collection of traffic statistics
  - Detection of congestions for automatic speed signaling
  - Detection of wrong way driving
- Intersections:
  - Detecting presence and volume of vehicles
- Smart parking:
  - Detect parked cars

1.2 Available object classes

The following object classes and object class filter are available:

- Person
- Vehicle
  - Bike
    - Bicycle
    - Motorbike
  - Car
  - Truck
    - Bus



Object classes are hierarchical. That means e.g. a bicycle is also a bike is also a vehicle, and a bus is also a truck is also a vehicle. Object class filters fully support this hierarchy, while visual class labels will only show the deepest level of classification, that is they will show person, bicycle, motorbike, car, truck and bus labels.

1.3 Limitations

- Setup is only possible via the Configuration Manager.
- Detection of persons, bicycles, motorbikes, cars, trucks, busses.
  - Persons, bicycles and motorbikes may be confused, especially when seen from the front.
  - Busses and trucks may be confused.
- Minimum object size: 16x16 pixel in 640x360 resolution
- Maximum object size: 500x500 pixel in 640x360 resolution
- Minimum object visibility: 50%. Objects occluded more than 50% may not be detected.
- 2D traffic tracking requires 50% overlap between two consecutive frames. Fast objects are only tracked well moving toward or away from the camera. Fast vehicles crossing the field of view will not be tracked properly.
- Speed, geolocation and color are only available in 3D traffic mode.
- Traffic Detector and Camera Trainer cannot run in parallel.
- Top-down views (birds eye views) are not supported.

1.4 Supported cameras

Traffic Detector is available on the following cameras:

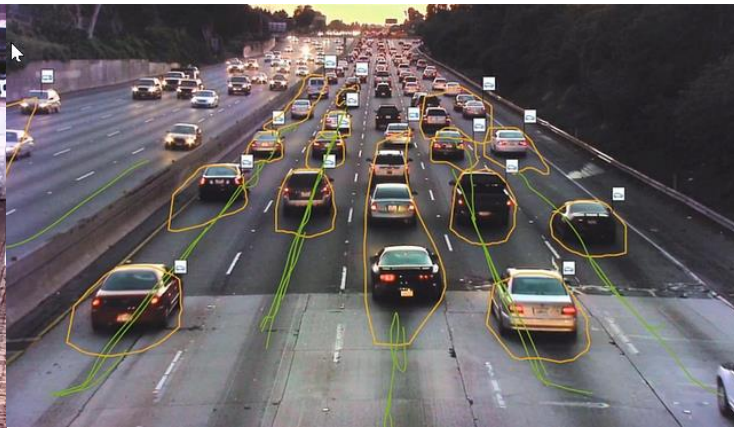
- AUTODOME inteox 7000i:
  - NPD-7602-Z30-OC
  - VG5-ITS1080P-30X7
- DINION inteox 7000i:
  - NBE-7604-AL-OC
- FLEXIDOME inteox 7000i:
  - NDE-7604-AL-OC
- MIC inteox 7000i:
  - MIC-7602-Z30BR-OC
  - MIC-7602-Z30WR-OC
  - MIC-7602-Z30GR-OC
  - MIC-7604-Z12BR-OC
  - MIC-7604-Z12WR-OC
  - MIC-7604-Z12GR-OC
  - MIC-ITS1080P-GE30X7
  - MIC-ITS1080P-WE30X7
  - MIC-ITS1080P-BE30X7
  - MIC-ITS1080P-B30X7
  - MIC-ITS1080P-W30X7
  - MIC-ITS1080P-G30X7
  - MIC-ITS4K-BE12X7
  - MIC-ITS4K-WE12X7
  - MIC-ITS4K-GE12X7

2 Intelligent Video Analytics, Camera Trainer and Traffic Detector

Intelligent Video Analytics was designed for intrusion detection and works well in scenes where objects are visually well separated. For traffic, that translated to low and medium traffic. In high-density traffic, where vehicles merge visually, Intelligent Video Analytics is not able to separate the vehicles anymore. Furthermore, Intelligent Video Analytics only detects moving objects.

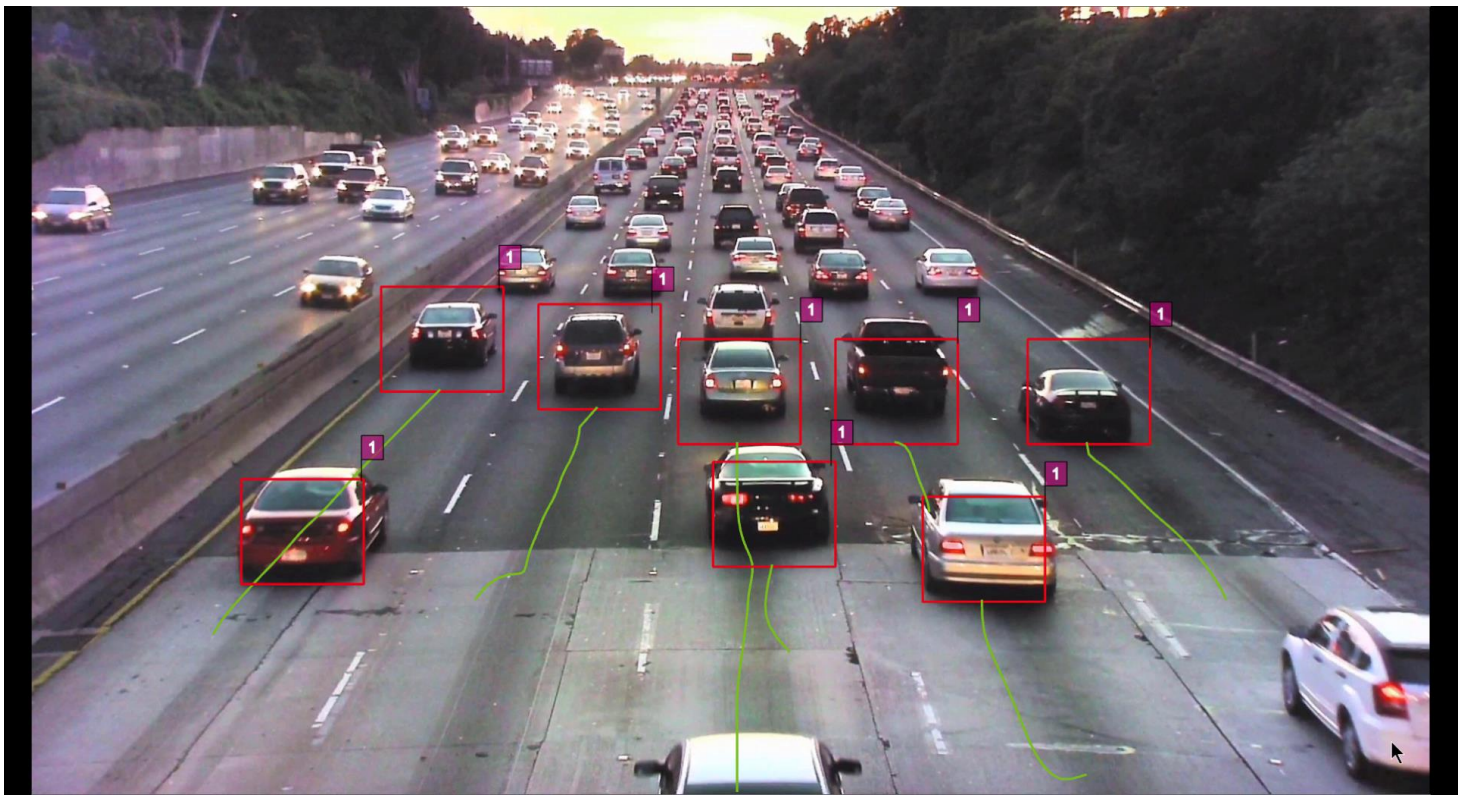


Intelligent video Analytics works well in low / medium traffic

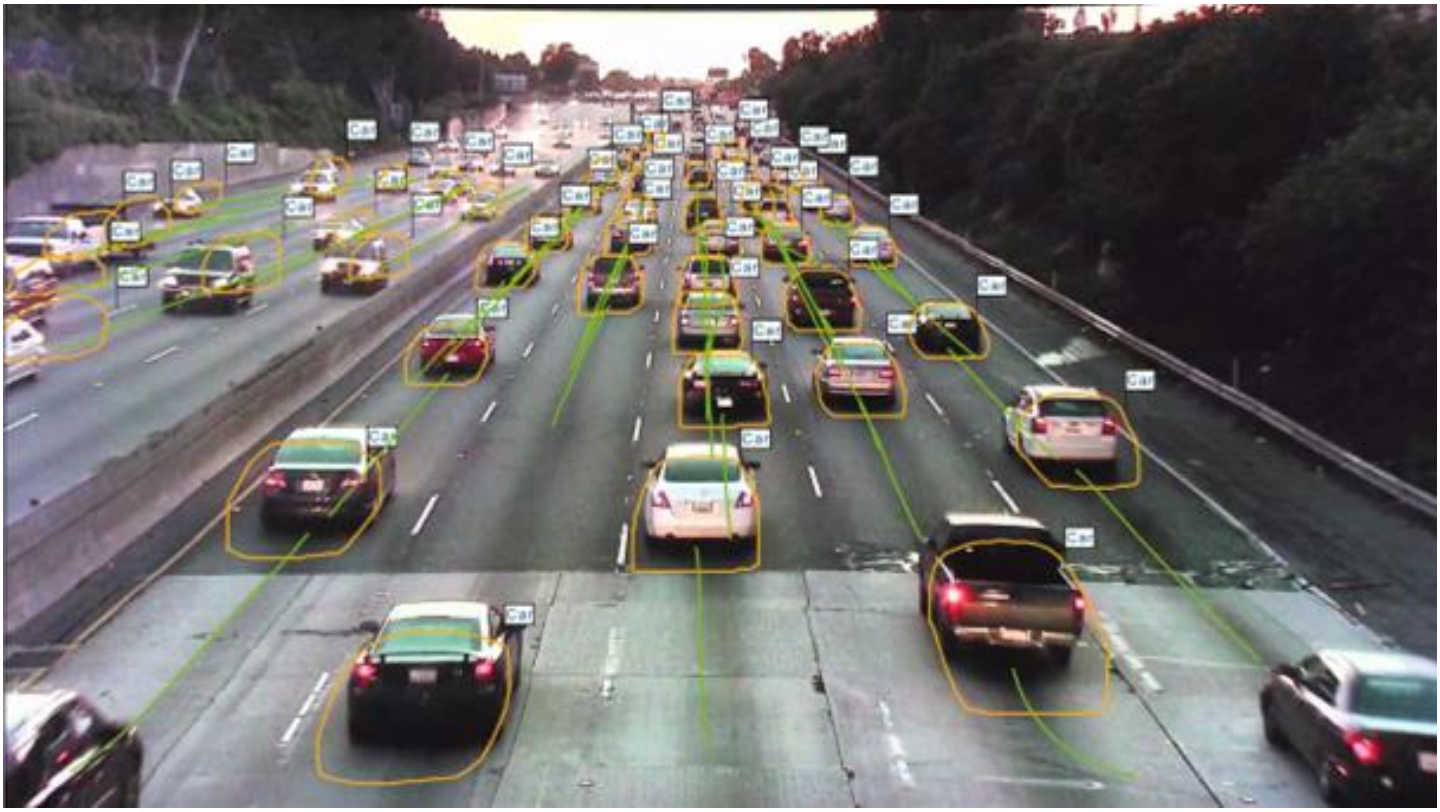


In high traffic, objects merge together

To separate vehicles in high traffic, or to detect parked vehicles, machine learning is needed. Camera Trainer was Bosch's initial solution for these applications. Camera Trainer's low processing power requirements made it ideal for use on Bosch IP cameras based on the CPP6, 7 and 7.3 common product platforms. However, Camera Trainer was limited to a short distance and needed to be trained by the user for every scene, resulting in high training effort. (The advantage of Camera Trainer is that any kind of rigid object can be trained. For more information please refer to the Camera Trainer Tech Note.)



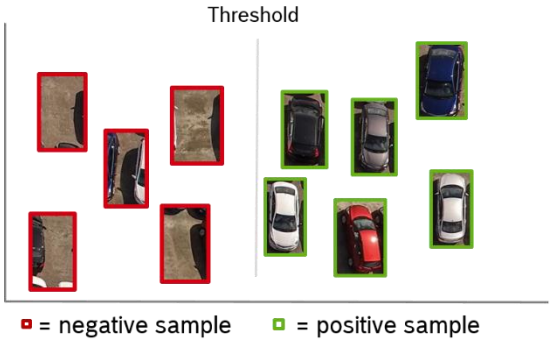
Traffic Detector is a pre-trained vehicle and person detector which also supports greater detection distances than Camera Trainer, though less than Intelligent Video Analytics. It separates persons, bikes, cars, trucks and busses even in dense congestion or traffic queues. Another benefit of traffic detector is that it is robust with regards to shadows or headlight beams.



3 Technical Background

3.1 Machine learning: Finding threshold between target object and world

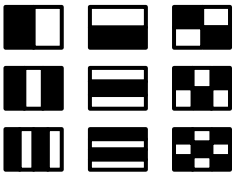
Machine learning for object detection is the process of a computer determining a good separation between positive (target) samples and negative (background) samples. In order to do that, a machine learning algorithm builds a model of the target object based on a variety of possible features, and of thresholds where these features do/don't describe the target object. This model building is also called the training phase or training process. Once the model is available, it's used to search for the target object in the images later on. This search in the image together with the model is called a detector.



Hand-crafted features typically describe edges and include:

- Haar features: Binary edge descriptors via black/white tiles
- Histogram of oriented Gradients (HoG): Histogram of quantized edge direction and contrast

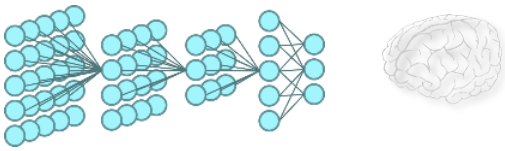
The resulting model will typically have around 2000 parameters. There are different methods of machine learning with hand-crafted features including support vector machines (SVNs), AdaBoost, and decision trees. Each of these methods has certain advantages, however, all of them result in similar performance levels. A detector based on these features can typically run in real-time in current network camera hardware.


Haar features


Histogram of oriented Gradients (HOG)

Camera Trainer is based on an SVN using Histogramms of oriented Gradients. For more details on Camera Trainer, see its own Tech Note.

Neural networks are based on the visual cortex and are able to learn descriptive features on their own. They utilize a neural network structure of optimization parameters instead of using the handcrafted features. Typically, neural networks for image processing will also learn edge features and combine them first to parts of the object, and then the full target object itself. Deep neural networks for image processing use roughly ~20 million parameters, and can deliver a performance boost of up to 30%. On the other hand, deep neural networks are a brute force approach that requires hundreds of times more computational power.


Neural networks: Layers of neurons. Each neuron combines several inputs to a decision.

Besides the model and training method, the samples of target objects and the background are very important. For a task like face or person detection, the positive samples need to show all possible variations including perspective and pose, while the background samples have to represent the full world. Therefore, machine learning needs tens of thousands of examples of the target object and billions of examples of what the rest of the world looks like. This is a huge effort both in collecting and preparing the sample images. For automated machine learning, either the target objects have to be marked in the image or the images restricted to show only the target object. Furthermore, modelling the complexity of the full world is one of the reasons machine learning is computationally expensive.

The new Bosch Traffic Detector is trained using a deep neural network.

3.2 Will machine learning, especially deep learning, replace current video analytics?

No. It will extend the possible applications.

Current intrusion detection has to cope with long detection distances, correspondingly few pixels on the target object, a wide variety of poses (walking, running, crawling, rolling) and needs to be able to deal with the unexpected. It can run on very low processing power. It is targeted towards moving objects, and will ignore all stationary ones.

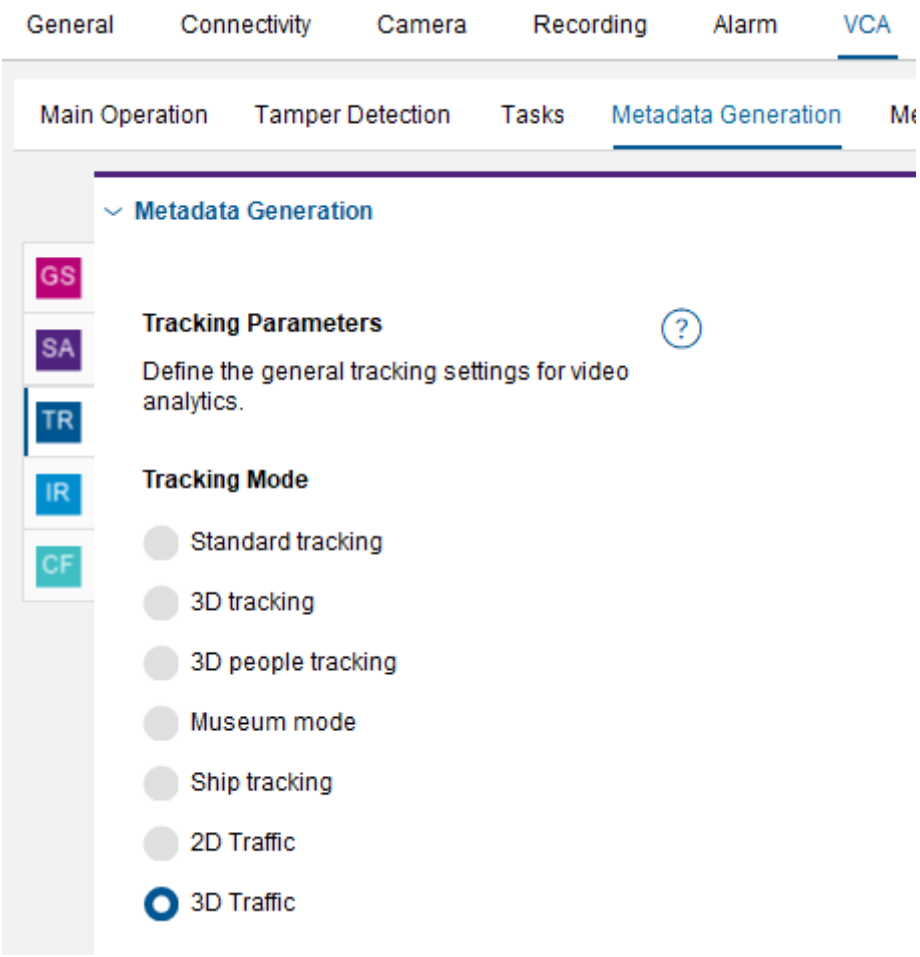
Machine learning, on the other hand, needs high resolution and will thus only work in near range. It also needs high processing power, which is one of the main reasons adoption is not more widespread yet. Furthermore, it can only detect trained and expected objects. But with machine learning, objects can be properly classified, and close objects can be separated well. Non-moving objects can still be detected, even after a very long time without motion. Therefore machine learning in general is a good technology for applications like parking lot occupancy, traffic monitoring, as well as for counting people without false counts from shopping carts or baggage.

Furthermore, a detector itself only detects an object in a single frame. For most application, on the other hand, the movement of the object over time is even more important, and a tracker is typically needed to combine the single detections, add robustness, and extract information like location, speed and direction. The best solutions arise when knowledge from several sources is combined, like detectors, trackers, optical flow, perspective information and more.

4 Configuration

4.1 Activating Traffic Detector

Activation of the Traffic Detector is available in VCA->Metadata Generation->Tracking Mode via 2D Traffic or 3D Traffic. Note that Camera Trainer will automatically be disabled when Traffic Detector is enabled and vice versa. Traffic Detector objects behave as all other video analytics objects, and the usual alarm and counting rules can be applied.



4.2 2D or 3D Traffic Mode?

2D Traffic is the choice for static applications like parking lot occupancy. A simple tracker combines the detector output over time, and objects need to overlap for 50% in two consecutive frames to be tracked properly. Thus, fast objects crossing the field of view may not be tracked properly.

- Only bounding boxes will be outputted.
- No color, speed, geolocation, nor direction filter
- No idle/removed object detection

3D Traffic is the choice whenever speed, map location / geolocation, and best tracking performance are needed. It requires the camera to be calibrated properly in order to understand the perspective in the scene and convert pixel into real-world size, speed and location. Once an object has been detected with the traffic detector, the tracker learns the appearance of the object and is able to follow it on its own.

- Bounding boxes for static objects, closer fitting shapes for moving objects
- Color, speed, geolocation, direction available
- Idle/Removed Object: only Stopped Object detection



Left: Bounding boxes are always used in 2D Traffic, and used for non-moving objects in 3D Traffic. Right: Once an object starts moving in 3D Traffic mode, the outline switches to a flexible shape. Visualization will switch back to bounding box only for cars, busses and trucks, once the objects stops long enough and is considered parking. Use VCA->MetadataGeneration->Idle/Removed->Stopped Object Debounce Time to determine when vehicles will switch to static objects handling. Recommendation: Set this longer than typical traffic light red phase in case of intersections. For persons and bikes, tracking will stop completely after this time.

Bosch Sicherheitssysteme GmbH
Robert-Bosch-Ring 5
85630 Grasbrunn
Germany
www.boschsecurity.com
© Bosch Sicherheitssysteme GmbH, 2019