

Which Bosch camera settings influence video analytics performance and why?

Whitepaper

Table of contents

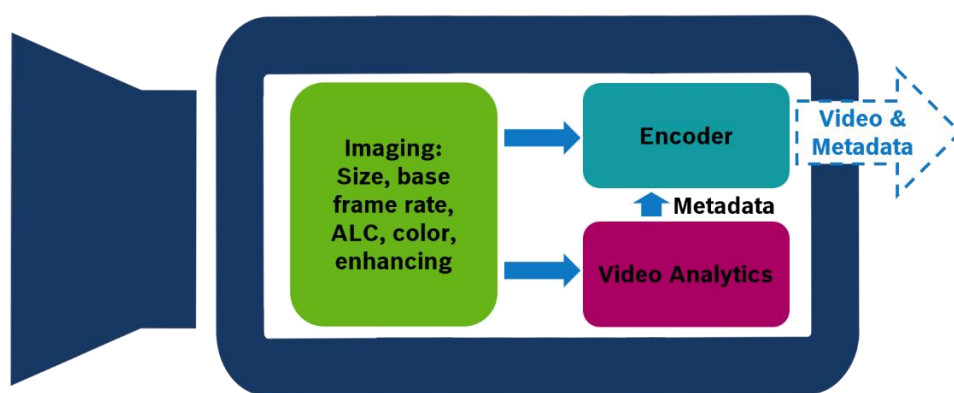
1 Introduction & Overview	3
2 Resolution & Base Frame Rate	3
2.1 Video analytics frame rate dropping?	4
3 Scene modes & image enhancement	4
3.1 Brightness, contrast, saturation, white balance	5
3.2 Automatic level control (ALC): ALC level, shutter, day/night.....	5
3.3 Advanced: Contrast, noise reduction, backlight compensation	5
4 Privacy masks & display stamping	6
4.1 Privacy masks	6
4.2 Display stamping.....	6
5 Encoder settings	6
6 Thermal camera	6

1 Introduction & Overview

Which camera settings influence video analytics performance in Bosch IP cameras, and what are the best practices for their settings? In general there are three different processing blocks:

- ▶ Imaging, which reads out the camera sensor and generates the video images
- ▶ Encoder, which takes these images and prepares them for transmission over the network
- ▶ Video analytics

In Bosch IP cameras, the following video analytics can be available: Intelligent Video Analytics (IVA) including IVA Pro, IVA Flow, Intelligent Tracking, Essential Video Analytics, and MOTION+. For simplicity, all those will be summarized as “video analytics” in this whitepaper wherever differentiation is not necessary. Those video analytics get the images to analyze directly from the imaging block and are completely independent of the encoder, which starts processing the images in parallel. Thus, all settings influencing imaging also influence video analytics, while the encoder settings are of no importance for video analytics performance. There are a view special cases like privacy masking, display stamping and thermal cameras, which will be covered here as well.



As the encoder sends out the video images as soon as possible, and video analytics are running in parallel, the video analytics results and metadata are typically send out 1-2 frames after the corresponding images. In live view, everything is displayed as soon as it is available to minimize the lag caused by encoding, transmission and decoding. Therefore no synchronization of video and metadata occurs, and the objects / changes detected by the video analytics seem to lag behind. When viewing recorded video, this is no longer visible as proper synchronization is done there.

2 Resolution & Base Frame Rate

The application variant of the camera and its basic settings determine the resolution of the video as well as the base frame rate. It is configured via Camera -> Configuration -> Installer Menu within the device's web page, or via General -> Initialization within Bosch's Configuration Manager. Please see the whitepaper on „VCA Capabilities per Device” to see what internal video analytics resolution and frame rate based on the video analytics type and settings chosen here are used. Note that the frame rates selectable in the encoder settings depend on the base frame rate but are independent of the frame rate the video analytics uses. Furthermore, the video analytics results and metadata will be delivered with the frame rate used for the video analytics itself. Thus, it is for example possible that the encoder delivers the video with only 1 frame per second (fps), but the video analytics results and metadata are actually delivered at 15 fps. It is also possible that the encoder delivers 60 fps for the video, but the video analytics results and metadata are only delivered at 15 fps. This will also be visible in live view as well as when viewing recorded video, as the display of the video analytics results and metadata is updated as often as new information is there.

2.1 Video analytics frame rate dropping?

Dropping frame rates are of concern for video analytics. Fast movements especially close to the camera may not be detected anymore, and because the initial detection of objects will be delayed. The drop of single frames does typically not degrade performance significantly, but if the frame rate drops too much, then objects and alerts will be missed completely. Usually, the video analytics runs with 15fps or 12.5 fps, depending on the camera base frame rate, while MOTION+ only uses 5fps. For a complete overview on video analytics frame rates, see the whitepaper on „VCA Capabilities per Device”.

There are special cases where video analytics frame rate may drop:

- ▶ Exposure time is set too large / gets too large. Then the imaging cannot provide the full frame rate to the video analytics. This can be avoided by ensuring that minimal frame rate in automatic exposure mode is not smaller than the specified video analytics frame rate. Further details can be found in section 3.1.
- ▶ There are too many objects in the field of view to be tracked in real time by the video analytics. Then the video analytics will process the whole frame even if it takes longer, the delivery of the resulting metadata is delayed, and the video analytics drops the next frame or frames to proceed with the next up-to-date image. Nevertheless, Intelligent Video Analytics / Essential Video Analytics will continue to track the objects. At the moment, this is how many objects can be tracked in real time for the different common product platforms (CPPs) and video analytics:
 - IVA Pro Perimeter: ~20
 - IVA Pro Buildings / Traffic: 64
 - Intelligent Video Analytics on CPP6 and higher: ~20
 - Intelligent Video Analytics on CPP4: ~10
 - Essential Video Analytics: ~10
- ▶ Overall processing load is too high and the video analytics does not get enough processing power. Depending on the common product platform (CPP) and architecture, the video analytics shares the processing power with the encoder. The overall processing load on the processor is then determined by whether and how many live streamings, recordings or device replays are taking place, and on the bitrates those use. In general, all have same priority and need to share processing power. (One exception is that JPEG streaming / snapshots have lower priority than video analytics, thus frame drops occur with the JPEG streaming.)
 - Intelligent Video Analytics on CPP4/6/7/7.3 uses a dedicated hardware accelerator unit for the full video analytics and is thus independent of the processing load on the encoder chip.
 - While CPP13/14 cameras have a dedicated neural network engine where object detection takes place with IVA Pro Buildings and IVA Pro Traffic, the tracking of the objects over time is on a processor shared with the encoder. Furthermore, the processor power inside the CPP14 cameras differs based on the camera model, with some models offering much less processing power per channel.
 - Essential Video Analytics runs on the processor shared with encoding.

MOTION+ and tamper detection are negligible from a processing power point of view, independent of where they are processed.

3 Scene modes & image enhancement

In general, an image pleasing and easy to interpret for a human is also an image that will provide a good video analytics performance. Video analytics don't cope well with high noise levels, low contrast or blurred objects, and cannot detect anything not visible in the image, e.g. because the whole scene is too dark. Therefore artificial lighting or a thermal sensor are always advisable for use of video analytics at night.

The available scene modes preset values which can then later be adjusted by the user. Instead of rating the user modes, the parameters they influence will be evaluated and described regarding their effect on video analytics.

3.1 Brightness, contrast, saturation, white balance

If the image is under- or overexposed, then details and contrast are reduced and video analytics performance drops. This effect can also be caused by setting extreme values of brightness, contrast, saturation or white balance. Therefore, best practice is to keep these values here at or near default, and white balance on auto.

3.2 Automatic level control (ALC): ALC level, shutter, day/night

The image brightness level is defined first of all by the amount of light reaching each pixel on the sensor. The exposure time, also called shutter time, defines the time span over which the light is aggregated. The larger the exposure time, the brighter the image will be. But a long exposure time also means that objects moving in the scene will become blurred, as they move over several pixels during the light aggregation time. On the other hand, the exposure time is restricted by the frame rate, as it can only aggregate the light for a single frame, and the next frame needs to be generated in time.

To always generate images with a good contrast and brightness, the exposure time needs to be adjusted to the illumination levels outside. The brighter the scene, the shorter the exposure time can and should be, to avoid overexposure where the whole image becomes too bright and details and contrast are lost. The darker the scene, the longer the exposure time needs to be to avoid underexposure and dark images with details and contrast lost. Video analytics performance will drop along with the loss of details and contrast. Though some of the exposure effects are compensated in the image enhancements later on, e.g. using the gain intensifier, best practice is to generate the best possible image here as all later compensations can decrease image quality. Therefore it is recommended to use automatic exposure.

To avoid frame rate dropping below the video analytics frame rate, the auto exposure can and should be restricted to a minimal frame equaling that of the video analytics used. This is important as exposure time can drop the frame rate down to 1 fps and below, causing objects and alerts to be missed completely with the video analytics.

In addition, a default shutter time can be set. This default shutter time is used as long as possible, though in extremely bright or dark scenes the automatic exposure will have to adjust it. Best practice is to use $\sim 1/30$.

Best practice is to keep the ALC level and saturation configurable there at zero.

During low-light situations e.g. at night, color information is no longer available, and to reduce noise level, switching to a grayscale image is recommended. Best practice is to keep the day / night switch on auto.

3.3 Advanced: Contrast, noise reduction, backlight compensation

A high sharpness level leads to exaggerated and thus better visible edges. Though this is visually often pleasing, this feature can increase image noise, which may become a problem especially at night or in low-light situations. It can also add other small image artefacts to which the optical flow reacts sensitive. Best practice is to keep it at the default setting of zero.

The aim of contrast enhancement is to provide the right amount of black and white in the image, and to boost the contrast of intensity values that occur often. Thus, low-contrast as well as camouflaged objects will be better visible to user and video analytics. Therefore, contrast enhancement should be active for best video analytics performance.

Temporal noise filtering smoothes each single pixel with the values occurring over several frames. While temporal noise filtering provides smooth images even in low light situations, the temporal smoothing done to achieve this can cause moving objects to blur, creating ghosting effects. Though feedback from video analytics can be used to suppress the temporal noise filter in areas of moving objects with Intelligent DNR, in extreme low-light situations the blurring can be so extensive that the video analytics is no longer able to detect the objects at all. Therefore, best practice is to keep temporal noise filtering on a low level and Intelligent DNR off for best video analytics performance, and add sufficient external illumination instead.

Spatial noise filtering smoothes each pixel with the values of the neighbouring region in the same frame and can be used to reduce image noise. As it also decreases contrast and blurs objects, it may decrease video analytics performance. Best practice is to keep it on the default value of zero.

Note that sharpness level, spatial noise filtering and temporal noise filtering have a value range from -15 to 15, where zero denotes the default value, which has been optimized for best overall visual performance. The actual “off” value can differ from camera to camera, and the lowest values may actually cause an inverted effect.

Intelligent defog should be kept on auto, as it improves image quality in low-contrast scenes.

The purpose of backlight compensation is to boost dark parts to brighter illumination values while minimizing their noise level, to generate good brightness and contrast levels in all parts of the image. However, bright areas thus get less contrast. Best practice is to use it whenever the scene has backlight, and keep it off otherwise.

4 Privacy masks & display stamping

4.1 Privacy masks

On most Bosch IP cameras, privacy masking is applied to the image before it is given to the video analytics, and no video analytics will occur in these regions. Note that when the camera is shaking, false detections may occur at the border to the privacy masks as the privacy masks itself will stay stable and thus generate a different motion pattern there.

On AUTODOME IP and MIC IP cameras based on the common product platform (CPP) 4, the privacy masking is applied to the image after being send to the video analytics. There, the video analytics also gets the information about the privacy masks to suppress processing there, but the user can actually choose to let the video analytics ignore the privacy masks to get alarms and metadata there. This is due to a different architecture on these devices, and cannot be transferred as a feature to the other cameras.

4.2 Display stamping

Display stamping, like camera name, current time or logo overlays, is done after the video analytics get the image and therefore does not influence video analytics performance. Nevertheless, as the video analytics will work behind the display stamping but the user cannot see behind it to confirm video analytics detection there, best practice is to disable video analytics in the areas covered by the display stamping to avoid questions. This disabling can be done via VCA masks.

5 Encoder settings

Bosch video analytics run in the camera and obtain the video images before they are passed to the encoder for compression and transmission over the network. Thus, all settings here are independent of and do not influence video analytic performance, up to and including frame rate settings (see section 1).

The only exception is Essential Video Analytics, as this has no hardware accelerator and thus needs to share processing power directly with the encoding. Here, high bitrates and additional loads by live streaming, recording and device replay may lead to dropping frame rates of Essential Video Analytics. JPEG streaming / snapshots have lower priority than Essential Video Analytics, thus frame drops occur with the JPEG streaming instead of the Essential Video Analytics metadata.

6 Thermal camera

In contrast to visible light cameras, the thermal cameras from Bosch (VOT 320, DINION IP thermal 8000) do not use the same image that is presented to the user, but are given a differently generated image with a dedicated contrast enhancement as well as tone mapping for optimal video analytics performance. Therefore, neither contrast enhancement nor tone mapping will influence video analytics performance on the thermal cameras.

Bosch Sicherheitssysteme GmbH

Robert-Bosch-Ring 5

85630 Grasbrunn

Germany

www.boschsecurity.com

© Bosch Sicherheitssysteme GmbH, 2023