WHITE PAPER

Unlocking the power of scene metadata

Elevating situational awareness, efficiency, and insights

January 2024

AXIS COMMUNICATIONS

Summary

In the context of video surveillance, metadata describes textually what is in the video. This can include which objects of interest are visible, or a high level description of the scene itself. It can also include attributes associated with the objects or the scene, such as colors of vehicles and clothing, exact locations, or direction of travel. The metadata is created in real time, directly in the camera or by another system component that is capable of running video analytics.

Metadata also provides context to events and allows large amounts of footage to be quickly sorted and searched. This enables functions that can be broadly categorized into three areas:

- **Post-event forensic search.** To search for objects or events of interest based on various search parameters that narrow down your search to a limited number of candidates. Object classification data enables searches that involve a broad range of details.
- **Real-time use.** To help operators respond quickly to situational changes, or to provide input to support decision making or enable automated action.
- **Identification of trends, patterns, and insights.** IoT and operational efficiency platforms for statistical reporting can rely on metadata for visitor counting, speed measurement, traffic flow data, and other types of automated data collection.

Some cameras can decode audio to retrieve audio metadata. Specific sound patterns can be detected and labeled in a similar manner as object classes are detected and labeled in video. An audio recognition system could, for instance, identify verbal aggression or detect glass breakage.

When you combine metadata from multiple inputs, such as visual, audio, activity-related, and process-related sources, you gain much more insights than you get from each input alone. Open protocols and industry standards are essential for seamless metadata integration.

Table of Contents

1 Introduction

Metadata is the foundation for gathering intelligence from video. It assigns digital meaning to video content by describing the key details in the scene. Using the metadata, you can quickly find, evaluate, and act on what is important in large amounts of video. This is why metadata has increasingly become an essential part of efficient security, safety, and business operations.

This white paper discusses metadata both in a surveillance context and an operational efficiency context. It details the benefits of metadata and how it is used in video management systems and other applications.

2 What is metadata?

Metadata is data about other data. In the context of video surveillance, metadata describes textually what is in the video, such as which objects of interest are visible, or a high level description of the scene itself. This can include attributes associated with the objects or the scene, such as colors of vehicles and clothing, exact locations, or direction of travel. The metadata is created in real time, either directly in the camera or by another component in the system capable of running video analytics.

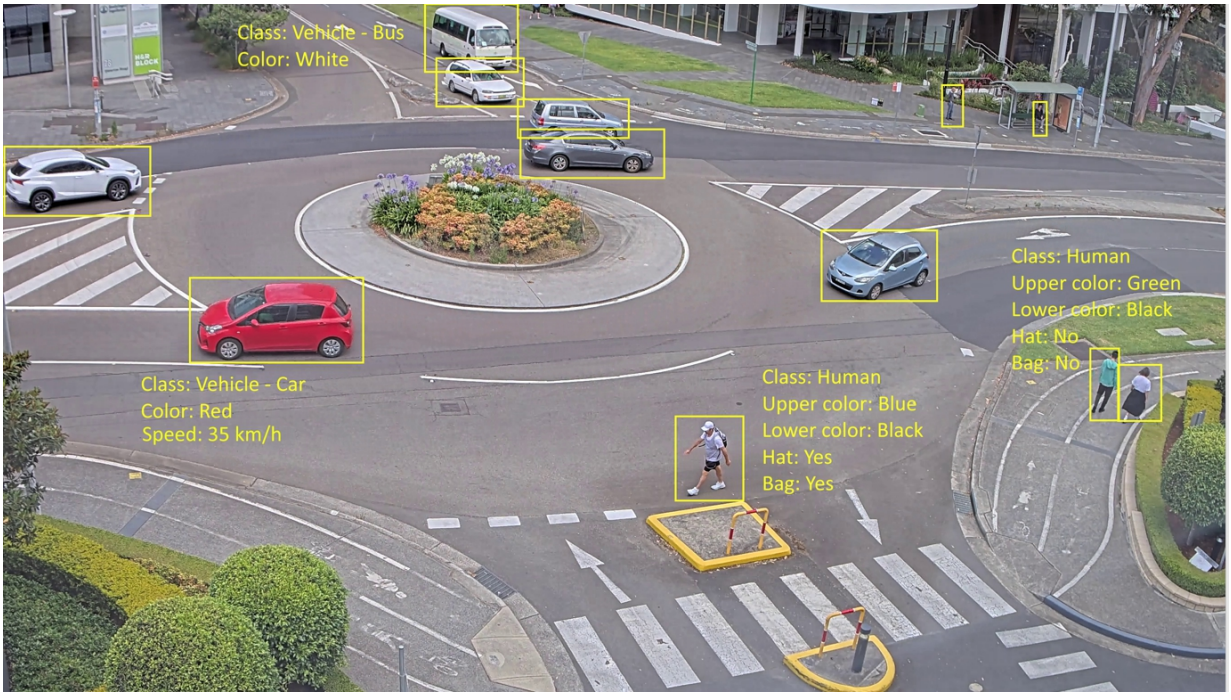

Figure 1. Example of a video frame where objects of interest are detected and analyzed to create metadata.

3 Generating metadata at the edge

High-performance video analytics used to be server based because they typically required more processing power than an edge device could offer. In recent years, algorithm development and increasing processing power of edge devices have made it possible to run advanced analytics at the edge. This means that the metadata is generated in the device and can be used directly in the device by other analytics. The video stream and metadata stream can also be transmitted to the VMS or another application for further processing.

Edge-based analytics have access to uncompressed video material with very low latency. This enables fast real-time applications, while also avoiding the additional cost and complexity of moving all video for processing elsewhere in the system. Edge based analytics also come with lower hardware and deployment costs since less server resources are needed in the system.

Generating metadata at the edge means extracting data from the video without losing any information in compression or transmission. This enables more accurate metadata and more exact analysis of the video content. The better the image quality, the better the metadata.

4 Use cases

Metadata not only provides details about objects in a scene. It also provides context to events and allows large amounts of footage to be quickly sorted and searched. This enables functions that can be broadly categorized into the areas of post-event forensic search, real-time use, and identification of trends, patterns, and insights.

4.1 Real-time use for instant action

Metadata can be used in real time to help operators respond quickly to situational changes. It can also provide valuable input to support decision making or enable automated action. Real time edge analytics that work with hiqh-quality metadata can help you secure people, sites, and buildings and protect them from intentional or accidental harm. You can rapidly detect, verify, and evaluate threats so they can be efficiently handled.

4.2 Forensic search

Metadata makes it possible to search efficiently and quickly for objects or events of interest. This can save investigators hours and hours of time, especially in searches on vast amounts of video from multiple video sources. You can search for objects, such as humans and vehicles, based on various search parameters to narrow down your search to a limited number of candidates. Search parameters may include, for example, motion, time, and object characteristics.

Table 4.1 Searches are facilitated using various metadata categories.

Metadata category	What is detected?	Examples
Motion	**How** an object moves	Direction, speed, other behavior
Time	**When** an object appears	Day of week, time of day, dwell time
Location	**Where** the object is	Place, field of view of camera
Object classification	**What** kind of object it is	People, vehicle (car, bus, truck, bicycle/motorcycle)
Object attributes	**What** characteristics it has	Clothing, accessories such as hats or bags, physical characteristics such as clothing color

Even if you have access to only one category of metadata, for example, time, this can prove to be crucial for finding the results you need.

Metadata about motion enables searches based on relative object speed and direction of movement. Object classification data enables searches that involve a broader range of details. Cameras with deep learning processing units (DLPU) can usually provide enriched metadata with more granular object classification, allowing you to search for, for example, a green truck or a person with a blue coat.

4.3 Identifying trends and patterns to gain insights

IoT and operational efficiency platforms for statistical reporting can rely on metadata for visitor counting, speed measurement, traffic flow data, and other types of automated data collection. The data is analyzed to generate actionable insights.

5 Where is metadata used?

The benefits are many of leveraging metadata to understand the characteristics and content of a scene. The main consumers of metadata can be categorized as follows.

Edge applications. Analytics running on the camera can apply logical filters and rules to the information about the object in the scene. Thereby, the analytics can trigger actions based on defined thresholds or specific behaviors, such as controlling a PTZ camera based on the detection and movement of a person in the scene.

Video management systems (VMS). In the context of video surveillance, metadata has commonly been used within a VMS to present visual overlays around potential objects of interest in the scene. With the development of more advanced object detection and classification algorithms, operators are now also capable of locating objects of interest based on specific characteristics, such as color of clothing. Having the ability to perform search queries using these data points greatly reduces the need to manually review large amounts of footage.

IoT platforms. Metadata can be gathered and presented visually in business intelligence platforms to generate actionable insights by analyzing real-time and historical trends. Statistical analyses based on customer flow or customer experience enable data-driven decision making to improve operations.

Second layer of analytics. Some applications require a combination of edge-based and server-based processing to perform more advanced analyses. Pre-processing can be performed on the camera and

further processing on a server. Such a hybrid system can facilitate cost-efficient scaling of analytics by streaming only relevant video and metadata to the server.

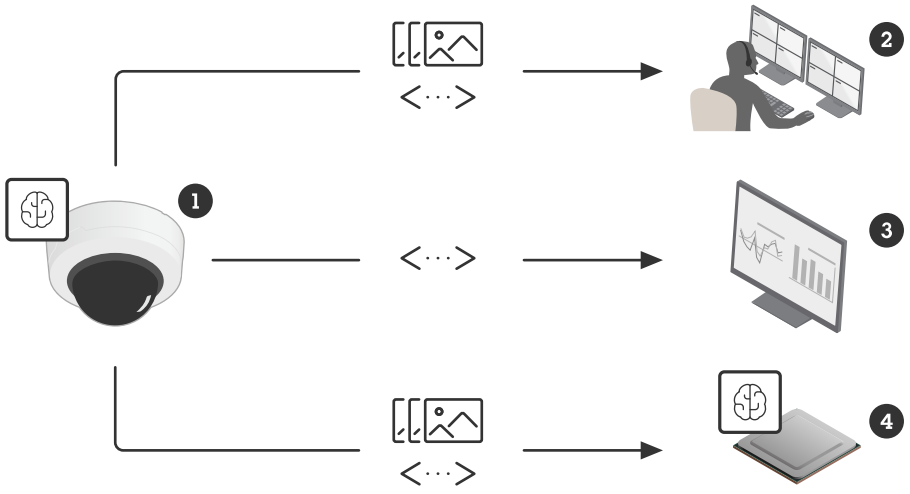

Figure 2. Metadata consumers

1 *Edge applications*
2 *VMS*
3 *IoT platforms*
4 *Second layer of analytics*

6 How is the metadata delivered?

The metadata generated can be delivered utilizing different approaches based on the intended use. In real-time applications, the metadata needs to be constantly streamed to the consumer on demand, as this is vital to ensure appropriate response and situational awareness. In other less critical applications where real-time action is not required, the metadata can be further consolidated, for example based on the

track of each specific object in the scene, before being delivered to the consumer. This reduces the total amount of data that needs to be stored and processed.

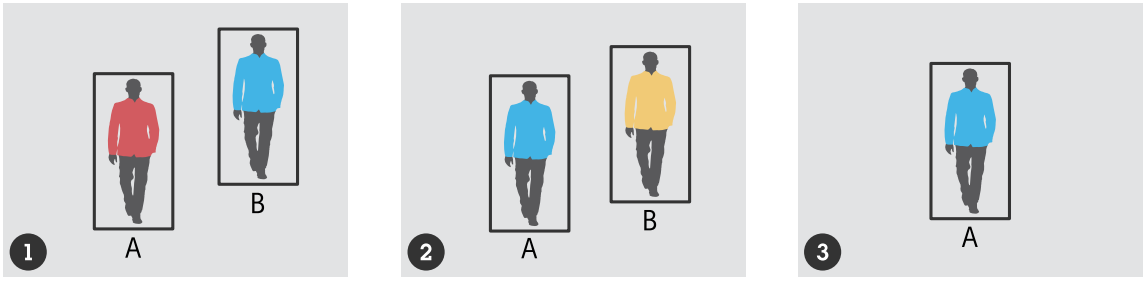

Figure 3. This figure illustrates the streaming of metadata, where continuous metadata frames from the camera provide real-time scene information. Each frame captures the scene at a specific moment, independent of past events.

1. *Frame 1 detects object A and B, classifying A as a human in red clothing and B as a human in blue clothing.*
2. *In frame 2, the camera updates the classification, determining that object A actually wears blue clothing, and object B wears yellow clothing. Although the objects remain the same as in frame 1, their color attributes change and this is reflected in the metadata.*
3. *Frame 3 shows the absence of object B, with the camera tracking only object A, still classified as a human in blue clothing.*

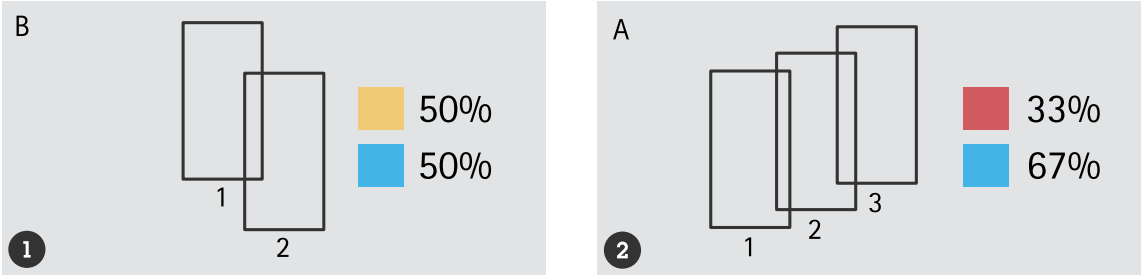

Figure 4. This figure demonstrates consolidated metadata delivery, where the camera provides information in a unified format based on the detected track of the objects in the scene. Frames for each object encompass all known details throughout the object's track lifetime.

1. *In the first frame, details about object B are presented, including its first and last detection, trajectory summary, and attributes detected during the track. Object B had a 50% likelihood of wearing yellow clothing and a 50% likelihood of wearing blue clothing.*
2. *The second frame mirrors this format for object A, revealing a 33% likelihood of red clothing and a 67% likelihood of blue clothing.*

The benefit of the consolidated method is that the camera significantly reduces the amount of data being sent to the consumer by delivering metadata only when there are objects present in the scene and

in that case it is summarized (consolidated) for easy interpretation. The streamed method delivers a complete description of the scene in every frame, even when there is no activity or objects present and the consumer needs to make sense of this data based on their specific need. As mentioned, the streamed method is beneficial for real-time use cases while consolidated is optimal for postprocessing when the consumer does not need to take immediate action.

Understanding the strengths and limitations of each approach is essential for designing the system architecture. For example, an IoT platform generating insights based on the metadata would benefit from receiving a post-incident summary of the objects in the scene, as these services are typically constrained with bandwidth and storage limitations.

In addition, the metadata could be delivered through a number of different communication protocols and file formats based on the specific needs and preferences of the intended consumer.

7 Audio metadata

Some cameras can decode audio to retrieve audio metadata. Audio recognition analytics can detect sound patterns and highlight sounds of interest in live and recorded audio. This way, audio recognition systems coupled with video surveillance devices can alert operators to ongoing potential incidents, guiding them to the relevant camera views. The system could, for instance, identify verbal aggression to prevent escalation and assault, detect glass breakage to prevent break-ins, or provide early warnings of patients in distress. By allowing operators to not only see but also hear what is happening in a scene, sound recognition systems may enable early detection, swift intervention, and in many cases, prevention of further escalation. Sound recognition can also serve as a secondary means of verification.

Analytics trained to recognize sound patterns typically listen for a combination of characteristics ranging from decibel level to the energy in different frequencies over time. Specific sound patterns can be detected and labeled in a similar manner as object classes are detected and labeled in video.

8 Combining metadata from multiple sources

The true potential of metadata is realized when applied to multiple inputs, such as visual, audio, activity-related, and process-related inputs. Data sources like RFID tracking, GPS coordinates, tampering alerts, meter readings (such as temperature or chemical levels), noise detection, and point of sale transactional data are valuable in the management of any site. Data from all the sources can be aligned based on their timestamps.

Combining metadata from different sources means gaining much more insights than one can ever get from each source alone. Open protocols and industry standards are essential for seamless metadata integration.

T10197886/EN/M1.11/2401

About Axis Communications

Axis enables a smarter and safer world by creating solutions for improving security and business performance. As a network technology company and industry leader, Axis offers solutions in video surveillance, access control, intercom, and audio systems. They are enhanced by intelligent analytics applications and supported by high-quality training.

Axis has around 4,000 dedicated employees in over 50 countries and collaborates with technology and system integration partners worldwide to deliver customer solutions. Axis was founded in 1984, and the headquarters are in Lund, Sweden

©2024 Axis Communications AB. AXIS COMMUNICATIONS, AXIS, ARTPEC and VAPIX are registered trademarks of Axis AB in various jurisdictions. All other trademarks are the property of their respective owners. We reserve the right to introduce modifications without notice.