An Introduction to IBM Video Analytics

Simplified visual recognition for object detection.

Learn how to uncover operational insights and manage compliance and risk with AI based video analytics.
An Introduction to IBM Video Analytics

The explosion in the number of video cameras capturing images around the clock makes it challenging for businesses to effectively use the massive amount of video content being collected. It’s become increasingly difficult for people to search and find a video (or a specific thing in a video) that they are interested in. And, when considering the monitoring of live cameras, increased volumes only exacerbate the human limitations. We have short attention spans, a tendency to miss details when something eye-catching is happening, and our eyes get fatigued when focusing on the same screen for a long time. Fortunately, these traits are not shared by computers and artificial intelligence. However, humans have better cognitive capabilities than computers. Humans are better at recognizing objects and applying contextual cues to make decisions as to what is and isn’t of interest.

To gain value from any technology, it must be applied in a way that allows humans to leverage it to solve problems. Solutions that combine the respective advantages of humans and computers are needed to deliver the value that video content alone cannot. Building such solutions is not easy, but IBM Video Analytics (IBMVA) is a product that is designed to reduce the effort and expertise needed to build them. IBM Video Analytics is capable of searching through libraries of on-demand video, or many simultaneous live cameras feeds, and can send real-time alerts when human attention is needed. It is a platform on which solutions can be built with minimal effort spent on required tooling and common tasks, allowing developers and integrators to focus their energy on solving the actual problems the solutions are meant to solve.

Video Analytics Value Propositions

Solutions that leverage video and video analytics can be used to deliver value in three basic ways:

- **Monitoring**
  Providing real-time situational awareness to allow users to monitor - and be alerted on, the most interesting live video streams

- **Forensics**
  Providing users the ability to quickly find the objects or activity they are looking for in recorded video

- **Data Analysis**
  Analyzing data over a time period that can be “mined” to discover trends and statistics
Functional requirements of a video analytics solution

Every video analytics solution has basic functional requirements. It must access video, analyze it, store the data generated by the analysis, make decisions based on the data, and provide insight to users based on these decisions. We'll call these basic functions:

- Acquisition
- Analysis
- Persistence
- Application

Acquisition

How live and recorded video content are made available to the solution

Standards used to encode, transport, store, and access video have facilitated the expansion of video content and are in a state of constant evolution. There are many that come into play and the possibility of new standards always exists as well. Some standards apply to all industries, while some are unique to a specific industry. A capable solution must maintain ongoing support for a wide variety of standards that are prevalent across various industries.

“The wonderful thing about standards is that there are so many of them to choose from.”
Grace Murray Hopper, computer pioneer and naval officer

Analysis

How video content is analyzed to produce data that describes what is happening in the video

Video Analytics is a term that describes a variety of computer algorithms that produce descriptive data, or metadata, about something that is observed within a video. Many algorithms have been developed over the last two decades of research in the field of Computer Vision. At the leading edge of video analytics, Deep Learning uses neural networking to produce results that are both more accurate and more precise than the previous generation of technology.

Analytics algorithms can be classified by the functions they perform:

- **Detection** - recognizing motion and objects. Detection metadata usually includes location and size, as well as confidence level (perhaps expressed by a percentage), and is output for each frame of video.

- **Classification** - describing qualities of objects such as color, age, and gender. Classification functions occur following detection, adding the additional metadata about the objects for each frame.

- **Feature Extraction** - generating numerical data sets that describe various points of interest of an object and their spatial relationships.
• **Tracking** - keeping track of per-image detections over time. Tracking functions take into account historical metadata from detection algorithms. This allows it to associate many independent detections into tracks of motion that start when an object is first detected and ends when the object leaves the scene. This allows behavioral attributes such as speed and direction of motion to be derived. In addition, temporal context allows for the summarization of many independent instances of metadata associated with a single object.

• **Alerting** - processing the combined output to detect when the metadata meets pre-defined rules. Alerting functions look at the combined output of detection, classification, feature extraction and tracking algorithms and essentially wait for the output to match pre-defined criteria. When there is a match, a real-time alert is generated.

A suitable platform must include a library of components that implement these algorithms and produce accurate metadata.

**Persistence**

How the data is stored and made available for use in decision making

The various types of metadata produced by the analytics functions must be stored and be readily accessible by decision-making processes. This access must allow for reasonably precise queries so the specific data relevant to a decision can be accessed quickly. Metadata can also take several forms including:

1) structured, textual or numeric data traditionally stored in relational databases
2) unstructured data such as JSON or XML documents
3) image files

The methods provided to access the metadata should support standard protocols so that application logic can be developed using commonly available libraries and skill sets.

**Application**

How the data is used to make decisions and how these decisions are presented to users

The possible solutions that can be built leveraging video, analytics, and the metadata they produce are virtually limitless. The video surveillance and security market was an early adopter of video analytics. However, other use cases have become more commonplace including video analytics solutions to help retailers understand customer behavior and mitigate fraud, solutions for better traffic management to help understand patterns in vehicle flow, solutions that help manufacturers detect anomalies in their production processes, and many more.

To facilitate the building of applications, a platform should provide a strong API and SDK and include libraries and samples that can be used by application developers to quickly understand how to build and reuse code that already exists.
Non-functional requirements of a video analytics solution

In addition to the functional requirements, a strong platform must be:

- **Extensible**
  Allowing new video codecs and streaming protocols, and new analytics algorithms, especially Deep Learning models, to be incorporated quickly and easily. In addition, as new developments in analytics may produce new forms of metadata, the persistence layer should be able to quickly accommodate them.

- **Scalable**
  Supporting deployment on smaller edge computing devices or servers, medium sized on-premises servers, and large clusters in data centers and clouds. This helps ensure that every conceivable deployment size, from a handful to thousands of video sources can be supported.

- **Secure**
  Providing all the de rigueur security functions of any information solution: encryption, authentication, authorization, and auditing.

- **Efficient**
  Minimizing the hardware resources required.

The IBM Video Analytics solution

IBMVA combines a long history of research and development in video analytics with new technology and ongoing research to deliver a solution that supports all of the requirements described above. Technology that originated in IBM Research in the 2000’s is still included with the product today and has been continually enhanced since then. In recent years, extensibility of the product became a major requirement for many clients. As a result, IBMVA technology is now highly componentized with strong interfaces for both internal and external integration. Video Analytics can be integrated into the current environment and receive the video feeds from existing analog and digital cameras from supported video management software. Organizations are not required to purchase the latest cameras or overhaul their existing video cameras or infrastructure.

Let’s look at what IBMVA includes “out of the box” to satisfy the functional requirements and how it can be extended in each of these areas.

**Acquisition**

IBMVA supports the most prevalent video streaming protocols in the industry today: Real Time Streaming Protocol (RTSP), which is the de facto standard in video surveillance and many other industries, and WebSockets which is supported by every modern web browser. It also includes SDK-based integrations with the two leaders in the Video Management (VMS) market: Milestone and Genetec, as well as other major VMS providers. For recorded video, it supports H264-encoded MP4 files, the predominant encoding and video container formats today, as well as .WMV and .AVI files that are common in Windows systems. Its video acquisition architecture is based on a video abstraction layer with an external interface that allows video acquisition components to be developed for new video sources with no impact to the other functional areas.
Analytics

IBMVA includes components that perform all the analytic functions and a library of Deep Learning detection and classification models. Its analytic architecture provides a plugin interface that allows for the deployment of new algorithms as components that can then be combined with existing components seamlessly. In addition, new Deep Learning models can be deployed by simply importing them into the system and including them in analytic workflows. Since Deep Learning is such a rapidly developing field, it also includes an inference adapter interface that allows adapters for new Deep Learning frameworks to be developed quickly.

Persistence

IBMVA includes a relational database for storage and easy retrieval of both configuration data and analytics metadata. It also stores image files in the operating system’s filing system in locations where they can be retrieved with simple HTTP calls. It associates the relational data with the image files so the images associated with an instance of metadata can be accessed and managed together with the metadata. It also includes content retention management functions that purge both data and image files on a scheduled basis to minimize the storage capacity required.

An essential part of its persistence architecture is a content type definition mechanism that allows for new types of metadata to be added to the system by simply defining new data models and importing them into the system. Once new data models have been deployed, all the functions of the persistence layer are applied to the new data. In addition, content definitions enable flexible views of the data and the implementation of custom logic to be applied specifically to different content types, making it much easier to satisfy different application requirements.

Application

IBMVA includes a fully functional web-based application that was originally developed for the public safety market. This application supports both Monitoring and Forensics use cases, and includes track summarization, heatmap, and histogram logic and visualizations that support Data Analysis use cases. While the IBMVA web application can be used as is for many engagements, it also serves as sample code that can be leveraged to build other applications.

The included web application accesses metadata from the persistence layer via a pseudo-RESTful API that supports the standard REST Create/Retrieve/Update/Delete operations and enables the triggering of custom logic that has been incorporated via content type definitions. This API is documented using OpenAPI, making it easy for developers to discover its capabilities.
Scalability

IBMVA is containerized, meaning it makes use of the latest trend in virtualization that allows software to be deployed virtually on a variety of platforms while operating in the same process space as software running natively in the OS. Containerization allows it to be deployed at the edge, on premises, and in data centers. In many cases, a hybrid topology is needed to optimally position different solution components. For example, components that analyze video are best deployed at the edge or on premises where they have sufficient network bandwidth to access video streams. In contrast, persistence and application components benefit from cloud deployment where operational costs can be minimized. In addition, IBMVA utilizes enterprise-class middleware that supports many of the largest computing solutions in existence, allowing it to scale to very large deployments. The system architecture allows IBMVA to scale horizontally, adding more hardware to support more live streams and faster video file processing. For example, a typical installation may involve supporting 2000 cameras or 10,000 video on-demand files, with the option of increasing to higher numbers being supported by adding additional hardware.

Security

IBMVA’s use of middleware also allows it to utilize all the built-in security mechanisms they provide. It supports authentication via both internal database or external LDAP directories, implements role-based access control for authorization, and can provide an audit trail of every transaction. This involves producing an audit trail for the system configuration as well as access to the metadata. Auditing options are available at various levels including authentications and operations for administrators or users, as well as highly granular access for each operation on individual content types. The IBMVA REST-based API is used for all access to configuration data and metadata. The default configuration requires HTTPS so all access to its data is encrypted.

Efficiency

In addition to the performance advantages provided by containerization, IBMVA’s most compute intensive processes are implemented in C++ with some low-level operations utilizing assembly code for maximum performance. This allows the analytics functions to support a large number of live video streams on a single enterprise-class server (up to one hundred per server on typical hardware configurations). While Deep Learning models provide the ultimate in analytics capabilities, they require high levels of compute performance. This means that they are typically deployed on GPUs, which are costly. IBMVA uses its library of traditional analytics, which require a lower level of computing power, in combination with its Deep Learning components, to offload the analytics functions that do not require Deep Learning to CPUs, thus lowering the overall hardware requirements of the system. The IBMVA persistence layer’s utilization of enterprise middleware was developed by IBM developers with tremendous experience in building large computing solutions. This combination has yielded the ability for it to support a very large volume of both incoming and outgoing data. A typical server can support well over a thousand cameras even in high activity environments.

Summary

Video Analytics is a powerful technology that can take video monitoring tasks to a whole new level and address the challenges created by the explosion in video cameras and content. IBMVA provides a complete solution suitable for many uses and provides video solution developers with a scalable, secure, extensible platform that enables to them to build video solutions with minimal effort. Video analytics can reduce time, money, and human effort in video monitoring tasks, offering more security, reliability, and consistency in video monitoring.
About IBM Video Analytics

IBM Video Analytics uses powerful algorithms and deep learning technology to rapidly identify and analyze security risks and anomalies captured in video content. By enabling enterprises to more effectively understand and respond to emerging threats and issues, Video Analytics dramatically lowers operational costs and risk.

To learn more, visit: https://www.ibm.com/us-en/marketplace/video-analytics