



The Open Platform Company



## Audio e video, l'annotazione automatica di A.I. Tech mediante metadata

*contributo di Gennaro Percannella, Sales Manager, Pasquale Foggia, Chief Software Architect, A.I. Tech srl*

### Metadati e videosorveglianza?

Per rispondere a questa domanda proviamo a pensare al seguente scenario: centinaia di telecamere di sorveglianza installate in un aeroporto. Qual è la probabilità che un operatore nella control room sia in grado di notare tempestivamente una persona che entra in un'area interdetta, o individuare un bagaglio abbandonato, o un comportamento sospetto di una persona? Studi scientifici dimostrano che dopo 20 minuti di osservazione continua un normale operatore non noterà oltre il 90% di eventi rilevanti. Si consideri inoltre il caso in cui, ore o giorni dopo che sia avvenuto un evento criminoso, si renda necessario ricercare all'interno di ore e ore di sequenze video catturate da decine di telecamere una persona vestita con colori specifici che passa attraverso una porta. Per ritrovare gli eventi d'interesse nelle sequenze non annotate l'operatore dovrebbe visualizzare tutti i video con l'elevato rischio già dopo pochi minuti di lavoro di perdere delle sequenze rilevanti. La video content analysis (anche detta analisi video intelligente o più brevemente analisi video) costituisce un insieme di tecniche alla frontiera con la ricerca scientifica nei campi della Intelligenza Artificiale e della Computer Vision che consentono ad un computer di analizzare un flusso video allo scopo di comprenderne il contenuto e di annotarlo automaticamente (i metadati) senza l'intervento umano. I sistemi di analisi video possono richiamare l'attenzione dell'operatore quando avviene qualche evento specifico nella scena inquadrata dalla telecamera o consentono di ridurre di diversi ordini di grandezza i tempi della ricerca consentendo all'operatore di trovare solo quelle sequenze video che soddisfano alcuni criteri specificati dall'operatore.



### I metadati nella piattaforma di analisi audio e video di A.I. Tech

In questo articolo cercheremo di dare una risposta alle seguenti domande: quali sono i tipi di metadati estratti automaticamente dalla piattaforma di analisi audio e video di A.I. Tech? Come sono rappresentati ed inviati ai sistemi deputati alla fruizione (Video Management Systems, piattaforme di business intelligence)?

La piattaforma di A.I. Tech incorpora un motore di analisi video basato su algoritmi avanzati di object detection e tracking e di filtraggio del rumore che consente di rilevare con accuratezza diverse tipologie di eventi (conteggio persone, heat-map, superamento di linea, rilevamento di intrusioni, riconoscimento di comportamento sospetti, segnalazione di oggetti rimossi/incustoditi, rilevamento di fumo e fiamme) anche in condizioni ambientali complesse sia indoor che outdoor.

A.I. Tech ([www.aitech.vision](http://www.aitech.vision)) è una società fondata nel 2010 che produce soluzioni avanzate di analisi audio e video per i mercati verticali del retail e della sicurezza. A.I. Tech è stata fondata da ricercatori universitari attivi da oltre 25 anni nella realizzazione di sistemi intelligenti basati sulla elaborazione di segnali audio e video.

La piattaforma di analisi è disponibile sia come applicazione lato server (si veda Fig. 1 per uno screenshot della piattaforma di analisi integrata con Milestone XProtect), o come una applicazione in modalità “edge” in grado di poter essere eseguita a bordo di telecamere Axis, Hikvision e Samsung che supportano l’elaborazione embedded. La piattaforma di A.I. Tech è ingegnerizzata in modo da consumare poche risorse di calcolo, consentendo di elaborare un elevato numero di flussi video contemporaneamente su server di fascia alta, o di elaborare a pieno frame rate un singolo flusso video a bordo della telecamera. Inoltre, è anche disponibile per dispositivi di calcolo a basse prestazioni e basso consumo energetico (quali Raspberry Pi o altre piattaforme embedded basate su Linux) consentendo di portare alla periferia (“to the edge”) l’elaborazione anche laddove non siano disponibili telecamere che supportano l’esecuzione embedded.

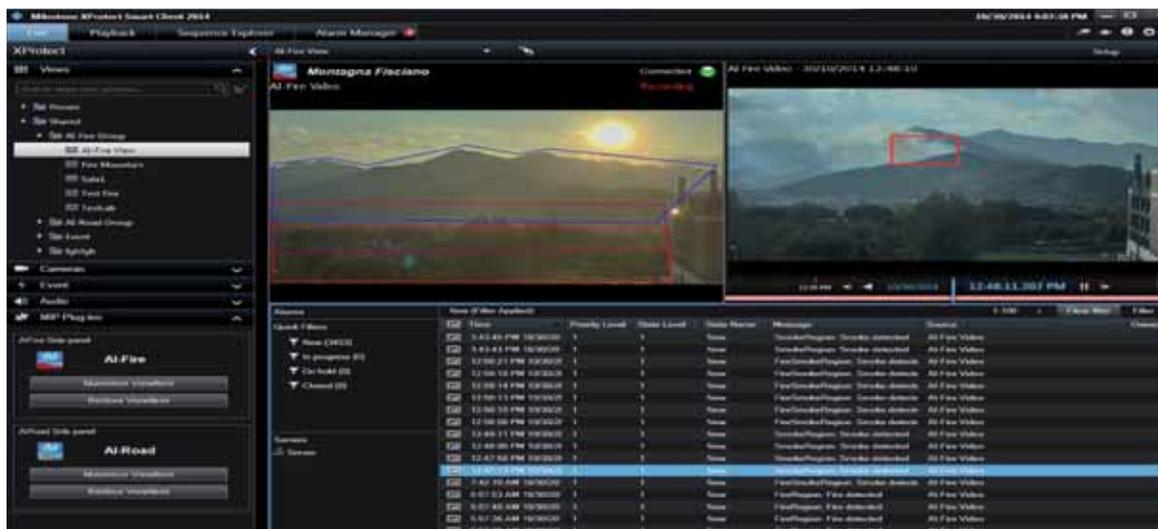


Fig. 1 - Funzioni di analisi video per il rilevamento fumo e fiamme integrate in Milestone XProtect

Inoltre, l’offerta di A.I. Tech si arricchisce con la disponibilità di prodotti per l’analisi audio che consentono la rilevazione ed il riconoscimento di eventi audio (quali urla, rottura di vetri, esplosioni) sia in ambienti affollati che sterili, fornendo così metadati aggiuntivi a quelli già forniti dalla analisi video (si veda la Fig. 2 per la schermata principale della applicazione di analisi audio di A.I. Tech).

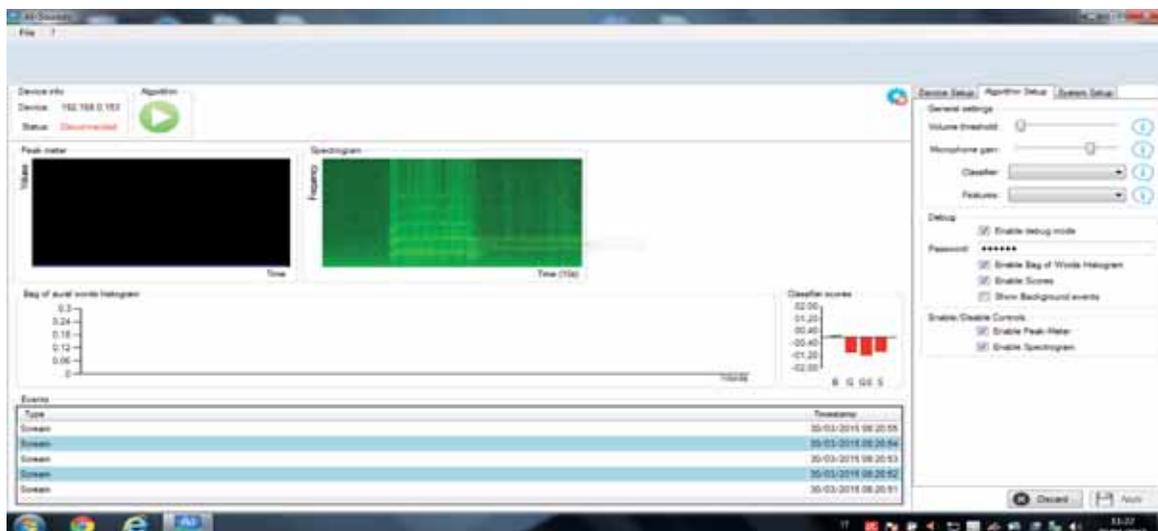


Fig 2 - Schermata principale della applicazione di analisi audio di A.I. Tech

L'informazione fornita attraverso i metadati dipende dalla specifica funzione di analisi: in generale, essa include l'evento individuato dall'applicazione, con il riferimento temporale utile per associarli alla porzione rilevante del flusso video, insieme con informazioni riguardanti gli oggetti o le entità coinvolti. Ad esempio, il bounding box degli oggetti (si veda in Fig. 3 per alcuni esempi), l'età stimata, il sesso e la razza della persona per eventi legati alla individuazione di volti.



Fig. 3 - Esempi di bounding box di oggetti rilevati mediante la piattaforma di analisi video di A.I. Tech.

Inoltre, ove appropriato, il sistema è in grado di fornire un'indicazione quantitativa del grado di confidenza della rilevazione, in modo da consentire all'applicazione che usa i metadati di filtrare ed elaborare e/o riportare all'utente solo quelli che siano ritenuti sufficientemente affidabili.

Nella Tabella 1 sono riassunti i metadati forniti dalla piattaforma di A.I. Tech per ogni specifico tipo di funzione di analisi audio/video.

METADATI	TIPO DI EVENTO
Timestamp	Tutti gli eventi audio/video
Bounding box dell'oggetto	Superamento linea Intrusione Oggetto abbandonato/rimosso Comportamento sospetto Fumo Fiamme Analisi volto
Traiettoria dell'oggetto	Intrusione Comportamento sospetto
Colore medio dell'oggetto Classe dell'oggetto (persona, veicolo, altro) Dimensioni dell'oggetto (altezza, larghezza) in cm Distanza dell'oggetto dalla telecamera Velocità dell'oggetto	Superamento linea Intrusione Oggetto abbandonato/rimosso Comportamento sospetto
Direzione dell'attraversamento linea	Superamento linea
Ingresso/uscita dall'area virtuale	Intrusione
Sesso (maschio/femmina) + confidenza Età (bambino, adolescente, adulto, anziano) + confidenza Razza (Bianco, Nero, Asiatico) + confidenza	Analisi volto
Classe audio (sparo, urlo, rottura vetro, altro) + confidenza	Analisi audio

Tab. 1 - Lista dei metadati correntemente supportati dalla piattaforma di analisi audio e video di A.I. Tech.

## Sistema di notifica dei metadati nella piattaforma di A.I. Tech

L'architettura del sistema di notifica dei metadati della piattaforma di analisi di A.I. Tech è realizzata sfruttando un approccio modulare basato su plug-in, come rappresentato nella Fig. 4.

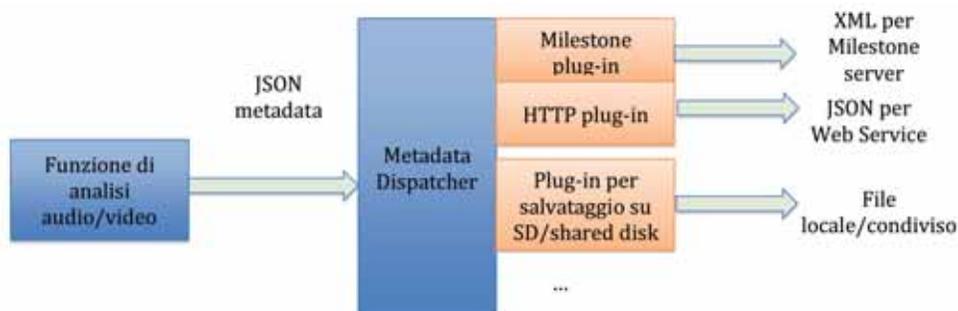


Fig. 4 · Architettura del sistema di notifica dei metadati nella piattaforma di A.I. Tech.

Tale architettura è stata progettata con lo scopo di garantire i seguenti vantaggi:

- Semplicità d'integrazione con Milestone XProtect, con supporto per il formato di eventi e metadati di Milestone.
- Semplicità d'integrazione con qualsiasi altra applicazione basata sul paradigma dei web service, mediante RESTful web services di semplice realizzazione.
- Possibilità di sviluppare in futuro altri plug-in per supportare altre piattaforme (ad esempio per business intelligence).
- Possibilità di immagazzinare i metadati localmente per uso off-line.
- Alta configurabilità, grazie alla possibilità di stabilire al momento della configurazione quali plug-in devono essere abilitati e quindi verso quali sistemi deve essere indirizzato il flusso di metadati.

La funzione di analisi audio/video della piattaforma di A.I. Tech genera i metadati in un formato interno basato sullo standard JSON (JavaScript Object Notation, standard ECMA-404), leggero e flessibile, in grado di rappresentare anche strutture dati complesse. JSON è solitamente utilizzato in applicazioni web e per RESTful web services, in quanto è più compatto e semplice da analizzare rispetto all'XML. Il formato JSON è anche popolare nei database NOSQL; ad esempio il database document-oriented MongoDB, usato spesso nel contesto della gestione di Big Data, salva i dati in formato JSON.

I metadati generati sono poi passati al Metadata Dispatcher, il componente della piattaforma di A.I. Tech, che si occupa della bufferizzazione dei metadati e del loro inoltramento ad un insieme di plug-in, ognuno dei quali può essere attivato e configurato tramite l'applicazione di configurazione di A.I. Tech. I plug-in sono responsabili per l'adattamento dei metadati a diversi formati ove richiesto e per il loro inoltramento ad applicazioni remote di fruizione. Ad esempio, il plug-in Milestone converte il formato JSON in XML, usando lo schema XML definito da Milestone per la sua linea di prodotti XProtect, ed invia le informazioni al Milestone XProtect Event Server in modo tale che l'informazione sia poi utilizzabile per operazioni di ricerca e filtraggio degli eventi.

Il plug-in HTTP invia i dati in formato JSON mediante una richiesta HTTP POST ad un indirizzo specificato in fase di configurazione; questo può essere semplicemente associato ad un web service basato sul paradigma REST allo scopo di consentirne la fruizione attraverso una applicazione utente. Questa rappresenta una soluzione molto conveniente per inviare i metadati ad applicazioni di terze parti basate su service oriented architecture(SOA).

Il plug-in di salvataggio aggiunge i metadati ad un file che può risiedere su un dispositivo di salvataggio locale (ad esempio una scheda SD nel caso in cui l'applicazione risieda sulla telecamera) or su un dispositivo di storage condiviso (ad esempio un server o un NAS). La prima opzione è estremamente utile quando la telecamera potrebbe non essere sempre collegata alla rete, rendendo possibile la raccolta differita dei metadati per un uso off-line. La seconda opzione può essere applicata quando non è possibile, o desiderabile, avere l'applicazione che deve consumare i metadati che gira su un server; in questo caso l'applicazione può verificare periodicamente i file da leggere e processare i metadati.