

Ricerca di testo libero in AXIS Camera Station Pro

Gennaio 2025

Sommario

AXIS Camera Station Pro è dotata di diversi strumenti di ricerca forense dei video preinstallati. Queste includono lo scrubbing della cronologia, la ricerca dei dati e la ricerca intelligente con oggetti preclassificati e ricerca a testo libero.

La ricerca a testo libero permette di cercare qualsiasi oggetto in movimento descrivendolo con parole proprie. La libertà di creare filtri di ricerca dettagliati con un'ampia gamma di attributi descrittivi consente di trovare più rapidamente i filmati pertinenti.

La funzione di ricerca a testo libero si basa sull'associazione di testo e immagine fornita da un modello fondativo open-source preaddestrato, ottimizzato da Axis per i casi d'uso della sorveglianza. La ricerca può essere applicata a una telecamera o a più telecamere contemporaneamente.

Una rappresentazione numerica della ricerca a testo libero viene confrontata con le rappresentazioni numeriche delle immagini degli oggetti in movimento rilevati. I risultati vengono visualizzati sotto forma di miniature, con il nome della telecamera, l'ora e la data, ordinati in base alla pertinenza del testo della ricerca.

Grazie alla ricerca a testo libero, utilizziamo l'intelligenza artificiale per aumentare l'accuratezza e l'efficienza delle nostre soluzioni di ricerca forense e, in ultima analisi, migliorare il processo decisionale umano. Per rispettare gli standard legali ed etici, la funzione di ricerca include una funzione di moderazione separata sviluppata da Axis che limita l'uso di parole offensive nel testo di ricerca. Tutte le ricerche sono inoltre registrate e visibili agli amministratori, consentendo di monitorare la situazione e intervenire in caso di uso improprio.

Indice

1	Introduzione	4
2	Premessa: ricerca forense in AXIS Camera Station Pro	4
3	Come funziona la ricerca a testo libero?	4
4	Scegliere le frasi di ricerca	5
	4.1 Linee guida per il testo da inserire	6
	4.2 Moderazione delle ricerche	6
	4.3 Registri del testo di ricerca	6
5	Corrispondenza testo-immagine	7
	5.1 Modello fondativo ottimizzato per la sorveglianza	7
	5.2 Vettori caratteristica	7
6	Panoramica del processo	7
7	Uso responsabile dell'AI	9

1 Introduzione

Lo strumento di ricerca a testo libero di AXIS Camera Station One consente di effettuare ricerche nelle registrazioni video utilizzando parole proprie anziché filtri predefiniti.

Questo documento tecnico illustra il funzionamento del metodo di ricerca e presenta alcune linee guida su come utilizzarlo. Descriviamo inoltre la funzione di moderazione e l'accesso al testo di ricerca per garantire la conformità agli standard legali ed etici.

2 Premessa: ricerca forense in AXIS Camera Station Pro

AXIS Camera Station Pro è dotata di diversi strumenti di ricerca video forense preinstallati, tra cui lo scrubbing della cronologia, la ricerca dei dati, la ricerca intelligente con oggetti preclassificati e la ricerca a testo libero.

La funzione di ricerca intelligente utilizza i metadati della scena generati dal dispositivo Axis. I metadati includono il tipo di oggetto (persona, veicolo o oggetto sconosciuto) per gli oggetti in movimento, insieme ad attributi come il colore degli abiti e del veicolo, le targhe, la velocità, l'ubicazione e i dati temporali.

Nei dispositivi con capacità di analisi limitate, la funzione di ricerca si basa sul rilevamento del movimento nel dispositivo combinato con la classificazione degli oggetti eseguita sul server AXIS Camera Station Pro. La ricerca forense in AXIS Camera Station Pro è quindi una soluzione ibrida in cui le capacità dei dispositivi edge vengono utilizzate il più possibile, ma integrate con i dati del server, se necessario.

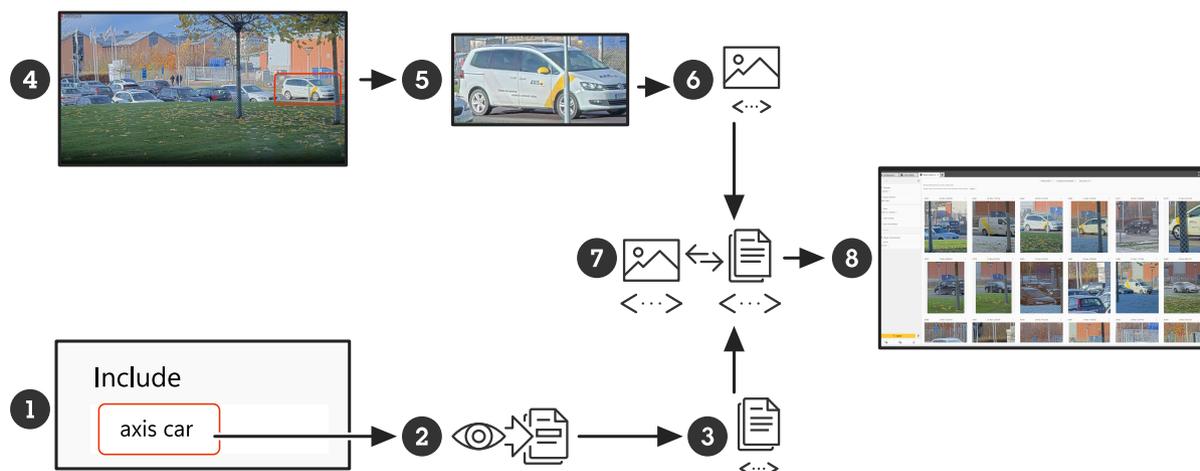
Tradizionalmente, le ricerche con i metadati della scena dovevano essere condotte utilizzando filtri di ricerca predefiniti. In questo modo, si scelgono descrittori fissi dell'oggetto da un elenco, tra cui il tipo di oggetto (come "veicolo"), il tipo di veicolo (se applicabile, ad es. "auto"), il colore (ad es. "blu") e altro. Il nuovo metodo di ricerca a testo libero permette invece di creare un filtro di ricerca personalizzato.

Sebbene la ricerca preclassificata fornisca risultati di grande precisione, non è in grado di rilevare nuovi tipi di oggetti non predefiniti. Per ovviare a questa limitazione, la ricerca a testo libero offre la libertà e la flessibilità di effettuare ricerche utilizzando le proprie parole. È possibile descrivere qualsiasi oggetto in movimento in modo più dettagliato con il linguaggio naturale e le associazioni per affinare la ricerca e ottenere più risultati.

3 Come funziona la ricerca a testo libero?

Una rappresentazione numerica della ricerca a testo libero viene confrontata con le rappresentazioni numeriche delle immagini degli oggetti in movimento rilevati. Il risultato di questa corrispondenza testo-immagine viene presentato e ordinato in base alla migliore corrispondenza. I risultati vengono

visualizzati sotto forma di miniature, con il nome della telecamera, l'ora e la data, ordinati in base alla pertinenza del testo della ricerca.



Panoramica semplificata del processo di ricerca a testo libero. I passaggi 4-6 avvengono continuamente, anche quando non si effettua la ricerca, per creare vettori caratteristica di tutti gli oggetti in movimento rilevati.

- 1 *Digita la tua domanda di ricerca a testo libero.*
- 2 *Un modulo di moderazione impedisce l'uso di parole malevole e non etiche.*
- 3 *Il modello fondativo crea una rappresentazione numerica (un vettore caratteristica) del testo di ricerca.*
- 4 *Una telecamera rileva il movimento in una scena.*
- 5 *La telecamera seleziona un'immagine ritagliata per rappresentare l'oggetto in movimento.*
- 6 *Il modello fondativo crea un vettore caratteristica dell'oggetto dopo averne analizzato la forma, i modelli, il colore e altro.*
- 7 *I due vettori caratteristica vengono confrontati.*
- 8 *Il risultato del confronto viene classificato in base alla migliore corrispondenza e presentato sotto forma di miniature.*

La ricerca a testo libero può essere applicata a una telecamera o a più telecamere contemporaneamente.

Per restringere la portata della ricerca a testo libero, è possibile combinarla con altre funzioni di ricerca intelligente, come la ricerca per similarità o la ricerca temporale, utilizzando un tipo di ricerca dopo l'altro.

4 Scegliere le frasi di ricerca

È possibile cercare qualsiasi oggetto in movimento e qualsiasi tipo di veicolo. Seguire le linee guida per ottenere risultati ottimali.

Si ricorda che bisogna cercare solo oggetti in movimento. La ricerca di oggetti fermi nella maggior parte dei casi non funziona.

Le frasi di ricerca vengono moderate e registrate per evitare comportamenti di ricerca non etici.

4.1 Linee guida per il testo da inserire

- Descrivete le situazioni come descrivereste un'immagine. Il modello è alimentato con immagini fisse, quindi la ricerca di azioni (come cadere, correre o rubare) può essere difficile, poiché richiede un contesto più ampio.
- Occorre descrivere gli oggetti utilizzando alcuni descrittori chiave: "a person in a red sweater" (una persona con un maglione rosso) o "a yellow pickup truck" (un furgone giallo). Come altri modelli multimodali, il modello di ricerca a testo libero si comporta bene con descrittori quali oggetti e colori, ma è meno adatto al conteggio ("tre persone"), al gergo o alle indicazioni emotive ("uomo dall'aria arrabbiata"). La descrizione dell'oggetto non deve essere soggettiva, troppo vaga né includere dettagli troppo specifici.
- Combinazione di più attributi di oggetti usando e: "person with red hat and backpack" (persona con cappello rosso e zaino).
- Descrizione di testi, loghi testuali o nomi di marchi: "van with text Axis" (furgone con testo Axis).
- Non puntare sulla descrizione degli ambienti. L'elaborazione viene effettuata su immagini ritagliate degli oggetti, il che significa che il modello potrebbe non vedere l'ambiente circostante. Descrittori generici della scena o dell'ambiente (come "città", "urbano", "parco", "giardino", "lago" o "spiaggia") potrebbero quindi non dare buoni risultati.
- Se il risultato non vi soddisfa, sperimentate frasi alternative.
- La ricerca con testo libero supporta solo la lingua inglese.

4.2 Moderazione delle ricerche

Abbiamo implementato una moderazione delle ricerche basata sulle pratiche comuni per garantire un filtro efficace. Il modello di moderazione è un modello di natural language processing (elaborazione del linguaggio naturale) che controlla la ricerca per limitare le parole offensive. Controlla le stringhe di testo intere alla ricerca di contenuti dannosi, inappropriati o malevoli. Inoltre, abbiamo potenziato queste funzionalità con misure proprietarie, tra cui un elenco personalizzato di categorie e parole di ricerca vietate. Se una ricerca contiene parole o frasi di questo elenco, la rifiutiamo per mantenere un ambiente di ricerca sicuro e garantire risultati etici. L'utente può fornire un feedback anonimo ad Axis se non è d'accordo con il blocco di una parola o se vuole suggerire il blocco di una parola.

4.3 Registri del testo di ricerca

AXIS Camera Station Pro mantiene una traccia di audit trail delle operazioni dell'utente. I registri di audit trail oltre a tenere traccia delle operazioni specifiche e dell'identità dell'utente, conservano anche tutti i dati utilizzati nelle operazioni. Ciò significa che tutte le ricerche dell'utente, comprese le richieste di ricerca, vengono registrate. Gli amministratori possono utilizzare i registri per identificare i comportamenti di ricerca inappropriati degli utenti, segnalare le richieste di ricerca non etiche e intraprendere azioni correttive.

Si noti che i dati video non vengono condivisi con Axis. I vostri dati rimangono sul vostro server.

5 Corrispondenza testo-immagine

La possibilità di cercare i metadati dei video utilizzando ricerche a testo libero amplia in modo significativo le capacità di ricerca da un elenco predefinito di attributi a criteri di ricerca quasi illimitati. In AXIS Camera Station Pro questa funzione si basa su un modello fondativo open-source, addestrato su miliardi di coppie immagine-testo e adattato da Axis ai casi d'uso di sorveglianza per migliorarne le prestazioni.

5.1 Modello fondativo ottimizzato per la sorveglianza

Il modello fondativo è un modello testo-immagine addestrato su grandi serie di dati di coppie testo-immagine. Si tratta di un modello chiamato zero-shot che abbina il testo a immagini pertinenti. Un modello zero-shot è un tipo di modello basato sull'intelligenza artificiale (AI) in grado di riconoscere e classificare oggetti o concetti senza integrazione di dati di addestramento precedenti. In altre parole, il modello è in grado di eseguire compiti senza averne visto alcun esempio in precedenza. Questa capacità è fondamentale per garantire prestazioni ottimali nella corrispondenza tra linguaggio naturale e immagini.

Il modello è stato addestrato su una grande quantità di combinazioni testo-immagine e opera su una rete neurale di oltre 2,5 miliardi di parametri. In Axis abbiamo utilizzato il nostro esclusivo materiale di addestramento per perfezionare ulteriormente questo modello, migliorandone la capacità di interpretare le immagini con le viste e gli oggetti tipici delle telecamere per la sorveglianza. Ciò significa che abbiamo ottimizzato il modello per i casi d'uso della sorveglianza.

5.2 Vettori caratteristica

Quando si effettua una ricerca a testo libero, il modello fondativo crea un vettore caratteristica del testo di ricerca.

Il modello fondativo genera anche descrizioni continue di ogni oggetto tracciato nelle scene e crea vettori caratteristica per rappresentarli. Ogni oggetto è rappresentato da un solo vettore caratteristica, che viene memorizzato nel nostro database. In questo modo la ricerca è veloce, poiché i vettori caratteristica sono già precalcolati e prontamente disponibili nel database.

Entrambi i tipi di vettori caratteristica vengono inseriti nel motore di confronto vettoriale per determinare la distanza di somiglianza tra il testo della ricerca e tutti i vettori caratteristica disponibili che rappresentano gli oggetti rilevati.

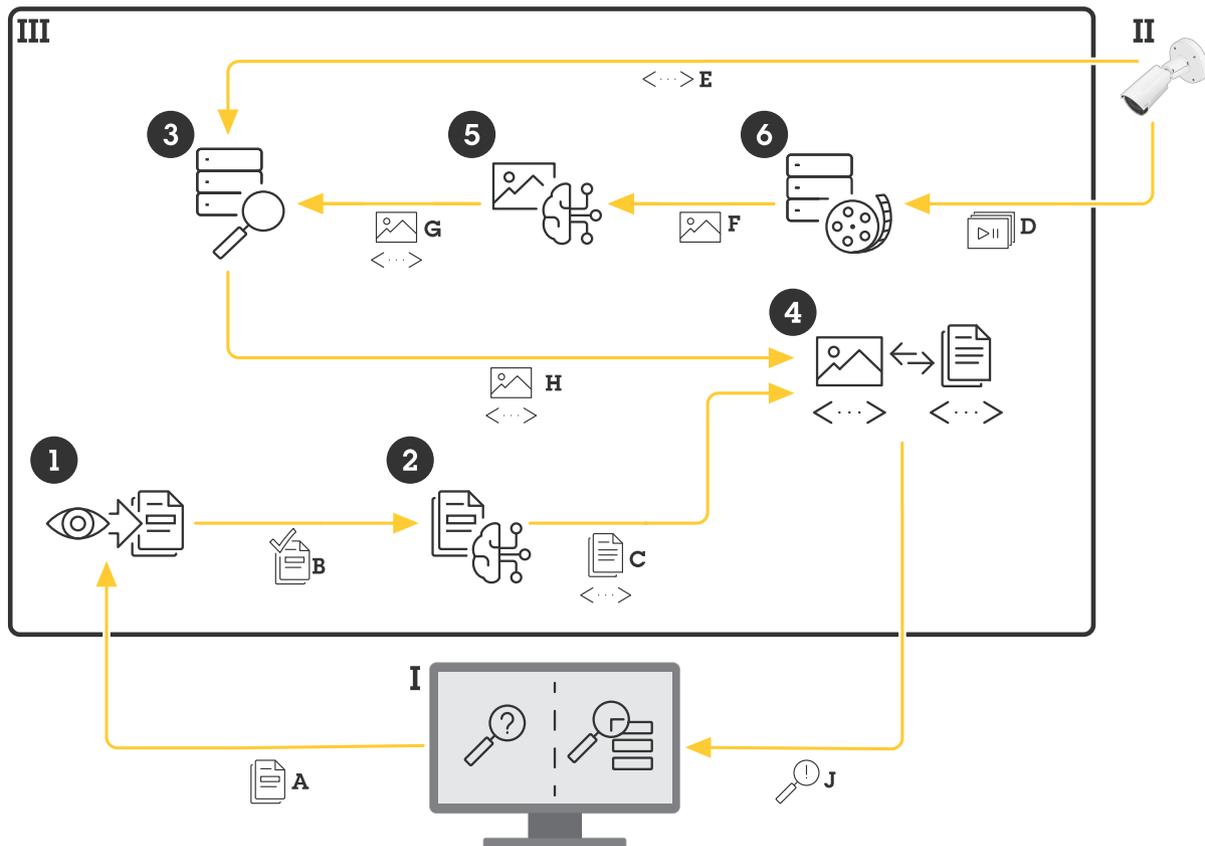
Un vettore caratteristica è una rappresentazione numerica di testo o immagini. I vettori caratteristica di persone o oggetti sono quindi solo rappresentazioni astratte dell'aspetto della persona o dell'oggetto. I vettori caratteristica non contengono nessuna informazione sulle caratteristiche interpretabili dall'uomo, ad esempio il colore dei capelli o dei vestiti, che possono essere mappate esplicitamente rispetto a una persona specifica o utilizzate per l'identificazione. I vettori caratteristica possono essere utilizzati solo per il confronto con altri vettori caratteristica.

6 Panoramica del processo

Lo schema con la panoramica del processo ne mostra le fasi principali, comprese le sedi in cui si svolgono le fasi e il tipo di dati prodotti da ciascuna fase.

Si noti che il ciclo superiore dello schema, che comprende la telecamera (II), l'archiviazione delle registrazioni (6), il modello fondativo (5) e il database di ricerca (3), è un processo che si svolge

continuamente per creare vettori caratteristica per tutti gli oggetti in movimento rilevati, e non solo quando si effettua una ricerca.



Sedi principali (I-III) per il processo di ricerca a testo libero

- I Client AXIS Camera Station Pro*
- II Telecamere*
- III Server AXIS Camera Station Pro*

Principali fasi del processo (1-6)

- 1 Moderazione del testo di ricerca*
- 2 Modello di fondativo (testo)*
- 3 Ricerca nel database*
- 4 Confronto dei vettori*
- 5 Modello di fondativo (immagine)*
- 6 Archiviazione delle registrazioni*

Tipo di dati o risultato (A-J)

- A Stringa di testo*
- B Stringa di testo*
- C Vettore caratteristica (testo)*
- D Video*
- E Metadati*
- V Immagini*
- G Vettori caratteristica (immagine)*
- H Vettori caratteristica (immagine)*
- J Risultati della ricerca*

(I) **Client AXIS Camera Station Pro**: Qui si digita la propria domanda di ricerca e si ricevono i risultati ordinati

(II) **Telecamera(e)**: la ricerca a testo libero funziona sulle telecamere Axis con AXIS OS 5.51 o successivo, ma migliori sono le telecamere, migliori sono i risultati ottenuti. I dispositivi più vecchi forniscono metadati meno precisi basati solo sul rilevamento del movimento. I dispositivi più recenti integrano AXIS Scene Metadata, che include la classificazione degli oggetti. Il rilevamento e il tracciamento degli oggetti in movimento della telecamera vengono utilizzati per trovare un'immagine rappresentativa di ogni oggetto rilevato, riducendo così il numero di immagini da analizzare sul server.

(III) **Server AXIS Camera Station Pro**: Qui vengono elaborati e memorizzati tutti i metadati e i dati video delle telecamere. Prima di effettuare una ricerca a testo libero, il server deve (per ogni oggetto in movimento rilevato) decodificare il video ed estrarre un'immagine dell'oggetto rilevato. Il modello fondativo elabora quindi questa immagine per creare il vettore caratteristica. Queste operazioni sono piuttosto costose in termini di capacità di elaborazione, quindi per migliorare le prestazioni i vettori caratteristica vengono salvati in un database consentendo la ricerca rapida in futuro. Se il proprio server dispone di capacità inutilizzate, è assolutamente consigliabile abilitare l'elaborazione in background dei video delle telecamere più importanti, in modo da velocizzare notevolmente la ricerca.

(1) **Moderazione del testo di ricerca**: Il modello di moderazione controlla il testo della ricerca per limitare i contenuti offensivi.

(2) **Modello fondativo (testo)**: il modello fondativo crea una rappresentazione numerica (vettore caratteristica) della stringa di testo della ricerca moderata.

(3) **Ricerca nel database**: Il database di ricerca contiene metadati completi provenienti da AXIS Scene Metadata o metadati creati dal server, compresi i dati di classificazione degli oggetti con attributi, ora, posizione e vettori caratteristica.

(4) **Confronto dei vettori**: La rappresentazione del vettore caratteristica della stringa di testo della ricerca viene confrontata con le rappresentazioni del vettore caratteristica delle immagini degli oggetti rilevati nel video.

(5) **Modello fondativo (immagine)**: il modello fondativo crea rappresentazioni numeriche (vettori caratteristica) di ogni traccia di oggetto nel video registrato. Si tratta di un processo continuo che avviene anche quando non si sta effettuando una ricerca.

(6) **Archiviazione delle registrazioni**: È qui che vengono memorizzati i video della telecamera e che il modello fondativo riceve le immagini.

7 Uso responsabile dell'AI

Grazie alla ricerca a testo libero, utilizziamo l'intelligenza artificiale per aumentare l'accuratezza e l'efficienza delle nostre soluzioni di ricerca forense e, in ultima analisi, migliorare il processo decisionale umano.

La responsabilità e l'affidabilità sono fondamentali per l'approccio di Axis all'intelligenza artificiale. Equivalgono a garantire che i sistemi di IA da noi creati aderiscano ai principi etici, rispettino le leggi e gestiscano efficacemente i rischi. Axis fornisce strumenti che permettono ai nostri clienti di essere sicuri dell'integrità delle proprie operazioni. La funzione di ricerca a testo libero di AXIS Camera Station Pro include un modello di classificazione del testo ottimizzato per la moderazione dei messaggi di testo. Abbiamo sviluppato questo modello per moderare il testo di ricerca e aiutare a prevenire l'uso non etico.

La ricerca a testo libero si collega ai servizi cloud di Axis una volta alla settimana per verificare se i modelli di intelligenza artificiale devono essere aggiornati per conformarsi a nuove normative o requisiti. Se la connessione si interrompe, le operazioni di ricerca a testo libero non saranno disponibili finché non verrà ristabilita.

Per rispettare ulteriormente gli standard legali ed etici nell'applicazione dell'AI, i nostri prodotti forniscono controlli degli accessi basati sulle credenziali di autenticazione degli utenti e sui permessi di accesso. Questo permette ai nostri clienti di imporre agli utenti la conformità alle procedure operative.

Informazioni su Axis Communications

Axis permette di creare un mondo più intelligente e sicuro grazie a soluzioni che migliorano la sicurezza e le prestazioni aziendali. In qualità di azienda leader nelle tecnologie di rete, Axis offre prodotti e servizi per la videosorveglianza, il controllo accessi, interfonni e sistemi audio, che supporta con applicazioni analitiche intelligenti e una formazione di alta qualità.

Axis ha oltre 4000 dipendenti in più di 50 paesi e collabora con partner tecnologici e integratori di sistemi in tutto il mondo per fornire soluzioni ai clienti. Fondata nel 1984, Axis è una società con sede a Lund, in Svezia