

WHITE PAPER

Pesquisa de texto livre no **AXIS Camera Station Pro**

Janeiro 2025

Resumo

O AXIS Camera Station Pro vem com várias ferramentas de pesquisa de vídeo forense pré-instaladas. Entre elas, depuração de linha do tempo, pesquisa de dados e pesquisa inteligente com objetos pré-classificados e pesquisa de texto livre.

A pesquisa de texto livre permite pesquisar qualquer objeto em movimento, descrevendo-o com suas próprias palavras. A liberdade de criar filtros de pesquisa detalhados com uma ampla faixa de atributos descritivos possibilita encontrar gravações relevantes mais rapidamente.

A função de pesquisa de texto livre é baseada na correspondência entre texto-imagem fornecida por um modelo de base de código aberto pré-treinado que foi otimizado pela Axis para casos de uso de monitoramento. A pesquisa pode ser aplicada a uma câmera ou a várias câmeras ao mesmo tempo.

Uma representação numérica de sua consulta de texto livre é comparada com representações numéricas de imagens de objetos em movimento detectados. Os resultados são exibidos como miniaturas, incluindo nome da câmera, hora e data, classificados por relevância à sua consulta de pesquisa.

Com a pesquisa de texto livre, usamos a IA para aumentar a precisão e a eficiência de nossas soluções de pesquisa forense e, em última análise, aprimorar a tomada de decisões humana. Para cumprir os padrões legais e éticos, a função de pesquisa inclui uma função de moderação separada desenvolvida pela Axis que restringe o uso de palavras ofensivas nas consultas de pesquisa. Todas as pesquisas também são registradas e ficam visíveis para administradores, possibilitando o acompanhamento e a adoção de medidas corretivas em caso de uso indevido.

Índice

1	Introdução	4
2	Contexto: pesquisa forense no AXIS Camera Station Pro	4
3	Como funciona a pesquisa de texto livre?	4
4	Construção de consultas de pesquisa	5
	4.1 Diretrizes do prompt	6
	4.2 Moderação de consultas	6
	4.3 Registro de consultas de pesquisa	6
5	Correspondência entre texto–imagem	6
	5.1 Modelo de base otimizado para monitoramento	7
	5.2 Vetores de características	7
6	Visão geral do processo	7
7	Uso responsável da IA	9

1 Introdução

A ferramenta de pesquisa de texto livre do AXIS Camera Station Pro permite pesquisar gravações de vídeo usando suas próprias palavras em vez de filtros predefinidos.

Este white paper descreve como funciona o método de pesquisa e apresenta algumas orientações sobre como usá-lo. Também descrevemos a função de moderação e o registro de consultas em vigor para garantir a conformidade com padrões legais e éticos.

2 Contexto: pesquisa forense no AXIS Camera Station Pro

O AXIS Camera Station Pro vem com várias ferramentas de pesquisa de vídeo forense pré-instaladas, incluindo depuração de linha do tempo, pesquisa de dados, bem como pesquisa inteligente com objetos pré-classificados e pesquisa de texto livre.

A função de pesquisa inteligente usa metadados de cena gerados pelo dispositivo Axis. Os metadados incluem o tipo de objeto (pessoa, tipo de veículo ou objeto desconhecido) para objetos em movimento, juntamente com atributos como cor da roupa e do veículo, placas de licença, velocidade, local e registro de data e hora.

Para dispositivos com recursos analíticos limitados, a função de pesquisa é baseada na detecção de movimento no dispositivo combinada com a classificação de objetos realizada no servidor do AXIS Camera Station Pro. A pesquisa forense no AXIS Camera Station Pro é, portanto, uma solução híbrida em que os recursos dos dispositivos de borda são usados tanto quanto possível, mas complementados com dados do servidor quando necessário.

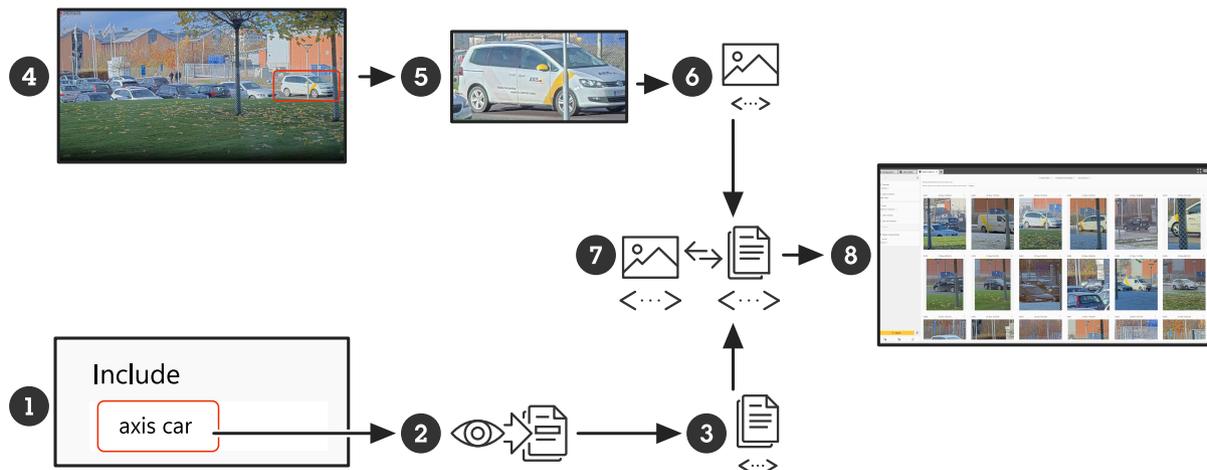
Tradicionalmente, pesquisas com metadados de cenas tinham que ser realizadas usando filtros de pesquisa predefinidos. Com eles, você escolhe descritores de objetos fixos em uma lista, incluindo o tipo de objeto (tais como "veículo"), o tipo de veículo (se aplicável, por exemplo, "carro"), a cor (por exemplo, "azul") e mais. Em vez disso, o novo método de pesquisa de texto livre permite que você crie seu próprio filtro de pesquisa.

Embora a pesquisa pré-classificada ofereça resultados de alta precisão, ela não é capaz de detectar novos tipos de objetos que não estejam predefinidos. Para resolver essa limitação, a pesquisa de texto livre lhe dá a liberdade e a flexibilidade de pesquisar usando suas próprias palavras. Você pode descrever qualquer objeto em movimento em mais detalhes com linguagem natural e associações para adequar sua pesquisa e obter mais resultados.

3 Como funciona a pesquisa de texto livre?

Uma representação numérica de sua consulta de texto livre é comparada com representações numéricas de imagens de objetos em movimento detectados. O resultado dessa correspondência entre texto e imagem

é apresentado e classificado de acordo com a melhor correspondência. Os resultados são exibidos como miniaturas, incluindo nome da câmera, hora e data, classificados por relevância à sua consulta de pesquisa.



Visão geral simplificada do processo de pesquisa de texto livre. Observe que as etapas 4-6 ocorrem continuamente, mesmo quando não se está pesquisando, para criar vetores de características de todos os objetos em movimento detectados.

- 1 Você digita sua consulta de pesquisa de texto livre.
- 2 Um módulo de moderação impede o uso de palavras tóxicas e antiéticas.
- 3 O modelo de base cria uma representação numérica (um vetor de características) da consulta de pesquisa.
- 4 Uma câmera detecta movimento em uma cena.
- 5 A câmera seleciona uma imagem cortada para representar o objeto em movimento.
- 6 O modelo de base cria um vetor de características do objeto após analisar sua forma, padrões, cor, entre outros.
- 7 Os dois vetores de características são comparados.
- 8 O resultado da comparação é classificado de acordo com a melhor correspondência e apresentado como miniaturas.

A pesquisa de texto livre pode ser aplicada a uma câmera ou a várias câmeras ao mesmo tempo.

Para restringir o escopo da pesquisa de texto livre, é possível combiná-la com outras funções de pesquisa inteligente, como a pesquisa por similaridade ou a pesquisa baseada no tempo, usando um tipo de pesquisa após o outro.

4 Construção de consultas de pesquisa

Você pode pesquisar qualquer objeto em movimento e qualquer tipo de veículo. Siga as diretrizes para obter os melhores resultados.

Observe que você só deve pesquisar objetos em movimento. A pesquisa de objetos fixos, na maioria dos casos, não funcionará.

As frases de pesquisa são moderadas e registradas para impedir comportamentos de pesquisa antiéticos.

4.1 Diretrizes do prompt

- Descreva as situações como você descreveria uma imagem. O modelo é alimentado com imagens estáticas, portanto, a pesquisa de ações (como cair, correr ou roubar) pode ser difícil, pois exigiria mais contexto.
- Descreva objetos usando alguns descritores-chave: "uma pessoa com um casaco vermelho" ou "uma caminhonete amarela". Como outros modelos multimodais, o modelo de pesquisa de texto livre tem bom desempenho com descritores como objetos e cores, mas é menos adequado com contagem de objetos ("três pessoas"), gírias ou pistas emocionais ("homem com aparência zangada"). A descrição do objeto não deve ser subjetiva, muito vaga nem incluir detalhes muito específicos.
- Combine vários atributos de objeto usando e: "pessoa com chapéu vermelho e mochila".
- Descreva texto, logotipos de texto ou nomes de marcas: "van com texto Axis".
- Não foque na descrição de ambientes. O processamento é feito em imagens cortadas de objetos, o que significa que o modelo pode não ver os arredores dos objetos. Descritores de cenas ou ambientes amplos (como "cidade", "urbano", "parque", "jardim", "lago" ou "praia") podem, portanto, não fornecer bons resultados.
- Experimente frases alternativas se não se satisfizer com um resultado.
- O prompt de texto livre é compatível apenas com o inglês.

4.2 Moderação de consultas

Implementamos a moderação de consultas com base em práticas comuns para garantir uma filtragem eficaz. O modelo de moderação é um modelo de processamento de linguagem natural que verifica a consulta para restringir formulações ofensivas. Ele verifica sequências de texto inteiras em busca de conteúdo nocivo, inadequado ou tóxico. Além disso, aprimoramos esses recursos com medidas proprietárias, incluindo uma lista personalizada de categorias e palavras de pesquisa proibidas. Quando uma consulta contém palavras ou frases dessa lista, rejeitamos a pesquisa para manter um ambiente de pesquisa seguro e garantir resultados éticos. Você pode fornecer feedback anônimo do usuário à Axis se não concordar com o bloqueio de uma palavra ou se quiser sugerir o bloqueio de uma palavra.

4.3 Registro de consultas de pesquisa

O AXIS Camera Station Pro mantém uma trilha de auditoria das operações dos usuários. As trilhas de auditoria não apenas monitoram as operações específicas e a identidade do usuário, mas também retêm todos os dados usados na operação. Isso significa que todas as pesquisas de usuários, incluindo prompts de pesquisa, são registradas. Os administradores podem usar os registros para identificar comportamentos de pesquisa inadequados entre usuários, sinalizar prompts de pesquisa antiéticos e adotar medidas corretivas.

Observe que dados de vídeo não são compartilhados com a Axis. Seus dados permanecem em seu servidor.

5 Correspondência entre texto-imagem

A possibilidade de pesquisar metadados de vídeo usando consultas de texto livre expande significativamente os recursos de pesquisa de uma lista predefinida de atributos para critérios de pesquisa quase ilimitados. No AXIS Camera Station Pro, essa função é baseada em um modelo de base de código aberto, treinado

em bilhões de pares de imagem-texto e aperfeiçoado pela Axis para casos de uso de monitoramento com o fim de melhorar o desempenho.

5.1 Modelo de base otimizado para monitoramento

O modelo de base é um modelo de texto para imagem treinado em grandes conjuntos de dados de pares de texto-imagem. É um modelo zero-shot que combina texto com imagens pertinentes. Um modelo zero-shot é um tipo de modelo de inteligência artificial (IA) que pode reconhecer e classificar objetos ou conceitos sem dados de treinamento prévio. Em outras palavras, o modelo pode executar tarefas sem ter visto nenhum exemplo da tarefa anteriormente. Essa capacidade é fundamental para garantir o desempenho ideal na correspondência de linguagem natural com imagens.

O modelo foi treinado em uma grande quantidade de combinações de texto-imagem e opera em uma rede neural com mais de 2,5 bilhões de parâmetros. Na Axis, usamos nosso próprio material de treinamento exclusivo para ajustar ainda mais esse modelo, melhorando sua capacidade de interpretar imagens com objetos e exibições típicas de câmeras de monitoramento. Isso significa que otimizamos o modelo para casos de uso de monitoramento.

5.2 Vetores de características

Quando você faz uma pesquisa de texto livre, o modelo de base cria um vetor de características da consulta de pesquisa.

O modelo de base também produz continuamente descrições de cada objeto rastreado nas cenas e cria vetores de características para representá-los. Cada objeto é representado por apenas um vetor de características, que é armazenado em nosso banco de dados. Isso agiliza a pesquisa, pois os vetores de características já foram pré-calculados e estão prontamente disponíveis no banco de dados.

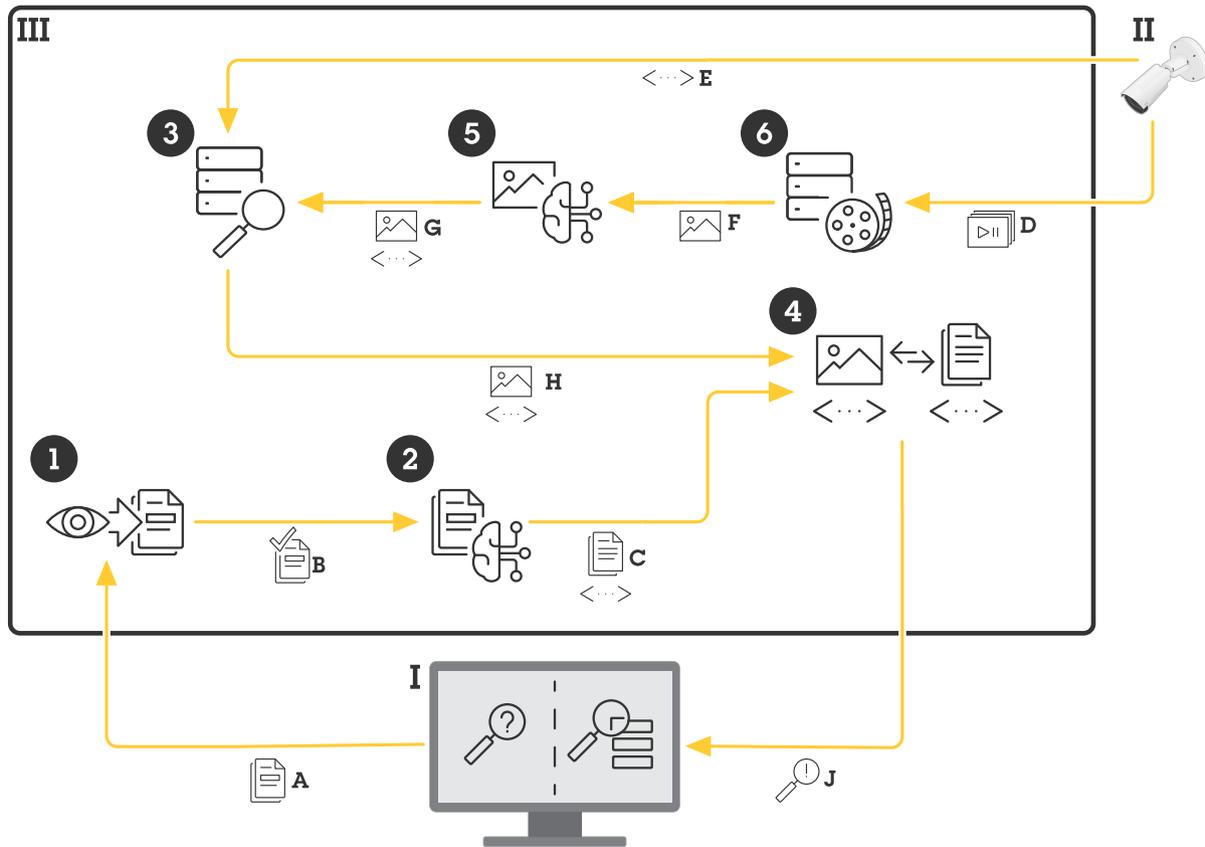
Ambos os tipos de vetores de características são alimentados no mecanismo de comparação de vetores para determinar a distância de similaridade entre sua consulta de pesquisa e todos os vetores de características disponíveis que representam objetos detectados.

Um vetor de características é uma representação numérica de texto ou imagens. Os vetores de características de pessoas ou objetos são, portanto, apenas representações abstratas da aparência da pessoa ou do objeto. Vetores de características não contêm nenhuma informação interpretável por humanos sobre características, como a cor do cabelo ou da roupa, que possa ser explicitamente mapeada para uma pessoa específica ou usada para identificação. Os vetores de características só podem ser usados para comparações com outros vetores de características.

6 Visão geral do processo

O diagrama de visão geral do processo mostra as principais etapas do processo, incluindo os locais em que as etapas ocorrem e o tipo de dados que cada etapa produz.

Observe que o loop superior no diagrama, incluindo câmera (II), armazenamento de gravações (6), modelo de base (5) e banco de dados de pesquisa (3), é um processo que ocorre continuamente para criar vetores de características de todos os objetos em movimento detectados, e não apenas quando você faz uma pesquisa.



Principais locais (I-III) para o processo de pesquisa de texto livre

- I Cliente do AXIS Camera Station Pro
- II Câmera(s)
- III Servidor do AXIS Camera Station Pro

Principais etapas do processo (1-6)

- 1 Moderação de consultas de pesquisa
- 2 Modelo de base (texto)
- 3 Banco de dados de pesquisa
- 4 Comparação de vetores
- 5 Modelo de base (imagem)
- 6 Armazenamento de gravações

Tipo de dados ou resultado (A-J)

- A Sequência de texto
- B Sequência de texto
- C Vetor de características (texto)
- D Vídeo
- E Metadados
- F Imagens
- G Vetores de características (imagem)
- H Vetores de características (imagem)
- J Resultado da pesquisa

(I) **Cliente do AXIS Camera Station Pro:** Aqui você digita sua consulta de pesquisa e recebe resultados de pesquisa ordenados

(II) **Câmera(s):** a pesquisa de texto livre funciona em câmeras Axis com o AXIS OS 5.51 ou posterior, mas quanto melhor for a câmera, melhores serão os resultados obtidos. Dispositivos mais antigos fornecem metadados menos granulares com base apenas na detecção de movimento. Dispositivos mais novos produzem os AXIS Scene Metadata, o que inclui a classificação de objetos. A detecção e o rastreamento de objetos em movimento da câmera são usados para encontrar uma imagem representativa de cada objeto detectado, reduzindo assim o número de imagens a serem analisadas no servidor.

(III) **Servidor do AXIS Camera Station Pro:** Aqui, todos os metadados e dados de vídeo das câmeras são processados e armazenados. Antes de fazer uma pesquisa de texto livre, o servidor deve (para cada objeto em movimento detectado) decodificar o vídeo e extrair uma imagem do objeto detectado. Em seguida, o modelo de base processa essa imagem para criar o vetor de características. Essas operações são bastante dispendiosas em termos de capacidade de processamento, portanto, para melhorar o desempenho, os vetores de características são salvos em um banco de dados, possibilitando a pesquisa rápida no futuro. Se o seu servidor tiver capacidade livre, recomendamos enfaticamente que você permita o processamento em segundo plano do vídeo de suas câmeras mais importantes, pois isso tornará a busca significativamente mais rápida.

(1) **Moderação de consultas de pesquisa:** O modelo de moderação verifica a consulta para restringir conteúdo ofensivo.

(2) **Modelo de base (texto):** o modelo de base cria uma representação numérica (vetor de características) da sequência de texto da consulta de pesquisa moderada.

(3) **Banco de dados de pesquisa:** O banco de dados de pesquisa contém metadados completos do AXIS Scene Metadata ou metadados criados pelo servidor, incluindo dados de classificação de objetos com atributos, hora, posição e vetores de características.

(4) **Comparação de vetores:** A representação do vetor de características da sequência de texto da consulta de pesquisa é comparada com as representações do vetor de características das imagens de objetos detectadas no vídeo.

(5) **Modelo de base (imagem):** o modelo de base cria representações numéricas (vetores de características) de cada trilha de objeto na gravação de vídeo. Esse é um processo contínuo que ocorre também quando você não está pesquisando.

(6) **Armazenamento de gravações:** É nesse local que o vídeo da câmera é armazenado e onde o modelo de base obtém suas imagens.

7 Uso responsável da IA

Com a pesquisa de texto livre, usamos a IA para aumentar a precisão e a eficiência de nossas soluções de pesquisa forense e, em última análise, aprimorar a tomada de decisões humana.

A responsabilidade e a transparência são fundamentais para a abordagem de IA da Axis. Isso envolve garantir que os sistemas de IA que criamos sigam os princípios éticos, cumpram as leis e administrem os riscos de forma eficaz. A Axis fornece ferramentas que permitem que nossos clientes tenham confiança na integridade de suas operações. O recurso de pesquisa de texto livre do AXIS Camera Station Pro inclui um modelo de classificação de texto aperfeiçoado para moderação de prompts de texto. Desenvolvemos esse modelo para moderar as consultas de pesquisa e, assim, ajudar você a impedir o uso antiético.

A pesquisa de texto livre se conecta aos serviços em nuvem da Axis uma vez por semana para verificar se os modelos de IA precisam de atualização para atender a novos regulamentos ou requisitos. Se a conexão falhar, as operações de pesquisa de texto livre ficarão indisponíveis até que a conexão seja restabelecida.

Para cumprir ainda mais os padrões legais e éticos na aplicação da IA, nossos produtos fornecem controles de acesso com base nas credenciais de autenticação do usuário e nas permissões de acesso. Isso permite que nossos clientes garantam a conformidade do usuário com políticas operacionais.

Sobre a Axis Communications

A Axis viabiliza um mundo mais inteligente e seguro, criando soluções que melhoram a segurança e o desempenho empresarial. Como uma empresa de tecnologia em rede e líder do setor, a Axis oferece soluções para sistemas de videomonitoramento, controle de acesso, interfone e áudio. Esses sistemas são aprimorados por meio de aplicativos de analíticos inteligentes e apoiados por treinamentos de alta qualidade.

A Axis conta com cerca de 4.000 funcionários dedicados, em mais de 50 países, e colabora com parceiros de tecnologia e integração de sistemas em todo o mundo para oferecer soluções aos clientes. A Axis foi fundada em 1984 e está sediada em Lund, na Suécia.