

WHITEPAPER

Das Potenzial von Szenen-Metadaten ausschöpfen

Mehr Situationsbewusstsein, Effizienz und Einblicke

Januar 2024

Zusammenfassung

Im Kontext der Videosicherheit liefern Metadaten eine Textbeschreibung des Inhalts eines Videos. Sie können etwa beschreiben, welche Objekte sichtbar sind, oder die Szene selbst allgemein beschreiben. Dies kann auch Attribute der Objekte oder der Szene umfassen, wie die Farben von Fahrzeugen und Kleidung, ihre genaue Position oder Fahrtrichtung. Die Metadaten werden in Echtzeit entweder direkt in der Kamera oder von einer anderen Komponente im System erzeugt, die Videos analysieren kann.

Metadaten liefern außerdem Kontext zu Ereignissen und ermöglichen das schnelle Sortieren und Durchsuchen großer Datenmengen. Dies ermöglicht Funktionen, die sich grob in drei Kategorien einteilen lassen:

- **Forensische Suche nach einem Ereignis.** Um anhand verschiedener Suchparameter nach Objekten oder Ereignissen zu suchen, die Ihre Suche auf eine geringe Zahl möglicher Kandidaten eingrenzen. Objektklassifizierungsdaten ermöglichen eine Suche nach vielen verschiedenen Details.
- **Echtzeitnutzung.** Um dem Betreiber zu helfen, schnell auf geänderte Situationen zu reagieren, oder um Input als Entscheidungsgrundlage zu bieten oder eine automatisierte Reaktion zu ermöglichen.
- **Identifizierung von Trends, Mustern und Erkenntnissen.** Plattformen für das IoT und betriebliche Effizienz für statistische Berichte können auf Metadaten für das Zählen von Besuchern, Geschwindigkeitsmessung, Daten zum Verkehrsfluss und andere Arten automatisch gesammelter Daten zurückgreifen.

Manche Kameras können Audio decodieren und Audio-Metadaten abrufen. Spezifische Schallmuster können auf ähnliche Weise erkannt und markiert werden wie Objektklassen in einem Video. Ein Audioerkennungssystem kann zum Beispiel verbale Angriffe oder zerbrechendes Glas erkennen.

Werden Metadaten aus verschiedenen Datenquellen kombiniert, z. B. aus visuellen, Audio-, aktivitäts- und prozessbezogenen Quellen, erhalten Sie viel mehr Einblicke als aus jeder einzelnen Datenquelle. Offene Protokolle und Branchenstandards sind für eine nahtlose Integration von Metadaten unverzichtbar.

Inhalt

1	Einführung	4
2	Was sind Metadaten?	4
3	Erzeugung von Metadaten „at the edge“	4
4	Anwendungsbeispiele	5
	4.1 Echtzeitnutzung für schnelle Reaktion	5
	4.2 Forensische Suche	5
	4.3 Identifizierung von aussagekräftigen Trends und Mustern	6
5	Wo werden Metadaten verwendet?	6
6	Wie werden die Metadaten bereitgestellt?	7
7	Audio-Metadaten	9
8	Kombination von Metadaten aus unterschiedlichen Quellen	9

1 Einführung

Metadaten liefern die Grundlage, um Informationen aus einem Video extrahieren zu können. Sie weisen den Videoinhalten eine digitale Bedeutung zu, indem sie die wichtigsten Details in der Szene beschreiben. Mithilfe der Metadaten können Sie in großen Videomengen schnell die wichtigsten Informationen finden, auswerten und entsprechend handeln. Deshalb sind Metadaten inzwischen zu einem wichtigen Teil effizienter Sicherheit und Geschäftsabläufe geworden.

Dieses Whitepaper zeigt, wie Metadaten im Sicherheitskontext und auch im Hinblick auf betriebliche Effizienz eingesetzt werden können. Es beschreibt die Vorteile von Metadaten und wie sie in Video Management Systemen und anderen Anwendungen eingesetzt werden.

2 Was sind Metadaten?

Metadaten sind Daten über andere Daten. Im Zusammenhang mit der Videoüberwachung beschreiben Metadaten in Textform, was auf dem Video zu sehen ist, z. B. welche Objekte von Interesse sichtbar sind, oder eine ausführliche Beschreibung der Szene selbst. Dazu können Merkmale gehören, die mit den Objekten oder der Szene assoziiert sind, z. B. die Farben von Fahrzeugen und Kleidung, der genaue Standort oder die Fahrtrichtung. Die Metadaten werden in Echtzeit erzeugt, entweder direkt in der Kamera oder von einer anderen Komponente im System, die Videos analysieren kann.

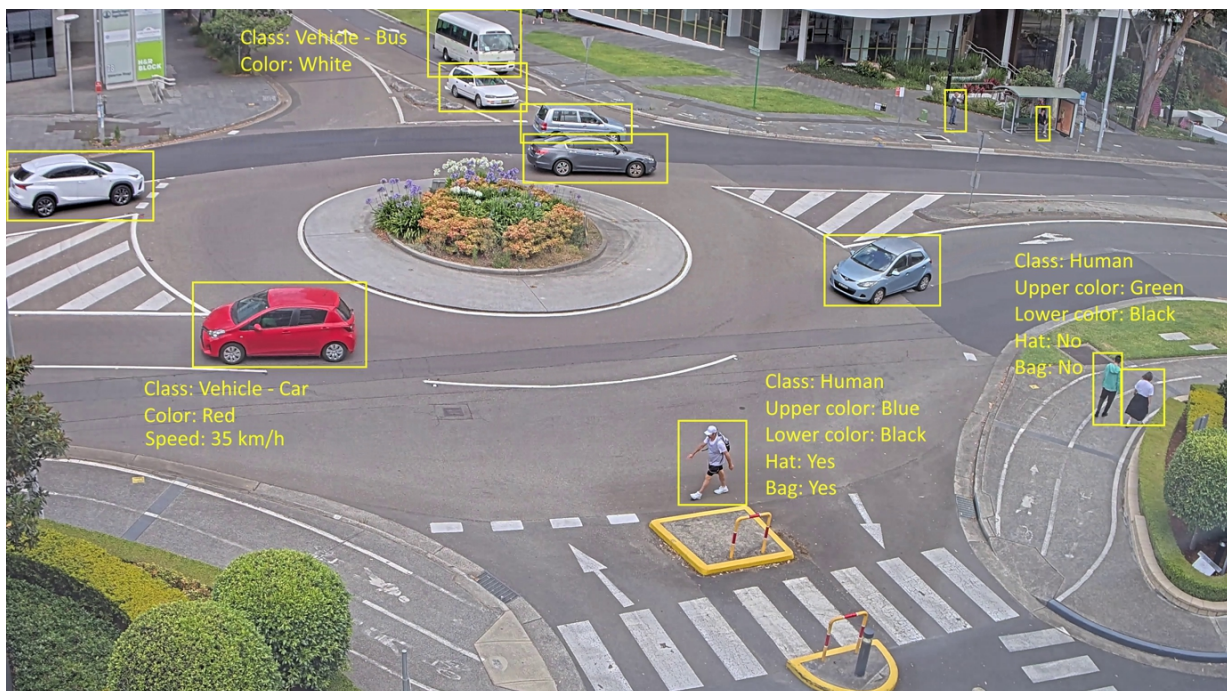


Figure 1. Beispiel eines Videobilds, in dem Objekte erkannt und analysiert werden, um Metadaten zu erzeugen.

3 Erzeugung von Metadaten „at the edge“

Leistungsstarke Videoanalysen waren bisher serverbasiert, da sie in der Regel mehr Rechenleistung benötigten, als die Videoanalyse direkt im Edge-Gerät bieten kann. Doch dank der Entwicklung neuer Algorithmen und der Erhöhung der Rechenleistung in den Edge-Geräten selber wurde es in den letzten Jahren möglich, moderne Analysefunktionen direkt in diesen Geräten auszuführen. Die Metadaten werden also direkt im Gerät erzeugt und können dort sofort von weiteren Analysefunktionen genutzt werden. Der

Videostream und Metadatenstream kann außerdem zur weiteren Verarbeitung an das VMS oder eine andere Anwendung weitergeleitet werden.

Edge-basierte Analysefunktionen haben mit sehr geringer Latenz Zugriff auf unkomprimiertes Videomaterial. Dies ermöglicht schnelle Echtzeitanwendungen ohne die zusätzlichen Kosten und die Komplexität, die bei der Verschiebung aller Videos zur Verarbeitung an einen anderen Punkt des Systems entstehen würden. Edge-basierte Analytik ist außerdem mit weniger Hardware- und Deployment-Kosten verbunden, da das System weniger Serverressourcen erfordert.

Beim Generieren von Metadaten „at the edge“ werden Daten aus dem Video extrahiert, ohne dass Daten durch Komprimierung oder bei der Übertragung verloren gehen. Dies ermöglicht spezifischere Metadaten und eine genauere Analyse der Videoinhalte. Je besser die Bildqualität, desto besser sind auch die Metadaten.

4 Anwendungsbeispiele

Metadaten liefern nicht nur Details zu Objekten in einer Szene. Sie bieten einen Kontext zu den Ereignissen und ermöglichen die schnelle Sortierung und Suche in großen Mengen von Filmmaterial. Dies macht Funktionen möglich, die grob in die Kategorien forensische Suche nach einem Ereignis, Echtzeitanwendungen und Identifizierung von Trends, Mustern und Einblicken eingeteilt werden können.

4.1 Echtzeitnutzung für schnelle Reaktion

Metadaten können in Echtzeit verwendet werden, um den Betreibern zu helfen, schnell auf Veränderungen der Situation zu reagieren. Sie können auch wertvolle Hinweise für die Entscheidungsfindung oder für eine automatische Reaktion liefern. Mit einer Edge-Analyse in Echtzeit, die mit hochwertigen Metadaten arbeitet, können Sie Personen, Standorte und Gebäude vor absichtlichem oder versehentlichem Schaden bewahren. Sie können Bedrohungen schnell erkennen, überprüfen und evaluieren, um sie dann effizient auszuräumen.

4.2 Forensische Suche

Metadaten ermöglichen eine effiziente und schnelle Suche nach Objekten oder Ereignissen. Dies kann Ermittlern eine Menge Arbeit ersparen, besonders, wenn viele Stunden Video aus mehreren Kameras durchsucht werden müssen. Sie können anhand von verschiedenen Suchparametern nach Objekten wie Personen oder Fahrzeugen filtern und brauchen dann nur eine kleine Anzahl von Treffern zu durchsuchen. Diese Suchparameter können beispielsweise Bewegung, Zeit oder Objektmerkmale sein.

Tabelle 4.1 Verschiedene Metadatenkategorien erleichtern die Suche.

Metadatenkategorie	Was wird erkannt?	Beispiele
Bewegung	Wie sich ein Objekt bewegt	Richtung, Geschwindigkeit, sonstiges Verhalten
Zeit	Wann ein Objekt erscheint	Wochentag, Tageszeit, Aufenthaltsdauer
Standort	Wo sich das Objekt befindet	Ort, Sichtfeld der Kamera

Tabelle 4.1. Verschiedene Metadatenkategorien erleichtern die Suche. (Fortsetzung)

Metadatenkategorie	Was wird erkannt?	Beispiele
Objektklassifizierung	Um welche Art von Objekt es sich handelt	Personen, Fahrzeuge (Auto, Bus, Lkw, Fahrrad/Motorrad)
Objekteigenschaften	Welche Merkmale es hat	Kleidung, Accessoires wie Hüte oder Taschen, physikalische Merkmale wie Kleidungsfarbe

Auch wenn Sie nur auf eine Kategorie von Metadaten Zugriff haben, wie etwa die Zeit, kann diese den entscheidenden Hinweis geben.

Metadaten zur Bewegung ermöglichen eine Suche anhand der relativen Geschwindigkeit und Bewegungsrichtung des Objekts. Objektklassifizierungsdaten ermöglichen eine Suche nach verschiedenen Details. Kameras mit Deep Learning Processing Unit (DLPU) können in der Regel erweiterte Metadaten mit einer genaueren Objektklassifizierung liefern, so dass Sie beispielsweise nach einem grünen Lkw oder einer Person mit einem blauen Mantel suchen können.

4.3 Identifizierung von aussagekräftigen Trends und Mustern

Plattformen für das IoT und betriebliche Effizienz für statistische Berichte können auf Metadaten für das Zählen von Besuchern, Geschwindigkeitsmessung, Daten zum Verkehrsfluss und andere Arten automatisch gesammelter Daten zurückgreifen. Die Daten werden analysiert, um aussagekräftige Erkenntnisse zu gewinnen.

5 Wo werden Metadaten verwendet?

Metadaten sind sehr wertvoll. Sie helfen unter anderem, die Merkmale und Inhalte einer Szene zu verstehen. Die wichtigsten Nutzer von Metadaten können in folgende Kategorien eingeteilt werden.

Edge-Anwendungen: In der Kamera ausgeführte Analysefunktionen können logische Filter und Regeln auf die Informationen zum Objekt in der Szene anwenden. Danach können sie anhand vordefinierter Schwellenwerte oder bestimmter Verhaltensweisen verschiedene Aktionen auslösen. Beispielsweise können sie eine PTZ-Kamera bewegen, sobald in der Szene eine Person in Bewegung erkannt wird.

Video Management System (VMS): Im Kontext der Videosicherheit wurden Metadaten bisher oft in einem VMS angewendet, um visuelle Overlays zu eventuell wichtigen Objekten in der Szene zu legen. Mit der Entwicklung verbesserter Algorithmen zur Erkennung und Klassifizierung von Objekten können Objekte jetzt auch anhand bestimmter Merkmale wie der Kleidungsfarbe gesucht werden. Dank der Möglichkeit von Suchanfragen anhand dieser Datenpunkte müssen nicht mehr viele Stunden Material manuell durchsucht werden.

IoT-Plattformen: Metadaten können gesammelt und visuell in Business-Intelligence-Plattformen präsentiert werden, um Echtzeit- und historische Trends zu analysieren und dadurch nützliche Erkenntnisse gewinnen zu können. Statistische Analysen auf der Grundlage von Kundenbewegungen oder Kundenerfahrungsdaten ermöglichen datengestützte Entscheidungen zur Verbesserung der Betriebsabläufe.

Zweite Analyseschicht: Manche Anwendungen erfordern eine Kombination aus Edge-basierter und Server-basierter Verarbeitung, um erweiterte Analysen anzustellen. Dann kann die Vorverarbeitung in der Kamera und die weitere Verarbeitung auf einem Server erfolgen. Ein solches Hybridsystem kann

eine kosteneffiziente Skalierung der Analysefunktionen ermöglichen, indem es nur relevante Video- und Metadaten an den Server überträgt.

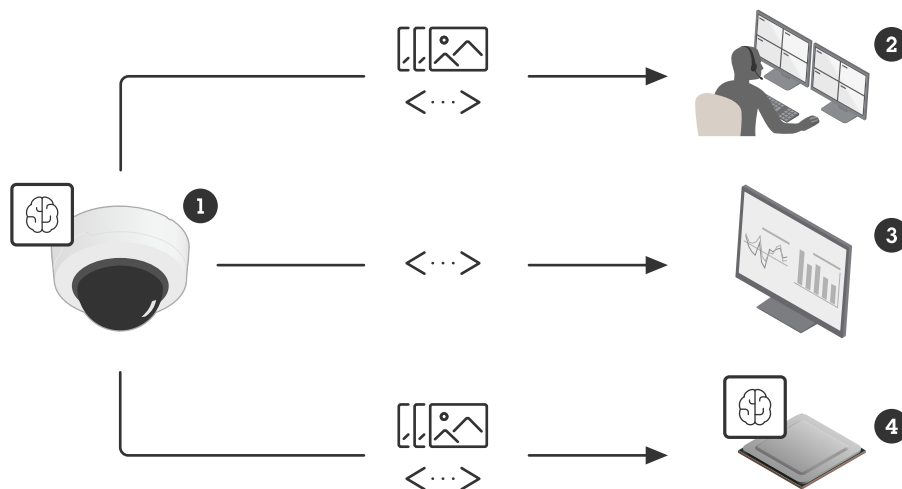


Figure 2. Nutzer von Metadaten

- 1 Edge-Anwendungen
- 2 VMS
- 3 IoT-Plattformen
- 4 Zweite Analyseschicht

6 Wie werden die Metadaten bereitgestellt?

Die erzeugten Metadaten können je nach Verwendungszweck auf unterschiedliche Weise bereitgestellt werden. Bei Echtzeitanwendungen müssen die Metadaten ständig auf Abruf bereitstehen, weil dies für eine angemessene Reaktion und den Überblick über die aktuelle Situation unverzichtbar ist. In anderen, weniger kritischen Anwendungen, bei denen kein sofortiges Handeln erforderlich ist, können die Metadaten vor der

Weiterleitung an den Nutzer weiter konsolidiert werden, etwa anhand des Wegs jedes Einzelobjekts in der Szene. Dies reduziert die Menge der Daten, die insgesamt gespeichert und verarbeitet werden müssen.

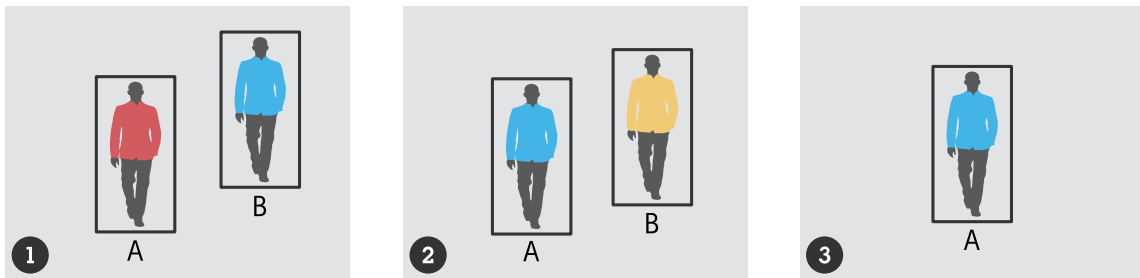


Figure 3. Diese Abbildung veranschaulicht das Streaming von Metadaten. Metadatenframes von der Kamera liefern ständig in Echtzeit Informationen zur Szene. Jedes Bild erfasst die Szene in einem bestimmten Moment, unabhängig von vergangenen Ereignissen.

- 1 Bild 1 erkennt die Objekte A und B und klassifiziert A als Mensch in roter Kleidung, B als Mensch in blauer Kleidung.
- 2 In Bild 2 aktualisiert die Kamera die Klassifizierung und stellt fest, dass Objekt A eigentlich blaue und Objekt B gelbe Kleidung trägt. Die Objekte sind zwar dieselben wie in Bild 1, aber ihre Farbattribute ändern sich, was von den Metadaten wiedergegeben wird.
- 3 In Bild 3 fehlt Objekt B, die Kamera verfolgt nur Objekt A, das weiterhin als Mensch in blauer Kleidung klassifiziert ist.

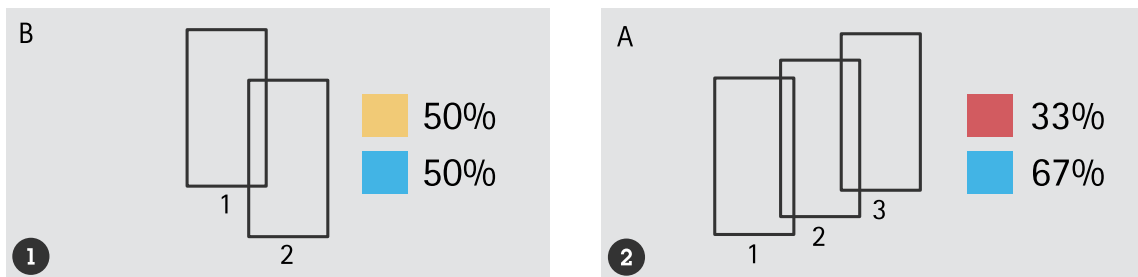


Figure 4. Diese Abbildung zeigt die Bereitstellung konsolidierter Metadaten. Dabei liefert die Kamera Informationen in einem vereinheitlichten Format, basierend auf dem erkannten Weg der Objekte in der Szene. Die Bilder zu jedem Objekt sind während der gesamten Sichtbarkeit des Objekts mit allen bekannten Details versehen.

- 1 Das erste Bild liefert Details zu Objekt B, einschließlich der Punkte, an denen es zuerst und zuletzt erkannt wurde, einer Zusammenfassung ihres Wegs sowie den in diesem Zeitraum erkannten Attributen. Objekt B trug mit 50 Prozent Wahrscheinlichkeit gelbe und mit 50 Prozent Wahrscheinlichkeit blaue Kleidung.
- 2 Das zweite Bild spiegelt dieses Format für Objekt A und zeigt eine Wahrscheinlichkeit von 33 Prozent für rote und 67 Prozent für blaue Kleidung.

Das konsolidierte Verfahren hat den Vorteil, dass die Kamera deutlich weniger Daten an den Nutzer überträgt, weil sie nur dann Metadaten liefert, wenn sich Objekte in der Szene befinden, und diese dann zur leichteren Interpretation zusammenfasst (konsolidiert). Das Streamingverfahren liefert in jedem Bild eine komplette Beschreibung der Szene, auch wenn keine Aktivitäten oder Objekte vorhanden sind, so dass der Nutzer diese Daten nach dem jeweiligen Bedarf analysieren muss. Das Streamingverfahren eignet sich wie erwähnt besonders für Echtzeitanwendungen, das konsolidierte Verfahren ist hingegen ideal für die Nachbearbeitung, wenn der Nutzer nicht sofort eingreifen muss.

Das Wissen um die Stärken und Schwächen der beiden Strategien ist beim Aufstellen der Systemarchitektur entscheidend. So wäre bei einer IoT-Plattform, die basierend auf den Metadaten Erkenntnisse generiert, eine spätere Zusammenfassung der in der Szene vorhandenen Objekte hilfreich, da diese Services meist nur über begrenzte Bandbreite und Speicherplatz verfügen.

Die Metadaten könnten außerdem über unterschiedliche Kommunikationsprotokolle und in unterschiedlichen Dateiformaten geliefert werden, je nach Bedarf und Vorlieben des jeweiligen Nutzers.

7 Audio-Metadaten

Manche Kameras können Audio decodieren und Audio-Metadaten abrufen. Audioerkennungsanalytik kann Klangmuster erkennen und bestimmte Klänge live und in Audioaufzeichnungen hervorheben. Auf diese Weise können Audioerkennungssysteme in Verbindung mit Videosicherheitsvorrichtungen den Bediener auf möglicherweise gerade stattfindende Vorfälle hinweisen und gezielt zur entscheidenden Kameraansicht leiten. Das System könnte zum Beispiel Streitereien erkennen, um eine Eskalation und Schlägereien zu verhindern, zerbrechendes Glas erkennen, um Einbrüche zu verhindern oder eine Frühwarnung auslösen, falls sich der Zustand eines Patienten verschlechtert. Indem sie den Betreiber nicht nur sehen, sondern auch hören lassen, was in einer Szene geschieht, können Klangerkennungssysteme eine frühzeitige Erkennung und schnelles Eingreifen ermöglichen und in vielen Fällen sogar eine weitere Eskalation verhindern helfen. Die Klangerkennung kann auch als zweite Verifizierungsmöglichkeit eingesetzt werden.

Analysefunktionen, die auf die Erkennung von Schallmustern trainiert sind, suchen meist nach einer Kombination aus Merkmalen wie die Lautstärke in Dezibel kombiniert mit der Energie unterschiedlicher Frequenzen im Laufe der Zeit. Spezifische Schallmuster können auf ähnliche Weise erkannt und markiert werden wie Objektklassen in einem Video.

8 Kombination von Metadaten aus unterschiedlichen Quellen

Das wahre Potenzial von Metadaten kommt zum Tragen, wenn sie auf mehrere Eingaben angewendet werden, z. B. auf visuelle, akustische, tätigkeitsbezogene und prozessbezogene Eingaben. Datenquellen wie RFID-Tracking, GPS-Koordinaten, Manipulationsalarme, Ableseergebnisse (wie z. B. Temperatur oder Chemikalienkonzentration), Geräuscherkennung und Transaktionsdaten an Kassenterminals sind für die Verwaltung jedes Standortes wertvoll. Die Daten aus allen Quellen können nach Zeitstempel sortiert werden.

Durch die Kombination der Metadaten aus unterschiedlichen Quellen lassen sich viel mehr Erkenntnisse gewinnen als aus einer einzelnen Quelle. Offene Protokolle und Branchenstandards sind für eine nahtlose Integration von Metadaten unverzichtbar.

Über Axis Communications

Axis ermöglicht eine intelligente und sichere Welt durch Lösungen zur Verbesserung der Sicherheit und Geschäftsperformance. Als Unternehmen für Netzwerktechnologie und Branchenführer bietet Axis Lösungen in den Bereichen Videosicherheit, Zutrittskontrolle sowie Intercoms und Audiosysteme. Sie werden verstärkt durch intelligente Analyseanwendungen und unterstützt durch gute Schulungen.

Axis beschäftigt rund 4.000 engagierte Mitarbeiter in über 50 Ländern und arbeitet weltweit mit Technologie- und Systemintegrationspartnern zusammen, um den Kunden Lösungen anbieten zu können. Axis wurde 1984 gegründet und der Hauptsitz befindet sich in Lund, Schweden