

白皮书

面向安防和安全的音频分析工具

二月 2024

概述

面向安防和安全的音频分析工具能够侦测声音模式，并突出实时音频中的异常声音。这些分析工具能够（比如）识别攻击性语言以防止事件升级和袭击事件，侦测玻璃破碎声以防止闯入，或者提供险境人员预警。

利用不同类型的传感器，比如振动和音频传感器（摄像机和麦克风），能够提高侦测结果的可靠性，提供更有助于行动决策的分析数据。

AXIS Audio Analytics是一个软件平台，供集成在安讯士设备软件中的基于前端的分析工具使用。这些分析工具在不保存原始音频流的情况下捕捉并侦测声音。这样就能够高效保护隐私安全，因为这些分析工具是基于前端的，并且提供的是音频元数据。

目录

1	引言	4
2	技术	4
2.1	侦测声音事件	4
2.2	捕捉并处理音频	4
2.3	基于前端或基于服务器的分析工具	5
2.4	元数据	6
3	AXIS Audio Analytics	6
3.1	隐私	7
3.2	应用场合	8
4	法律限制	11
5	免责声明	11

1 引言

面向安防和安全的音频分析工具能够侦测声音模式，并突出实时音频中的异常声音。这些分析工具能够（比如）识别与打斗相关的声音、侦测玻璃破碎声、或提供险境人员预警。

在与视频监控相结合的情况下，音频分析工具能够向操作人员警告正在发生的潜在事件，并引导他们查看相关的摄像机画面。这可有助于提早侦测、快速干预，并且在许多情况下，有助于防止事态进一步升级。

本白皮书介绍了音频分析工具可如何应用于安防和安全领域。我们讨论音频的捕捉和处理技术，并着重介绍实时前端分析，比如直接在摄像机或麦克风中进行的基于AI的声音分类。我们还展示前端的音频分析工具如何以多种不同的方式通过音频元数据来保护隐私安全。

本白皮书不提供法律建议。在安装任何监控系统之前，您需要研究当地以及具体的应用场合适用哪些法律法规。系统所有者应负责确保符合当地法律、法规和建议。

2 技术

2.1 侦测声音事件

一个声音事件即一个音频片段，人类通常将其识别为某个特殊的概念，例如，尖叫或玻璃破碎的概念。这些类型的概念声音可以被侦测到并进行标记，如同在视频分析工具中侦测和标记对象类别一样。

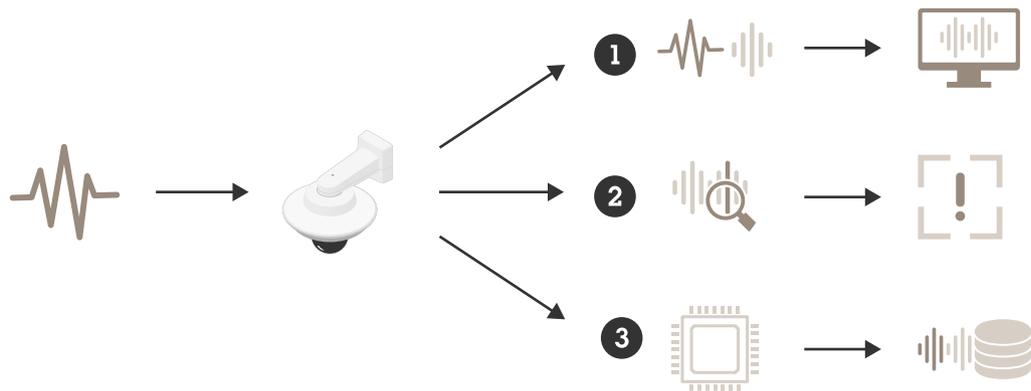
经过训练来识别声音模式的分析工具通常可以监听一系列特征，包括从分贝级别到随时间变化的不同频率上的能量。在侦测到特定声音模式后，系统可以通过视觉警报或通过触发报警的方式，向工作人员发送自动通知。

如果使用基于AI的算法，则可以通过大量的数据来训练这些工具。例如，在以数千次的相应声音进行训练之后，算法能够可靠侦测人类尖叫声。

2.2 捕捉并处理音频

音频分析工具使用所捕捉的音频数据，并分析相关的声音特征，从而生成非音频输出。捕捉音频基本上意味着，对音频进行数字化处理，并使其可在软件中使用。其具体实现方式是，使用麦克风拾取空气中的声音振动，将这些模拟信号转换为数字信号，然后将这些信号传送到处理单元。如果所捕捉的音频未置于闪存、硬盘等永久存储介质上，那么便不会记录这些音频。在安讯士摄像机中，默认关闭音频的流传输和记录功能。

在初始音频捕捉之后，需要对捕捉的信息进行预处理，以便再进行后续处理。不同的预处理可以同时或单独进行。



音频预处理

1 转换

2 实时前端分析

3 处理和编码以便进行流传输或存储——如果使用安讯士摄像机，除非主动开启音频流传输功能，否则便不会对音频进行流传输或存储。

- **转换。** 声音被提取并转换成（比如）视觉信息，以图形形式显示声谱。这个过程不可逆：无法再通过声谱图获取原始声音。

- **实时前端分析。**

如果在前端处理声音，可以使用**声音分类器**。这将得到描述声音特征的元数据。无法通过元数据再造原始声音。

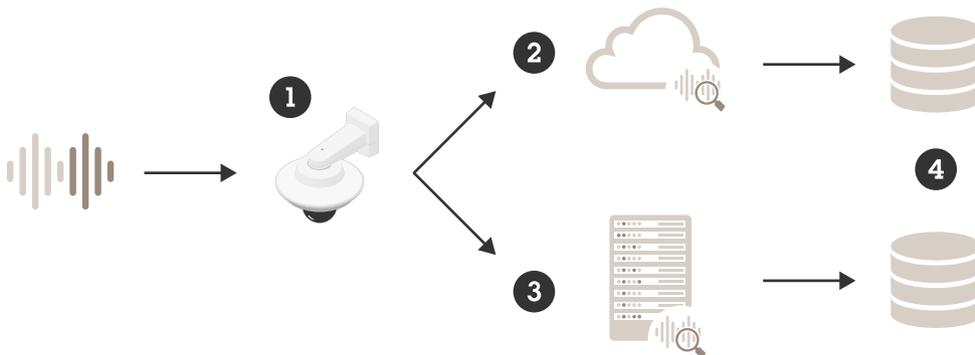
可以使用**声音检测器**来识别声音形式、声级或频率，并提供状态信息。它同样也无法还原原始声音。

- **处理和编码。** 如要使用原始（未经转换或分析的）音频，通常需要执行某些处理和编码，由此对音频数据进行预处理，以供预期应用场合之用。这些应用场合涉及在前端存储音频数据、将数据流传送到外部客户端以供（在服务器或云端）进一步处理、或者外部存储。在拥有安讯士摄像机的情况下，您需要先主动开启音频流传输功能，出于隐私保护原因（音频隐私管控），此功能为默认关闭状态。

2.3 基于前端或基于服务器的分析工具

出于多方面的原因，分析引擎在系统中的位置非常重要。尤其是在保护隐私和遵守有关个人数据的法律法规方面，它涉及到软件算法在哪里分析音频数据。有时，可能无法通过网络发送音频数据，那么在本地分析所捕捉（但未存储）的音频数据的就变得非常重要。如

果所用的算法需占用大量计算资源，进而导致其无法在前端运行，则可能需要将数字音频数据发送到云或服务器。



分析工具的潜在运行位置

1 前端

2 云端

3 服务器

4 存储——只有在激活音频流传输的情况下，才能存储原始音频。

- **前端分析。**当分析工具在前端运行时，音频流不需要离开摄像机。仅来自分析工具的输出（即，元数据或触发信息）会被送出摄像机。AXIS Audio Analytics内的分析工具是基于前端的分析工具。
- **服务器分析。**当分析工具在服务器上运行时，需要将音频数据传出到摄像机之外。如果在摄像机中已进行预处理，这些数据就只能是经提取或去个性化处理的元数据。服务器通常是封闭式系统（由系统所有者控制）的组成部分，因此能够管理所传输音频的隐私保护。但也必须遵守相应的法律法规。
- **云端分析。**也可以将音频数据传输到云端服务器。跟服务器分析的情况一样，可以将音频信息预处理成元数据。云应用通常是分布式应用，因此需要更谨慎地对待隐私问题，确保符合相关法律法规。

2.4 元数据

音频分析工具可生成音频响度数据的恒定元数据流。它还能够根据分类器侦测到的事件以及自适应音频侦测来生成相关元数据。在前端运行的分析工具能够在摄像机内分析音频信息。它们不需要将实际音频流传输到别处——只能传输元数据，以供分析场景中的当前事态。

元数据流允许您查看音频包络，例如，在可以统一校验和查看音频事件和视频事件的面板中显示音频包络以供查看。这样，元数据就有助于高效快速地搜索特定事件和异常声音。这能够节省调查时间以及对大量影像进行搜索的时间。

3 AXIS Audio Analytics

AXIS Audio Analytics是一个软件平台，将持续引入新的和升级的分析工具和功能。分析工具集成在设备软件AXIS OS中，免费随附于摄像机和其他安讯士设备。

平台中的分析工具是基于前端的分析工具，算法直接在摄像机上运行。这就保证了良好的扩展能力、较少的数据传输以及出色的隐私保护。仅来自分析工具的输出（即，元数据或

触发信息) 会被保存。不会记录音频, 也不会将音频流传输到摄像机之外, 并且也无法通过元数据重新创建原始声音。

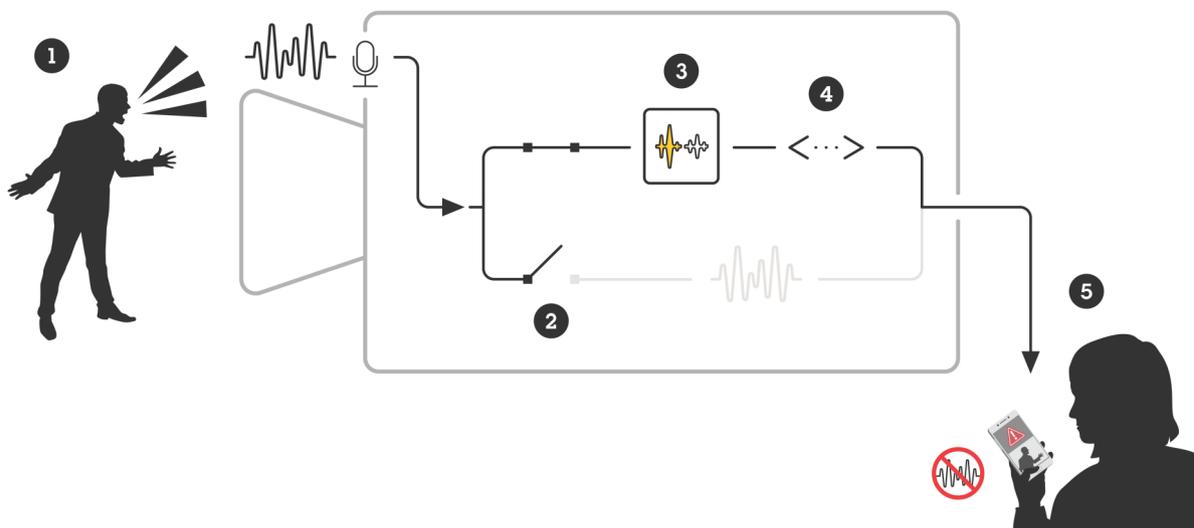
- **音频分类 (尖叫声和咆哮声)** 它是一种基于AI的声音分类器, 能够侦测并分析特定声音, 如尖叫声和咆哮声。未来将会加入更多的音频分类, 比如玻璃破碎声。通过分类分析工具的分析, 将得到描述声音特征的元数据。
- **自适应音频侦测。** 它是一种声音侦测器, 能够在音频响度突然变化时, 创建相关事件。它侦测不同类型的峰值声音, 能够根据环境噪音自行调整, 即使在噪音响度发生变化时, 也适应良好。

3.1 隐私

音频分析工具通常不记录传入的音频, 也不将音频传出到摄像机之外。它们仅处理声音, 以便在接收系统中搜索特定事件、模式或声音响度, 例如, 在显示面板中搜索以供后续调查, 或者在视频管理软件中搜索以便警告操作人员。音频数据无法被重建, 也不会记录私人谈话。这是因为, 这些分析工具是基于前端的, 并且提供的是音频元数据。

AXIS Audio Analytics既不会记录音频, 也不会流传输音频, 而是仅传输元数据。出于隐私保护原因, 安讯士设备中还默认关闭音频流传输功能 (音频隐私管控), 这就意味着, 音频既不会被流传输、记录, 也无法被重新创建。如有需要, 您可以开启音频流传输, 但即使在此功能关闭的情况下, 如果发生某些相关事件, 也会向您发出通知。

下图概括地显示AXIS Audio Analytics如何在与音频隐私管控相结合的情况下拾取声音并使用元数据来创建警告。



安讯士摄像机中的音频隐私管控和AXIS Audio Analytics, 从尖叫声 (1) 到警报 (5)。

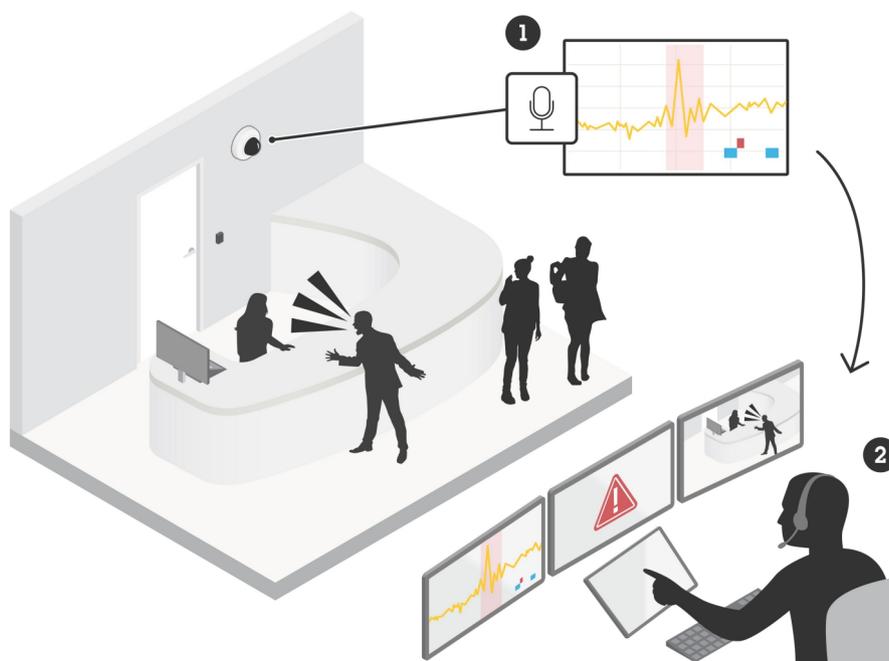
- 1 麦克风拾取声音。
- 2 音频流传输功能默认关闭。
- 3 AXIS Audio Analytics的音频分类器在传入的音频中侦测到尖叫声或咆哮声。
- 4 音频分类器生成元数据 (包括事件通知)。
- 5 利益相关者接收到基于事件通知和元数据的警报。他们可以通过核查视频流来进行确认。不会提供音频流。

3.2 应用场合

虽然基于AI的分析工具能够较好地筛除不相关的噪音，但在背景噪音非常多的情况下，它们也可能提供假警报。雨滴敲打窗的声音、雷声、鸣笛声、音乐或人来人往的繁忙场景都可能触发假警报。因此，它们的典型应用场合是安静区域，如银行和前台，但也可以是下班后的不同室内空间，如商店、餐厅、楼梯间或办公室。

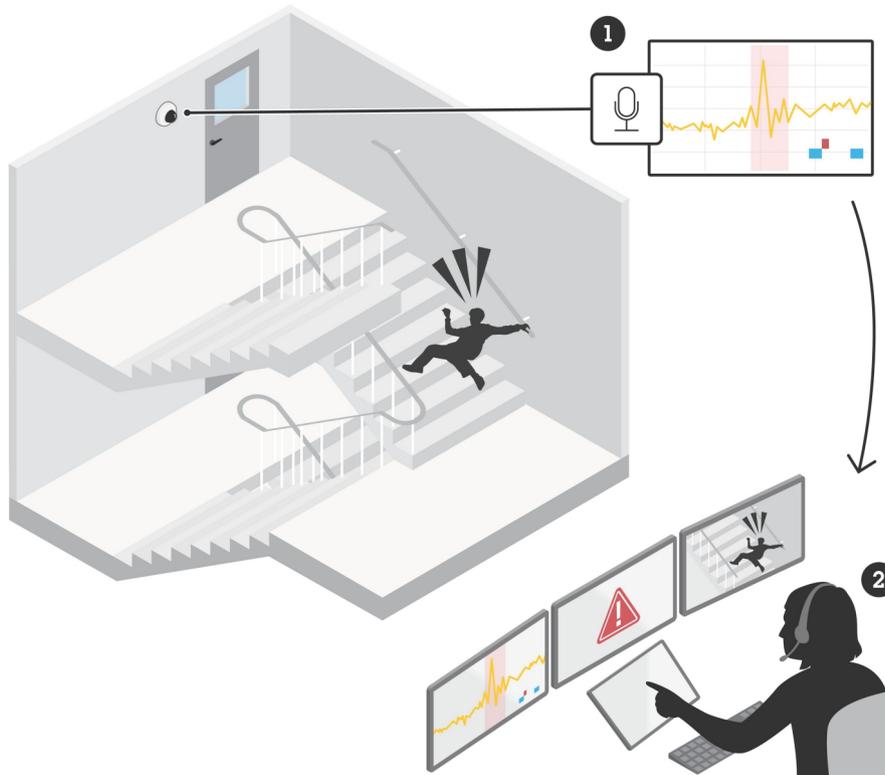
3.2.1 声音侦测与警报

在银行或前台，借助声音分类分析工具，可以监视相关区域，并侦测与打斗或玻璃破碎相关的声音。在侦测到这些声音后，系统的事件数据和音频元数据通过视觉警报或通过触发报警的方式，向工作人员发送自动通知。这能够提供预警，确保快速响应和干预。

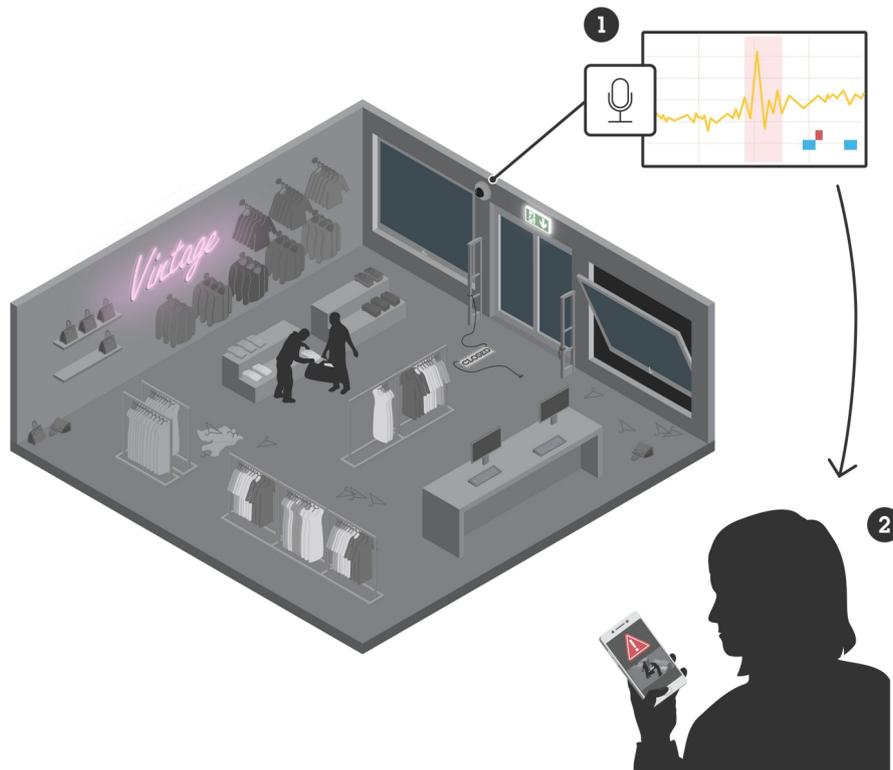


- 1 配备声音分类分析工具的摄像机侦测到前台的尖叫声或咆哮声。
- 2 操作人员收到警报，并能够核查视频流以进行确认，然后再相应地采取后续行动。

自适应音频侦测分析工具可用于在办公时间之外侦测异常声响。该分析工具能够分析环境声音，并在侦测到说话声、破窗声或其他突然且短暂的噪音时，做出响应。当侦测到相关事件时，该分析工具会转发元数据，向操作人员发出相应通知。



- 1 配备声音分类分析工具的摄像机在下班后侦测异常声响。
- 2 操作人员收到警报，并能够核查视频流以进行确认，然后再相应地采取后续行动。



- 1 配备声音分类分析工具的摄像机在下班后侦测到店内声响。
- 2 店主收到警报，并能够核查视频流以进行确认，然后再相应地采取后续行动。

3.2.2 与传感器相结合，进一步发掘监控系统的潜力

监控系统通常采用多种类型的传感器。当然，摄像机的图像传感器就是其中之一，它用于寄存场景视觉画面。此外，通常也会使用非视觉传感器，比如，基于雷达技术或红外辐射的移动侦测器。非视觉传感器能够增加其他类型的信息输入，成为摄像机系统的补充。

通过在监控系统中纳入音频传感器（麦克风），能够增强大多数应用场合的监控效果。在非音频系统中加入音频能力和音频分析工具，可实现多传感器交互。如果您已在使用视频分析工具，那么在加入音频分析工具后，则能够提高侦测可靠性。在视频分析工具因（比如）低照度条件而难以发挥作用的情形下，尤其如此。

您可以在（比如）视频管理软件中设置系统，使得仅在视频分析工具和音频分析工具均做出响应时，才触发相关操作。例如，音频分析工具检测到尖叫声，并且视频分析工具检测到摄像机视野中有人。在某些环境中，这种组合能够提供高效的安防保障。

3.2.3 将输入提供到显示面板

可以将音频元数据输入到分析工具面板，如VSaaS（视频监控即服务）、IoT（物联网）或商业信息平台，在那里集中显示这些元数据。这些面板会分析实时趋势和历史趋势，由此

生成即时概览和可执行分析数据。基于顾客流或客户体验的统计分析可以促使由数据驱动的决定，从而提升运营效益。

借助这些面板，您能够在不监听实际音频或者无法获取原始声音的情况下，查看结果。您能够通过（例如）计数事件，获取可执行分析数据，而且还能够很好地保障数据隐私。请注意，在音频的记录或仅捕捉方面，可能有不同的法律限制。

4 法律限制

许多人对视频监控中的麦克风使用存在担忧。这些担忧通常涉及语音和视频材料的记录。但在使用音频分析工具的情况下，通常不会记录或流传输任何音频。不同国家和地区的监控法律各不相同，因此，在视频监控系统中使用音频功能之前，务必要知道哪些是允许的。

出于某些原因，国家法律法规或各类地方法规条例可能会禁止捕捉和记录音频，或者要求满足特殊要求。尽管某个区域或环境可能允许音频捕捉，但它也可能禁止录音。企业也可能禁止在其场所范围内使用音频监控。

5 免责声明

本文档及其内容经由安讯士提供，与本文档或其中所涉及的任何知识产权（包括但不限于其中的商标、商业名称、徽标以及类似标志）有关的所有权利均受到法律保护，本文档或其中所涉及的任何知识产权中的或关联的所有权限、权利和/或权益都并且应都归属于Axis Communications AB。

请注意，本文档系“按原本”提供，不包含任何类型的保证，仅供参考之用。本文档中提供的信息不构成且不意在构成法律建议。本文档不意在构成且不应构成Axis Communications AB和/或其任何附属公司的任何法律义务。Axis Communications AB和/或其任何附属公司的与任何安讯士产品相关的义务仅遵从在安讯士与直接从安讯士购买此产品的实体之间的协议中所规定的条款和条件。

为避免疑问，与本文档的使用、结果和效用有关的一切风险均由本文档的使用者承担，安讯士在法律允许的限度范围内否认并排除一切保证，无论是法定的、明确的还是隐含的保证，这其中包括但不限于适销性、对特定用途的适用性、权益和非侵权方面的隐含保证以及产品责任、或者因与本文档相关的任何提议、规格指定或样本所致的任何保证。

关于 Axis Communications

Axis 通过打造解决方案，不断提供改善以提高安全性和业务绩效。作为网络技术公司和行业领导者，Axis 提供视频监控解决方案，访问控制、对讲以及音频系统的相关产品和服务。并通过智能分析应用实现增强，通过高品质培训提供支持。

Axis 在 50 多个国家/地区拥有约 4,000 名敬业的员工 并与全球的技术和系统集成合作伙伴合作 为客户带来解决方案。Axis 成立于 1984 年，总部在瑞典隆德