



A decentralized Privacy-sensitive Video Surveillance Framework

T.Senst, V.Eiselein, A. Badii, M. Einig, I. Keller, T. Sikora

Communication Systems Group | DSP 2013



Motivation

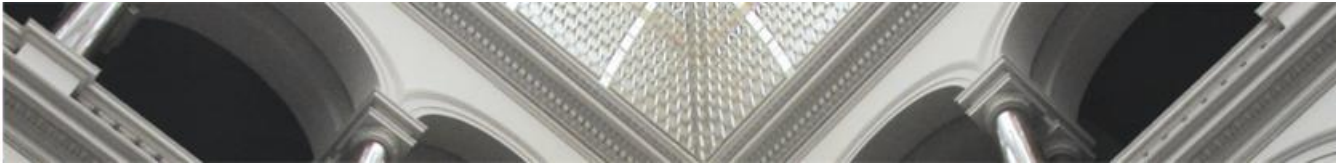
Safety, privacy and security needs are rapidly growing:

The compliance with a global standard will be a key of success of any surveillance product in the near future

ONVIF defines a common protocol for the exchange of information between IP-based video devices, including :

- automatic device discovery
- video streaming
- intelligence metadata

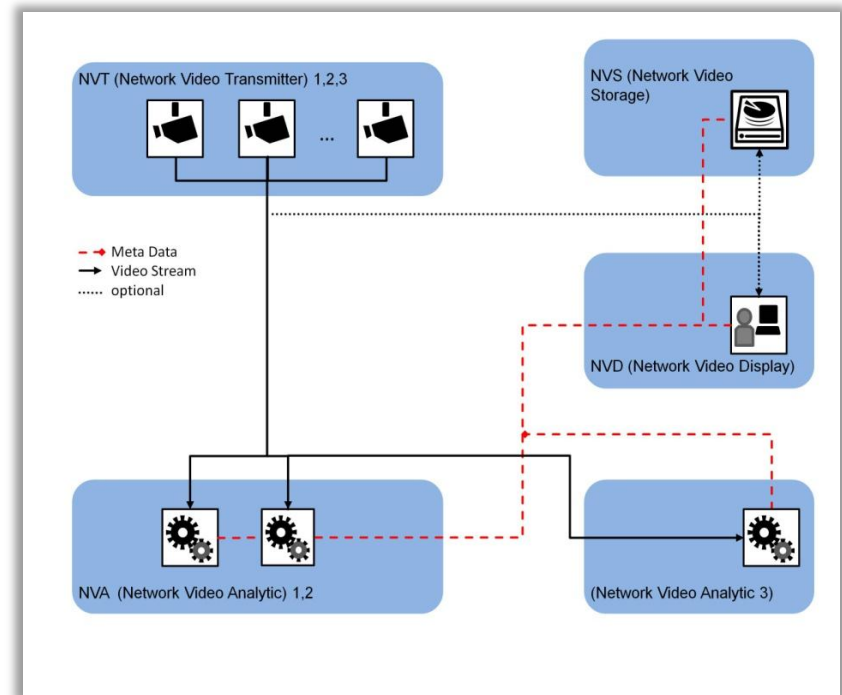
ONVIF uses web services and provides a formal conformance process



Proposed System

Analyze video sequences captured by IP video cameras and send information of the activity observed to network nodes which collect information and makes it available to the user interface.

Privacy protection is inherent to the system as the most-viewed scene representation is an abstract 3D model of the site which provides the user with a fast and comfortable overview of the current events.

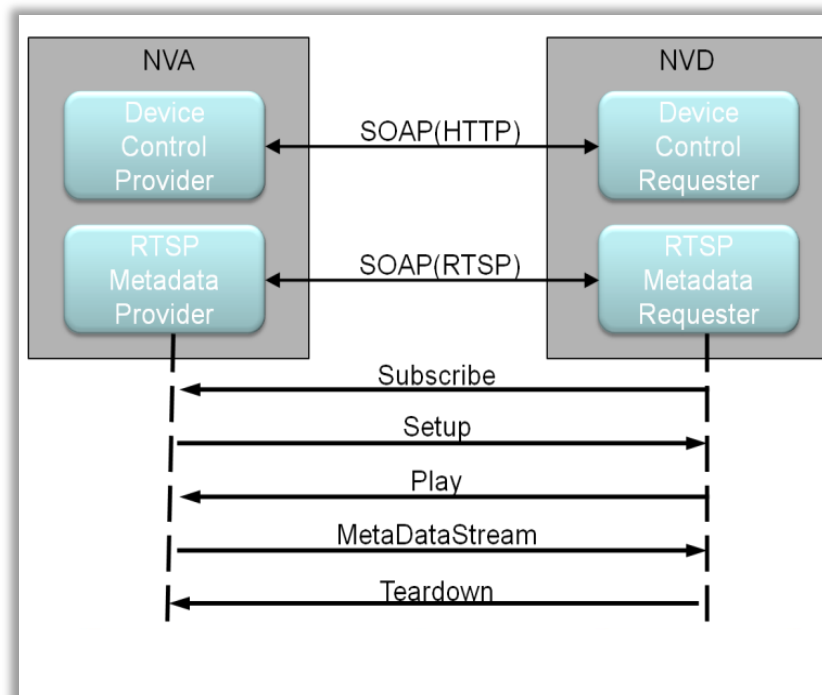




Communication

The proposed system uses two ways to transfer data between devices:

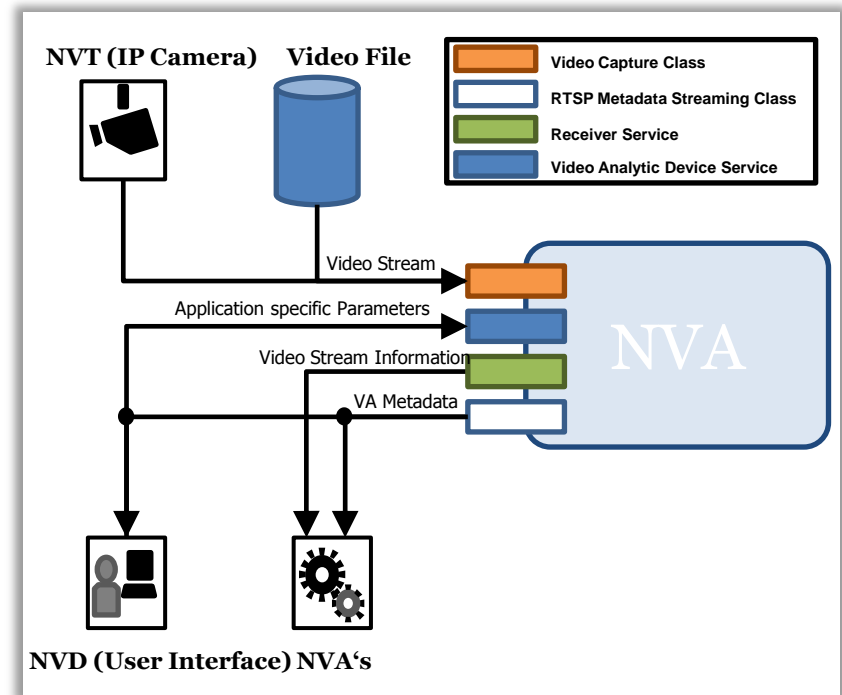
- Device control data such as configuration and parameterization is transmitted via web service interfaces. (HTTP/SOAP/ONVIF)
- Real-time capable data such as metadata from video analytics e.g. object trajectories and intrusion events. (RTSP/SOAP/ONVIF)
- Web service definition: WSDL
- Data structure definition: XSD
- Data structure and exchange: SOAP



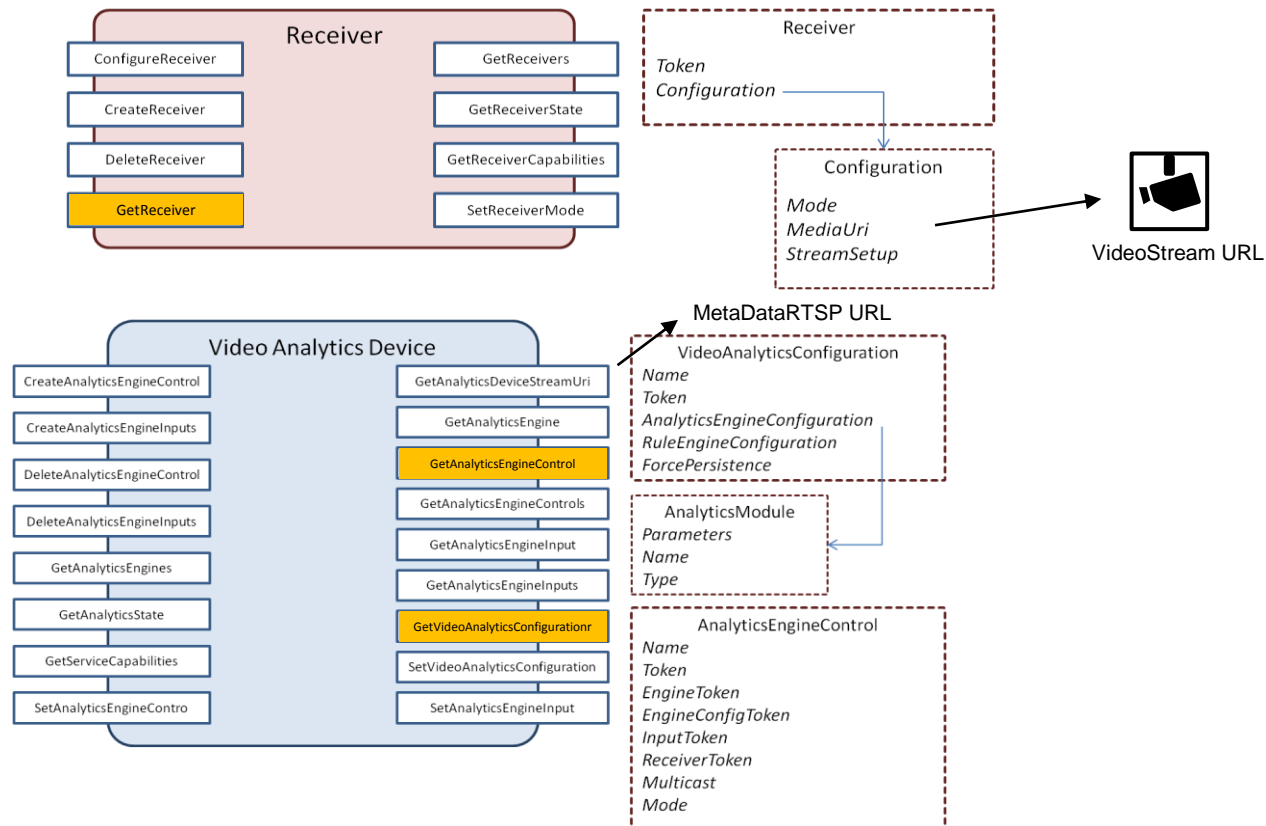
Communication

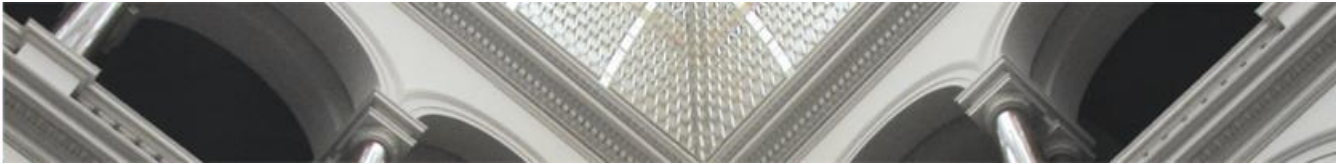
The proposed system uses two ways to transfer data between devices:

- Device control data such as configuration and parameterization is transmitted via web service interfaces. (HTTP/SOAP/ONVIF)
- Real-time capable data such as metadata from video analytics e.g. object trajectories and intrusion events. (RTSP/SOAP/ONVIF)
- Web service definition: WSDL
- Data structure definition: XSD
- Data structure and exchange: SOAP



Communication I / Configuration, Parametrization





Communication II / VA Metadata

Video analytic results are serialized using the ONVIF SOAP specification and encoded as XML string. The XML string is transmitted via the metadata streaming over RTSP.

ONVIF metadata description specifies two different kinds of analytic structures to assemble the analysis results: the EventStream type and the VideoAnalyticsStream type.

Metadata is assembled in a **Frame-** and **Object-**based manner. Each frame contains **Source** and **Time Identifier**. Each object contains a VA-unique identifier.

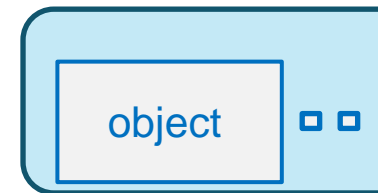
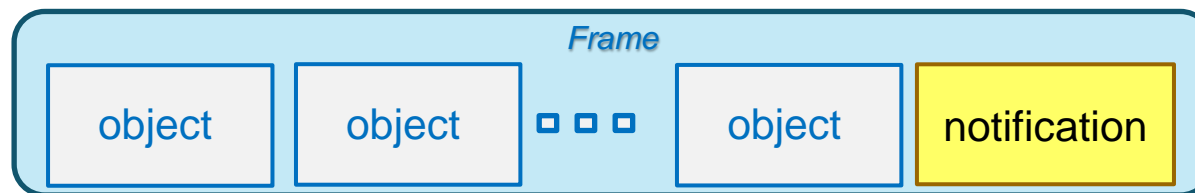
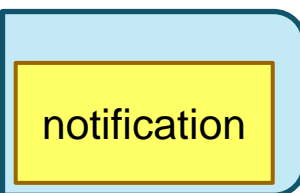
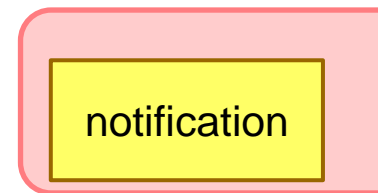
```
<?xml version="1.0" encoding="UTF-8"?>
<tt:Message xmlns:tt="http://www.onvif.org/ver10/schema" ...
  <tt:VideoAnalytics xsi:type="tt:VideoAnalyticsStream">
    <tt:Frame UtcTime="1970-01-01T01:00:00+01:00" xsi:type="tt:Frame">
      <tt:Object ObjectId="1" xsi:type="tt:Object">
        <tt:Appearance xsi:type="tt:Appearance">
          <tt:Shape xsi:type="tt:ShapeDescriptor">
            <BoundingBox bottom="19" left="32" right="74" top="19" ...
          </tt:Shape>
          <tt:Extension xmlns:sv="http://www.nue.org/schema" ...
            <3DPosition x="12" xsi:type="sv:Point3F" y=" 64" z="0"/></3DPosition>
          </tt:Extension>
        </tt:Appearance>
        <tt:Behaviour xsi:type="tt:Behaviour"></tt:Behaviour>
      </tt:Object>
      <tt:Extension xsi:type="sv:WsntNotificationFrameExtension">
        <wsnt:Topic
          Dialect="www.videosense.eu"xsi:type="wsnt:TopicExpressionType">
            vtp:PeopleTracking</wsnt:Topic>
          <sv:Source xsi:type="tt:ItemList">
            <tt:SimpleItem Name="ReceiverToken" Value="tkVS1" />
            <tt:SimpleItem Name="VideoAnalyticsConfigurationToken"
              Value="tkVA1"/>
          </sv:Source>
          <sv:UTCTime>2013-Jun-13 14:51:04.741887</sv:UTCTime>
        </tt:Extension>
      </tt:Frame>
    </tt:VideoAnalytics>
  </tt:Message>
```

Communication II / VA Metadata

Object
Identifier
3DPosition
2DBoundingBox
... Appearance

Notification
UTC timestamp
Topic
ReceiverToken
VideoAnalyticsConfigurationToken

Message
Keys
Data



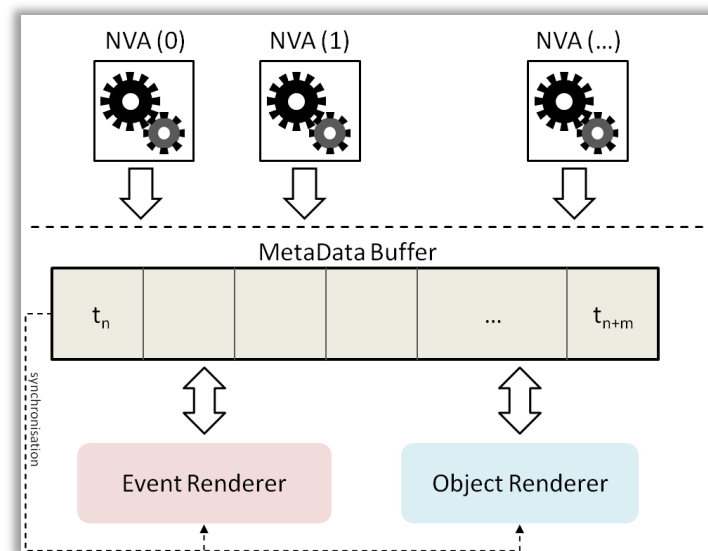


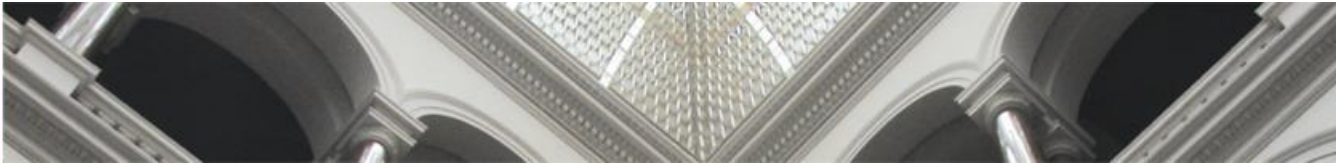
Communication II / VA Metadata

To transmit deterministic data as people tracks, these have to be disassembled by the NVA. Due to the RTSP and UDP real-time data streaming the reliability of the metadata transmission is not given

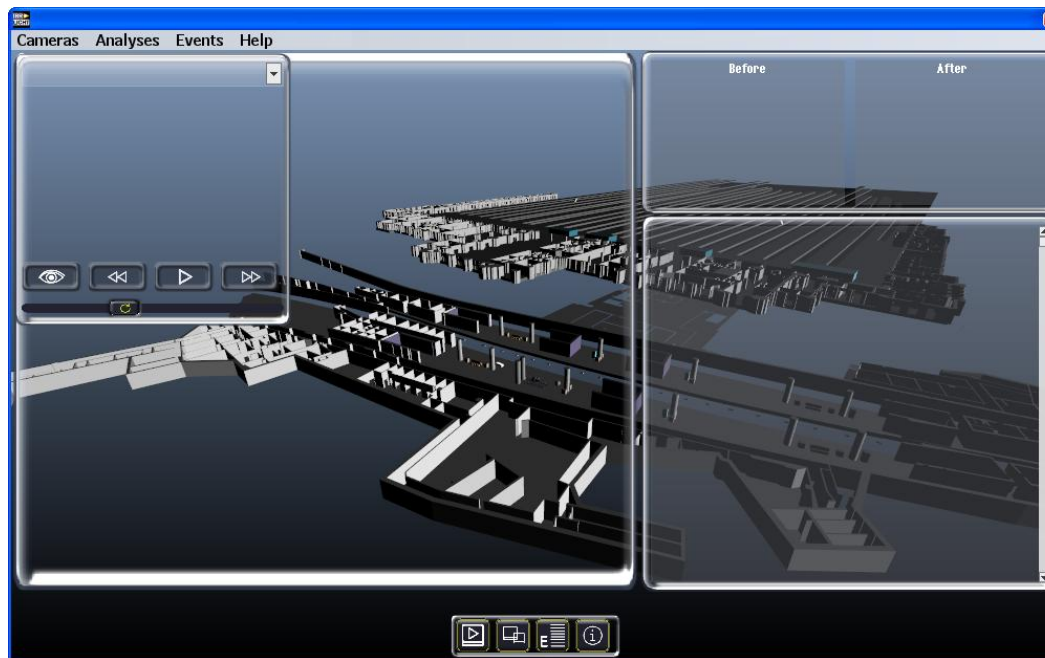
A central metadata buffer manages the connected NVAs. Tracks could be assembled by global identifiers composed of **Object-ID**, **ReceiverToken** and **VideoAnalyticsConfigurationToken**.

A User-role-based access model supports the overall privacy protection by applying renderer filters for specific sets of cameras or analytic results.

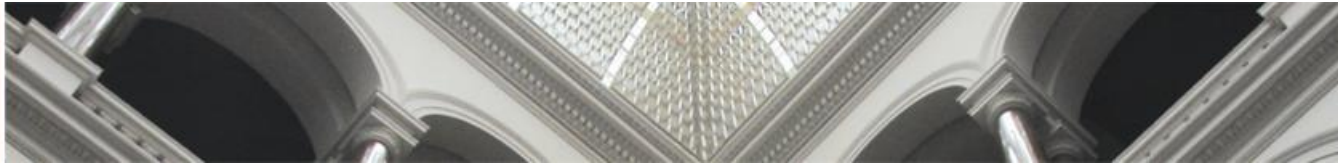




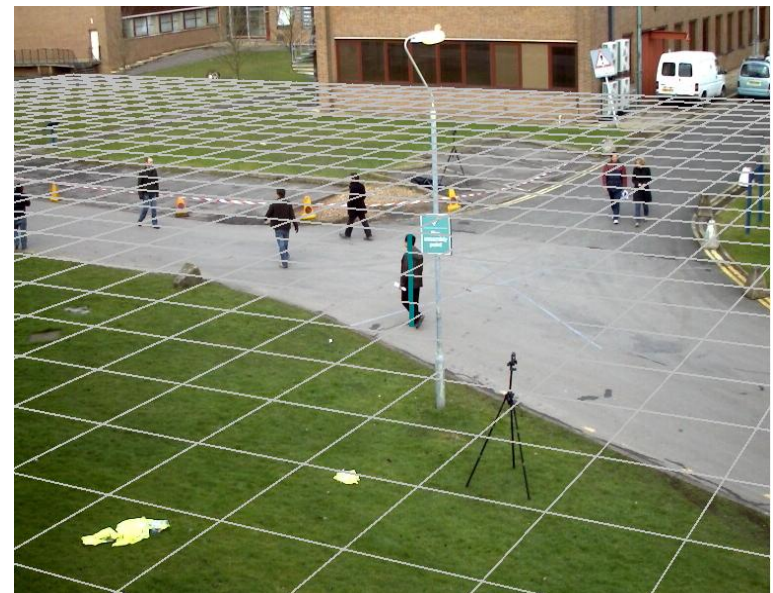
User Interface



- ✓ *fast and comfortable overview*
- ✓ *several degrees of information detail as overlays*
- ✓ *humming bird navigation*



Automatic Camera Calibration

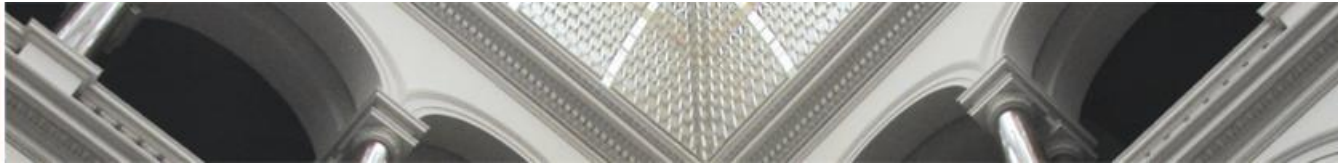


*Worapan Kusakunniran, Hongdong Li, and Jian Zhang, "A direct method to self-calibrate a surveillance camera by observing a walking pedestrian," in DICTA, 2009, pp. 250–



Event Detection





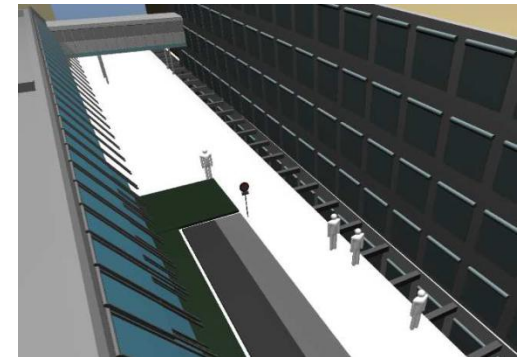
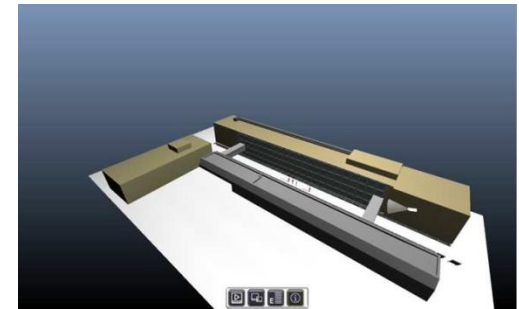
Conclusion

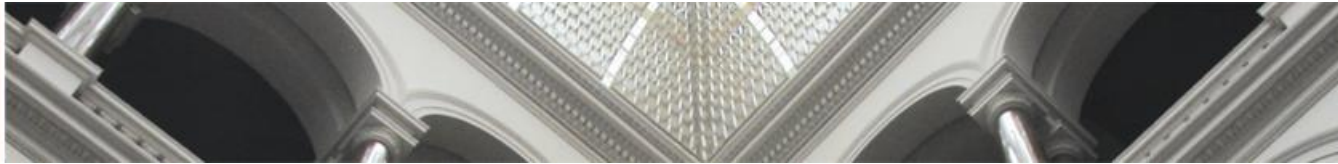
Clearly defined interfaces allow interchange of devices and analysis tools

Distribution of analysis tasks allowed us to run several algorithms in parallel with high computational demands which could not have been run at a centralized processing unit

3D model representation allows a fast and comfortable management of the surveilled area

Automated calibration system is sufficient for showing symbolical content of the observed scenarios in the 3D model





Thank you for your attention!

Tobias Senst

The research leading to these results has received funding from the European Communitys FP7 under grant agreement number 261743 (NoE VideoSense).