

AI Face Detection/AI People Detection/AI Vehicle Detection  
External Interface Specification

V1.01

i-PRO Co., Ltd.

Change log

VER.	Date	Item no.	Comment	Revise trigger
1.00	2022/4	All	First edition	-
1.01	2022/5	2.2.2	*Delete <Gray><Orange><Purple><Pink> from <Shoes><Color>.  *Delete <Gray> from <Bag><Color>. (Supported from app version V1.20)	Software fix
		2.1.2	*Changed the definition of <tt: Color> (Supported from app version V1.20) *Added Table 1.	
		2.1	*Fixed Meta Stream Example * Modify ObjectId to an integer value (Supported from app version V1.20)	
		2.1.3 2.2.3	Changed <Suv> to <SUV>	Document fix
		3.1	Changed trackID trackID=3 -> trackID=4	

## Contents

1. Introduction .....	3
1.1. Functional specifications.....	3
2. Onvif Meta Stream .....	3
2.1. Analytics Stream.....	4
2.1.1. AI Face Detection.....	4
2.1.2. AI People Detection.....	6
2.1.3. AI Vehicle Detection .....	12
2.2. Event Stream.....	14
2.2.1. AI Face Detection.....	14
2.2.2. AI People Detection.....	19
2.2.3. AI Vehicle Detectoin .....	28
3. Appendix .....	32
3.1. Transmission sequence of ONVIF Meta Stream.....	32

## 1. Introduction

This document describes the specifications for the external interface of AI Face Detection, AI People Detection and AI Vehicle Detection.

### 1.1. Functional specifications

WV-XAE204W (hereinafter referred to as XAE204) detects the face and determines the best shot. In addition, for the thumbnail determined to be the best shot, feature amount extraction for face matching and gender age estimation processing are performed. Thumbnails and meta information (gender and age information) can be acquired as Onvif meta information. (Features for face matching are not disclosed)

WV-XAE205W (hereinafter referred to as XAE205) detects a person and determines the best shot. In addition, the person attribute estimation process is performed on the thumbnail determined to be the best shot. Thumbnails and meta information (attribute information) can be acquired as Onvif meta information.

WV-XAE206W (hereinafter referred to as XAE206) detects two-wheeled vehicles / four-wheeled vehicles and determines the best shot. In addition, vehicle type / vehicle color estimation processing is performed on the thumbnail determined to be the best shot. Thumbnails and meta information (attribute information) can be acquired as Onvif meta information.

## 2. Onvif Meta Stream

There are two types of ONVIF meta information as follows.

[1] Analytics stream : The detection frame information is sent regularly. See below for the transmission cycle.

Network Camera other than Multi-Sensor Camera :

(When the camera's imaging mode is set to 30fps) 10fps

(When the camera's imaging mode is set to 25fps) 8.3fps

Multi-Sensor Camera :

(When the camera's imaging mode is set to 15fps/30fps) 3.75fps

(When the camera's imaging mode is set to 12.5fps/25fps) 3.1fps

[2] Event stream : Send thumbnails and meta information. The transmission interval is different for each applications.

## 2.1. Analytics Stream

The format of Analytics Stream is different for each applications.

### 2.1.1. AI Face Detection

#### 2.1.1.1. Parameter description

Parameter	Value	Description
UTC time	Date and Time(UTC)	UTC of video frame
AplSource	AIFaceDetection	Fixed value
ObjectId	4 byte integer	Detected object ID
BoundingBox	-1 ~ 1	Rectangle area of the detected object (coordinate of upper-left position and lower-right position)
CenterOfGravity	-1 ~ 1	Midpoint of BoundingBox
Class	- Face(object) - 0 ~ 1(Likelihood)	Kind of object and likelihood
HumanFace - Age - Min, Max	(Min, Max)= (0,10),(11,20),(21,30),(31,40), (41,50),(51,60),(61,-)	Age * This is applied when the best shot is confirmed. * This is not applied when the value of HumanFace -> Accessory -> Mask -> Wear is true.
HumanFace - Gender	Male, Female	Gender * This is applied when the best shot is confirmed.
HumanFace - Accessory - Mask - Wear	false, true	Whether or not human is wearing a mask. * This is applied when the best shot is confirmed.

### 2.1.1.2. Meta Stream Example

```
<?xml version="1.0" encoding="utf-8"?>
<tt:MetadataStream xmlns:tt="http://www.onvif.org/ver10/schema"
xmlns:fc="http://www.onvif.org/ver20/analytics/humanface">
  <tt:VideoAnalytics>
    <tt:Frame UtcTime="2020-01-20T10:00:08.20Z" AplSource="AIFaceDetection">
      <tt:Object Objectid="12345">
        <tt:Appearance>
          <tt:Shape>
            <tt:BoundingBox left="-0.20" top="0.99" right="0.83" bottom="-0.78" />
            <tt:CenterOfGravity x="0.73" y="0.105" />
          </tt:Shape>
          <tt:Class><tt:Type Likelihood="0.8">Face</tt:Type></tt:Class>
          <tt:HumanFace>
            <fc:Age>
              <tt:Min>11</tt:Min>
              <tt:Max>20</tt:Max>
            </fc:Age>
            <fc:Gender>Male</fc:Gender>
            <fc:Accessory>
              <fc:Mask>
                <fc:Wear>>false</fc:Wear>
              </fc:Mask>
            </fc:Accessory>
          </tt:HumanFace>
        </tt:Appearance>
      </tt:Object>
    </tt:Frame>
    <tt:Frame UtcTime="2020-01-20T10:00:18.20Z">
      ... (another frame information)
    </tt:Frame>
  </tt:VideoAnalytics>
</tt:MetadataStream>
```

## 2.1.2. AI People Detection

### 2.1.2.1. Parameter description

Parameter	Value	Description
UTC time	Date and Time(UTC)	UTC of video frame
ApiSource	AIPeopleDetection	Fixed value
ObjectId	4 byte integer	Detected object ID
BoundingBox	-1 ~ 1	Rectangle area of the detected object (coordinate of upper-left position and lower-right position)
CenterOfGravity	-1 ~ 1	Midpoint of BoundingBox
Class	- Human(object) - 0 ~ 1(Likelihood)	Kind of object and likelihood
HumanFace - Age - Min, Max	(Min, Max)= (0,10),(11,20),(21,60), (61,-)	Age * This is applied when the best shot is confirmed.
HumanFace - Gender	Male, Female	Gender * This is applied when the best shot is confirmed.
HumanFace - Hair - length - Color	- Long, Short  - (X,Y,Z) : RGB values	Hairstyle and hair color * This is applied when the best shot is confirmed. * Refer to Table 1 for an XYZ (RGB) value and the corresponding table of color.
HumanFace - FacialHair - Beard	false, true	Whether or not human has a beard. * This is applied when the best shot is confirmed.
HumanFace - Accessory - Optionals - Wear	false, true	Whether or not human is wearing sunglasses. * This is applied when the best shot is confirmed.
HumanFace - Accessory - Hat - Wear	false, true	Whether or not human is wearing a hat. * This is applied when the best shot is confirmed.
HumanFace - Accessory	false, true	Whether or not human is wearing a mask. * This is applied when the best shot is confirmed.

- Mask - Wear		
HumanBody - Clothing - Tops - Category - Color	- LongSleeve, ShortSleeve  - (X,Y,Z) : RGB values	Type and Color of tops * This is applied when the best shot is confirmed. * Refer to Table 1 for an XYZ (RGB) value and the corresponding table of color.
HumanBody - Clothing - Bottoms - Category - Color	- Trousers, Shorts  - (X,Y,Z) : RGB values	Type and Color of bottoms * This is applied when the best shot is confirmed. * Refer to Table 1 for an XYZ (RGB) value and the corresponding table of color.
HumanBody - Clothing - Shoes - Color	- Black, Brown, White, Gray, Green, Red, Blue, Yellow, Orange, Purple, Pink	Color of shoes * This is applied when the output setting is set to "On".
HumanBody - Belonging - Bag - Color	- Black, Brown, White, Gray, Green, Red, Blue, Yellow, Orange, Purple, Pink	Color of bag * This is applied when the output setting is set to "On"
Properties - DirectionNamed	Up/Right/Down/Left/UpRight/UpLeft/ DownRight/DownLeft	Direction of movement

**Table 1: XYZ (RGB) value and the corresponding table of color**

Color	X	Y	Z
Black	0	0	0
Brown	165	42	42
White	255	255	255
Gray	128	128	128
Green	0	255	0
Red	255	0	0
Blue	0	0	255
Yellow	255	255	0
Orange	255	165	0
Purple	128	0	128
Pink	255	192	203
Gold * Only as for the hair color	255	215	0

### 2.1.2.2. Meta Stream Example

```
<?xml version="1.0" encoding="utf-8"?>
<tt:MetadataStream xmlns:tt="http://www.onvif.org/ver10/schema"
xmlns:fc="http://www.onvif.org/ver20/analytics/humanface"
xmlns:bd="http://www.onvif.org/ver20/analytics/humanbody">
  <tt:VideoAnalytics>
    <tt:Frame UtcTime="2020-01-20T10:00:08.20Z" AplSource="AIPeopleDetection">
      <tt:Object ObjectId="12345">
        <tt:Appearance>
          <tt:Shape>
            <tt:BoundingBox left="-0.20" top="0.99" right="0.83" bottom="-0.78" />
            <tt:CenterOfGravity x="0.73" y="0.105" />
          </tt:Shape>
          <tt:Class><tt:Type Likelihood="0.8">Human</tt:Type></tt:Class>
          <tt:HumanFace>
            <fc:Age>
              <tt:Min>11</tt:Min>
              <tt:Max>20</tt:Max>
            </fc:Age>
            <fc:Gender>Male</fc:Gender>
            <fc:Hair>
              <fc:Length>Short</fc:Length>
              <fc:Color>
                <tt:ColorCluster>
                  <tt:Color X="0" Y="255" Z="0"
                    Colorspace="http://www.onvif.org/ver10/colorspace/RGB"/>
                </tt:ColorCluster>
              </fc:Color>
            </fc:Hair>
            <fc:FacialHair>
              <tt:Beard>true</tt:Beard>
            </fc:FacialHair>
            <fc:Accessory>
              <tt:Opticals>
                <tt:Wear>>false</tt:Wear>
              </tt:Opticals>
            </fc:Accessory>
          </tt:HumanFace>
        </tt:Appearance>
      </tt:Object>
    </tt:Frame>
  </tt:VideoAnalytics>
</tt:MetadataStream>
```

```
<tt:Hat>
  <tt:Wear>true</tt:Wear>
</tt:Hat>
<tt:Mask>
  <tt:Wear>>false</tt:Wear>
</tt:Mask>
</fc:Accessory>
</tt:HumanFace>
<tt:HumanBody>
  <bd:Clothing>
    <bd:Tops>
      <bd:Category>LongSleeve</bd:Category>
      <bd:Color>
        <tt:ColorCluster>
          <tt:Color X="0" Y="255" Z="0"
Colorspace="http://www.onvif.org/ver10/colourspace/RGB"/>
        </tt:ColorCluster>
      </bd:Color>
    </bd:Tops>
    <bd:Bottoms>
      <bd:Category>Trousers</bd:Category>
      <bd:Color>
        <tt:ColorCluster>
          <tt:Color X="0" Y="255" Z="0"
Colorspace="http://www.onvif.org/ver10/colourspace/RGB"/>
        </tt:ColorCluster>
      </bd:Color>
    </bd:Bottoms>
  <bd:Shoes>
    <bd:Color>
      <tt:ColorCluster>
        <tt:Color X="0" Y="255" Z="0"
Colorspace="http://www.onvif.org/ver10/colourspace/RGB"/>
      </tt:ColorCluster>
    </bd:Color>
  </bd:Shoes>
```

```
</bd:Clothing>
<bd:Belonging>
  <bd:Bag>
    <bd:Color>
      <tt:ColorCluster>
        <tt:Color X="0" Y="255" Z="0"
Colorspace="http://www.onvif.org/ver10/colospace/RGB"/>
      </tt:ColorCluster>
    </bd:Color>
  </bd:Bag>
</bd:Belonging>
</tt:HumanBody>
</tt:Appearance>
<tt:Extension xmlns="">
  <Properties>
    <Property name="DirectionNamed">Up</Property>
  </Properties>
</tt:Extension>
</tt:Object>
</tt:Frame>
<tt:Frame UtcTime="2020-01-20T10:00:18.20Z">
  ... (another frame information)
</tt:Frame>
</tt:VideoAnalytics>
</tt:MetadataStream>
```

### 2.1.3. AI Vehicle Detection

#### 2.1.3.1. Parameter description

Parameter	Value	Description
UTC time	Date and Time(UTC)	UTC of video frame
ApISource	AIVehicleDetection	Fixed value
ObjectId	4 byte integer	Detected object ID
BoundingBox	-1 ~ 1	Rectangle area of the detected object (coordinate of upper-left position and lower-right position)
CenterOfGravity	-1 ~ 1	Midpoint of BoundingBox
Class	- Vehicle(object) - 0 ~ 1(Likelihood)	Kind of object and likelihood
VehicleInfo	- Pickup, Truck, Bus, SUV, Van, Sedan - 0 ~ 1(Likelihood)	Kind of vehicle and likelihood
Color	Gray, White, Red, Black, Blue, Green, Brown, Yellow, Purple, Pink	Color of vehicle
Properties - DirectoinNamed	Up/Right/Down/Left/UpRight/ UpLeft/DownRight/DownLeft	Direction of movement

### 2.1.3.2. Meta Stream Example

```
<?xml version="1.0" encoding="utf-8"?>
<tt:MetadataStream xmlns:tt="http://www.onvif.org/ver10/schema">
  <tt:VideoAnalytics>
    <tt:Frame UtcTime="2020-01-20T10:00:08.20Z" AplSource="AIVehicleDetection">
      <tt:Object ObjectId="12345">
        <tt:Appearance>
          <tt:Shape>
            <tt:BoundingBox left="-0.20" top="0.99" right="0.83" bottom="-0.78" />
            <tt:CenterOfGravity x="0.73" y="0.105" />
          </tt:Shape>
          <tt:Class><tt:Type Likelihood="0.8">Vehicle</tt:Type></tt:Class>
          <tt:VehicleInfo>
            <tt:Type Likelihood="0.8">Sedan</tt:Type>
          </tt:VehicleInfo>
          <tt:Color>Black</tt:Color>
        </tt:Appearance>
        <tt:Extension xmlns="">
          <Properties>
            <Property name="DirectionNamed">Up</Property>
          </Properties>
        </tt:Extension>
      </tt:Object>
    </tt:Frame>
    <tt:Frame UtcTime="2020-01-20T10:00:18.20Z">
      ... (another frame information)
    </tt:Frame>
  </tt:VideoAnalytics>
</tt:MetadataStream>
```

## 2.2. Event Stream

The format of Event Stream is different for each applications.

### 2.2.1. AI Face Detection

The transmission interval of thumbnails and meta information is 1 second or 2 seconds (initial value: 1 second) interval, which can be changed by setting.

The maximum number of faces transmitted at one time is 6 faces at 1-second intervals and 12 faces at 2-second intervals. If more than the maximum number of faces are shown, the face information with the smallest number of faces to be sent is given priority for transmission.

#### 2.2.1.1. Parameter description

Parameter	Value	Description
UTC time	Date and Time(UTC)	UTC of video frame
ApISource	AIFaceDetection	Fixed value
Objectld	4 byte integer	Detected object ID
GUID	Hex numbers	UUID
BoundingBox	-1 ~ 1	Rectangle area of the detected object * Coordinates (-1 ~ 1) of best shot for the entire angle of view.
CenterOfGravity	-1 ~ 1	Midpoint of BoundingBox
Class	- Face(object) - 0, 1(Likelihood)	Kind of object and likelihood
Image	Base64 encoded	JPEG image
HumanFace - FeatureValue - L2Norm - feature-value-version	Non-disclosure	Non-disclosure
HumanFace - start-time	Date and Time(UTC)	Date and Time when the face is detected for the first time.
HumanFace - bs-frame	- X coordinate, Y coordinate, Width, Height Decimal numbers (0 ~320)	Coordinates(upper left) of the frame of the detected face of the best shot. * QVGA

HumanFace - thumb-frame	- X coordinate, Y coordinate, Width, Height Decimal numbers (0 ~320)	Coordinates with Upper left of thumbnail image of face as the origin.
HumanFace - face-info	Decimal numbers 0 fixed,0 fixed, Score of face detection x100	Face-likeness of the frame of the detected face.
HumanFace - bs-score	Decimal numbers 0 fixed, 0~3000	Score of best shot
HumanFace - send-premise	Hex numbers  - Upper 4bit : Timing of transmitting the best shot 1: The timing when the tracking is finished 2 : The timing when the best shot is transmit periodically  - Upper 5-8bits : Reason of transmitting best shot 0 : Best shot is unrenewed 1 : Best shot is renewed	This parameter shows the timing and reason for transmitting thumbnails, and is classified as follows.  - Timing of periodic transmission(Best shot is renewed) - Timing when tracking is completed(Best shot is renewed) - Timing of periodic transmission(Best shot is unrenewed) - Timing when tracking is completed(Best shot is unrenewed)
HumanFace -Mask	false, true	Whether or not human is wearing a mask.
HumanFace -Age - <Range min="0" max="10"> - <Range min="11" max="20"> - <Range min="21" max="30"> - <Range min="31" max="40"> - <Range min="41" max="50"> - <Range min="51" max="60"> - <Range min="61">	0 ~ 1	Likelihood of age  * This is not applied when the value of HumanFace -> Wear is true.
HumanFace -Gender - Male	0 ~ 1	Likelihood of gender

- Female		
Properties - RecommendedSize	false, true	If the width of the frame of detected human meets the recommended detection size, the value will be true. If not satisfied, the value will be false.
Properties - DetectPixelWidth	0 ~ Maximum width of camera resolution	Width of face [pixel] * It is not the size of the best shot, but the size of the detected object before resizing.
Properties - DetectPixelHeight	0 ~ Maximum height of camera resolution	Height of face [pixel] *It is not the size of the best shot, but the size of the detected object before resizing.

### 2.2.1.2. Meta Stream Example

```
<?xml version="1.0" encoding="utf-8"?>
<tt:MetadataStream xmlns:tt="http://www.onvif.org/ver10/schema">
  <tt:VideoAnalytics>
    <tt:Frame UtcTime="2020-01-20T10:00:08.20Z" AplSource="AIFaceDetection">
      <tt:Object ObjectId="107" GUID="7CE62E87D1BB845F4A256728FED42A53">
        <tt:Appearance>
          <tt:Shape>
            <tt:BoundingBox left="-0.20" top="0.99" right="0.83" bottom="-0.78"/>
            <tt:CenterOfGravity x="0.73" y="0.105" />
          </tt:Shape>
          <tt:Class><tt:Type Likelihood="0.8">Face</tt:Type></tt:Class>
          <tt:Image>(base64 image data)</tt:Image>
          <tt:Extension xmlns="">
            <HumanFace>
              <FeatureValue> ((*snip*)) </FeatureValue>
              <L2Norm> ((*snip*)) </L2Norm>
              <start-time>2020-01-20T10:00:03.60Z</start-time>
              <bs-frame>50,34,18,21</bs-frame>
              <thumb-frame>107,114,16,15</thumb-frame>
              <face-info>0,0,870</face-info>
              <bs-score>0,1234</bs-score>
              <send-premise>21000000</send-premise>
              <feature-value-version>0.1</feature-value-version>
              <Mask>>false</Mask>
              <Age>
                <Range min="0" max="10">0.08</Range>
                <Range min="11" max="20">0.63</Range>
                <Range min="21" max="30">0.12</Range>
                <Range min="31" max="40">0.1</Range>
                <Range min="41" max="50">0.03</Range>
                <Range min="51" max="60">0.02</Range>
                <Range min="61">0.02</Range>
              </Age>
            </HumanFace>
          </tt:Extension>
        </tt:Appearance>
      </tt:Object>
    </tt:Frame>
  </tt:VideoAnalytics>
</tt:MetadataStream>
```

```
<Gender>
  <Male>0.37</Male>
  <Female>0.63</Female>
</Gender>
</HumanFace>
</tt:Extension>
</tt:Appearance>
<tt:Extension xmlns="">
  <Properties>
    <Property name="RecommendedSize">true</Property>
    <Property name="DetectPixelWidth">120</Property>
    <Property name="DetectPixelHeight">120</Property>
  </Properties>
</tt:Extension>
</tt:Object>
</tt:Frame>
<tt:Frame UtcTime="2020-01-20T10:00:18.20Z" AplSource="AIFaceDetection">
  ...(Another frame)
</tt:Frame>
</tt:VideoAnalytics>
</tt:MetadataStream>
```

## 2.2.2. AI People Detection

Thumbnails and meta information are sent at 1 second intervals. The maximum number of people that can be transmitted at one time is 10.

### 2.2.2.1. Parameter description

Parameter	Value	Description
UTC time	Date and Time(UTC)	UTC of video frame
AplSource	AIPeopleDetection	Fixed value
ObjectId	4 byte integer	Detected object ID
GUID	Hex numbers	UUID
BoundingBox	-1 ~ 1	Rectangle area of the detected object * Coordinates (-1 ~ 1) of best shot for the entire angle of view.
CenterOfGravity	-1 ~ 1	Midpoint of BoundingBox
Class	- Human(object) - 0, 1(Likelihood)	Kind of object and likelihood
Image	Base64 encoded	JPEG image
HumanFace - Age - <Range min="0" max="10"> - <Range min="11" max="20"> - <Range min="21" max="60"> - <Range min="61">	0 ~ 1	Likelihood of age
HumanFace - Gender - Male, Female	0 ~ 1	Likelihood of gender
HumanFace - Hair - length - Long, Short_Bald, Hat - Color - Black, Brown, White, Gray, Gold	0 ~ 1	Likelihood of hairstyle and hair color

HumanFace - FacialHair - Beard, NoBeard	0 ~ 1	Likelihood of whether or not human has a beard.
HumanFace - Accessory - Opticals - Sunglass, NoGlass	0 ~ 1	Likelihood of whether or not human is wearing sunglasses.
HumanFace - Accessory - Mask - Mask, NoMask	0 ~ 1	Likelihood of whether or not human is wearing a mask.
HumanBody - Clothing - Tops - Category - LongSleeve, ShortSleeve - Color - Black, Brown, White, Gray, Green, Red, Blue, Yellow, Orange, Purple, Pink	0 ~ 1	Likelihood of type and Color of tops
HumanBody - Clothing - Bottoms - Category - Trousers, Shorts - Color - Black, Brown, White, Gray, Green, Red, Blue, Yellow, Orange, Purple, Pink	0 ~ 1	Likelihood of type and Color of bottoms
HumanBody - Clothing - Shoes - Color - Black, Brown, White, Green, Red, Blue, Yellow	0 ~ 1	Likelihood of color of shoes

HumanBody - Belonging - Bag - Bag - NoBag - Color - Black, Brown, White, Green, Red, Blue, Yellow, Orange, Purple, Pink	0 ~ 1	Likelihood of color of bag
UpperBodyCoordinate	- left, right : 0 ~ Maximum width of best shot  - top, bottom : 0 ~ Maximum height of best shot	Detection frame information of head in the beset shot [pixel]
FaceCoordinate	- left, right : 0 ~ Maximum width of best shot  - top, bottom : 0 ~ Maximum height of best shot	Detection frame information of face in the beset shot [pixel]
Properties - DirectionNamed	Up/Right/Down/Left/UpRig ht/UpLeft/DownRight/Down Left	Direction of movement
Properties -Color	false, true	True for color images. False for black-and-white images.
Properties - Wholebody	false, true	True if the whole body is detected, false if not detected.
Properties - Upperbody	false, true	True if the head is detected, false if not.
Properties - Face	false, true	If the face is detected, it is true, and if it is not detected, it is false.

Properties - RecommendedSize	false, true	If the width of the frame of detected human meets the recommended detection size, the value will be true. If not satisfied, the value will be false.
Properties - DetectPixelWidth	0 ~ Maximum width of camera resolution	Width of human [pixel] * It is not the size of the best shot, but the size of the detected object before resizing.
Properties - DetectPixelHeight	0 ~ Maximum height of camera resolution	Height of human [pixel] *It is not the size of the best shot, but the size of the detected object before resizing.
Properties - Shoes	0 ~ 1	Likelihood of whether or not human puts on shoes.
Properties - Bag	0 ~ 1	Likelihood of whether or not human has a bag.

### 2.2.2.2. Meta Stream Example

```
<?xml version="1.0" encoding="utf-8"?>
<tt:MetadataStream xmlns:tt="http://www.onvif.org/ver10/schema">
  <tt:VideoAnalytics>
    <tt:Frame UtcTime="2020-01-20T10:00:08.20Z" ApiSource="AIPeopleDetection">
      <tt:Object ObjectId="101" GUID="28FED42A7BB845F4A2567CE62E87D153">
        <tt:Appearance>
          <tt:Shape>
            <tt:BoundingBox left="-0.20" top="0.99" right="0.83" bottom="-0.78" />
            <tt:CenterOfGravity x="0.73" y="0.105" />
          </tt:Shape>
          <tt:Class><tt:Type Likelihood="0.8">Human</tt:Type></tt:Class>
          <tt:Image>(base64 image data)</tt:Image>
          <tt:Extension xmlns="">
            <HumanFace>
              <Age>
                <Range min="0" max="10">0.2</Range>
                <Range min="11" max="20">0.63</Range>
                <Range min="21" max="60">0.1</Range>
                <Range min="61">0.07</Range>
              </Age>
              <Gender>
                <Male>0.37</Male>
                <Female>0.63</Female>
              </Gender>
              <Hair>
                <length>
                  <Long>0.1</Long>
                  <Short_Bald>0.8</Short_Bald>
                  <Hat>0.1</Hat>
                </length>
                <Color>
                  <Black>0.6</Black>
                  <Brown>0.1</Brown>
                  <White>0.1</White>
                </Color>
              </Hair>
            </HumanFace>
          </tt:Extension>
        </tt:Appearance>
      </tt:Object>
    </tt:Frame>
  </tt:VideoAnalytics>
</tt:MetadataStream>
```

```
<Gray>0.1</Gray>
<Gold>0.1</Gold>
</Color>
</Hair>
<FacialHair>
  <Beard>0.1</Beard>
  <NoBeard>0.9</NoBeard>
</FacialHair>
<Accessory>
  <Opticals>
    <Sunglass>0.1</Sunglass>
    <NoGlass>0.9</NoGlass>
  </Opticals>
  <Mask>
    <Mask>0.1</Mask>
    <NoMask>0.9</NoMask>
  </Mask>
</Accessory>
</HumanFace>
<HumanBody>
  <Clothing>
    <Tops>
      <Category>
        <LongSleeve>0.9</LongSleeve>
        <ShortSleeve>0.1</ShortSleeve>
      </Category>
      <Color>
        <Black>1</Black>
        <Brown>0</Brown>
        <White>0</White>
        <Gray>0</Gray>
        <Green>0</Green>
        <Red>0</Red>
        <Blue>0</Blue>
        <Yellow>0</Yellow>
```

```
<Orange>0</Orange>
<Purple>0</Purple>
<Pink>0</Pink>
</Color>
</Tops>
<Bottoms>
  <Category>
    <Trousers>0.9</Trousers>
    <Shorts>0.1</Shorts>
  </Category>
  <Color>
    <Black>1</Black>
    <Brown>0</Brown>
    <White>0</White>
    <Gray>0</Gray>
    <Green>0</Green>
    <Red>0</Red>
    <Blue>0</Blue>
    <Yellow>0</Yellow>
    <Orange>0</Orange>
    <Purple>0</Purple>
    <Pink>0</Pink>
  </Color>
</Bottoms>
<Shoes>
  <Color>
    <Black>1</Black>
    <Brown>0</Brown>
    <White>0</White>
    <Green>0</Green>
    <Red>0</Red>
    <Blue>0</Blue>
    <Yellow>0</Yellow>
```

```

        </Color>
    </Shoes>
</Clothing>
<Belonging>
    <Bag>
        <Bag>0.63</Bag>
        <NoBag>0.37</NoBag>
        <Color>
            <Black>1</Black>
            <Brown>0</Brown>
            <White>0</White>
            <Green>0</Green>
            <Red>0</Red>
            <Blue>0</Blue>
            <Yellow>0</Yellow>
            <Orange>0</Orange>
            <Purple>0</Purple>
            <Pink>0</Pink>
        </Color>
    </Bag>
</Belonging>
</HumanBody>
</tt:Extension>
<UpperBodyCoordinate left="10" top="12" right="40" bottom="120" />
<FaceCoordinate left="15" top="18" right="35" bottom="110" />
</tt:Appearance>
<tt:Extension xmlns="">
    <Properties>
        <Property name="DirectionNamed">Up</Property>
        <Property name="Color">true</Property>
        <Property name="Wholebody">true</Property>
        <Property name="Upperbody">true</Property>
        <Property name="Face">true</Property>
        <Property name="RecommendedSize">true</Property>
    </Properties>

```

```
<Property name="DetectPixelWidth">120</Property>
<Property name="DetectPixelHeight">500</Property>
<Property name="Shoes">0.82</Property>
<Property name="Bag">0.63</Property>
</Properties>
</tt:Extension>
</tt:Object>
<tt:Object ObjectId="102" GUID="4A2567CE62E87D15328FED42A7BB845F">
  ...
</tt:Object>
</tt:Frame>
<tt:Frame UtcTime="2020-01-20T10:00:18.20Z" ApiSource="AIPeopleDetection">
  ...(Another frame)
</tt:Frame>
</tt:VideoAnalytics>
</tt:MetadataStream>
```

### 2.2.3. AI Vehicle Detectoin

Metadata is sent after the vehicle has been tracked or when the tracking is time-out (\*). The maximum number of vehicles that can be transmitted at one time is 20.

\* The time-out conditions are shown below.

Condition 1: When the detection period of ID reaches 10 seconds from the start of tracking.

Condition 2: When the detection period of ID reaches 1 hour after determining that it has timed out.

#### 2.2.3.1. Parameter description

Parameter	Value	Description
UTC time	Date and Time(UTC)	UTC of video frame
ObjectId	4 byte integer	Detected object ID
GUID	Hex numbers	UUID
BoundingBox	-1 ~ 1	Rectangle area of the detected object * Coordinates (-1 ~ 1) of best shot for the entire angle of view.
CenterOfGravity	-1 ~ 1	Midpoint of BoundingBox
Class	- Vehicle(object) - 0, 1(Likelihood)	Kind of object and likelihood
Image	Base64 encoded	JPEG image
VehicleInfo - Type - Sedan, Van, SUV, Truck, Bus, PickupTruck, TwoWheels	0 ~ 1	Likelihood of model of car
VehicleInfo - Color - Black, Brown, White, Gray, Green, Red, Blue, Yellow, Purple, Pink	0 ~ 1	Likelihood of color of car
Properties - DirectionNamed	Up/Right/Down/Left/UpRight /UpLeft/DownRight/DownLeft	Direction of movement

Properties - Color	false, true	True for color images. False for black-and-white images.
Properties - RecommendedSize	false, true	If the width of the detection frame meets the recommended detection size, the value will be true. If not satisfied, the value will be false.
Properties - DetectPixelWidth	0 ~ Maximum width of camera resolution	Width of car [pixel] * It is not the size of the best shot, but the size of the detected object before resizing.
Properties - DetectPixelHeight	0 ~ Maximum height of camera resolution	Height of car [pixel] *It is not the size of the best shot, but the size of the detected object before resizing.

### 2.2.3.2. Meta Stream Example

```
<?xml version="1.0" encoding="utf-8"?>
<tt:MetadataStream xmlns:tt="http://www.onvif.org/ver10/schema">
  <tt:VideoAnalytics>
    <tt:Frame UtcTime="2020-01-20T10:00:08.20Z" ApiSource="AIVehicleDetection">
      <tt:Object ObjectId="105" GUID="BB845F4A256728FED42A7CE62E87D153">
        <tt:Appearance>
          <tt:Shape>
            <tt:BoundingBox left="-0.20" top="0.99" right="0.83" bottom="-0.78" />
            <tt:CenterOfGravity x="0.73" y="0.105" />
          </tt:Shape>
          <tt:Class><tt:Type Likelihood="0.8">Vehicle</tt:Type></tt:Class>
          <tt:Image>(base64 image data)</tt:Image>
          <tt:Extension xmlns="">
            <VehicleInfo>
              <Type>
                <Sedan>0.37</Sedan>
                <Van>0.63</Van>
                <SUV>0</SUV>
                <Truck>0</Truck>
                <Bus>0</Bus>
                <PickupTruck>0</PickupTruck>
                <TwoWheels>0</TwoWheels>
              </Type>
              <Color>
                <Black>1</Black>
                <Brown>0</Brown>
                <White>0</White>
                <Gray>0</Gray>
                <Green>0</Green>
                <Red>0</Red>
                <Blue>0</Blue>
                <Yellow>0</Yellow>
                <Purple>0</Purple>
                <Pink>0</Pink>
              </Color>
            </VehicleInfo>
          </tt:Extension>
        </tt:Appearance>
      </tt:Object>
    </tt:Frame>
  </tt:VideoAnalytics>
</tt:MetadataStream>
```

```
        </Color>
      </VehicleInfo>
    </tt:Extension>
  </tt:Appearance>
  <tt:Extension xmlns="">
    <Properties>
      <Property name="DirectionNamed">Up</Property>
      <Property name="Color">>true</Property>
      <Property name="RecommendedSize">>true</Property>
      <Property name="DetectPixelWidth">120</Property>
      <Property name="DetectPixelHeight">500</Property>
    </Properties>
  </tt:Extension>
</tt:Object>
<tt:Object ObjectId="105" GUID="BB845F4A256728FED42A7CE62E87D153">
  ...
</tt:Object>
</tt:Frame>
<tt:Frame UtcTime="2020-01-20T10:00:18.20Z" ApiSource="AIVehicleDetection">
  ...(Another frame)
</tt:Frame>
</tt:VideoAnalytics>
</tt:MetadataStream>
```

### 3. Appendix

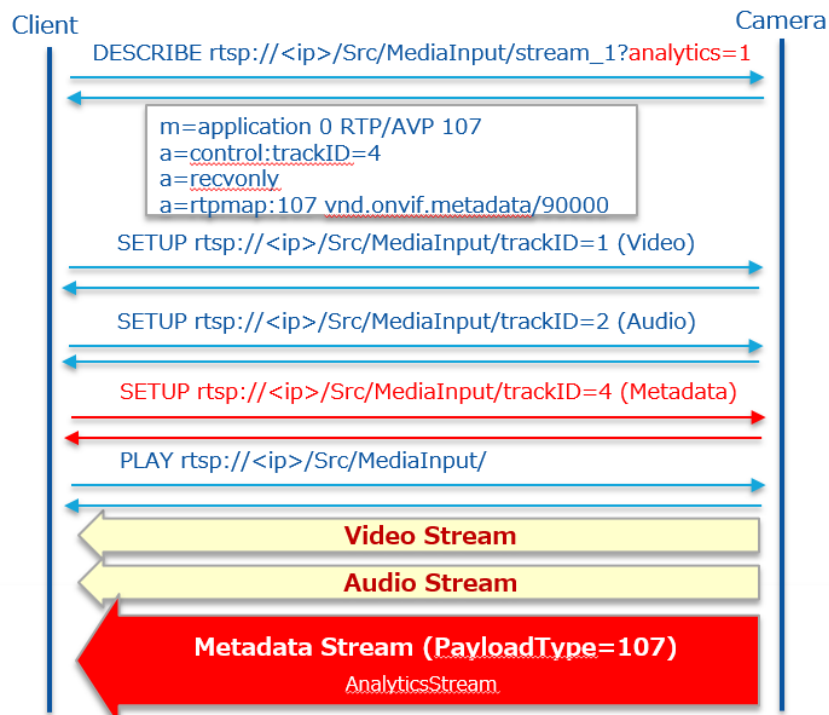
#### 3.1. Transmission sequence of ONVIF Meta Stream

[RTSP URL]

\*Send request(RTSP URL) with “analytics=1” in case of requesting Analytics stream.

\*Send request with “event=1” in case of requesting Event stream.

\*Send request with “analytics=1&event=1” in case of requesting Analytics stream and Event stream.



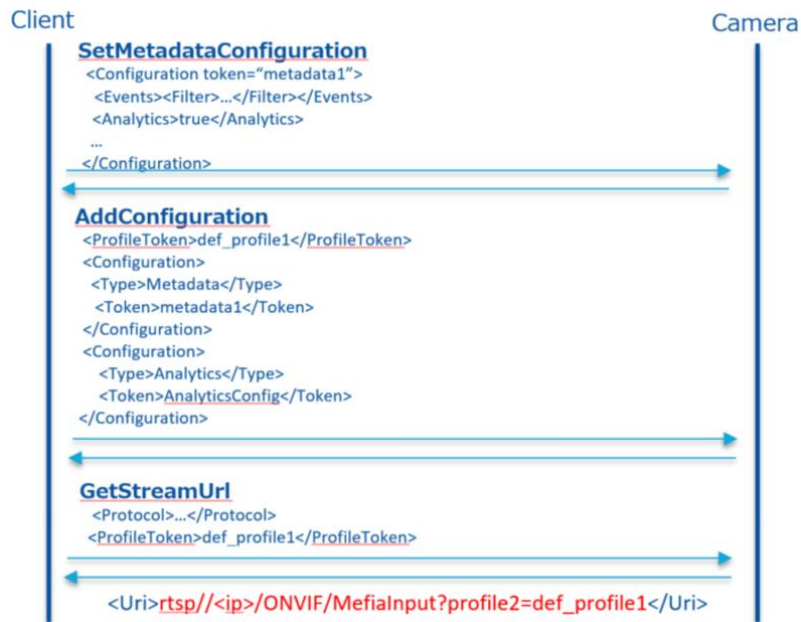
[ONVIF]

\* Configure by ONVIF commands

- SetMetadataConfiguration(Event filter, analytics flag)

- AddConfiguration(Add “metadata1” and “AnalyticsConfig” at “MediaProfile”)

\* Get RTSP URL by ONVIF commands(GetStreamUrl)



- \* Streaming by URL got by ONVIF commands
- Event stream is also sent by streaming analytics

