

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

V1.04

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	
1.02	2022/7	2.2.2 2.2.3	Delete <Property name="Color"> (Supported from app version V1.21)	Software version upgrade
1.03	2023/2	2.1.3	Added "Motorcycle" to VehicleInfo	Document fix
		2.2.3	·Corrected the description of the timeout condition. ·Corrected the description of VehicleInfo	
1.04	2023/7	2.2.2	Re-added <Property name="Color"> (Supported from app version V1.70)	Software version upgrade
		2.2.3		
		3.	New addition	

Contents

1. Introduction	4
1.1. Functional specifications.....	4
2. Onvif Meta Stream	4
2.1. Analytics Stream.....	5
2.1.1. AI Face Detection.....	5
2.1.2. AI People Detection.....	7
2.1.3. AI Vehicle Detection	13
2.2. Event Stream.....	15
2.2.1. AI Face Detection.....	15
2.2.2. AI People Detection.....	20
2.2.3. AI Vehicle Detectoin	29
3. Interface commands (CGI): Download the CSV files	33
3.1. AI People Detection	33
3.1.1. CSV file format	33
3.1.2. Getting the CSV file.....	37
4. Appendix	44
4.1. Transmission sequence of ONVIF Meta Stream.....	44

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

AI Face Detection 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)

AI People Detection 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. Meta information can be saved as csv and acquired by CGI.

AI Vehicle Detection 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
Objectld	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, Motorcycle - 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 video, false for black-and- white video.
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>
        <Property name="DetectPixelWidth">120</Property>
        <Property name="DetectPixelHeight">500</Property>
    </Properties>

```

```
        <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 duration of the same ID reaches 10 seconds from the start of tracking.

Condition 2: When the detection duration of the same ID reaches 1 hour after judging that it has timed out. (Metadata for parked vehicles will be sent every hour.)

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, Pickup, 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 video, false for black-and-white video.
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. Interface commands (CGI): Download the CSV files

3.1. AI People Detection

AI People Detection performs personal attribute estimation processing on the thumbnail determined to be the best shot. Meta information (attribute information) can be saved as csv and acquired by CGI.

The maximum retention period and number of meta information are as follows. Delete old meta information if any of the following limits are exceeded.

- Maximum retention period: 92 days
- Maximum saved number: 50,000

[CGI to enable csv save function]

http://192.168.0.10/cgi-bin/adam.cgi?methodName=sendDataToAdamApplication&appName=HumanBestshotApp&s_appDataType=2&s_appData=e3thcHBNZXRob2Q6UE9TVH0se3NhdmluZ19jc3ZfZW5hYmxlOjF9fQ==

[CGI to disable csv save function (initial setting)]

http://192.168.0.10/cgi-bin/adam.cgi?methodName=sendDataToAdamApplication&appName=HumanBestshotApp&s_appDataType=2&s_appData=e3thcHBNZXRob2Q6UE9TVH0se3NhdmluZ19jc3ZfZW5hYmxlOjB9fQ==

3.1.1. CSV file format

One meta information of the person is saved at the timing of transmission of ONVIF meta information. However, if the same person continues to be detected across the save timing of the csv file, two pieces of meta information for one person may be saved. See section 2.2.2 for ONVIF meta information transmission timing.

When the maximum retention period or the maximum number of stored items is exceeded, the oldest meta information is deleted.

The csv file format and meta information content are described below.

localtime, utctime, timezone, summertime, age, gender, [No.7 to 55 in the table below]
localtime, utctime, timezone, summertime, age, gender, [No.7 to 55 in the table below]
localtime, utctime, timezone, summertime, age, gender, [No.7 to 55 in the table below]
...
localtime, utctime, timezone, summertime, age, gender, [No.7 to 55 in the table below]

Meta information list

No.	Data	Format	Description
1	localtime	YYYY/MM/DD HH:mm:ss ss	Display the time stamp of the YC image corresponding to the detection frame in "2 digits after the decimal point of year, month, day, hour, minute, second, second" (e.g.)2011/04/21 0:24:36 01
2	utctime	YYYY/MM/DD HH:mm:ss ss	UTC (e.g.)If localtime= 2011/04/21 0:24:36 01 in Japan, 2011/04/20 15:24:36 01
3	timezone	-12:00~ +12:00	Displays the setting value of the time zone set on the camera. (e.g.)In case of Japan, +09:00
4	summertime	IN, OUT	Displays the daylight saving time setting set on the camera. IN: Daylightsaving time (Summertime) OUT: Not daylight saving time
5	age	0~6,255	Displays the age determination result 0 : 0~10 years old 1 : 11~20 years old 2 : 21~30 years old 3 : 31~40 years old 4 : 41~50 years old 5 : 51~60 years old 6 : 61 and over 255 : Undecidable * When score_mask is 0.2 or more, the accuracy of age identification decreases. Therefore, when score_mask is 0.2 or more, it is

			recommended not to use age determination results.
6	gender	0, 1, 255	Display the gender determination result. 0 : Male 1 : Female 255 : Undecidable
7	score_age_0	0.00~1.00	Likelihood of age. The age represented by each parameter name is shown below. score_age_0 : 0~10 years old score_age_1 : 11~20 years old score_age_2 : 21~30 years old score_age_3 : 31~40 years old score_age_4 : 41~50 years old score_age_5 : 51~60 years old score_age_6 : 61 and over * If the judgment result is null, set it to -1.0.
8	score_age_1		
9	score_age_2		
10	score_age_3		
11	score_age_4		
12	score_age_5		
13	score_age_6		
14	score_male	0.00~1.00	Likelihood of gender * If the judgment result is null, set it to -1.0.
15	score_female		
16	score_hair_long	0.00~1.00	Likelihood of hairstyle * If the judgment result is null, set it to -1.0.
17	score_hair_short		
18	score_hair_hat		
19	score_hair_color_black	0.00~1.00	Likelihood of hair color * If the judgment result is null, set it to -1.0.
20	score_hair_color_brown		
21	score_hair_color_white		

22	score_hair_color_gray		
23	score_hair_color_gold		
24	score_beard	0.00~1.00	Likelihood of whether or not human has a beard.
25	score_nobeard		* If the judgment result is null, set it to -1.0.
26	score_sunglass	0.00~1.00	Likelihood of whether or not human is wearing sunglasses.
27	score_noglass		* If the judgment result is null, set it to -1.0.
28	score_mask	0.00~1.00	Likelihood of whether or not human is wearing a mask.
29	score_nomask		* If the judgment result is null, set it to -1.0.
30	score_tops_long sleeve	0.00~1.00	Likelihood of type of tops
31	score_tops_short sleeve		score_tops_long sleeve : Long score_tops_short sleeve : Short * If the judgment result is null, set it to -1.0.
32	score_tops_color_black	0.00~1.00	Likelihood of color of tops * If the judgment result is null, set it to -1.0.
33	score_tops_color_brown		
34	score_tops_color_white		
35	score_tops_color_gray		
36	score_tops_color_green		
37	score_tops_color_red		
38	score_tops_color_blue		
39	score_tops_color_yellow		
40	score_tops_color_orange		
41	score_tops_color_purple		
42	score_tops_color_pink		
43	score_bottoms_trousers	0.00~1.00	Likelihood of type of bottoms
44	score_bottoms_shorts		score_bottoms_trousers : Long score_bottoms_shorts : Short

			* If the judgment result is null, set it to -1.0.
45	score_bottoms_color_black	0.00~1.00	Likelihood of color of bottoms * If the judgment result is null, set it to -1.0.
46	score_bottoms_color_brown		
47	score_bottoms_color_white		
48	score_bottoms_color_gray		
49	score_bottoms_color_green		
50	score_bottoms_color_red		
51	score_bottoms_color_blue		
52	score_bottoms_color_yellow		
53	score_bottoms_color_orange		
54	score_bottoms_color_purple		
55	score_bottoms_color_pink		

■About saving meta information

The storage period of the meta information in the internal memory is one hour. Unsaved meta information will not be retained if the app operation stops or if the power to the camera is turned off.

3.1.2. Getting the CSV file

3.1.2.1. Get storing period (mode:range)

[Command interface]

Method: GET

[CGI URL]

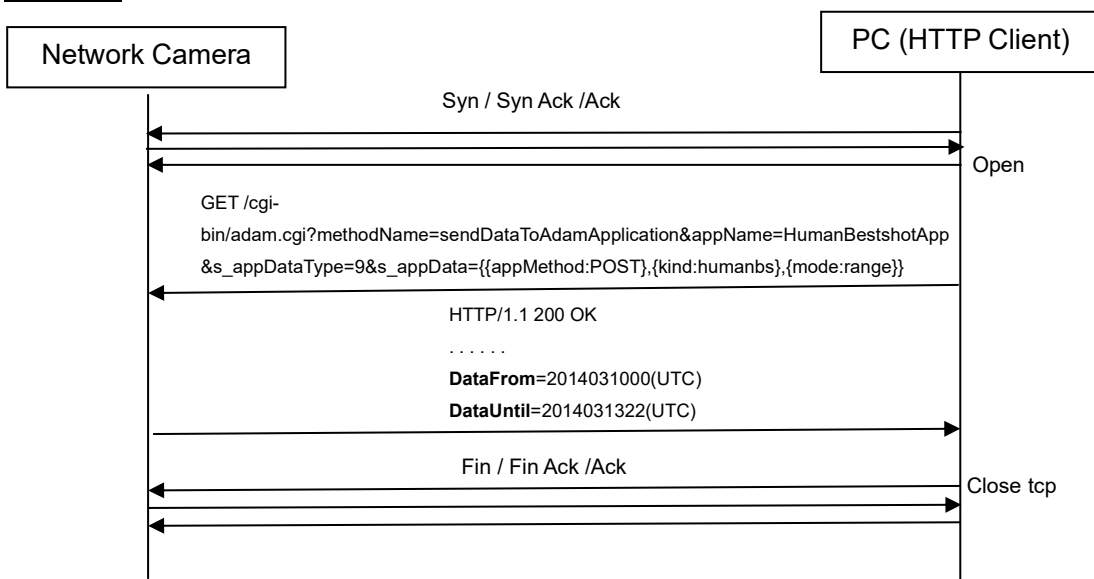
http://192.168.0.10/cgi-bin/adam.cgi?methodName=sendDataToAdamApplication&appName=HumanBestshotApp&s_appDataType=9&s_appData={{appMethod:csv},{kind:humanbs},{mode:range}}

*Apply base 64 encode to the preference value that follows "s_appData=".

[Request parameters]

Parameter	Value	Description
appMethod	csv	Set the method.
kind	humanbs	Kind of the csv data humanbs: Meta information of AI People Detection *This parameter can't be omitted
mode	range	Kind of response range: Get storing period *This parameter can't be omitted

Sequence



Response format

```

HTTP1.1 200OK [CR][LF]
Status 200[CR][LF]
....
Content-Length: xxxxx[CR][LF]
DataFrom=YYYYMMDDHHmm(UTC)[CR][LF]

```

Response data

Data	Format	Description
DataFrom	YYYYMMDDHHmm(UTC) YYYY: year(4 columns) MM: month(2 columns) DD: day(2 columns) HH: hour(2 columns) mm: minute(2 columns)	Time and date of the oldest csv file (UTC time)
DataUntil	YYYYMMDDHHmm(UTC) YYYY: year(4 columns) MM: month(2 columns) DD: day(2 columns) HH: hour(2 columns) mm: minute(2 columns)	Time and date of the latest csv file (UTC time)

The response when there are multiple pieces of meta information is

DataFrom=YYYYMMDDHH(UTC) (time and date of the oldest meta information)

DataUntil=yyyymmddhh(UTC) (time and date of the latest meta information)

to the client as a response to the CGI.

```
HTTP/1.1 200 OK[CR][LF]
Status: 200[CR][LF]
...
...
Content-length: xxxx[CR][LF]
DataFrom=2014031000(UTC)[CR][LF]
DataUntil=2014031322(UTC)[CR][LF]
```

When there is only one piece of meta information, the response is sent to the client as a response to the CGI with DataFrom and DataUntil set to the same date and time.

3.1.2.2. Get csv files by a date (mode:multi)

[Command interface]

Method: GET

[CGI URL]

http://192.168.0.10/cgi-bin/adam.cgi?methodName=sendDataToAdamApplication&appName=HumanBestshotApp&s_appDataType=9&s_appData={{appMethod:csv},{kind:humanbs},{mode:multi},{year:YYYY},{month:MM},{date:DD},{hour:HH},{days:d}}

*Apply base 64 encode to the preference value that follows "s_appData=".

[Request parameters]

Parameter	Value	Description
appMethod	csv	Set the method.
kind	humanbs	Kind of the csv data humanbs: Meta information of AI People Detection *This parameter can't be omitted
mode	multi	Kind of response multi: Get csv/index files by a date. *This parameter can't be omitted
year	(numerical value) (4 columns)	Date of the file to acquire (Year) *This parameter can't be omitted when the 'mode' parameter set to 'multi'.
month	1 - 12	Date of the file to acquire (Month) *This parameter can't be omitted when the 'mode' parameter set to 'multi'.
date	1 - 31	Date of the file to acquire (Date) *This parameter can't be omitted when the 'mode' parameter set to 'multi'.

hour	0 - 23	The time of file to acquire. *This parameter can't be omitted when the 'mode' parameter set to 'multi'.
days	1 - 6	The days of the file to acquire *This parameter can't be omitted when the 'mode' parameter set to 'multi'.

[Command examples]

Getting csv files of 4/11/2023 for one day.

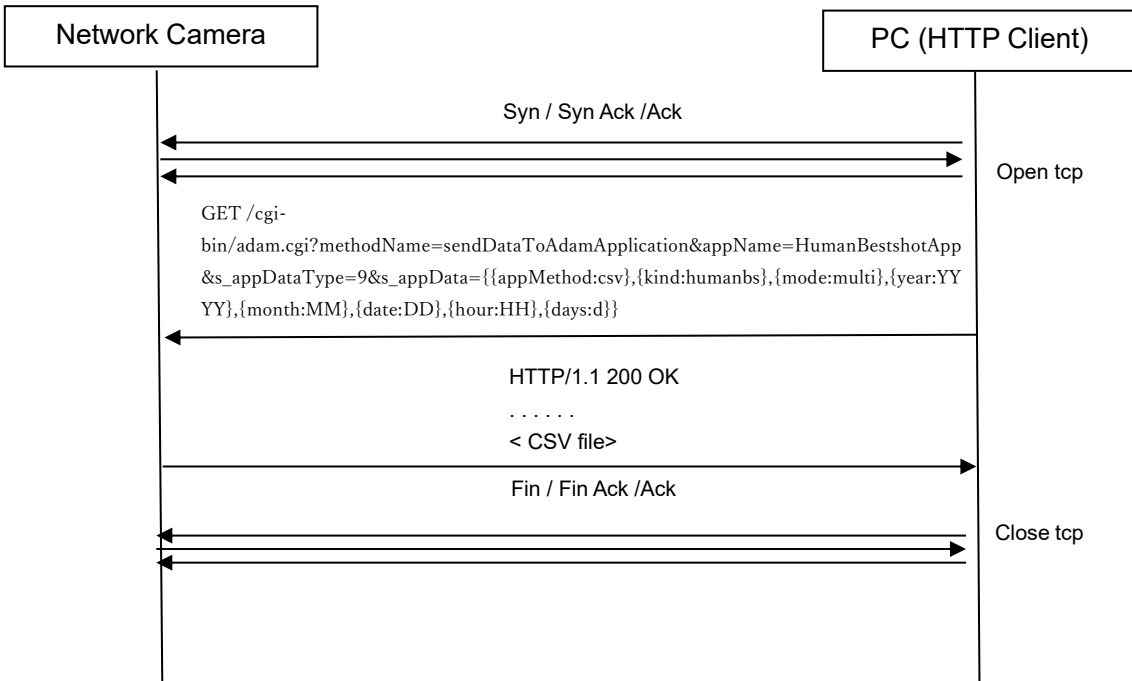
In case of the UTC time difference is +9 hours, specify {date:10},{hour:15} which is 9 hours before 4/11 00:00.

http://192.168.0.10/cgi-bin/adam.cgi?methodName=sendDataToAdamApplication&appName=HumanBestshotApp&s_appDataType=9&s_appData={{appMethod:csv},{kind:humanbs},{mode:multi},{year:2023},{month:04},{date:10},{hour:15},{days:1}}

*Apply base 64 encode to the preference value that follows "s_appData=".

If meta information exists within the number of days to be acquired from the date specified by CGI, all existing meta information is sent as csv files.

Sequence



[csv file name]

human_bs_YYYYMMDDHHMM2_yyyymmddhhmm2.csv

human_bs_ : fixed prefix

YYYY :

The year (4 digits) of the oldest meta information

MM :

The month (2 digits) of the oldest meta information

DD :

The day (2 digits) of the oldest meta information

HH :

The hour (2 digits) of the oldest meta information

MM2 :

The minutes (2 digits) of the oldest meta information

yyyy :

The year (4 digits) of the latest meta information

mm :

The month (2 digits) of the latest meta information

dd :

The day (2 digits) of the latest meta information

hh :

The hour (2 digits) of the latest meta information

mm2 :

The minutes (2 digits) of the latest meta information

e.g.) Get csv files for 3 days from 7:00:00 on April 20, 2011.

When the oldest meta information's localtime is April 20, 2011 7:30:00 and the latest meta information's localtime is April 23, 2011 5:00:00, csv file name is

human_bs _201104200730_201104230500.csv

Response format

```
HTTP/1.1 200 OK[CR][LF]
```

```
Status: 200[CR][LF]
```

```
Connection: close[CR][LF]
```

```
Content-Disposition:form-
data;name="data"filename="human_bs_YYYYMMDDHH_yyyymmddhh.csv"[CR][LF]
Content-Type: text/plain[CR][LF]
Content-Length: xxxx[CR][LF][CR][LF]
(meta data(csv))
```

3.1.2.3. Error response

Error response

```
HTTP1.1 200OK [CR][LF]
Status 200[CR][LF]
....
Content-Length: xxxxx[CR][LF]
xxxxxxxxxx[CR][LF]
[CR][LF]
```

Kind of the error	Content of xxxxxxxxxxxx
csv files doesn't exist.	No Data.
Abnormal CGI parameters (Value is out of range, required parameter is not assigned, etc.)	Parameter Error:xxxx *Give xxxx the parameter name that caused the error

4. Appendix

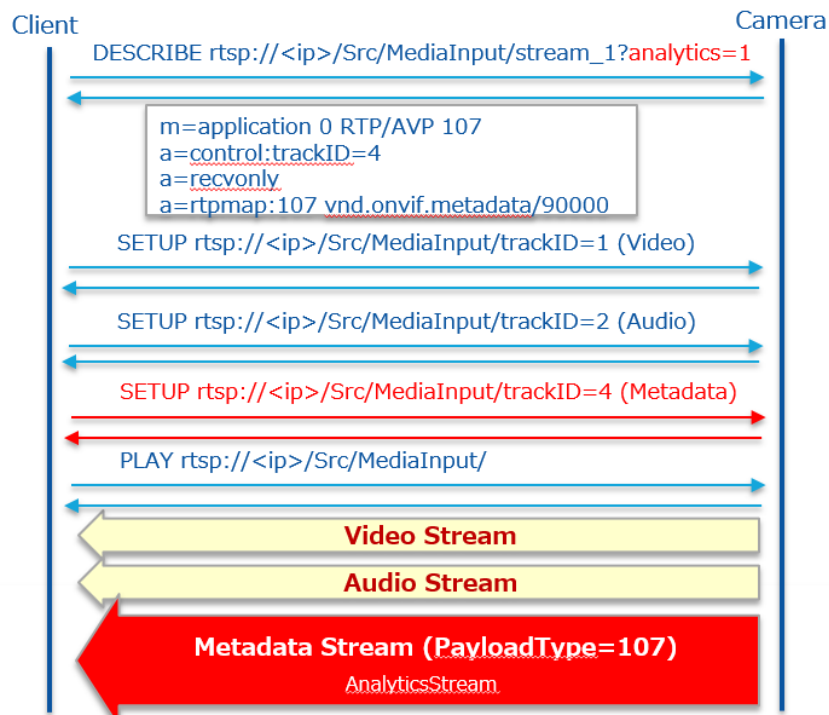
4.1. Transmission sequence of ONVIF Meta Stream

[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.



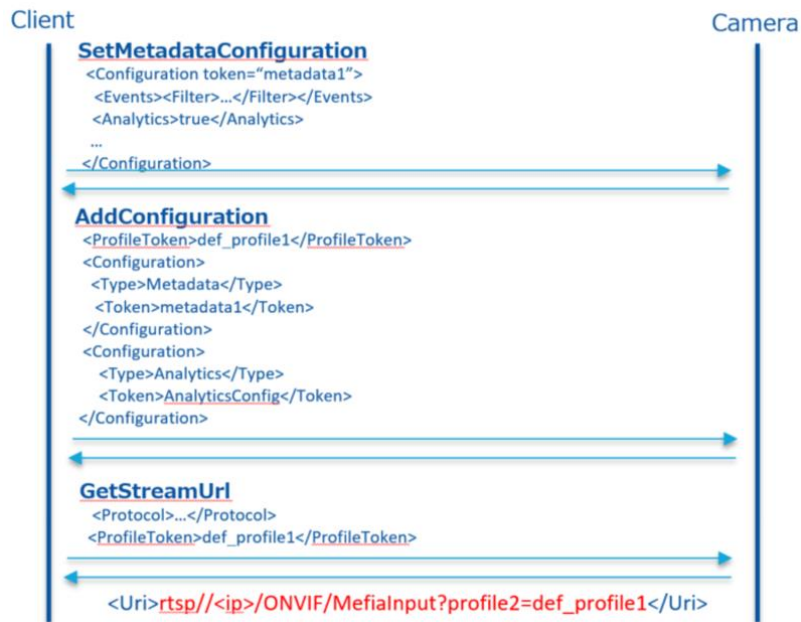
[ONVIF]

* 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

