

# Working Document

## Output Formats for Video Understanding

**Date: 11/10/2010**

---

**Version: 1.0**  
**Author: INRIA**

# OUTPUT FORMATS FOR VIDEO UNDERSTANDING

## 1 INTRODUCTION

This document describes the data output format (tracking output, fusion output, event output) used at INRIA, and describes also the required files and formats used for ViSEvAL tool developed at INRIA. The data transmitted from a module to another can be translated into an xml file to allow offline tests. This document describes each tag and attribute in this xml schema.

## 2 COMMON TO ALL OUTPUTS

Here is the structure of all the formats (xml1, 2, 3, 4).

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Cofriend version="2.0">
  Objects described in later chapters
</Cofriend>
```

## 3 TRACKING OUTPUT

The **frame to frame tracking output** is **tracking.id1**. It can be translated into xml format and named xml1 format.

**Example of tracking output:**

```
<?xml version="1.0" encoding="ISO-8859-1"?>
<Cofriend version="2.0">
<MobFrame frameID="34" timeYear="2008" timeMonth="11" timeDay="3" timeHour="16" timeMin="26"
timeSec="2" timeMs="810" camID="4" nbMobiles="1">
  <Mobile mobileID="1" confidence="2.5" type="PERSON1" subType="0">
    <TimeStart frameID="34" timeYear="2008" timeMonth="11" timeDay="3" timeHour="16" timeMin="26"
timeSec="2" timeMs="810" camID="4"/>
    <Track numParents="1">
      <ParentMobile mobileID="1" frameID="34" timeYear="2008" timeMonth="11" timeDay="3"
timeHour="16" timeMin="26" timeSec="2" timeMs="810" camID="4" confidence="1.00"/>
    </Track>
    <Blob>
      <CameraViews>
        <Info2D camID="Unknown" xCenter="628" yCenter="701" xCog="281" yCog="356"
width="664.00" length="318.50" Orientation="180" nbMovPix="866" xSpeed="45" ySpeed="61" frameID="34"
timeYear="2008" timeMonth="11" timeDay="3" timeHour="16" timeMin="26" timeSec="2" timeMs="810"/>
      </CameraViews>
      <Info3D x="-6.22" y="11.72" z="0.00" w3D="3.59" h3D="2.82" l3D="0.00" orientation="31"
xSpeed="5" ySpeed="65" zSpeed="14"/>
      <Uncertainty3D m11="0" m12="0" m13="0" m21="0" m22="0" m23="0" m31="0" m32="0"
m33="0" />
    <Properties>
      <Property name="ColorMean" value="126" />
      <Property name="Histogram" value="126,12,78,6787;78,47,9,85" />
    </Properties>
  </Blob>
</Mobile>
</MobFrame>
</Cofriend>
```

Each tag is described in the following:

## &lt;MobFrame

**frameID: identifier of the frame depending of the camera** [unsigned long]**timeYear: the year** [integer, from 0 to current year]**timeMonth: the month** [integer, from 1 to 12]**timeDay: the day** [integer, from 1 to 7]**timeHour: the hours** [integer, from 0 to 23]**timeMin: the minutes** [integer, from 0 to 59]**timeSec: the seconds** [integer, from 0 to 59]**timeMs: the milliseconds** [integer, from 0 to 999]**camID: the camera identifier** [integer, from 01 to 13 (see under)]**nbMobiles: the number of mobile objects in the frame** [integer, greater than 0]

&gt;

## &lt;Mobile

**mobileID: identifier of the mobile depending of the frame** [unsigned long]**confidence: indicator of the confidence of the detection: from bad confidence (mobile on the image borders, noise mobile, split or merged mobile, occlusion) to good confidence (well segmented and classified mobile)** [float, from 0 to 1]**type: type of the mobile object according to the defined types of the expected physical objects** [string: containing a list of possible types (GROUP\_OF\_PERSONS, UNKNOWN, PERSON1, PERSON2, LUGGAGE) and their confidence in percent ex: GROUP\_OF\_PERSONS:78- UNKNOWN:11-PERSON1:4-PERSON2:2- LUGGAGE:1]**subType: sub-type of the mobile object according to the defined subtypes of the expected physical objects types** [string: containing the list of possible subtypes (not applied)]

&gt;

## &lt;TimeStart

**frameID: first frame on which the mobile was detected previously** [unsigned long]**timeYear: the year** [integer, from 0 to current year]**timeMonth: the month** [integer, from 1 to 12]**timeDay: the day** [integer, from 1 to 7]**timeHour: the hours** [integer, from 0 to 23]**timeMin: the minutes** [integer, from 0 to 59]**timeSec: the seconds** [integer, from 0 to 59]**timeMs: the milliseconds** [integer, from 0 to 999]**camID: the camera identifier** [string, from 01 to 17 or “Unknown”]

&gt;

## &lt;Track

**numParent: number of the parent mobiles of the mobile (a parent is a mobile corresponding to the same physical object in a previous frame, the aim of the tracker is to create links between the same mobiles in successive frames)** [integer, from 0 to 100]

&gt;

## &lt;ParentMobile

**mobileID: identifier of the parent mobile** [unsigned long]**frameID: ID of the frame where this parent mobile was detected** [unsigned long, lesser than previous frame ID]**timeYear: the year** [integer, from 0 to current year]**timeMonth: the month** [integer, from 1 to 12]**timeDay: the day** [integer, from 1 to 7]**timeHour: the hours** [integer, from 0 to 23]**timeMin: the minutes** [integer, from 0 to 59]**timeSec: the seconds** [integer, from 0 to 59]**timeMs: the milliseconds** [integer, from 0 to 999]**camID: the camera identifier** [string, from 01 to 17 or “Unknown”]**confidence: indicator of the confidence of the link: from bad confidence (the parent mobile has a bad**

**matching with the current mobile) to good confidence (the parent mobile has a good matching with the current mobile)** [float, from 0 to 1]

>

<Blob

<CameraViews>

<Info2D

**camID: the camera identifier** [string, from 01 to 17 or “Unknown”]

**xCenter: x coordinate of the center of the 2D bounding box in the image** [float, from 0 to image size on X]

**yCenter: y coordinate of the center of the 2D bounding box in the image** [float, from 0 to image size on X]

**xCog: x coordinate of the center of gravity of the mobile (moving pixels density center) in the image** [float, from 0 to image size on X]

**yCog: y coordinate of the center of gravity of the mobile (moving pixels density center) in the image** [float, from 0 to image size on X]

**width: width of the 2D bounding box in the image** [float]

**length: length of the 2D bounding box in the image** [float]

**orientation: direction of the 2D bounding box along x axis** (optional) [float, from 0 to 179°]

**nbMovPix: number of moving pixels in the blob** [unsigned long, from 0 to bounding box area]

**xSpeed: horizontal velocity along X axis** [float, in pixels/sec]

**ySpeed: vertical velocity along Y axis** [float, in pixels/sec]

**frameID: identifier of the frame depending of the camera** [integer, greater than 0]

**timeYear: the year** [integer, from 0 to current year]

**timeMonth: the month** [integer, from 1 to 12]

**timeDay: the day** [integer, from 1 to 7]

**timeHour: the hours** [integer, from 0 to 23]

**timeMin: the minutes** [integer, from 0 to 59]

**timeSec: the seconds** [integer, from 0 to 59]

**timeMs: the milliseconds** [integer, from 0 to 999]

>

>

<Info3D

**x: x coordinate in centimeter of the bottom-centered point of the mobile in the 3D referential of the apron (point on the ground)** [float, from -n to n]

**y: y coordinate in centimeter of the bottom-centered point of the mobile in the 3D referential of the apron (point on the ground)** [float, from -n to n]

**z: z coordinate in centimeter of the bottom-centered point of the mobile in the 3D referential of the apron (point on the ground)** [float, from -n to n]

**w3D: width of the bounding box in the real world in centimeters, computed using calibration** [integer, from 0 to n]

**h3D: height of the bounding box in the real world in centimeters, computed using calibration** [integer, from 0 to n]

**l3D: length of the bounding box in the real world in centimeters, computed using calibration** [integer, from 0 to n]

**orientation: direction of the 3D bounding box along X axis** [float, from 0 to 359°]

**xSpeed: horizontal velocity along X axis** [float, in m/sec]

**ySpeed: horizontal velocity along Y axis** [float, in m/sec]

**zSpeed: vertical velocity along Z axis** [float, in m/sec]

>

<Uncertainty3D

( m11 m12 m13 )

( m21 m22 m23 )

( m31 m32 m33 )

**3D matrix of uncertainty**

```
>
<Properties
<Property
name: name of the property [string, can be anything]
value: value of the property [string, the separators have to be chosen by the writer and reader of this
property]
>
>
```

### Examples of properties:

```
<Properties>
  <Property name="motionOrientation" value="120" />
  <Property name="meanColor" value="122;45;200"/>
  <Property name="histogram" value="54;12;68;64;...;97" />
  <Property name="anyMatrixFirstRow" value="54;12;68;64;...;97" />
  <Property name="anyMatrixSecondRow" value="54;12;68;64;...;97" />
  ...
  <Property name="anyMatrixNthRow" value="54;12;68;64;...;97" />
</Properties>
```

The following image shows how we represent a 2D bounding box in the communication process.



Figure 1: 2D bounding box

## 4 FUSION OUTPUT

The **data fusion output** is the same as Tracking output, tracking.idl. It can be translated into xml format and named xml2 format. The tags are the same than in xml1. The only differences are that there can be several **Info2D** tags, one for each camera fused, and the attribute **bestCamID** is added to the **Mobile** tag.

## Example of fusion output:

```

<?xml version="1.0" encoding="UTF-8"?>
<Cofriend version="2.00">
<MobFrame frameID="3541" timeYear="2008" timeMonth="11" timeDay="11"
  timeHour="15" timeMin="29" timeSec="31" timeMs="960" nbMobiles="10">
  <Mobile mobileID="1" confidence="0.82" type="VEHICLE:98"
    subType="GROUND_POWER_UNIT:55-TRUCK:25-CONTAINER:12">
    <TimeStart frameID="42" timeYear="2008" timeMonth="11" timeDay="11"
      timeHour="14" timeMin="54" timeSec="29" timeMs="840"/>
    <Track numParent="1">
    <ParentMobile mobileID="1" frameID="42" timeYear="2008" timeMonth="11"
      timeDay="11" timeHour="15" timeMin="29" timeSec="31" timeMs="40"
      confidence="0.00"/>
    </Track>
  <Blob>
    <CameraViews>
    <Info2D camID="3" xCenter="604.00" yCenter="296.00" xCog="604.50"
      yCog="296.00" width="80.00" length="212.00" orientation="0.00"
      nbMovPix="5896" xSpeed="0.00" ySpeed="-0.00" frameID="21309"
      timeYear="2008" timeMonth="11" timeDay="11" timeHour="15" timeMin="29"
      timeSec="31" timeMs="640"/>
    <Info2D camID="4" xCenter="501.00" yCenter="227.00" xCog="501.00"
      yCog="227.00" width="112.00" length="190.00" orientation="0.00"
      nbMovPix="1271" xSpeed="0.00" ySpeed="-0.00" frameID="21296"
      timeYear="2008" timeMonth="11" timeDay="11" timeHour="15" timeMin="29"
      timeSec="31" timeMs="760"/>
    <Info2D camID="5" xCenter="556.00" yCenter="175.00" xCog="556.00"
      yCog="175.50" width="238.00" length="216.00" orientation="0.00"
      nbMovPix="2046" xSpeed="0.00" ySpeed="-0.00" frameID="21286"
      timeYear="2008" timeMonth="11" timeDay="11" timeHour="15" timeMin="29"
      timeSec="31" timeMs="850"/>
    <Info2D camID="7" xCenter="228.00" yCenter="312.00" xCog="228.00"
      yCog="312.00" width="96.00" length="166.00" orientation="0.00"
      nbMovPix="2798" xSpeed="0.00" ySpeed="-0.00" frameID="21280"
      timeYear="2008" timeMonth="11" timeDay="11" timeHour="15" timeMin="29"
      timeSec="31" timeMs="590"/>
    </CameraViews>
    <Info3D x="2.95" y="7.45" z="0.00" w3D="1.32" h3D="1.57" l3D="7.42"
      orientation="90.00" xSpeed="0.00" ySpeed="-0.00" zSpeed="0.00"/>
    <Uncertainty3D m11="320.63" m12="0.00" m13="0.00" m21="0.00"
      m22="320.63" m23="0.00" m31="0.00" m32="0.00" m33="0.00"/>
    <Properties/>
  </Blob>
</Mobile>
<Mobile mobileID="44" confidence="0.58" type="AIRCRAFT:77-VEHICLE:22"
  subType="AIRCRAFT:100|TRUCK:48-PASSENGER_BOARDING_BRIDGE:19-
  BULK_LOADER:16">
  <TimeStart frameID="749" timeYear="2008" timeMonth="11" timeDay="11"
    timeHour="15" timeMin="0" timeSec="45" timeMs="480"/>
  <Track numParent="1">
    <ParentMobile mobileID="44" frameID="749" timeYear="2008"
timeMonth="11"
    timeDay="11" timeHour="15" timeMin="29" timeSec="31" timeMs="40"
    confidence="0.00"/>
  </Track>
</Mobile>

```

```

</Track>
<Blob>
<CameraViews>
  <Info2D camID="2" xCenter="587.00" yCenter="408.00" xCog="587.98"
    yCog="408.50" width="156.00" length="288.00" orientation="0.00"
    nbMovPix="51" xSpeed="0.00" ySpeed="-0.00" frameID="21321"
    timeYear="2008" timeMonth="11" timeDay="11" timeHour="15" timeMin="29"
    timeSec="31" timeMs="960"/>
  <Info2D camID="3" xCenter="293.00" yCenter="396.00" xCog="293.50"
    yCog="396.00" width="170.00" length="472.00" orientation="0.00"
    nbMovPix="97" xSpeed="0.00" ySpeed="-0.00" frameID="21309"
    timeYear="2008" timeMonth="11" timeDay="11" timeHour="15" timeMin="29"
    timeSec="31" timeMs="640"/>
  <Info2D camID="4" xCenter="252.00" yCenter="400.00" xCog="252.50"
    yCog="400.00" width="228.00" length="456.00" orientation="0.00"
    nbMovPix="55" xSpeed="0.00" ySpeed="-0.00" frameID="21296"
    timeYear="2008" timeMonth="11" timeDay="11" timeHour="15" timeMin="29"
    timeSec="31" timeMs="760"/>
  <Info2D camID="5" xCenter="448.00" yCenter="456.00" xCog="448.50"
    yCog="456.50" width="164.00" length="258.00" orientation="0.00"
    nbMovPix="27" xSpeed="0.00" ySpeed="-0.00" frameID="21286"
    timeYear="2008" timeMonth="11" timeDay="11" timeHour="15" timeMin="29"
    timeSec="31" timeMs="850"/>
  <Info2D camID="7" xCenter="566.00" yCenter="418.00" xCog="566.50"
    yCog="418.50" width="186.00" length="250.00" orientation="0.00"
    nbMovPix="632" xSpeed="0.00" ySpeed="-0.00" frameID="21280"
    timeYear="2008" timeMonth="11" timeDay="11" timeHour="15" timeMin="29"
    timeSec="31" timeMs="590"/>
  <Info2D camID="8" xCenter="343.00" yCenter="479.00" xCog="343.17"
    yCog="479.52" width="190.00" length="424.00" orientation="0.00"
    nbMovPix="368" xSpeed="0.00" ySpeed="-0.00" frameID="21285"
    timeYear="2008" timeMonth="11" timeDay="11" timeHour="15" timeMin="29"
    timeSec="32" timeMs="40"/>
</CameraViews>
<Info3D x="-3.06" y="-14.90" z="0.00" w3D="12.37" h3D="5.72" l3D="17.66"
  orientation="255.00" xSpeed="-0.01" ySpeed="-0.01" zSpeed="0.00"/>
<Uncertainty3D m11="320.63" m12="0.00" m13="0.00" m21="0.00"
  m22="320.63" m23="0.00" m31="0.00" m32="0.00" m33="0.00"/>
<Properties/>
<Blob>
</Mobile>
</MobFrame>
</Cofriend>

```

## 5 EVENT OUTPUT

The **event output** is event.idl. The frame to frame output is named xml3, and the event output is xml4.

**Example event output:**

```

<SUVideoFrame frameID="187" timeYear="2008" timeMonth="11" timeDay="3" timeHour="16" timeMin="11"
timeSec="18" timeMs="410">
<ListMobileObjects>
  <Mobile mobileID="1" confidence="2.5" type="PERSON1:93-UNKNOWN:5" subType="0">
    <TimeStart frameID="34" timeYear="2008" timeMonth="11" timeDay="3" timeHour="16" timeMin="26"
timeSec="2" timeMs="810" camID="4"/>
    <Track numParents="1">
      <ParentMobile mobileID="1" frameID="34" timeYear="2008" timeMonth="11" timeDay="3"
timeHour="16" timeMin="26" timeSec="2" timeMs="810" camID="4" confidence="1.00"/>
    </Track>
    <Blob>
      <CameraViews>
        <Info2D camID="04" xCenter="628" yCenter="701" xCog="281" yCog="356" width="664.00"
length="318.50" Orientation="180" nbMovPix="866" xSpeed="45" ySpeed="61" frameID="34" timeYear="2008"
timeMonth="11" timeDay="3" timeHour="16" timeMin="26" timeSec="2" timeMs="810"/>
      </CameraViews>
      <Info3D x="-6.22" y="11.72" z="0.00" w3D="3.59" h3D="2.82" l3D="0.00" orientation="31"
xSpeed="5" ySpeed="65" zSpeed="14"/>
      <Uncertainty3D m11="0" m12="0" m13="0" m21="0" m22="0" m23="0" m31="0" m32="0"
m33="0" />
      <Properties>
        <Property name="ColorMean" value="126" />
        <Property name="Histogram" value="126,12,78,6787;78,47,9,85" />
      </Properties>
    </Blob>
  </Mobile>
<ListActivities>
  <Activity ID="292" name="Vehicle_Enters_Zone" confidence="1.00" type="PRIMITIVE_STATE"
AType="URGENT" AText="Zone_Scenario">
    <TimeStart frameID="34" timeYear="2008" timeMonth="11" timeDay="3" timeHour="16"
timeMin="26" timeSec="2" timeMs="810"/>
    <TimeEnd frameID="0" timeYear="1970" timeMonth="1" timeDay="1" timeHour="0" timeMin="59"
timeSec="59" timeMs="0"/>
    <ListActivityPhysicalObjects>
      <ActivityPhysicalObject type="OBJECT" ID="5" name="V_05" dynamicObjectID="0"/>
      <ActivityPhysicalObject type="STATIC_ZONE" ID="7" name="Right_FWD_TS_Zone"
dynamicObjectID="0"/>
    </ListActivityPhysicalObjects>
    <ListActivityCameraCtrl>
      <CameraControl controlType="2" controlPriority="3" controlDuration="652" objectID="5"
objectType="OBJECT" specificVisionTask="Refueling" camID="10"/>
    </ListActivityCameraCtrl>
  </Activity>
</ListActivities>
</SUVideoFrame>

```



Each tag is described in the following:

### <SUVideoFrame

frameID: [identifier of the frame depending of the camera](#) [unsigned long]  
 timeYear: [the year](#) [integer, from 0 to current year]  
 timeMonth: [the month](#) [integer, from 1 to 12]  
 timeDay: [the day](#) [integer, from 1 to 7]  
 timeHour: [the hours](#) [integer, from 0 to 23]  
 timeMin: [the minutes](#) [integer, from 0 to 59]  
 timeSec: [the seconds](#) [integer, from 0 to 59]  
 timeMs: [the milliseconds](#) [integer from 0 to 999]  
 camID: [the camera identifier](#) [string, from 01 to 17 or “Unknown” (see under)]  
 nbMobiles: [the number of mobile objects in the frame](#) [integer, greater than 0]

> <ListMobileObjects> <

### Mobile

The tags are identical to those in Tracking output.

>

### <ListActivities>

#### <Activity

ID: [activity instance identifier](#) [unsigned long]  
 name: [activity name](#) [string]  
 confidence: [confidence in the activity](#) [float, from 0 to 1]  
 type: [activity type](#) [string, one of {PRIMITIVE\_STATE, COMPOSITE\_STATE, PRIMITIVE\_EVENT, COMPOSITE\_EVENT}]  
 AType: [activity alarm type, defines the priority of the activity, allows to output only activities superior to a threshold](#) [string, one of {NOTURGENT, URGENT, VERYURGENT, SUPERURGENT}]  
 AText: [activity alarm type, allows to output a message](#) [string]

>

#### <TimeStart

frameID, timeYear, timeMonth, timeDay, timeHour, timeMin, timeSec, timeMs: [frame and time of the first detection of the activity.](#)

>

#### <TimeEnd

frameID, timeYear, timeMonth, timeDay, timeHour, timeMin, timeSec, timeMs: [frame and time of the last detection of the activity. Put to 0 while the activity is still occurring \(the timestamp 0 is in 1970\). A state is considered ended when not recognized for a number of frames superior to a threshold.](#)

>

### <ListActivityPhysicalObjects>

#### <ActivityPhysicalObject

type: [type of parameter for this activity](#) [string, one of {OBJECT, STATIC\_ZONE, DYNAMIC\_ZONE}]  
 ID: [identifier of the parameter](#) [unsigned long]  
 name: [name of the parameter](#) [string]  
 dynamicObjectID: [if the attribute type is “DYNAMIC\\_ZONE”, identifier of the zone it refers to](#) [unsigned long]

>

### <ListActivityCameraCtrl>

#### <CameraControl

controlType: [type of control to take on the camera](#) [unsigned int, equiv to enum {CENTER, CENTER\_ZOOM, FIXE, TRACK, TRACK\_ZOOM, ...}].

controlPriority: [priority of the control to take on a camera](#) [unsigned int, equiv to enum {LOW, MEDIUM, HIGH,...}]

controlDuration: [time to block the PTZ camera, i.e. time allowed to the specific vision task. If 0, waits for an event to stop \(i.e. an activity with higher priority, or the end of the specific vision task\)](#) [float, in sec]

objectID: [identifier of the object to focus on](#) [unsigned long]

objectType: [type of the object to focus on](#) [string, one of {OBJECT, STATIC\_ZONE, DYNAMIC\_ZONE}]

specificVisionTask: [name of the specific vision task to run](#) [string]

camID: [identifier of the camera to run the specific vision task on. "00" is if the camera has to be chosen by a strategy](#) [string, from 00 to 17]

>

>