|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ITU logo | INTERNATIONAL TELECOMMUNICATION UNION  **TELECOMMUNICATION**  **STANDARDIZATION SECTOR**  STUDY PERIOD 2022-2024 | | | SG12-Cn |
| STUDY GROUP 12 |
| Original: English |
| **Question(s):** | | [19/12] | [San Mateo, 26-27 June 2023] | |
| **CONTRIBUTION** | | | | |
| **Source:** | | NTT | | |
| **Title:** | | Draft Terms of Reference (ToR) P.obj-recog | | |
| **Contact:** | | Masanori Koike  Japan  NTT | Tel: +81-422-59-4453  Fax: +81-422-59-6364  E-mail: masanori.koike@ntt.com | |
| **Contact:** | | Kazuhisa Yamagishi  Japan  NTT | Tel: +81-422-59-4397  Fax: +81-422-59-6364  E-mail: kazuhisa.yamagishi@ntt.com | |
| **Contact:** | | Noritsugu Egi  Japan  NTT | Tel: +81-422-59-4008  Fax: +81-422-59-6364  E-mail: noritsugu.egi@ntt.com | |

Abstract

The Terms of Reference for the P.obj-recog model is specified in this document.

1. Scope

P.obj-recog provides an object-recognition-ratio estimation model to check whether the observer of the monitoring center can achieve object recognition by using the surveillance video.

To ensure the safety of autonomous driving, objects that interfere with driving need to be automatically recognized. To do that, an object-recognition system is needed to support actions of autonomous driving, such as autonomous braking and passing. In general, object recognition is performed automatically in autonomous driving systems using cameras mounted on an autonomous car. Under certain conditions, such as on a highway, autonomous driving systems work well without any human support. On the other hand, current autonomous driving systems are challenging to use on local streets because there are various objects, such as traffic signs or pedestrians.

To use the current autonomous driving systems on local streets, a remote monitoring system has also been studied in which an observer at a monitoring center recognizes objects and brakes remotely. In some countries, object recognition is needed for autonomous driving to ensure safety. In Germany, legislation on autonomous driving has been developed, indicating that remote monitoring is necessary [1]. Japan revised its road traffic law, allowing unattended autonomous vehicles to operate under remote supervision on specific routes in depopulated areas [2]. In France, the transport law was revised to specify means for monitoring the safety of the unattended autonomous driving systems [3]. Moreover, remote autonomous driving of transportation vehicles (e.g., bus and taxi) requires interior vehicle monitoring to watch for a passenger behaving suspiciously or falling over.

The primary application for P.obj-recog is the monitoring in which a surveillance video with a high object-recognition ratio is always being provided to a monitoring center. To do that, an object-recognition-ratio estimation model needs to be developed to estimate the object-recognition ratio of surveillance video at the remote monitoring center. The output of P.obj-recog model is the object recognition ratio (from 0 to 100 %) for a target distance assuming that the object is located away from a surveillance camera. Considering the purpose of the monitoring, the target distance is the distance at which the autonomous car can be stopped short of the object after detection. Note that even when there is no object in the surveillance video, remote monitoring providers want to know whether a surveillance video with a high object-recognition ratio is always provided.

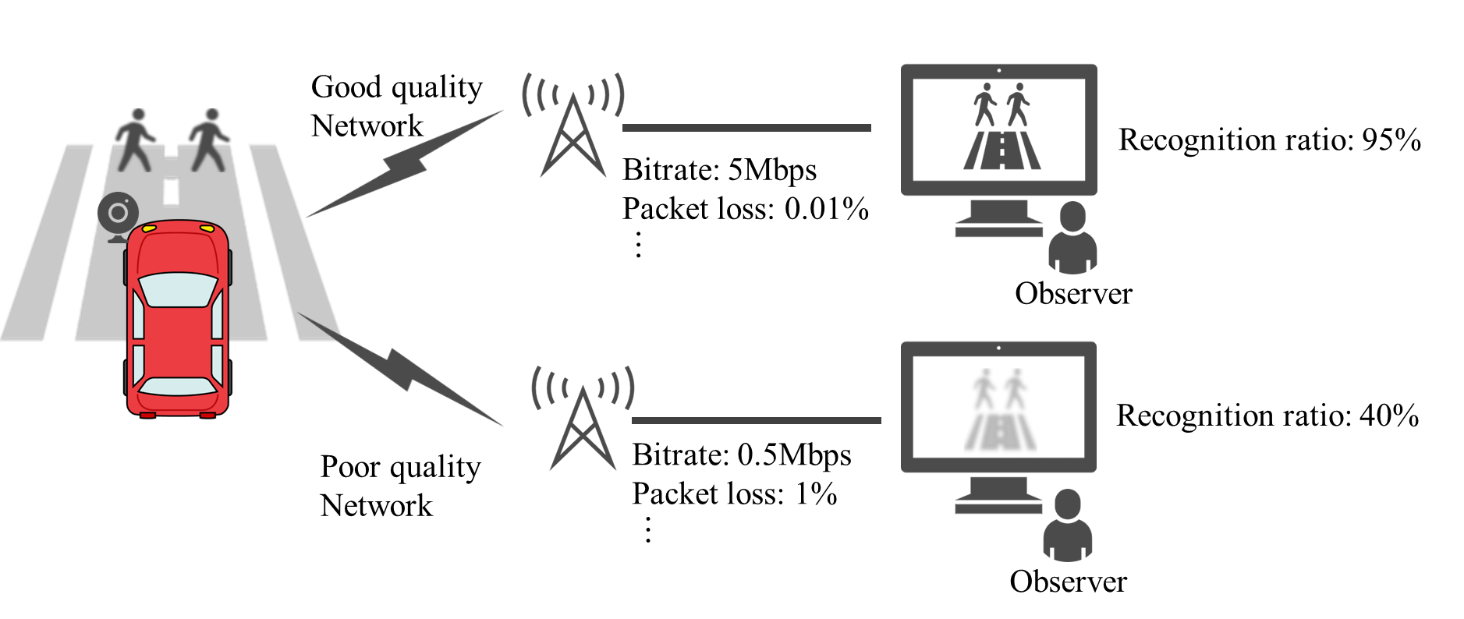
1. General application area

The application area for P.obj-recog is:

* A remote monitoring system with a remote monitoring center in which surveillance video in an autonomous car is encoded and transmitted via radio access networks. Note that an autonomous car includes one or more surveillance cameras and transmits the surveillance video.
  + A remote monitoring system is based on the protocols TCP/IP, RTP/UDP/IP, UDP/IP. Note that the model is agnostic to the specific network delivery method (HTTP or DASH or other), except that it assumes reliable delivery (TCP/IP or QUIC/UDP/IP).

## 2.1 General use case

P.obj-recog is an object-recognition-ratio estimation model for a monitoring center based on codec, network, car, and source-related parameters calculation. P.obj-recog is meant to be helpful in monitoring a surveillance video with a high object-recognition ratio for assessing the effects of quality degradation of surveillance video, network degradation, the autonomous car’s velocity, and the surrounding environment of the autonomous car. The example use case of this monitoring system is shown in Figure 1. Figure 1 shows two patterns of network quality. When the network quality is good, the surveillance video is transmitted in high quality, and the object-recognition ratio is high, which can be confirmed at the monitoring center. However, if the network quality is poor and the video quality is not good, the object-recognition ratio is low.



**Figure 1: Use case of the monitoring system**

1. Application range to which the model shall be applicable

| **Application information** | **Value range, unit** | **Comments** |
| --- | --- | --- |
| Display size and resolution | Screen size is larger than 14 inches and smaller than 50 inches, with device resolution equal or not much higher than 1080p. | In the actual operation, multiple surveillance videos from multiple autonomous cars may be displayed on one display. In this contribution, the case where one video is displayed on one display is covered. |
| The number of transmitted videos | One | In the actual operation, multiple surveillance videos from cameras in multiple directions may be transmitted. In this contribution, the case where one video is transmitted is covered. |
| Interior and exterior surveillance camera | Exterior | In this contribution, the surveillance video for outside is covered. |
| Video size and frame rate | Up to 1080p,  Up to 60 fps |  |
| Video codec | H.265: Main Profile |  |
| Codec video bitrate | 200 kbps – 5 Mbps |  |
| Group of pictures (GoP) | Variable, fixed adaptive length. |  |
| Packet-loss rate | Up to 10 % |  |
| Packet-loss type | Freezing with skipping, block noise |  |
| Autonomous car’s velocity | Up to 40 km/h | This is only to specify the range of qualities that will probably be tested in this project. |
| Unit time for video and network information acquisition | 0.1 to 10 seconds |  |

**Table 1: Overview of the application range of P.obj-recog**

## 3.1 Modes of operation

P.obj-recog model will consist of two modes of operation for P.obj-recog, which are defined Table 2.

Table 2: P.obj-recog modes of operation

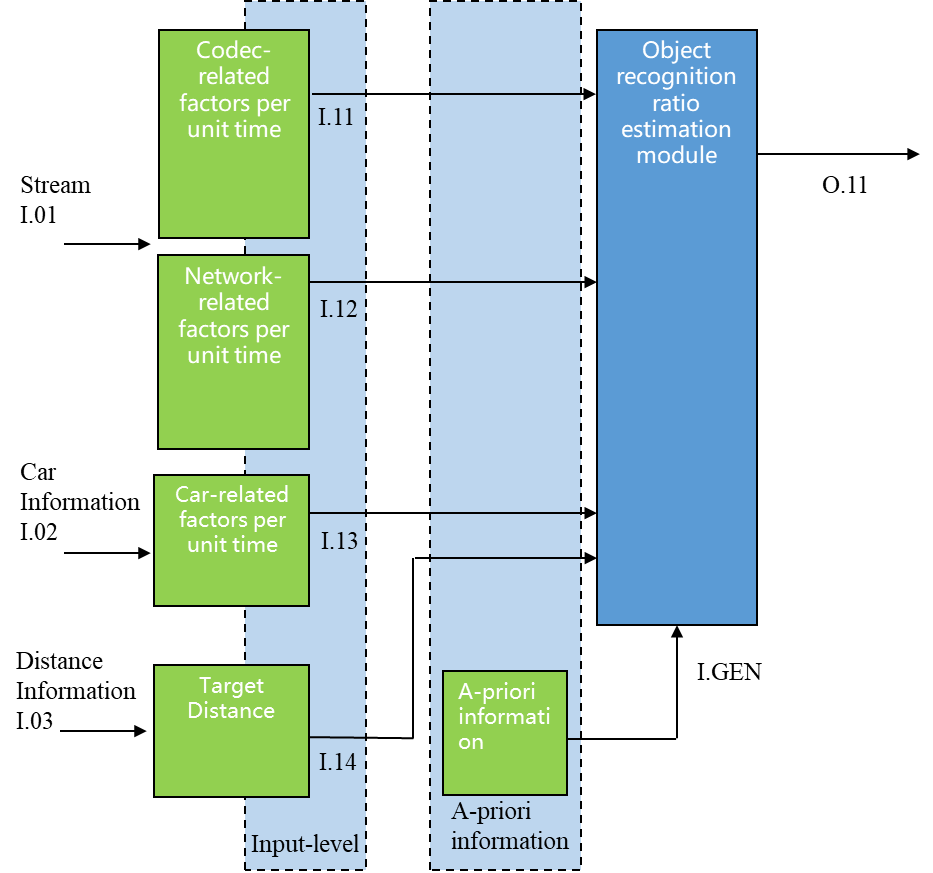
|  |  |
| --- | --- |
| **Mode** | **Input** |
| 0 | Meta-data related to network and video. |
| 1 | Autonomous car’s velocity as well as input mode 0. |

## 3.2 Application range to be applied in the future

This technique covers cases where one video is displayed on one display and one video is transmitted. Also, this technique targets the exterior surveillance camera to monitor outside of the autonomous car. In the following technique version, the cases need to be considered where multiple surveillance videos from multiple autonomous cars may be displayed on a display and multiple surveillance videos from cameras in multiple directions are transmitted. Moreover, interior vehicle monitoring in the case of a passenger behaving suspiciously or falling over will be covered in the following version of this technique.

1. Building Blocks

The module layout of the P.obj-recog is depicted in Figure 1 below.



**Figure 1: Building blocks of the P.obj-recog model**

The main factors affecting the object recognition ratio are codec-related factors, network-related factors such as packet-loss ratio and error concealment, car-related factors such as the car’s velocity, and source-related factors such as color, size, background of objects, weather, and day and night. The surveillance video differs from autonomous driving situations, and the kinds of surveillance video displayed and kinds of objects recognized are different. Therefore, P.obj-recog uses codec-related, network-related, and car-related factors as input information in the estimation and uses source-related factors as a priori information. Target distance assuming that the object is located away from a surveillance camera is also input.

## 4.1 Model Inputs

The model receives the following input signal:

### I.11: Codec-related factors

Video resolution (pels); video bitrate (kbps); video frame rate (fps) per unit time.

### I.12: Network-related factors

Video packet-loss ratio per unit time.

### I.13: Car-related factors

Autonomous car’s velocity per unit time. Note that this factor is used only in mode 1.

### I.14: Target distance

Target distance assuming that the object is located away from a surveillance camera.

### I.GEN: A priori information for estimation module

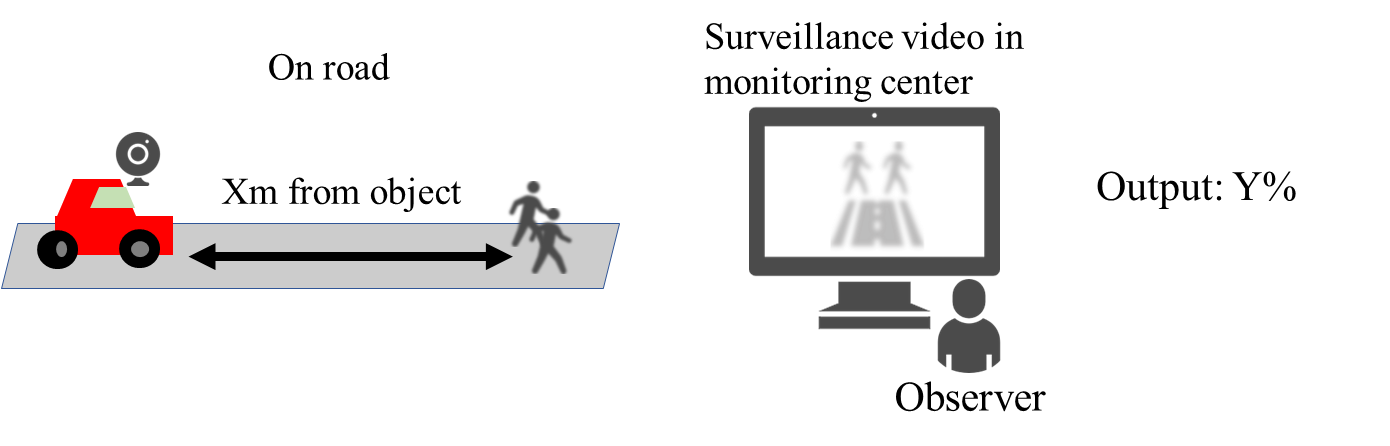
This input is a priori information for the estimation module. This information includes the following:

* Packet-loss error concealment
  + Freezing with skipping
  + Block noise
* Source-related factors
  + Object color
  + Object size
  + Object background
  + Weather
  + Day and night

## 4.2 Model outputs

### O.11: Output object recognition ratio

The recognition ratio varies depending on the distance from the object. Therefore, the object-recognition ratio is the probability of recognizing an object Y% of the time, assuming that the object is located X meters away from a surveillance camera. The output image of the object recognition ratio is shown in Figure 2. The output of the O.11 is object recognition ratio Y %. Recognition probability in the subjective assessment methods for the recognition task described in P.912 [4] can be applied.



**Figure 2: Output object recognition ratio**

1. References

[1] Entwurf eines Gesetzes zur Änderung des Straßenverkehrsgesetzes und des Pflichtversicherungsgesetzes – Gesetz zum autonomen Fahren, 2021

[2] Initiatives for Realization of Automated Driving by Japan Police, ITS World Congress, 2022

[3] The French strategy for the development of automated road mobility 2020-2022, 9th GRVA, 2021

[4] ITU-T P.912, Subjective video quality assessment methods, 2016

\_\_\_\_\_\_\_\_\_\_\_\_\_\_