|  |  |
| --- | --- |
|  | Moving Picture, Audio and Data Coding  by Artificial Intelligence  www.mpai.community |

|  |  |
| --- | --- |
| **Public Document** | |
| **N570** | 2022/02/23 |
| **Source** | Requirements (OSD) |
| **Title** | MPAI-OSD report and plans |
| **Target** | MPAI Members |

[1 Introduction 1](#_Toc96266332)

[2 Description of Use Cases 1](#_Toc96266333)

[2.1 Audio Tape Irregularity 1](#_Toc96266334)

[2.1.1 MPAI-CAE-ARP: Audio Recording Preservation 1](#_Toc96266335)

[2.2 Identify object in a human’s hand 2](#_Toc96266336)

[2.2.1 MPAI-MMC-MQA: Multimodal Question Answering. 2](#_Toc96266337)

[2.2.2 MPAI-CAV-HCI: Human-CAV Interaction 2](#_Toc96266338)

[2.2.3 Conversation About a Scene 2](#_Toc96266339)

[2.3 Detecting emotion and meaning in human face 2](#_Toc96266340)

[2.3.1 MPAI-MMC-CWE: Conversation with Emotion 2](#_Toc96266341)

[2.4 MPAI-CAV 2](#_Toc96266342)

[2.4.1 MPAI-MCS: Mixed-reality Collaborative Spaces 3](#_Toc96266343)

[2.5 Tracking video game player’s movements 3](#_Toc96266344)

[2.6 Correct Posture 3](#_Toc96266345)

[2.7 Integrative genomic/video experiments (animals) 3](#_Toc96266346)

# Introduction

Visual object and scene description (MPAI-OSD) is an MPAI project at the Use Case stage. The current goal of the project is to collect Use Cases sharing the goal of describing visual objects and, in some cases, locate them in the space.

By scene description we mean the usual description of objects and their attributes in a scene and the semantic description of the objects.

AIMs in the MPAI-OSD area have already been requested in Conversation with emotion, Mul­timodal Conversation, Audio Recording Preservation.

New use cases are constantly identified that require new AIMs falling in the MPAI-OSD scope.

This document does not contain new elements compared to the document of MPAI-14.

# Description of Use Cases

## Audio Tape Irregularity

This belongs to the family of generic object description.

### MPAI-CAE-ARP: Audio Recording Preservation

MPAI is using this component for Audio Tape Irregularity Detection.

1. Receive the video signal of a camera pointing to the magnetic reading head of a traditional audio tape.
2. Detect the images that show irregularities on the tape.
3. Provide as output, if an image shows an irregularity:
   1. The image
   2. The type of irregularity
   3. The time code

## Identify object in a human’s hand

This belongs to the family of generic object description.

### MPAI-MMC-MQA: Multimodal Question Answering.

It is designed to:

1. Receive the picture of an object.
2. Recognise the type of object.
3. Provide the object identifier as output.

### MPAI-CAV-HCI: Human-CAV Interaction

Similar to the above. A CAV passenger holds an object in their hand and the CAV recognises it and uses its identifier for further actions.

### Conversation About a Scene

A human is part of a scene containing objects. The human asks a machine questions about an object using their finger.

## Detecting emotion and meaning in human face

This belongs to the family of human description.

### MPAI-MMC-CWE: Conversation with Emotion

It is designed to:

1. Receive a video of the face of a human.
2. Identify the type and intensity of the emotion in the face of the human.
3. Provide as output:
   1. The type of emotion out of a finite set of codified emotions.
   2. The intensity (grade) of the emotion.

Version 2 of MMC-CWE considers the time stamp a new emotion appears and the duration of the transition between the earlier and current emotion.

## MPAI-CAV

A Connected Autonomous Vehicle (CAV) performs the following visual functions in several instances:

1. In the Human-CAV Interaction Subsystem
   1. Outside of CAV: Captures the scene outside of CAV, separates the human object from the rest of the scene, separates the face object, recognises the identity of the human object.
   2. Inside the CAV: Captures the scene in the cabin, separates the human objects from the rest of the scene, locates the passengers, recognises their identity, extracts emotion, direct the gaze of the avatar to the particular passenger it is holding a dialogue with.
2. In the Environment Sensing Subsystem:
   1. Captures visual information using Cameras, Lidar, Radar, Ultrasound (potentially only some of them).
   2. Receives visual information from on-line maps.
   3. Receives position information from GNSS and Motion Activation Subsystem.
   4. Creates object-based description of each captured scene.
   5. Fuses object-based descriptions into Basic World Representation.
3. In the Autonomous Motion Subsystem:
   1. Receives Basic World Representation from Environment Sensing Subsystem.
   2. Receives Basic World Representations from CAVs in range.
   3. Fuses Basic World Representations into its Full World Representation.

### MPAI-MCS: Mixed-reality Collaborative Spaces

MPAI-MCS is currently addressing the following scenario where:

1. Participants use MCS clients to send an MCS server their data, including their visual descriptors of head, face, arms, hands, fingers.
2. The MCS server distributes the data to the MCS clients.
3. MCS clients create local virtual 3D audio-visual spaces populated by avatars whose face and head are animated by the received descriptors.
4. Participants can watch their virtual spaces from a viewpoint of their choice.

## Tracking video game player’s movements

This Use Case belongs to the family of human description.

It is a system designed automatically to understand the game player’s physical movements in a video game. The features of the movements are:

1. The human object is largely static, and only hand/arm and finger movements are detected.
2. The types of movements are limited in number.
3. The system should understand the movements fast and accurately.

The system is designed to:

1. Receive a video.
2. Compute descriptors of the human.
3. Understand the inten­tion expressed by the movements from the descriptors.

## Correct Posture

This Use Case belongs to family of human description.

It is a system designed to advise the user by suggesting how they should correct their pose.

The main features of this Use Case are

1. The human using the application walks in a restricted environment.
2. Very specific type movements must be detected with high accuracy.
3. The detected movements are compared with reference movements.

The system is designed to:

1. Receive a video.
2. Compute descriptors of the human.
3. Compares the descriptors with reference descriptors.
4. Provide suggestions about movement corrections.

## Integrative genomic/video experiments (animals)

This belongs to the family of animal description (however, see later).

MPAI is using this component in several MPAI-GSA Use Cases.

It is a system designed to

1. Receive a sequence of images containing laboratory animals with specified genomic data of which the effects on behavioural patterns are assessed.
2. Compute behavioural patterns of living organisms, e.g., measures of the parameters of animal activity, everywhere and/or in specified ROI, such as:
   1. Distance.
   2. (average) velocity.
   3. Acceleration.
   4. Time spent.
   5. Time spent near walls.
   6. Trajectories.
   7. Turning speed.

There is another component that belongs to the family of plant description.