May 31, 2019

A SMART TASK ASSISTANT TO INSTRUCT USERS IN HOW TO PERFORM A TASK BASED ON MACHINE LEARNING TECHNIQUES BY VOICE

HP INC

A Smart Task Assistant to instruct users in how to perform a task based on Machine Learning techniques by voice.

Abstract

A voice assistant that guide users with different levels of knowledge through an operation or repair. The proposed solution is a camera that can be attached to the user's security glasses or in a headset that will transmit the Point of View (POV) and receive instructions about what is the next action that the user needs to perform to complete a given task. A machine learning algorithm is responsible to process all the data input and give instructions to the user about how to perform an action. The instructions are transmitted to the user's ear-piece.

This technology streamlines the maintenance processes and could be used to give instruction to a user adapting to the current problem. One example of application is the maintenance process of GSB printers. Currently there is the necessity of knowledge pass from a highly specialized technician (that sometimes may not be available or is too expensive) to another so it can be capable of fixing a specific issue. With our technology anyone with minimal training would able to fix almost any issue since it will be guided through the process by a virtual personal assistant or even a real expert in the subject.

Description

The system has a hardware piece to receive and transmit the information to a middleware and software that will be responsible to make predictions about what is the best next action to solve that specific task in that specific moment. The figure 1 explains the hardware that the user will use to receive and transmit data.

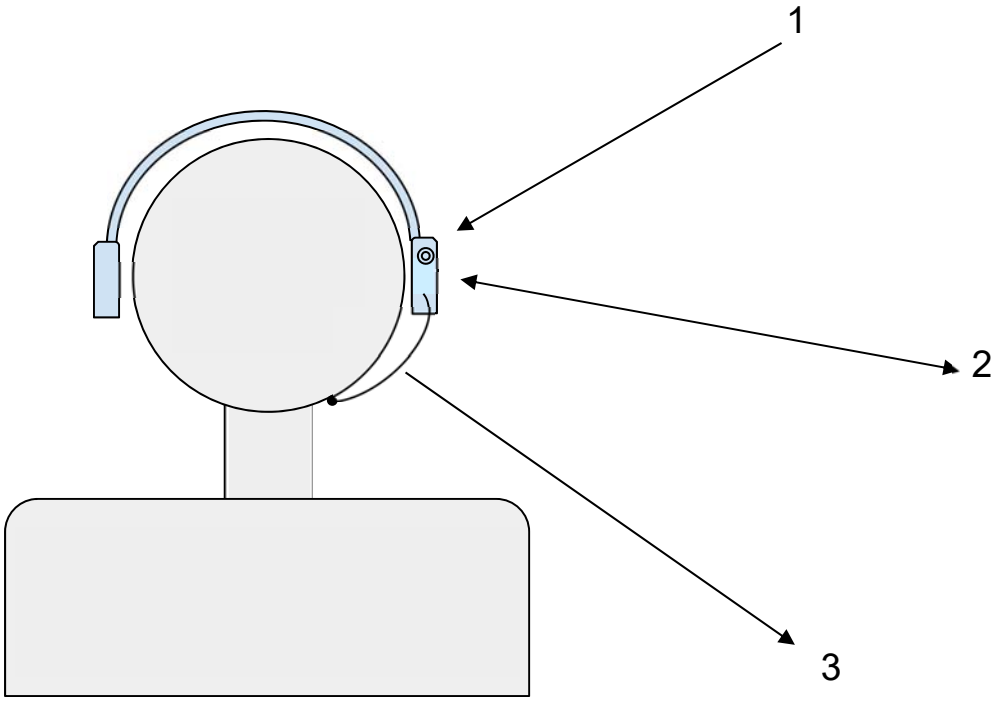

Figure 1 - Example of a possible implementation with a headset.

The device would have to have 3 major components that would handle the communication and the input and output of the processed data.

1- Camera capturing what the user is seeing. The camera should have a view angle similar to the user.

2- Headset connected to the server (cloud or local) to process the video inputs and the process the next action for the user.

3- Microphone to receive user voice inputs.

This solution requires a hardware to capture the video and audio streaming, and an audio device to send the instructions to the user. The hardware will attach or build to a device for example: security glasses, headsets or a dedicated device. The camera is remotely connected to a middleware and then to an earpiece. The camera sends its feed to the middleware, that is local computer or cloud one or even the "headset" itself (considering the recent advances in machine learning at the edge field). The middleware is responsible for all the instructions that are generated for the user, in case of error the middleware will find a countermeasure and inform the user. All this information will be relayed to the ear-piece, lastly the ear-piece will output the information via a small speaker.

The middleware will accommodate a neural network that can be trained with the routines that will be performed, a routine is a set of instructions that the user will follow through. The model will need to be trained to understand the context that the user is so it's able to identify if an instruction was already completed. The context consists but it's not limited to: where the user is, what is the state of the task right now, what was the user's last action, what part the user is holding (if holding any), where and how the user is trying to fit the aforementioned part. If the user followed the instruction the next instruction is passed, in case the user makes a mistake the middleware will then infer what the user did wrong for

example, picking the wrong part, putting the part on the wrong place, the part is loose and then will generate a solution for the user before passing to the next step.

The method to decide what is the best instruction that the user needs to perform next is based on neural networks that use action-prediction. So, the idea is to create a neural network structure and do separated training for each task, for example, if we want to create a "training" to replace a cartridge in a printer: We would record experts doing this task with a camera attached in the same angle as the real device. After collecting several experts doing these tasks in different printers and models and with different strategies, we would use that video data to train our neural network to predict the next action. Based on the network previous experiences (with the experts) it will know what is the best action that the user should perform if he was an expert. And this will be our recommendation of action based on what the user just did.

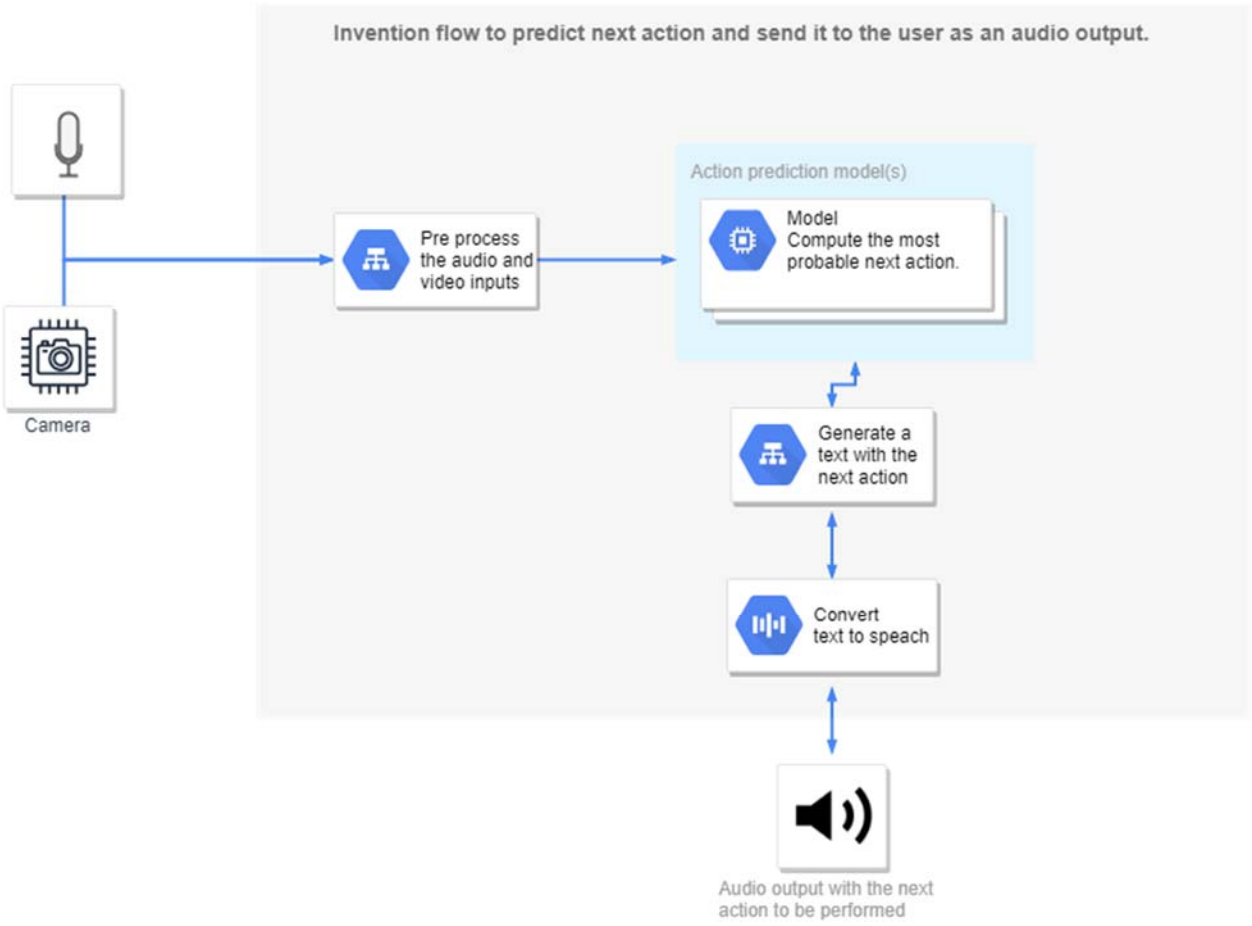

Figure 2 - Description of the dataflow and how the prediction model is integrated in the flow.

(Optional) All the maintenance process is recorded through the user's camera. This information can then be used to further train the neural network to increase its accuracy and for any maintenance audition.

Disclosed by Carlos Leão, Lucio Cossio, Rodrigo Ramos and Matheus Bauer, HP Inc.