January 2023

Automatic Video Conference Session Transfer Based on Presence Detection and Wearable Device Identification

D Shin

Automatic Video Conference Session Transfer Based on Presence Detection and Wearable Device Identification

ABSTRACT

When a video conference participant using a desktop or other fixed device temporarily leaves their location, they lose access to the full conferencing experience, unless they take manual action to join the video conference via a mobile or wearable device. This disclosure describes techniques that, with user permission, automatically detect when a user leaves a physical location from where they have joined a video conference. The techniques further detect mobile and/or wearable devices available to the user and automatically configure such devices for video conferencing. A link to join the video conference via such a device is provided to the user, and upon user selection, the video conference session is seamlessly transferred to the available device.

KEYWORDS

- Video conferencing
- Interruption
- Person detection
- Object detection
- Motion detection
- Wearable device

BACKGROUND

A video conference participant may occasionally need to temporarily leave the physical space, e.g., work desk, from where they participate in a video conference. Such participants lose access to the full conferencing experience. For example, a user that has walked to a different room, being away from the video conferencing device, may not be able to view or listen to other participants via the device display/speakers. A user that has moved to a remote corner of the same room may not be able to provide input since they are far from the device microphone. While a participant can temporarily join the conference from a mobile device to retain access, there is no easy method to transfer a video conference to an appropriate device based on the user's location when a user temporarily leaves their desk.

DESCRIPTION

This disclosure describes techniques, implemented with user permission, to automatically detect user presence based on a presence sensor, e.g., a device camera. The techniques further detect whether a user has access to a wearable device (e.g., earbuds or other head-worn device) or a mobile device (e.g., a smartphone) that has the capability to join the video conference.

For example, consider a scenario where a user first joins a video conference from a desktop computer with fixed microphone and speakers. The user temporarily leaves the location (e.g., to brew a cup of coffee) carrying their smartphone. The described techniques detect that the user has left the location but is carrying the smartphone and automatically surface a link to the video conference on the smartphone that the user can select to join the video conference seamlessly. Upon the user selecting the link, the display, microphone, and speaker of the smartphone are used to enable the user to participate in the video conference. If the smartphone

is paired with earbuds worn by the user, the earbuds are used for the audio of the video conference. The user can thus continue to seamlessly participate in the video conference from a different device, without interruption. A similar approach can be utilized if the user is detected as carrying a tablet, a wrist-worn device, a head-worn device (e.g., an augmented reality headset), or another mobile device.

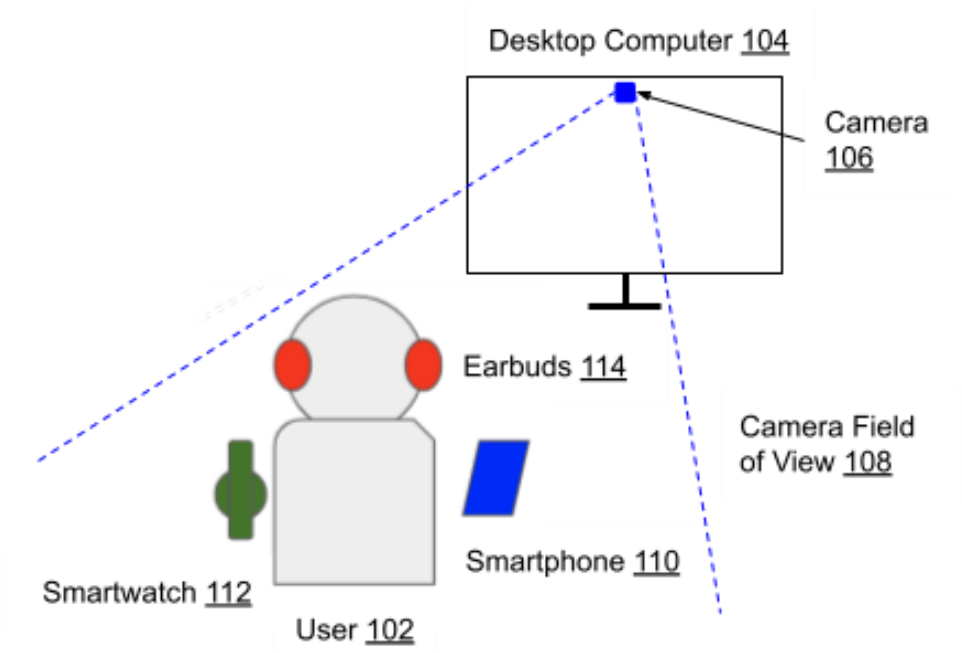

Fig. 1: User participating in a video conference

Fig. 1 illustrates an example of implementation of the techniques to automatically initiate transfer of a video conference. A user (102) participates in a video conference via a desktop computer (104). In this instance, the user is using earbuds (114) to listen to the audio and a desktop microphone to provide speech input to the video conference.

With user permission, a camera (106) of the desktop computer (or coupled to the desktop computer) is used to detect whether the user is present within the field of view (108). The raw camera feed from the desktop, while being used to generate the video for the video

conference, is analyzed using computer vision techniques. For example, each camera frame is analyzed using person detection techniques (e.g., a person detection neural network) that is responsible for understanding human pose and cues around movement. A cue for "scene exit" by a person can be determined from skeletal keypoint models generated from pose networks used for human detection. For example, it may be detected that the user is standing up.

After determining that the user is likely to exit the scene, on-device techniques (e.g., a device detection neural network) can be used to detect wearables (e.g., smartwatch 112) or mobile devices (e.g., smartphone 110) that the user is carrying during exit. The user can provide information regarding their devices (e.g., as part of a multi-device ecosystem). This information can be used to match the detected wearables to the specific set of devices used by the user. Once visual detection and identification of devices carried by the user is completed, the input and output of the video conference is reassigned to available devices based on user preferences.

With reference to the example illustrated in Fig. 1, an audio prompt is provided via the earbuds, asking the user to confirm whether the user would like to continue participating in the video conference via the smartphone (paired with the earbuds). Upon user confirmation, the video conference is initiated on the smartphone (after the user leaves the vicinity of the desktop). After the initiation, the earbuds are automatically switched to both listening and speaking mode, where a microphone on the earbuds is used to detect the user's speech input. Further, if the user then sets the smartphone on a stand and is detected to be within a field of view of the smartphone camera, the user's video stream to the video conference can be resumed with user confirmation.

A listing of device features for each available device can be generated automatically

- **Smartwatch**
  - Audio input: Yes, using on-watch microphones.
  - Audio output: Yes, using on-watch speakers.
  - Video input: No. Typically N/A -- as smartwatches don't have cameras.
  - Video output: Yes, using on-watch display.
- **Earbuds**
  - Audio input: Yes, using on-bud microphones.
  - Audio output: Yes, using bud speakers.
  - Video input: No. Earbuds don't have cameras.
  - Video output: No. Earbuds don't have displays
- **Phone**
  - Audio input: Yes, using phone microphones.
  - Audio output: Yes, using phone speakers.
  - Video input: Yes, using the phone camera.
  - Video output: Yes, using the phone display.

The process of video conference transfer to a different device can be performed any number of times and with appropriate devices. Because of differences in capabilities of different devices, when the user leaves the desktop with different combinations of devices, video conference input/output can be automatically reassigned to the available combination of devices based on user preference. The user can provide their preferences for the use of each device for input/output during the video conference. The user-specified preferences can be used to break ties amongst the available options. For example, if the user does not pick up the smartphone, one of the below reassignments can be utilized to continue the video conference via the smartwatch and earbuds.

- **Combination 1**
  - Audio input: Watch
  - Audio output: Watch
  - Video input: N/A
  - Video output: Watch
- **Combination 2**
  - Audio input: Earbuds
  - Audio output: Earbuds
  - Video input: N/A
  - Video output: Watch
- **Combination 3**
  - Audio input: Watch
  - Audio output: Earbuds
  - Video input: N/A
  - Video output: Watch
- **Combination 4**
  - Audio input: Earbuds
  - Audio output: Watch
  - Video input: N/A
  - Video output: Watch

Further to the descriptions above, a user may be provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable collection of user information (e.g., information about a user's devices, a user's video conferences, a user's preferences, or a user's current location), and if the user is sent content or communications from a server. In addition, certain data may be treated in one or more ways before it is stored or used, so that personally identifiable information is removed. For example, a user's identity may be treated so that no personally identifiable information can be determined for the user, or a user's geographic location may be generalized where location information is obtained (such as to a city, ZIP code, or state level), so that a particular location

of a user cannot be determined. Thus, the user may have control over what information is collected about the user, how that information is used, and what information is provided to the user.

CONCLUSION

This disclosure describes techniques that, with user permission, automatically detect when a user leaves a physical location from where they have joined a video conference. The techniques further detect mobile and/or wearable devices available to the user and automatically configure such devices for video conferencing. A link to join the video conference via such a device is provided to the user, and upon user selection, the video conference session is seamlessly transferred to the available device.