June 2020

ARTIFICIAL INTELLIGENCE BASED HYPER-REALISTIC DEEP FAKES FOR PARTICIPANT VIDEO IN COLLABORATION MEETINGS

Rajarathinam Chidambaram

Karthikeyan Muthusamy

ARTIFICIAL INTELLIGENCE BASED HYPER-REALISTIC DEEP FAKES FOR PARTICIPANT VIDEO IN COLLABORATION MEETINGS

AUTHORS:
Rajarathinam Chidambaram
Karthikeyan Muthusamy

ABSTRACT

Techniques are described for Artificial Intelligence (AI) based deep fakes for a participant view in the meeting. A deep fake video may be built from poor quality and noisy/low-light video. A participant can eliminate others in the video that the participant did not intend to see but are in the camera field of view. Quality of video from poor quality cameras can also be eliminated. Moreover, a deep fake algorithm is described to build a high quality hyper realistic video from a low quality/no video scenario from the audio. This can be achieved by capturing only the face metadata on the meeting client using previously learned AI models and sending it to the cloud. The meeting server uses the AI based deep fake to create a high-resolution video from a low-resolution video using AI algorithms previously learned from previously shared videos/metadata from the user. The apparent location of the participant may also be changed using deep fakes (e.g., bedroom to work room, car to office room, etc.), in order to provide an optimal meeting experience for the remote participant.

DETAILED DESCRIPTION

Current video collaboration experiences are poor and limited. In low bandwidth situations a video cannot be streamed. Blurred low-resolution video is provided on current-generation large 4K/8K displays. Participants may not have enough ambient light to provide good quality video using device cameras (e.g., smart phone), may not be dressed or groomed to attend a professional video call, and may not be in the right place to attend a video call (e.g., car, supermarket, parking lot, home, bedroom, etc.). Furthermore, unpleasant/unintentional views of the participant personal space may be shown.

Accordingly, techniques are described herein to provide a unique video collaboration experience with Artificial Intelligence (AI) even in situations where high-quality video conferencing cannot be established properly. Manual triggers from the user

may be used to produce an AI generated video view for the participant. The user may trigger the AI generated video view if the user perceives that the video quality generated is sufficient, the user does not have a high definition camera, the user does not want to show a user live feed, or there is a known low network bandwidth. The AI generated video view may be automatically triggered if the video quality had degraded beyond an acceptable level, the meeting server detects that there is a low light or low resolution video, the communications manager decides the video communication cannot be established or switched over to audio, the meeting server detects low bandwidth/jitter/latency/packet loss after a video call has been established, or the meeting client detects low bandwidth/jitter/latency/packet loss and notifies the meeting server via Session Initiation Protocol (SIP) or Hypertext Transfer Protocol Secure (HTTPS) to switch to the AI generated video.

A notification pipe may obtain HTTPS feedback from the telepresence endpoint and send a SIP notification from audio/video clients.

The meeting client may provide a button to signal a user intention to switch to the AI generated video. The meeting client captures only the face metadata using previously learned AI models, sends it to cloud, and notifies the meeting server regarding a user intention that a simulated video need to be presented to the audience.

The meeting server identifies users who are enrolled for the feature and trains the meeting video AI engine. This can be performed by walking the participants through previously configured instructions. The meeting server continuously records participant video and generates deep learning AI based three-dimensional (3D) models of the participant based on past meetings joined by the participant. The meeting server may then generate the AI-based hyper realistic video based on any one or more of the aforementioned triggers.

Figure 1 below illustrates an example system overview.

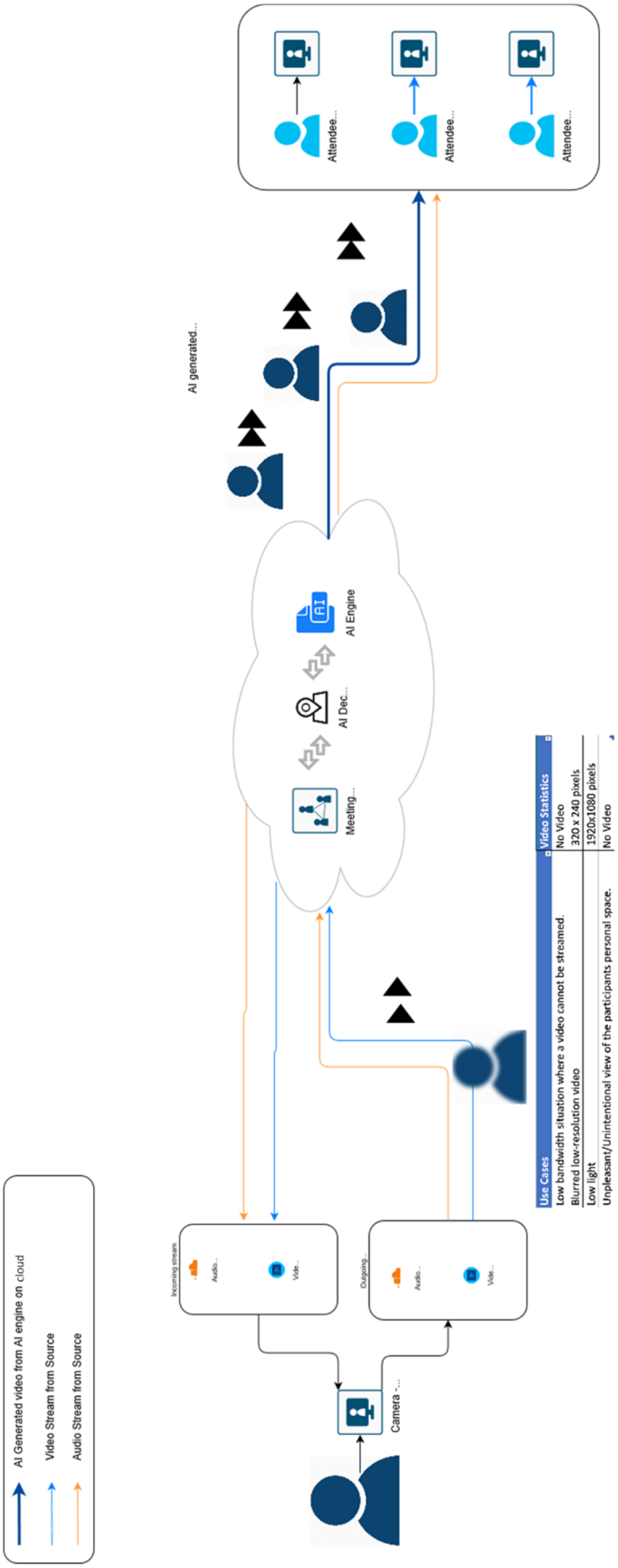
AI Generated video from AI engine on cloud
Video Stream from Source
Audio Stream from Source
Camera -...
Incoming stream
Audio...
Vide...
Outgoing
Audio...
Vide...
Meeting...
AI Dec...
AI Engine
AI generated...
Attendee...
Attendee...
Attendee...
Use Cases
Low bandwidth situation where a video cannot be streamed.
Blurred low-resolution video
Low light
Unpleasant/Unintentional view of the participants personal space.
Video Statistics
No Video
320 x 240 pixels
1920x1080 pixels
No Video

Figure 1

Figure 2 below illustrates an example processing flow.

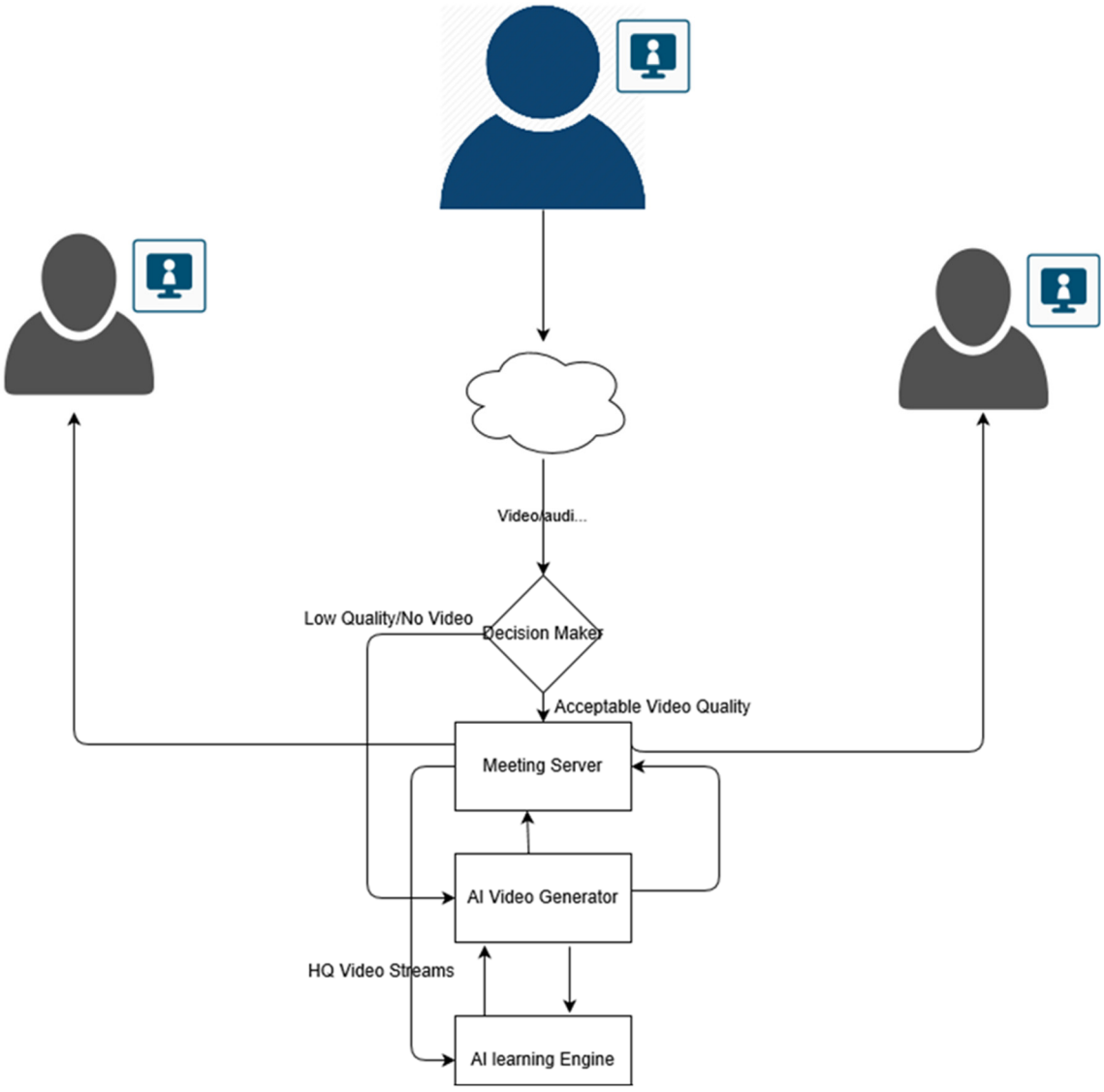

Figure 2

Currently, without the deep learning AI, participants might view a dark unpleasant video, or a blurred low-resolution video might be shown on a large display. Instead of a video collaboration, only audio might be provided with a static profile picture. Furthermore, the audio and video may not be in synchronization. Unpleasant views of the participants personal space may also be shown unintentionally.

By contrast, the deep learning AI may provide a bright, high-quality, automatically generated video of the participant. The deep learning AI may provide a high-definition video experience for attendees via a high-resolution large display, such as a large

telepresence endpoint. It may also automatically generate high-quality video for a participant transmitting only audio but no video. The audio and video may be synchronized, and participant personal spaces may be hidden.

In summary, techniques are described for AI based deep fakes for a participant view in the meeting. A deep fake video may be built from poor quality and noisy/low-light video. A participant can eliminate others in the video that the participant did not intend to see but are in the camera field of view. Quality of video from poor quality cameras can also be eliminated. Moreover, a deep fake algorithm is described to build a high quality hyper realistic video from a low quality/no video scenario from the audio. This can be achieved by capturing only the face metadata on the meeting client using previously learned AI models and sending it to the cloud. The meeting server uses the AI based deep fake to create a high-resolution video from a low-resolution video using AI algorithms previously learned from previously shared videos/metadata from the user. The apparent location of the participant may also be changed using deep fakes (e.g., bedroom to work room, car to office room, etc.), in order to provide an optimal meeting experience for the remote participant.