METHOD OF PASSIVE PARTICIPANT TRACKING IN VIDEO CONFERENCE

Morgan Fang
Yasi Xi
Zhaocai Wang
Janus Wang

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation
Fang, Morgan; Xi, Yasi; Wang, Zhaocai; and Wang, Janus, "METHOD OF PASSIVE PARTICIPANT TRACKING IN VIDEO CONFERENCE", Technical Disclosure Commons, (December 21, 2018)
https://www.tdcommons.org/dpubs_series/1809

This work is licensed under a Creative Commons Attribution 4.0 License.
This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.
METHOD OF PASSIVE PARTICIPANT TRACKING IN VIDEO CONFERENCE

AUTHORS:
Morgan Fang
Yasi Xi
Zhaocai Wang
Janus Wang

ABSTRACT
An innovative technique is proposed to highlight video streams of participants mentioned by the current active speaker in a video conference. By immediately highlighting mentioned participants, valuable visual information can be delivered and perceived by meeting attendees so that meeting productivity is improved.

DETAILED DESCRIPTION
The following experience is typical in an online video conference. When a meeting host Alice introduces Bob in the meeting by saying "Bob Smith is a new member of our team", the active speaker is still Alice, and Bob's video stream will not be highlighted, centered, or enlarged in the meeting. When the meeting host Alice says "Now, Bob will make a speech on his recent research on artificial intelligence (AI). Let's give warm applause", the active speaker is still Alice, and Bob's video stream will not be highlighted. However, when Alice makes an introduction of Bob or invites Bob to make a speech, as meeting attendees, Bob's video stream is highlighted so that others may immediately recognize Bob.

Here, techniques are provided to make the video conference user experience more natural. By leveraging facial recognition, speech to text (STT), and AI techniques, when someone, e.g., Bob, has been mentioned in an online video conference, Bob's video stream can be highlighted so that the meeting attendees can recognize Bob immediately.

In existing video conferences, the user experience is that when Bob is mentioned in a meeting, Bob's video stream will not be highlighted until he himself begins to speak in the meeting. However, for a more natural and considerate user experience, when Bob
has been mentioned, his video stream should be highlighted because meeting attendees will be more interested in Bob's video stream than that of the current active speaker.

The descriptions and figures below illustrate what the user experience will look like in a typical video conference. It is proposed that when someone has been mentioned in a meeting, his/her video stream will be highlighted accurately and immediately. For example, in FIG. 1, which is a bottom floating layout, when Bob and Charlie have been mentioned by the active speaker, Alex, their video streams are highlighted. There are many ways to highlight a video stream, including:

1. Render an attractive rectangle around the video stream
2. Place an icon or symbol on top of the mentioned participant
3. Enlarge the video grid of the mentioned participant
4. Place a video grid in a dedicated area for attraction
5. Make the video grid flash

For example, FIG. 1 illustrates some examples of how to highlight a participant (using methods 1, 2, 3 above). For simplicity, we render a yellow rectangle for the mentioned participant.

![FIG. 1](image)

FIG. 1

FIG. 2 illustrates another scenario in a Hollywood square layout. Participant A is the active speaker. Participants B through O are in current visible grids and participant P is not visible. When participants P and J are mentioned by active speaker A, P is moved...
to the grid, at the position used to accommodate participant B, and J is moved to the grid, at the position used to accommodate participant C. The others are moved backward accordingly, for example:

\[
\begin{align*}
B & \Rightarrow D \\
C & \Rightarrow E \\
D & \Rightarrow F \\
\ldots \\
N & \Rightarrow O
\end{align*}
\]

FIG. 2

FIG. 3 illustrates a bottom floating layout. If Bob and Charlie's video stream are not displayed in the floating video panel (due to limited space), they will be re-arranged to be included in the display area. Attractive yellow rectangles are rendered around grids of Bob and Charlie's video streams.

When a video stream has been re-arranged to prioritize video streams, the grid of this video stream will stay at the foremost place in the grid list or grid array. When a second
video stream has also been highlighted, the grid of the second stream is positioned at the second place. This is repeated in a similar manner for the 3rd, 4th and so forth highlighted video streams. These grids always stay at foremost places until they are no longer highlighted.

Similarly, in a side-by-side layout (FIG. 4), vertical-split layout (FIG. 5), Hollywood square layout (FIG. 6) or top right floating layout (FIG. 7), video streams of mentioned participants are reprioritized and highlighted.
FIG. 6

FIG. 7
N-in-1 scenario

If there are multiple participants at one endpoint (e.g., Iris, Jack and Hans join a meeting in a TP room (FIG. 8)), when Jack is mentioned, his video stream will be re-prioritized and highlighted. Meanwhile, since Jack's facial rectangle can be detected by facial recognition techniques and only Jack (excluding Iris and Hans) has been mentioned, this video stream will zoom in on Jack to highlight him.

Referring to FIG. 9, another example is provided regarding prioritizing video streams. Alex, Bob, Charlie, Daisy, Hans, Iris, Jack and Karl are in an online video conference. Iris, Jack and Karl join the meeting in a telepresence room. Alex is currently the active speaker. The meeting session may have a general participant list, wherein the name under each video stream in listed in FIG. 9:

Alex
Bob
Charlie
Daisy
Hans
Iris
Using facial recognition techniques, the meeting session becomes aware of a more accurate participant list that coordinates with each participant (as shown by the name above each video stream in FIG. 9):

Alex  
Bob  
Charlie  
Daisy  
Hans  
Iris, Jack and Karl (grouped)

Thus, a context-aware module of a meeting session is aware of: the general participant list, the accurate participant list by facial recognition, and meeting conversation context.

When the active speaker Alex, mentions someone, e.g., Bob and Jack, in the meeting, the system deduces by meeting session that Bob and Jack have just been mentioned by the active speaker in the meeting. Then, the meeting session broadcasts signals to each attendee in the meeting:

To Bob: Highlight Jack (Please note that Bob himself is excluded)  
To Charlie, Daisy, ..., and Hans: Highlight Bob and Jack  
To the group of Iris, Jack and Karl: Highlight Bob (Please note that Jack himself is excluded)

Once signals are received from meeting session, each meeting client re-prioritizes grids of video streams and highlights those included in the signal. Since there are multiple participants at Jack's side, when highlighting Jack, the video stream zooms to Jack, which is illustrated by FIG. 8.
These techniques may leverage AI, STT, facial recognition, etc. By leveraging these techniques, when someone has been mentioned in a video conference, the mentioned participant's video stream is immediately highlighted for the meeting attendees' immediate examination.

Video streams of mentioned participants are prioritized when mentioned and return to normal mode after a set time (e.g., an empirical duration of 3 seconds). During the period of highlighting, attendees have an immediate examination of the faces, reactions, etc. of the mentioned participants and may perceive the most timely, accurate and subtle visual information which is very helpful to improve video conference user experience and collaboration productivity.

For participants with the same first name, the meeting session may use an AI algorithm to analyze the meeting conversation context to decide which Bob to highlight. If determination is beyond a pre-defined threshold of certainty, the system may ignore the
mentioned Bob. When the highlighting ends, the video stream will no longer remain 
prioritized and will return to normal. For an N-in-1 scenario (FIG. 8), e.g., zooming to Jack, 
the system will cancel zooming in and will return to normal mode.

The system may also handle cases in which the participant is off video or absent. 
If a participant chooses to turn off video or is out of sight of the camera's field of view, the 
mentioned participant may be ignored. A message may be prompted at Bob's meeting client 
stating that "Bob, we can't see you. Turn on your camera, please" or "Bob, we can't see 
you. Move closer to your camera, please". Bob can accept, ignore or suppress the 
suggestion. If a mentioned participant is not in the participant list or is not identified by 
facial recognition, the system may ignore the mentioned participant. Meanwhile, a message 
is prompted to each participant, e.g., Bob's meeting client may state "Bob, is Larry at your 
side? We can't see him. Ask him to move closer to your camera, please". Bob can accept, 
ignore or suppress the suggestion.

In summary, an innovative technique is provided to highlight video streams of 
participants mentioned by the current active speaker in a video conference. With this 
method, the user experience of online video conference can be more natural and 
considerate. By immediately highlighting mentioned participants, valuable visual 
information can be delivered and perceived by meeting attendees so that meeting 
productivity is improved.