March 2022

Automatically Generating Descriptive Annotations for Participants in a Virtual Meeting

Daniel V. Klein

Igor S. Ramos

Automatically Generating Descriptive Annotations for Participants in a Virtual Meeting

ABSTRACT

The effectiveness of virtual meetings can be negatively affected when participants in a video conference are unfamiliar with each other. This disclosure describes techniques, implemented with appropriate user permissions, for automatic generation of descriptive text and annotations for participants in a virtual meeting. The descriptive text and annotations are generated based on combining various available pieces of information about the participant obtained with permission from a variety of information sources and/or via computational analysis of the video and audio from the live feed of the meeting. The information obtained can be used to determine and verify the identities of the meeting attendees prior to generating the descriptive text. The techniques can be incorporated within any video or audio conferencing software used for virtual meetings in corporate or personal settings.

KEYWORDS

- Virtual meeting
- Online meeting
- Participant description
- Participant identification
- Face recognition
- Voice recognition
- Voiceprint
- Corporate directory
- Speaker indicator
- User profile

BACKGROUND

People increasingly participate in virtual meetings held via video conferencing software. In many cases, the participants may not have met each other in person or may not be familiar with each other prior to the meeting. In such cases, it is difficult for a user to recognize the other participants in the meeting as well as to determine whether everyone attending the meeting is a legitimate attendee. For instance, it might not be straightforward to determine from the video feed whether unfamiliar persons are who they claim to be. Further, since many video conference tools provide links that allow anyone to join by clicking a shared link, or non-invited individuals may inadvertently be in meeting rooms during a meeting, it may be difficult to determine whether those present are all authorized to be in the meeting. This issue is exacerbated when some participants have their videos turned off or join the meeting via a phone call. In these cases, the only cues available to determine participant identity and authorization is their screen name and display picture (if set) or their phone number.

Even when the video feed is on and the participants are sufficiently familiar with each other to recognize each other, some individuals may have trouble associating each person with their names. While screen names shown with each individual's video feed can help associate their faces with their names, the mechanism does not work when multiple people share a single video feed, such as a group of co-located attendees joining the meeting from a meeting room, or when an individual attendee's identification is the name/number of the room they are in.

Addressing the issue of not recognizing participants in a virtual meeting by explicitly asking the attendees for their names and/or for authentication can be socially awkward as well as inefficient. Yet, not addressing the issue can also create awkwardness during the meeting and can have security and confidentiality implications, particularly in a business setting.

DESCRIPTION

This disclosure describes techniques for automatic generation of descriptive text annotations for participants in a virtual meeting held via video conferencing software, implemented with permission of the participants. The descriptive text is derived based on combining different pieces of information about the participant obtained with permission from a variety of information sources and/or via automated computational analysis of the video and audio from the live feed of the meeting.

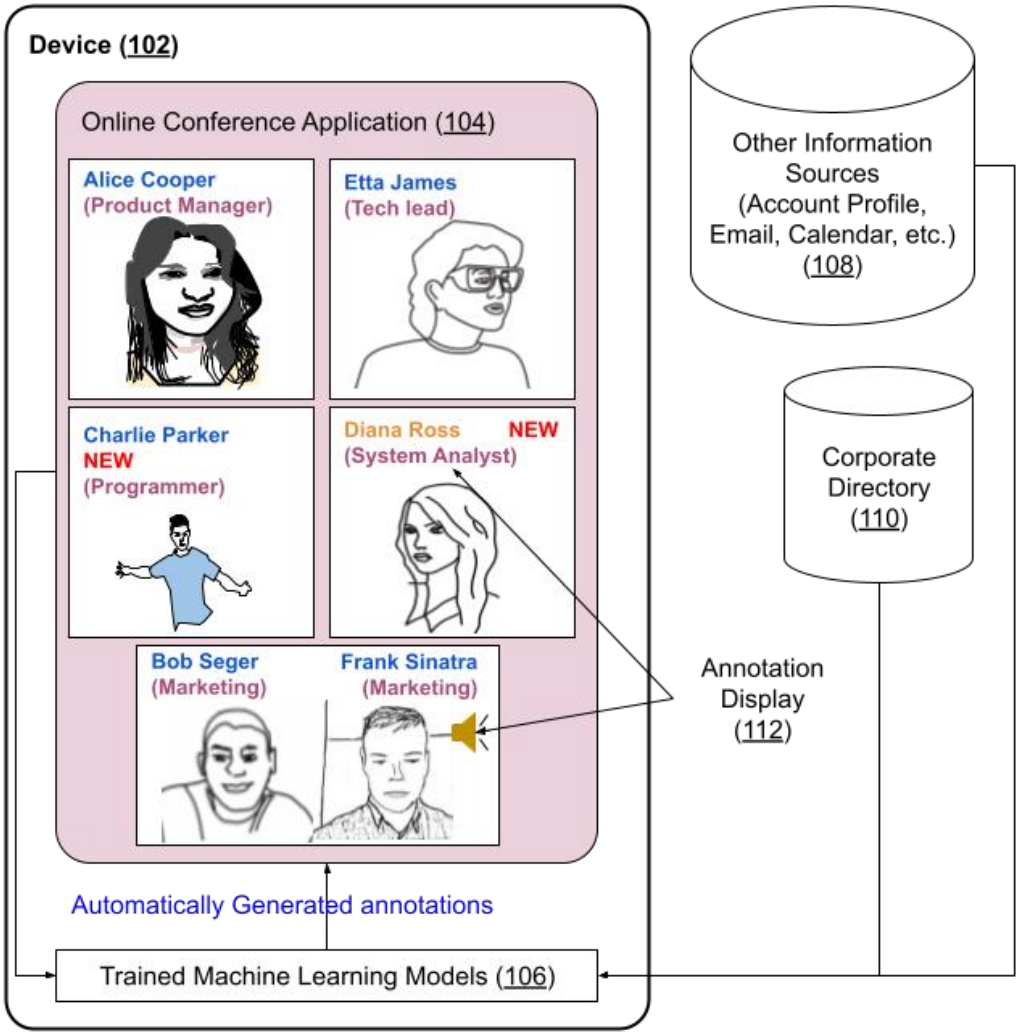

Fig. 1: Marking participants in a virtual meeting with automatically generated annotations

Fig. 1 shows an example of operational implementation of the techniques described in this disclosure. A user is participating in a business meeting via an online video conferencing application (104) on a device (102). With permission of the participants as well as from video conferencing devices in meeting rooms (e.g., the room from where two participants Bob and Frank have joined), data from various information sources (108), the corporate directory (110), and the live feed of the meeting is accessed and analyzed, e.g., via suitably trained machine learning models (106) for automatic generation of annotations that identify and describe the participants and authenticate their presence in the meeting.

As seen in Fig. 1, the annotations (112) are displayed within each participant's video feed showing information, such as name, job function, etc. Participants who are likely unfamiliar to the user are marked with the word "NEW" in red. The person who is currently speaking is indicated with a speaker icon. As shown in Fig. 1, the color of a participant's name indicates whether a person in the video matches the information from other sources (e.g., the person's photo in the corporate directory), with amber indicating a lack of a certain match.

Relevant data to match the faces of meeting participants with their identities can be obtained with specific user permission from one or more information sources, such as permitted/ public portions of stored user profiles, corporate directory entry, caller ID of the phone dialing into the meeting, voice fingerprint, email messages, calendar events, personal pages on the web or social media, text and instant messages, contacts in address books, microphone audio, location, etc. For instance, with user permission, user profiles can be accessed to obtain information such as name, preferred pronouns, email address, phone number, photograph, job title, biography, etc. If users permit, information from the profiles can be appropriately cross-checked for verification. For instance, with user permission, computer vision techniques can be

used to ensure that a user's profile photo is not a non-human object or a celebrity picture and/or to match it with the official photo in the corporate directory.

The identities of those who are authorized to be in a given meeting can be determined in a variety of ways. For instance, the content of communications such as email and text or instant messages can be programmatically analyzed with permission to detect relevant information such as new team members based on welcome messages, introductions, meeting invitations and links, etc. Similarly, calendar entries for individuals and meeting rooms, accessed with permission, can provide information about invited participants for a particular virtual meeting.

The information regarding those who are authorized to be in a meeting can be compared with the identities of those who are present in the meeting. The identities of the participants present can be determined by cross-checking user-permitted data such as: status setting match in video conferencing, messaging, or presentation software; co-location with other meeting attendees; match between caller ID and the phone number in the corporate directory and/or in people's address books; etc. Further, the presence of a person in a meeting room that they are scheduled to be in (e.g., as indicated by their calendar) can be confirmation of their identity. When multiple participants are present in a meeting room, detection of the same audio at the same time by microphones of multiple devices can be utilized to determine identity of the participants. For example, if a person's voice is detected on a microphone and also on their phone, which are co-located (in the same meeting room), it is an indication that an as-yet unidentified face belongs to the person.

Once identities of the meeting participants have been determined and verified, as described above, the information can be used to annotate their presence in the meeting in one or more of ways, such as:

- Brief biographical descriptions for each person, such as name, title, job function, etc.
- Reasons for the presence of each attendee in the meeting.
- Names shown alongside phone numbers of those dialed into the meeting via phone.
- Labels showing whether persons in videos match their respective photos in their respective profiles or corporate directory entries.
- Tags (e.g., bubbles, etc.) or other appropriate visual cues (e.g., arrows, highlights, etc.) to help identify each person within video feeds that include multiple people and/or to help mark the person who is speaking at any given time.
- Indicators (e.g., arrows, borders, etc.) to provide information about speakers who are off camera.
- Markers that distinguish attendees with whom a given user has a prior history of interaction (e.g., *M* meetings in the previous *N* days) from those that are new or unusual from that user's perspective.

Apart from the information extracted with user permission as described above, the annotations can be generated based on the user-permitted live analysis of the video and audio feeds of the meeting with relevant machine learning techniques such as recognition of faces and voices, detection of facial cues (such as mouth movement), etc. With permission, those previously deemed as new or unusual from a given user's perspective may cease to be marked as such based on appropriate criteria, such as a threshold number of shared meetings, threshold time period after the first interaction, frequency of shared meetings, explicit setting by the user marking someone as "recognized," etc.

The techniques described in this disclosure can be utilized for additional purposes, such as a meeting map akin to a dynamic organizational chart that summarizes attendee information,

in order to improve or replace the initial round of introductions typical in meetings in which not all participants know each other. Further, if users permit, the techniques can be used to augment the live captioning capabilities of video conferencing applications by adding the speaker's name to the captions for their speech.

With user permission, the techniques can be incorporated within any video or audio conferencing software used for meetings in corporate and/or personal settings. The features can be provided for all users of the application or all personnel of an organization. Alternatively, or in addition, with appropriate permissions, users can choose to avail themselves of the described functionality locally on the user device to obtain annotations and/or captions for the meetings they attend. Implementation of the techniques can facilitate better and more effective identification, verification, and labeling of participants in a virtual meeting, thus enhancing the user experience (UX) and security of participating in such meetings.

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 profile, social network, messaging communications, a user’s calendar, social actions or activities, profession, 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

The effectiveness of virtual meetings can be negatively affected when participants in a video conference are unfamiliar with each other. This disclosure describes techniques, implemented with appropriate user permissions, for automatic generation of descriptive text and annotations for participants in a virtual meeting. The descriptive text and annotations are generated based on combining various available pieces of information about the participant obtained with permission from a variety of information sources and/or via computational analysis of the video and audio from the live feed of the meeting. The information obtained can be used to determine and verify the identities of the meeting attendees prior to generating the descriptive text. The techniques can be incorporated within any video or audio conferencing software used for virtual meetings in corporate or personal settings.

REFERENCES

1. ‘Face-blindness’ disorder may not be so rare. Available at: https://news.harvard.edu/gazette/story/2006/06/face-blindness-disorder-may-not-be-so-rare/ Accessed: 2022-02-23.