

# Technical Disclosure Commons

---

Defensive Publications Series

---

August 2020

## Automatically Augmenting Video Conference With Related Content Overlays

Peter Lewis

Stephen Lewis

Andrew Lewis

Follow this and additional works at: [https://www.tdcommons.org/dpubs\\_series](https://www.tdcommons.org/dpubs_series)

---

### Recommended Citation

Lewis, Peter; Lewis, Stephen; and Lewis, Andrew, "Automatically Augmenting Video Conference With Related Content Overlays", Technical Disclosure Commons, (August 28, 2020)  
[https://www.tdcommons.org/dpubs\\_series/3553](https://www.tdcommons.org/dpubs_series/3553)



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

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.

## **Automatically Augmenting Video Conference With Related Content Overlays**

### **ABSTRACT**

During a video conference (VC), users often want to share information with other participants in the VC. For example, a user may refer verbally to concepts or show objects to the other participants. For example, a teacher in a video conference may hold up a book and request students to turn to a particular page. Currently, such actions interrupt the flow of presentation, since the presenting user is required to provide information such as URLs for the students to access. This disclosure describes techniques to automatically share within a video conference (VC) expanded information about concepts or objects referenced by a user within the VC. With specific user permission, image recognition is performed on the video stream of the VC to identify objects and/or speech recognition is performed to identify concepts in the user's voice stream. Additional information about the recognized objects/concepts is made available to other participants in the VC.

### **KEYWORDS**

- Video conferencing
- Video meeting
- Video call
- Content overlay
- Object recognition
- Video augmentation

## BACKGROUND

During a video conference (VC), users often want to share information with other participants in the VC. For example, a user may refer verbally to concepts or show objects to the other participants. For example, a teacher may hold up a book with the title visible in their video stream. In another example, a teacher may utter “turn to page 123.”

However, such a process is not frictionless in current video conferences. For example, to refer to a particular book or page, the user may need to send a link in a text chat of the video conference which other participants can select to view that portion. Unlike in a physical gathering of people, where a user can simply continue speaking and simply pass the referenced book around to participants at the physical gathering, having to send a link separately interrupts the flow of presentation. Further, the user may need to halt the presentation to search for and then send additional information.

## DESCRIPTION

This disclosure describes techniques to automatically share within a video conference (VC) expanded information about concepts or objects referenced by a user within the VC. With specific user permission, image recognition is performed on the video stream of the VC to identify objects, e.g., a type of tool the user is holding that is used for crafting, a particular book, etc. and/or speech recognition is performed to identify concepts in the user’s voice stream, e.g., “refer to the manual for the tool.”

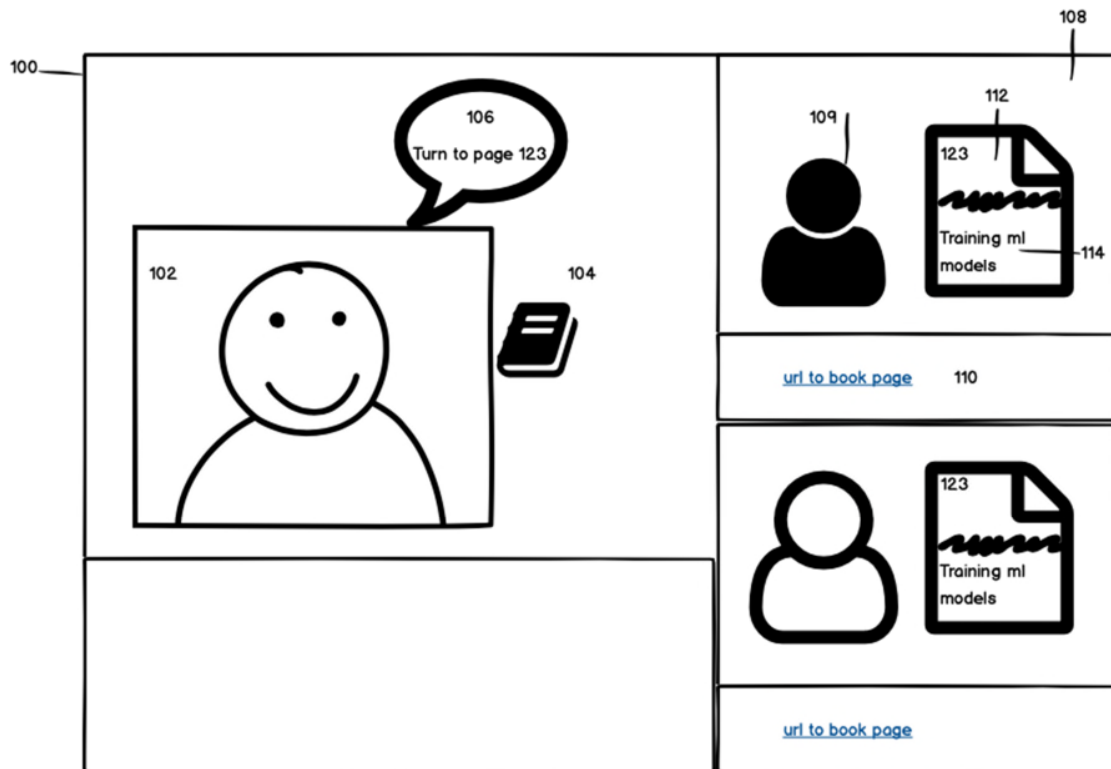
Once the object or concept referred to by the user is identified, the objects or concepts are automatically searched for in a knowledge base that is a distilled knowledge representation of structured information that relates to the concepts. Information obtained from the knowledge base is automatically provided to the other participants, without any additional input from the

user presenting on the VC. Such additional information can include, e.g., URLs, lookups of page mappings between different versions of a book, a link to purchase an identified book, or other related material. The provided information can persist in the user interface provided in the user interface to the other participants, e.g., until they access it with touch input, or can be provided in a fleeting manner, e.g., as a notification that briefly appears and is then removed from the user interface.

For example, consider a situation where the presenting user holds up a store card, a card that gives frequent shoppers of the particular store access to benefits such as points, discounts, special promotions, etc. The described techniques enable identification of information from the card using image recognition (e.g., the name of the store) and search for more information about the store in a knowledge base. The participants are then automatically provided, e.g., in a chat window in the VC, additional information about the store, such as a link to the online store, suggested items for purchase, store address, etc.

Similarly, with user permission, image and voice recognition can be applied to augment the presenting user's displayed video with additional information. For example, if the presenting user sips a cup of coffee and then puts it down, a speech bubble overlay with the word "aaah" is automatically added and shown as if emanating from the presenting user's head. In another example, the presenting user yawns and an overlay graphic that shows "zzzzz"s emanating from the presenting user's head is displayed.

The described techniques can also be activated by invoking a hotword. For example, before a presenting user starts to describe a concept or show an object, they can invoke the augmentation by saying a particular hotword.



**Fig. 1: User interface providing additional information in a video conference**

Fig. 1 illustrates automatically providing additional information in an example video conference user interface (100), per the techniques of this disclosure. A presenting user (102), for example a teacher, is presenting to multiple participants (109), e.g., students, shown on the right side of the figure. In this example, the presenting user speaks a phrase, “turn to page 123” (106) that references a book (104) that is in the video stream of the presenting user, e.g., held by the presenting user, seen on a desk or a bookshelf, etc.

Content identification and/or augmentation of the video conference is performed with specific permission from the presenter and/or other participants. The spoken phrase “turn to page 123” is detected via speech recognition, and the book title for the book (104) is identified using object recognition. A URL (110) for the referenced book page is automatically provided to each

participant in their respective chat portion of their user interface (108), without any additional input from the presenting user.

Upon selecting the URL, the other participants are automatically provided with a copy of the referenced page 123 of the book. The specific page can also be automatically displayed in the VC user interface without any user input.

Further to the descriptions above, a user is provided with controls allowing the user to make an election as to both if and when systems, programs or features described herein may enable use of user information (e.g., user's speech and/or video streams, user's location, user profile, user preferences), 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. Thus, the user has 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 to automatically share within a video conference (VC) expanded information about concepts or objects referenced by a user within the VC. With specific user permission, image recognition is performed on the video stream of the VC to identify objects and/or speech recognition is performed to identify concepts in the user's voice stream. Additional information about the recognized objects/concepts is made available to other participants in the VC. Such automatic recognition, and augmentation of the video conference can reduce friction in the experience of sharing content in a video conference.

## REFERENCES

1. Barakonyi, István, Werner Frieb, and Dieter Schmalstieg. "Augmented reality videoconferencing for collaborative work." In *Proceedings of the 2nd Hungarian Conference on Computer Graphics and Geometry*. 2003.