January 2021

ENHANCING SHARED VIDEO PRESENTATIONS WITH INTERACTIVE ELEMENTS

Arsenii Astashkin

Will Reed

Matt Frey

ENHANCING SHARED VIDEO PRESENTATIONS WITH INTERACTIVE ELEMENTS

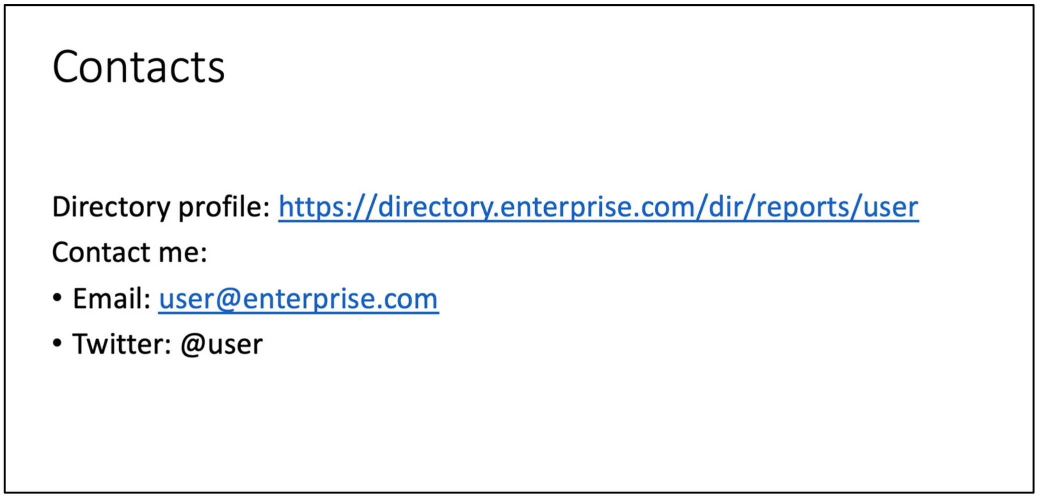
AUTHORS:
Arsenii Astashkin
Will Reed
Matt Frey

ABSTRACT

It is common for presentations shared over a video stream to include rich information such as web links, email addresses, social media handles, and phone numbers. However, because the presentation is a video rather than a slide deck, a viewer is often limited to typing content, if they want to utilize it, in another application, for instance, a browser or email client, which can be error-prone or even impossible if the presentation progresses. Techniques herein provide the viewer with the ability to select a piece of information from a video presentation and trigger an associated action like opening a web link or sending an email seamlessly without typing or performing copy-paste operations.

DETAILED DESCRIPTION

Imagine, for example, that a meeting participant is viewing a video presentation in which a presenter is sharing their contact information via a presentation slide, as shown in Figure 1, below.



Contacts

Directory profile: https://directory.enterprise.com/dir/reports/user

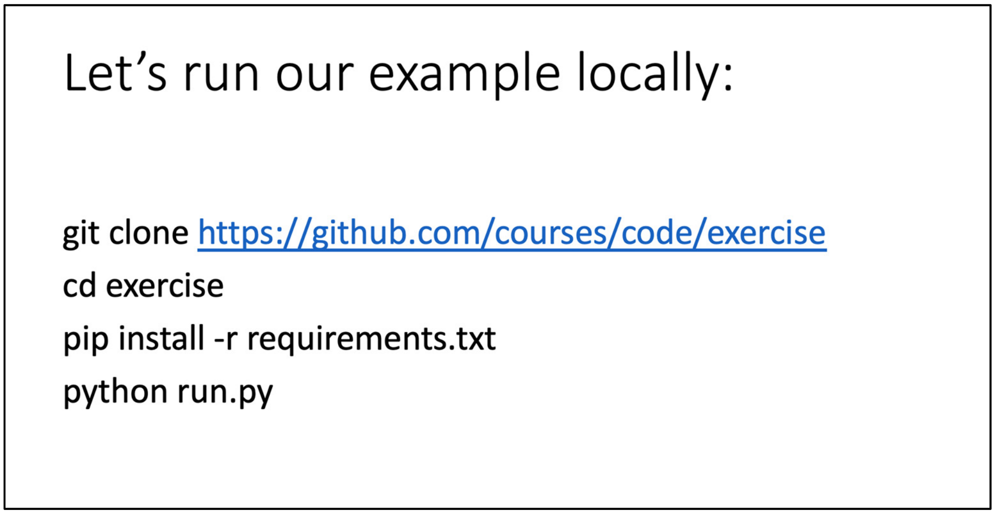
Contact me:

- Email: user@enterprise.com
- Twitter: @user

Figure 1: Contact Information Slide Example

If, for instance, the meeting participant desires to briefly take a look at what is behind the profile link or send a message to the email address, there are limited options available. One option may include interrupting the presenter to request the link or email address via a chat dialog, however it will interrupt flow of the presentation for other participants. Another option may include typing the link into a browser, but it might be challenging, especially for long or difficult links, e.g.: https://example.enterprise.com/open.html?c=8da1f8a8-1f29-4e56-bc09-eb43175a3bba. An alternative may be to wait until the end of the presentation and request the slides from the presenter, though it requires a delay in accessing the content, a viewer can potentially forget the intended action through the course of the presentation. So, none of the currently available options are particularly desirable to the meeting participant in many circumstances.

Apart from contact or reference information sharing, an important use case may involve live or prerecorded learning courses and online education in general in which an instructor expects a meeting participant to execute commands illustrated in a slide, such as:



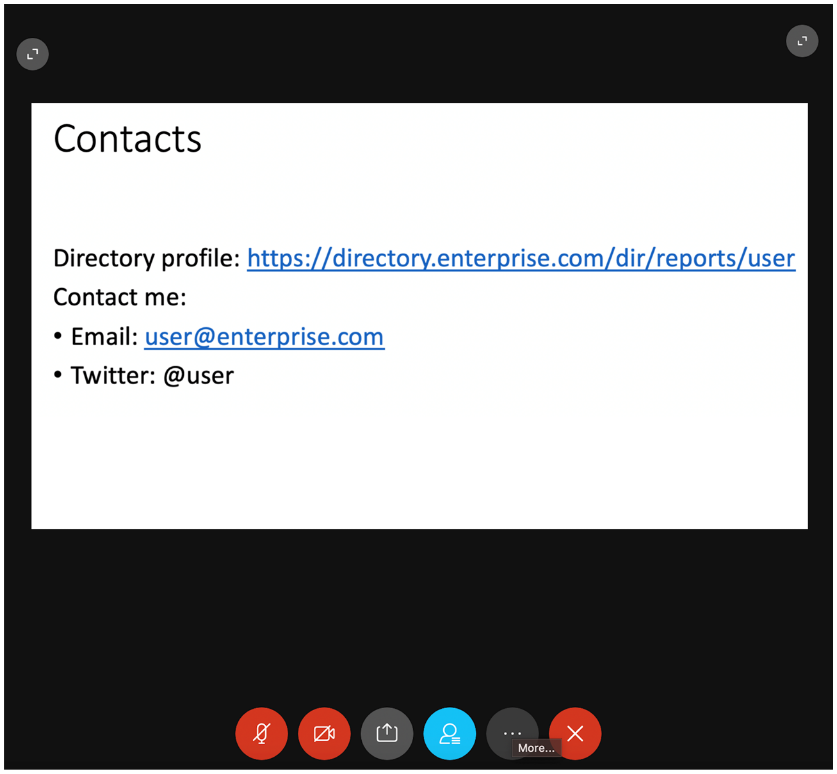
Let’s run our example locally:

```
git clone https://github.com/courses/code/exercise
cd exercise
pip install -r requirements.txt
python run.py
```

Like in the previous example, typing the commands shown above may be difficult during a presentation and unnecessary for a learning purpose.

The enhanced video presentation proposal provides techniques to equip a video presentation with image detection and text recognition functionality that can be seamlessly integrated into a video teleconferencing application, in essence, providing a video presentation web-page capability. A meeting participant using this feature can click on an interesting part of a shared presentation causing the system to identify an associated bounding box of the clicked element, convert the information in the box into text, match it against a known set of entities (e.g., email, link, tag, phone number, etc.), and trigger an action associated with a matched entity. Some possible actions are discussed in further detail, herein.

Consider an operational example illustrated through user interface mockups in Figures 2-7. It should be noted for the slides below that the colored boxes around the information of interest are provided for visualization purposes only and may not be seen by users during operation. Figure 2, below, illustrates an example screenshot of a video teleconference interface in which a meeting presenter is sharing a slide with a web link, but it cannot be opened or even copy-pasted, since it is a screen involving video sharing.



Contacts

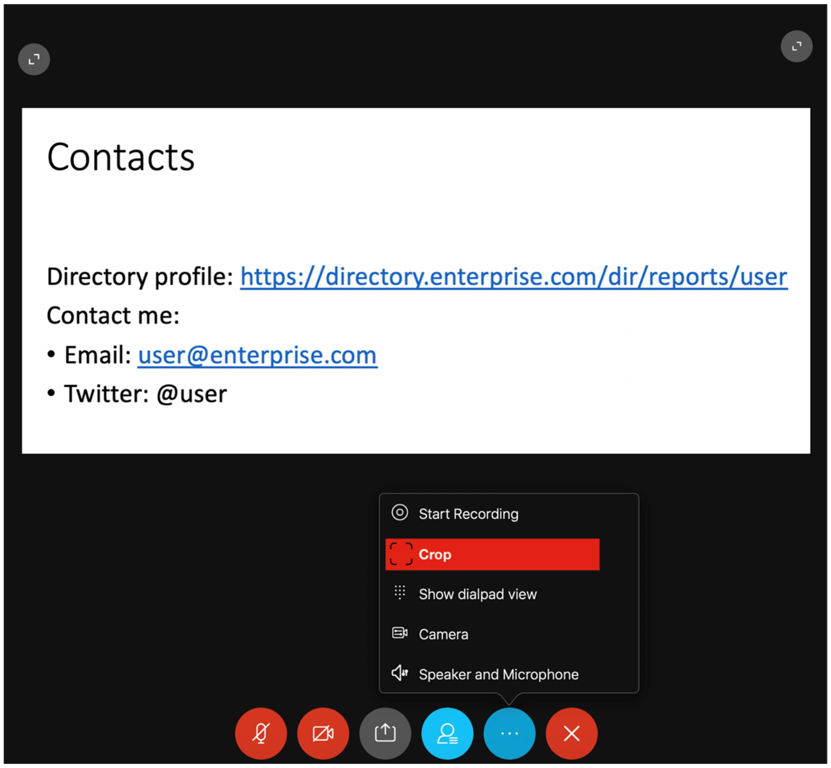
Directory profile: https://directory.enterprise.com/dir/reports/user

Contact me:

- Email: user@enterprise.com
- Twitter: @user

Figure 2: Meeting Interface Example

To trigger the enhanced video presentation feature a user can switch to the crop mode by activating the corresponding option, as shown below in Figure 3.



Contacts

Directory profile: https://directory.enterprise.com/dir/reports/user

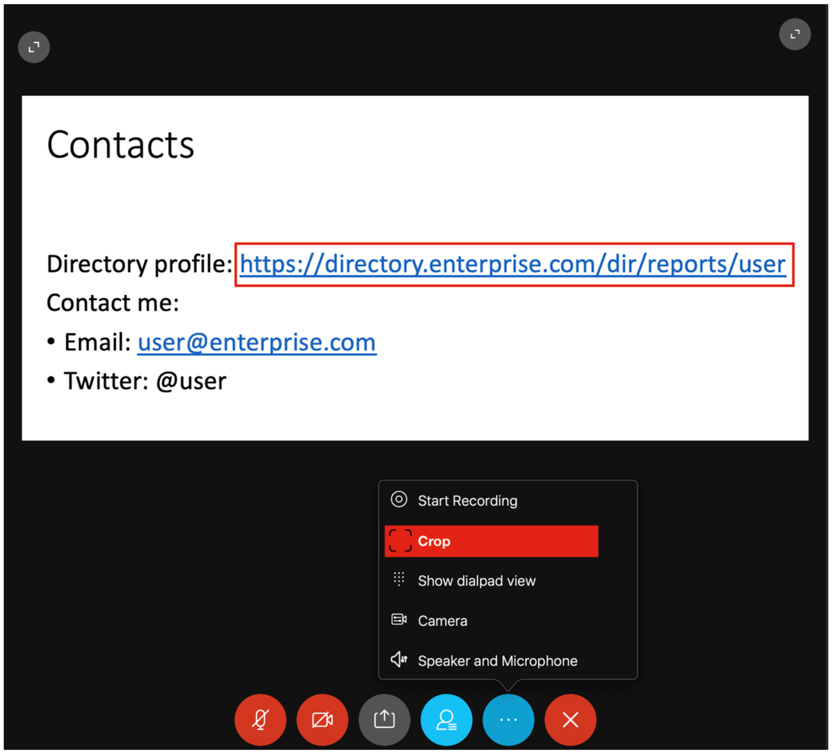
Contact me:

- Email: user@enterprise.com
- Twitter: @user

Figure 3: Switching to Enhanced Presentation Mode

Upon switching to the crop mode, the user is able to click on the link. The system will detect a bounding box around the selected choice (as shown in Figure 4) and convert the information inside the selection into text. Simple heuristics and regular expressions can then be used to recognize and extract entities of interest.

As a result, in the example below, a web link has been identified in a bounding box around the user selection. Based on this information, the system can open the link in a browser, as shown in Figure 5.



Contacts

Directory profile: https://directory.enterprise.com/dir/reports/user

Contact me:

- Email: user@enterprise.com
- Twitter: @user

Figure 4: Link Selection Example

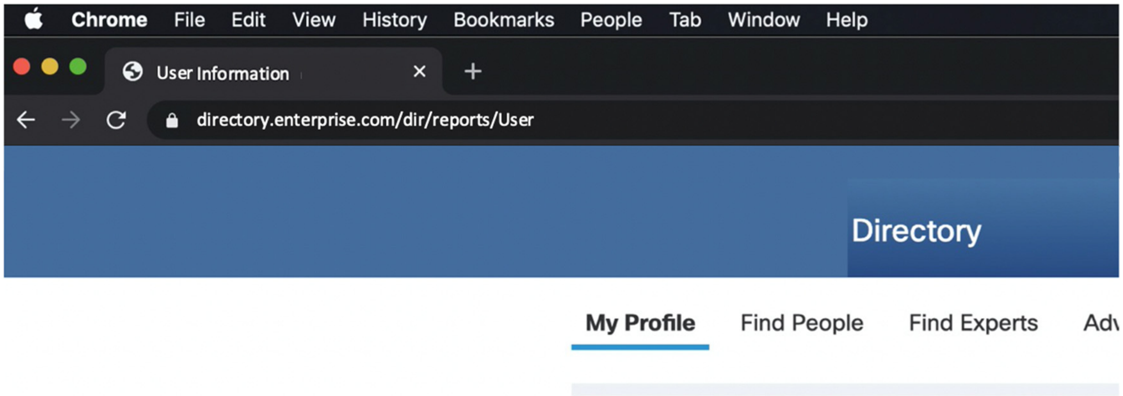
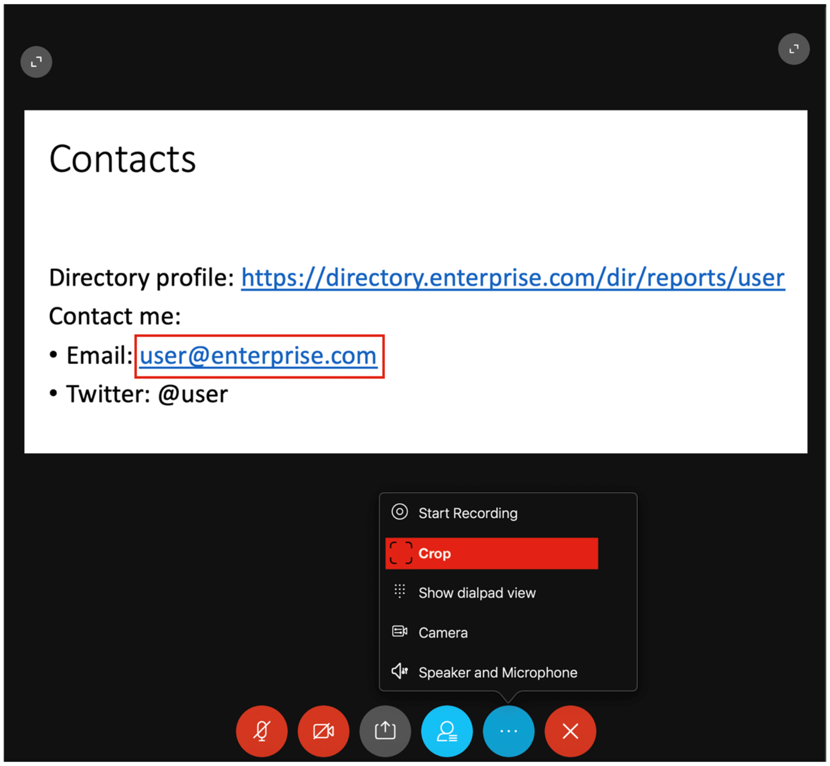


Figure 5: Navigation to User Information Web Link

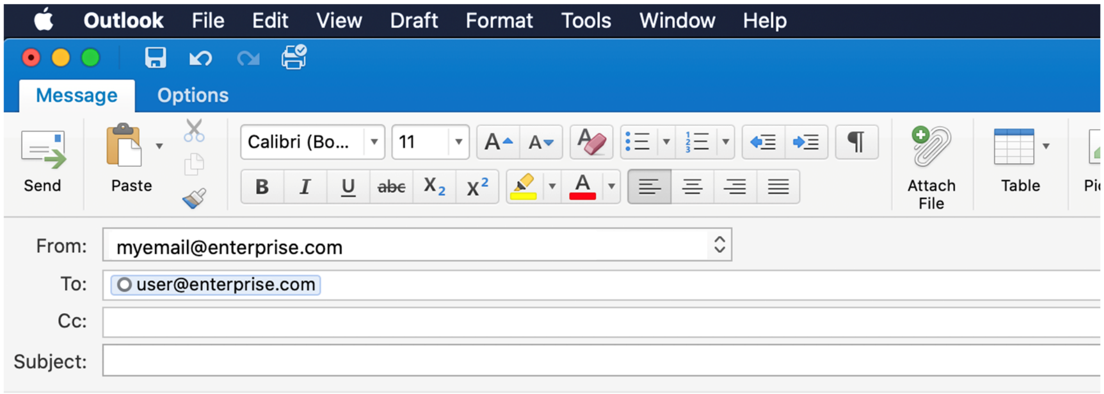
Figure 6 shows an example of an interaction with an email address. A user clicks on the email, which triggers the system to open a mail client to compose correspondence, as shown in Figure 7.



Contacts

Directory profile: https://directory.enterprise.com/dir/reports/user
Contact me:
- Email: user@enterprise.com
- Twitter: @user

Figure 6: Email Address Selection Example



From: myemail@enterprise.com
To: user@enterprise.com
Cc:
Subject:

Figure 7: Opening an Email Client

Thus, in various implementations, this solution can be used for seamlessly interacting with any of the following interactive elements:

- Web links (e.g., for opening in a browser);

- Email addresses (e.g., for opening a mail client with the email destination);
- Social media tags / IDs (e.g., for opening a social media page); and/or
- Telephone numbers (e.g., for saving a contact in a phone book or even making a call).

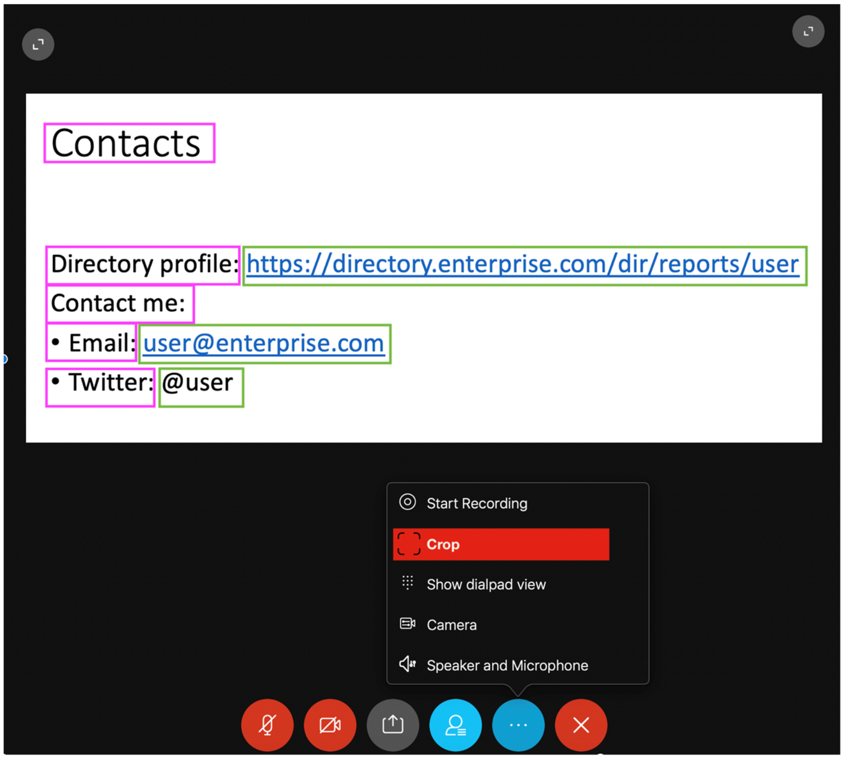
Further, techniques herein boost user experience by providing the ability to interact with shared content without:

- Performing unnecessary actions;
- Interrupting a presenter (asking the presenter to send information directly);
- Bothering other participants (sending information to a group chat and triggering notifications for all people in the chat); and/or
- Waiting for the presentation materials after the presentation.

The image to text converter may be based on an optical character recognition (OCR) open-source library with the performance sufficient for being applied in the "lazy" processing scenario, where the detection and recognition are triggered only by a user click-signal saving computational resources by avoiding each frame processing on the background. In background processing mode though, the converter will be run constantly and highlight links, emails, and text information for a user proactively, so that precomputed entities may be utilized by the system of this proposal. However, in these settings, extensive fine-tuning should be applied to the OCR configuration or custom light-weighted system should be developed to make the processing seamless for an end-user.

Although the computational load in the background processing case will be higher, it can still be partially mitigated by applying the converter only to different video stream frames and skipping an image to text conversion if a previously processed frame is "similar" to a current frame. In some instances, a similarity threshold may be based on a mixture of K-Nearest Neighbor based algorithms, mean structural similarity, and more common metrics, such as Mean Squared Error (MSE), Root MSE (RMSE), etc.

To illustrate the proactive background processing scenario let us take a look at the shared video slide shown below in Figure 8. The frame has been preprocessed on the clientside by a video teleconferencing client such that all entities were identified and recognized even before the user saw the slide.



Contacts

Directory profile: https://directory.enterprise.com/dir/reports/user

Contact me:

- Email: user@enterprise.com
- Twitter: @user

Figure 8: Preprocessing Example

By clicking on the screen part that is inside a green “interactive" bounding box a presentation viewer can immediately trigger a corresponding action associated with the content of the box as opposed to the lazy processing scenario where a slightly longer delay is anticipated.

In summary, techniques of this proposal provide for the ability to select a piece of information from a video presentation in order to trigger an associated action to be executed seamlessly without typing or performing copy-paste operations by a user viewing the video presentation.