How Do I Look? An Evaluation of Visual Framing Feedback in Desktop Video Conferencing

April Slayden Mitchell, Mary G. Baker, Chen Wu, Ramin Samadani, Dan Gelb

HP Laboratories
HPL-2010-175

Keyword(s):

Video Conferencing, Feedback, Framing

Abstract:

When do you want to see a mirror view of yourself in a video conference? All of the time...or never...or only when there is a problem? You don't need to worry about the other person not seeing you when you talk face-to-face, but this can happen in a video conference if you move outside the camera's view angle. We describe a desktop video conferencing system that detects when a participant is out of frame of his camera and provides real-time visual feedback so he can adjust his position until others see him properly framed. We also report on our user study to determine which of several types of visual feedback users prefer. The study reveals that most participants prefer visible feedback only when they are out of frame, rather than the continuously visible mirror view provided by many current systems. Also, feedback preferences vary between one-on-one and multi-party conferences.

External Posting Date: January 6, 2011 [Fulltext] Approved for External Publication

Internal Posting Date: November 8, 2010 [Fulltext]

How Do I Look? An Evaluation of Visual Framing Feedback in Desktop Video Conferencing

April Slayden Mitchell, Mary G. Baker, Chen Wu, Ramin Samadani, & Dan Gelb
Hewlett-Packard Labs, Stanford University
1510 Page Mill Rd. MS 1181 Palo Alto, CA 94304
april.mitchell@hp.com, mary.baker@hp.com, chenwu@stanford.edu,
ramin.samadani@hp.com, dan.gelb@hp.com

ABSTRACT

When do you want to see a mirror view of yourself in a video conference? All of the time...or never...or only when there is a problem? You don't need to worry about the other person not seeing you when you talk face-to-face, but this can happen in a video conference if you move outside the camera's view angle. We describe a desktop video conferencing system that detects when a participant is out of frame of his camera and provides real-time visual feedback so he can adjust his position until others see him properly framed. We also report on our user study to determine which of several types of visual feedback users prefer. The study reveals that most participants prefer visible feedback only when they are out of frame, rather than the continuously visible mirror view provided by many current systems. Also, feedback preferences vary between one-on-one and multi-party conferences.

Author Keywords
Video Conferencing, Feedback, Framing

MOTIVATION

When we make eye contact with another person in a face-to-face conversation, we usually assume that the other person can see us. For standard desktop video conferencing, this assumption does not hold true, as it is possible to see the remote parties clearly but not be fully visible to them. This is due to the fact that their view of us is determined by our position within the field of view of our local camera. Ensuring proper video framing of meeting participants is a critical goal in video conferencing as it will help eliminate one of the important differences between face-to-face communication and many video conferences [1].

Several solutions exist for ensuring that the local user is properly framed within the field of view of his camera. Some high-end conferencing systems control the physical environment to ensure that users can only sit in locations properly framed by cameras [2]. This solution does not work well for desktop video conferencing where we have little control over the physical environment. Some desktop cameras [3] use face-tracking to follow the user as he shifts positions, but this results in the remote party seeing a constantly shifting view of the local user which can be distracting. Other systems overlay a "mirror view" window of the video captured by the local user's camera onto the incoming video from the remote party. This solution constantly occludes a portion of the remote video and can fail to draw the local user's attention when necessary since the mirror view is always present. Having a mirror view always present can also be distracting to the local user if he ends up focusing on it instead of the incoming video.

To help address these problems, we created a desktop video conferencing system that automatically detects when participants are out of frame and provides visual feedback to the user only when he begins to move out of frame. When the local user adjusts his position and is once again in frame, the feedback fades away. We experimented with parameters such as how long the user should be out of frame before the feedback appears and how long it takes for the feedback to fade after the user returns to proper framing. In early informal testing, we found that this form of feedback is intuitive to users engaged in one-on-one conferences. This led us to implement other types of feedback and to conduct a study to test which feedback alternative - including the standard mirror window in a corner - users preferred in one-on-one as well as multi-party desktop video conferences.

SOLUTION

Technology

To enable our tests we built out-of-frame *detection* and out-of-frame *feedback* modules on top of a real-time video conferencing software system [4]. Our custom out-of-frame detection algorithm uses real-time video analysis to determine how far out of frame a participant is, and in which direction. This information is then used by the out-of-frame feedback module.

The goal of the out-of-frame detector is to provide accurate in-frame and out-of-frame indicators in real time, with acceptable levels of processor usage. A core component of the detection algorithm is a Viola-Jones like face detector [5,6]. Since face detectors are not 100% reliable under general lighting conditions, we combine the face detector with a feature tracker based on Harris corners [7] and a motion detection module that provides a fall-back operation in those instances when the face detector completely fails to detect the participant. We find this system more robust and faster than a standalone face-detection based system. To keep processing needs to an acceptable level, we only run

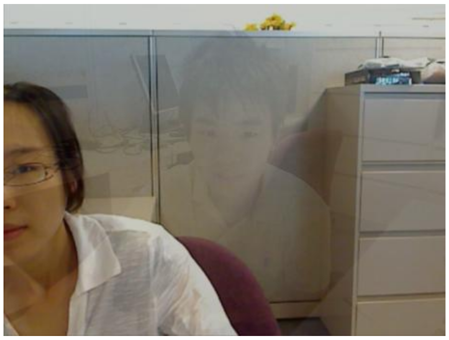

Figure 1. AlphaBlend - all feedback types, except Always On, are activated only when user moves out of the camera's view

Figure 2. AlphaCorner

the face detector on every tenth frame, and twice as often if feature tracking has failed. More details on our custom out-of-frame detector will be provided in a future publication.

The out-of-frame feedback module uses GPU processing to provide different pixel blending and geometric effects, which are used to build the different feedback methods that we describe next. Running the detection and feedback modules uses roughly 7% of the CPU on an Intel Xeon X5560 2.80GHz quad core processor with an NVIDIA Quadro FX 4800 graphics adaptor, and we continue to receive a full 60 frames per second.

Feedback Alternatives

While there are many ways to provide a user feedback about their framing (auditory, tactile, visual), we chose visual feedback, since many users have experience with it already. In most standard desktop conferencing systems, the inset mirror view window can be toggled off by the user but by default is on all of time. We wanted to experiment with alternative types of feedback that we hoped would be less distracting and more effective. Our goal for this user test was to determine which of the following types of visual feedback users prefer:

- AlphaBlend: An alpha blend of the user's local video (mirror image) at the same relative size as the remote user is shown on top of the video received from the remote user whenever the local user moves out of frame. The blended overlay fades to transparent when the local user moves back in frame. See Figure 1.

Figure 3. Slide Window - shown during multi-person conference

Figure 4. AlphaCenter

- AlphaCorner: An alpha blend of the user's local video at approximately ¼ of the relative size of the remote user is shown in a window at the bottom left corner on top of the video received from the remote user. The blended overlay fades to transparent when the local user moves back in frame. See Figure 1 for an example of what the local user would see when they move out of frame. See Figure 2.
- SlideWindow: The remote image slides in the opposite direction whenever the local user moves out of the frame. This feedback mimics the concept of parallax and is similar to how you might view people through a door or window aperture. The further the local user moves past the boundary of the aperture, the less of the remote user they are able to see. See Figure 3.
- AlphaCenter: This is similar to the AlphaCorner feedback, but the feedback window is located in the center of the video (see Figure 4). We only tested this feedback alternative for multi-party conferences, and we chose the center position as being least likely to interfere with the grid layout of our multiple remote users. Other locations would need to be tested for different layouts.
- Always On: This feedback method displays the user's local video at all times in the bottom left corner on top of the video received from the remote user.

METHOD

Participants

We recruited participants through a site-wide email at our workplace. All respondents completed an online survey that asked them to indicate their prior experience with desktop video conferencing, describe a typical setup for a desktop video conference (one on one vs. multi-party, as well as how many people on each side), and provide their age/gender. From the respondents, we selected a total of 14 participants to include a broad range of ages ranging from 20-59) and split equally by gender.

We selected only participants who had some prior desktop video conferencing experience, since only a few respondents had none at all. Skype was the most commonly used desktop video conferencing platform used by our participants (>80%) followed by Google Video Chat (50%) and others (each <25%). While conferencing for personal use was most common, 50% of our users indicated that they had used desktop video conferencing for work purposes. The typical past conferencing experience for the majority of our participants involved a single user on a laptop connected with just one remote party for personal communication.

Design

The testing environment consisted of an office cube with a PC placed under the desk and a monitor/keyboard/mouse located on the desk. A chair was available for the user which placed them directly in front of a USB camera located on top of the monitor.

A moderator met each participant at the beginning of the session and introduced them to the study by explaining that during the half hour we would ask them to evaluate different types of visual feedback provided during two different conference setups. Each participant was informed that not all of the feedback types would be visible 100% of the time. The moderator indicated that at any time if the participant did not see any feedback on the screen, they should reach over and touch a sticky-note located about 3' to the left of the monitor. They were informed that this action (or any action that would move them out of the frame of the camera) would cause the missing feedback to appear. During the first test, the user was connected to a remote person sitting in a different cube. They were able to see and hear and respond to the remote person (through audio available on the PC speakers). The remote party in this 1-1 test was not a facilitator and interaction with the remote party was only an introduction, thus making the experiment more easily repeatable. The user was then allowed to step through four different types of feedback self-paced. These included the standard corner "always on" feedback as well as the first three types of feedback described in the previous section. After seeing all of the types of feedback for the one-on-one conference, we asked each user to choose a favorite and least favorite type of feedback. We also asked users to describe their ideal feedback for this scenario.

We followed the same steps for the multi-party conference scenario which included all of the previous types of feedback as well as AlphaCenter. For this test, we used still images to represent the four remote participants in lieu of live video feeds. This limitation was due to lack of additional people resources during the testing; however, we feel the experiment is still valid because our aim was to understand users' preferences among many feedback alternatives and not to evaluate their overall conference experience. We always performed this test second, and the order of the feedback types for both the one-on-one and the multi-party conferences were randomized for each user. Each per-user session lasted approximately 30 minutes.

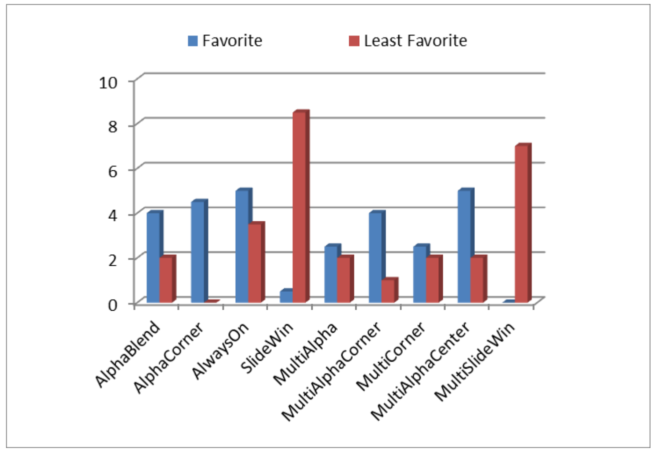

Figure 5. Number of users who chose each feedback as favorite or least favorite

RESULTS

Desirable Alternatives to Always On Exist

As shown in Figure 5, over 60% of our users preferred the standard type of feedback (AlwaysOn) to be off. Overall, they expressed a desire to see the other person and not themselves and preferred one of the alpha blending feedback types to the "always on" feedback.

[in regards to "always on"] "It provides no additional feedback, it's always there." – M, 30-39

Two of the users that wanted the feedback always visible stated that during their typical video chat something they were holding or wearing was always the focus of the conversation, so they wanted to control the view of it received by the remote party.

"Maybe if I'm showing her [Mom] my dress…I can adjust myself. If it's with my Dad, I wouldn't need to see it." F, age 30-39

Different defaults for one-on-one vs. multi-party

Over 70% of our users had a different preference for their favorite feedback method for one-on-one versus multi-party conferences. Without prompting, users' comments indicated that they tend to associate the one-on-one conferences with personal or home use and the multi-party with professional or work use. These associations influenced their choices and often led users to prefer a different type of feedback as their favorite for the multi-party "work-like" setting.

"I'm thinking of it [AlphaCenter] in a corporate sense… in a personal engagement I may feel differently." – F, 40-49

The majority of users preferred the AlphaCenter feedback (see **Error! Reference source not found.**) with a slightly smaller feedback window for the multi-party scenario because it combined the subtle blending and wouldn't occlude any users' faces.

Faces are important

Overall, many participants mentioned that in either type of conference it was important not to obscure the face of the remote party. This helps explain the ranking of SlideWindow as the least favorite feedback type overall. Among its faults, users primarily complained that they never wanted the remote party to move off the screen partially or entirely (as can happen with the SlideWindow multi-party scenario).

"If I'm talking to him, I can't see him. You're losing half your audience." M, age 50-59

Several participants also mentioned that they did not like to see themselves "looking down." In our setup (which is similar to many existing conferencing setups), the camera was located at the top of the monitor and the feedback window appeared at the bottom of the screen in a lower corner. Whenever participants looked at the feedback window, they could see their eyes focused down as opposed to straight ahead (see Figure 6). They wanted to be able to see themselves looking straight ahead, as if they were looking at the other participants even when they were looking at themselves.

"When I'm looking at myself, they see me looking down." M, age 20-29

When describing their unique solution, several users also mentioned wanting to make their own feedback window smaller as it was more important to see the remote party. While our tests focused only on visual feedback alternatives, our users did mention other forms of feedback as part of their ideal solution. Half of our users mentioned using an audio alert or "bing" as they called it to draw attention to them moving out of frame.

CONCLUSIONS & FUTURE WORK

Based on our findings, we recommend that by default feedback should only become visible in response to users being out of frame but that users have an easy option to toggle the feedback to "always on." The visual feedback location during a multi-party conference should be closely tied to the layout of faces on the screen; therefore it might not be in the same location as during a one-one-one conference. Whenever possible, the feedback window should be located near the position of the camera to ensure the appearance of good eye contact with remote participants even when a user is looking at the visual feedback window. Making the feedback window moveable and resizable should accommodate these different preferences.

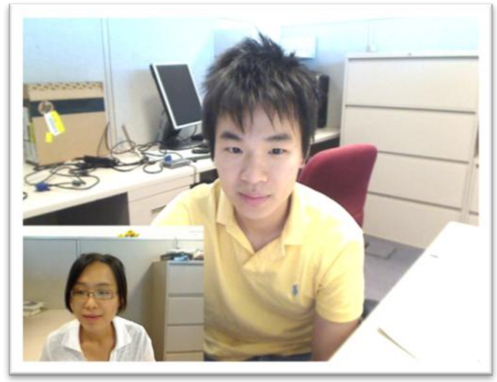

Figure 6. Users' eyes are focused down when the feedback window is not located near the camera

While these results can have broad application across many services that offer conferencing solutions on desktop and laptop platforms, there are more opportunities to understand how varying screen size and mobility can affect a user's need for feedback. For instance, with a mobile device such as a phone, making sure you are properly framed is even more important as the camera will likely move during the conversation. We plan to expand this research to include alternative feedback options for mobile devices and those with limited screen real estate. We would also like to test how effective our feedback methods are at improving participant framing in live video conferences from the point of view of the remote party.

REFERENCES

1. David Nguyen and John Canny, "More than Face-to-Face: Empathy Effects of Video Framing." *CHI 2009*, April 2009.
2. HP Halo, URL: http://www.hp.com/go/Halo, last accessed 24 September 2010.
3. Logitech Orbitz motorized tracking, URL: http://www.logitech.com/en-us/webcam-communications/webcams/devices/3480, last accessed 24 September 2010.
4. Donald Tanguay, Dan Gelb, and Harlyn Baker, "Nizza: A Framework for Developing Real-time Streaming Multimedia Applications, *HP Labs Technical Report HPL-2004-132, HP Laboratories*, August 2004.
5. Paul Viola and Michael Jones, "Robust Real-time Face Detection." *International Journal of Computer Vision*, vol. 57, no. 2, 2004.
6. Darryl Greig, "Video Object Detection Speedup Using Staggered Sampling," *IEEE Workshop on Applications of Computer Vision*, 2009.
7. J. Shi and C. Tomasi, "Good Features to Track." *9th IEEE Conference on Computer Vision and Pattern Recognition*, June 1994.