Avoiding Nostril-cam and Postage-stamp People in Mobile Video Conferences

Mary Baker, Ramin Samadani, Ian Robinson, Mehmet Yilmaz, Kean Wong, Matthew Hornyak

HP Laboratories
HPL-2011-53

Keyword(s):

mobile, video conference, teleconference, user experience

Abstract:

We would like to provide high-quality video conferencing to make it possible for people to communicate with each other anywhere, anytime. There are now several applications on mobile platforms that bring us closer to achieving the "anywhere, anytime" goal, but these mobile conferences are often of poor quality. This is due to the many challenges presented by mobile devices, such as constrained networks, limited processing power, small displays, and uncontrolled view angles and lighting. These challenges mean that simply porting existing video conferencing solutions to portable devices will not provide the best user experience. Fortunately, these devices also have many advantages which we can exploit to enable better-quality portable video conferences. In this paper we describe how we use the devices' mobility and rich set of embedded sensors to detect and fix two user experience problems: bad view angles and too-tiny views of people and content, especially in multi-party conferences.

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

Avoiding Nostril-cam and Postage-stamp People in Mobile Video Conferences

Mary Baker, Ramin Samadani, Ian Robinson, Mehmet Yilmaz, Kean Wong, Matthew Hornyak

Hewlett-Packard Laboratories, Hewlett-Packard Personal Systems Group, Stanford University

1501 Page Mill Rd., Mail Stop 1183, Palo Alto, CA 94304

mary.baker@hp.com

Abstract

We would like to provide high-quality video conferencing to make it possible for people to communicate with each other anywhere, anytime. There are now several applications on mobile platforms that bring us closer to achieving the "anywhere, anytime" goal, but these mobile conferences are often of poor quality. This is due to the many challenges presented by mobile devices, such as constrained networks, limited processing power, small displays, and uncontrolled view angles and lighting. These challenges mean that simply porting existing video conferencing solutions to portable devices will not provide the best user experience. Fortunately, these devices also have many advantages which we can exploit to enable better-quality portable video conferences. In this paper we describe how we use the devices' mobility and rich set of embedded sensors to detect and fix two user experience problems: bad view angles and too-tiny views of people and content, especially in multi-party conferences.

The problems

We address two core user experience problems that affect video conferencing on portable devices. The first problem is bad view angles of conference participants, and the second is tiny displays of participants and documents (or other content) in conferences that involve more than two or three people.

Bad View Angles

Users often hold portable devices in their hands so that the video camera capture view angle is not constrained or fixed, and this can result in a poor angle of view for video capture. Commonly, users hold the phones too low, even if they are propped on a table, resulting in a distorted and unappealing upward view of nostrils and chin, or what we call "nostril cam." We see these participants looking down their noses at us, with chins tucked toward necks so that even the slimmest participants have multiple chins (Figure 1). Instead, we would like to see people at their best and be assured others see us from a good view angle as well.

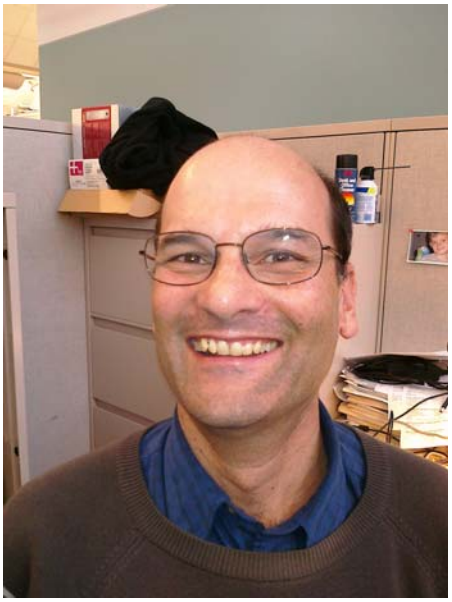

Figure 1a: a good view angle

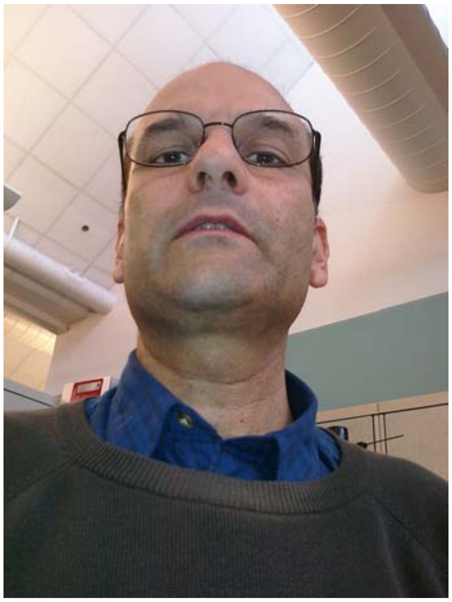

Figure 1b: a bad view angle

Postage Stamp People

For a comfortable user experience in video conferencing, it is important to see the other participants well enough to judge mood and expression, and to have them arranged such that they appear to maintain their positions with respect to each other during the conference. For instance, on a large enough display, the local user might always see participant Alice next to and to the left of participant Bob. This allows the local user to participate with the same set of expectations regarding positioning as if he were at a physical table with the remote participants physically present. Likewise, the user should have a good view and comfortable sense of the position of any document or other shared artifact being viewed in the video conference. Portable devices such as smart phones do not have large enough displays to render multiple remote participants and shared artifacts at sufficient size. To have a large enough image to judge mood and facial expression, we might only be able to view one remote participant at a time. Instead, we would like to be able to view other participants and documents with sufficient size and a comfortable sense of their positions.

Our solutions

Our current solutions to these problems exploit the mobility of the portable devices and their embedded sensors.

Bad View Angle Detection and Feedback

There are two parts to solving the bad view angle problem: the first is to detect the bad angle and the second is to fix it. Instead of using unreliable computer vision techniques, we reliably determine the angle of the phone dynamically during the conference using sensors that are already found in many smart phones and slates. For instance, three-axis accelerometers are commonly used to sense orientation of the display screen (portrait or landscape), but in our prototype application we use them for a different purpose. The accelerometers sense gravity direction and the components of "gravity acceleration" in the local coordinates of the accelerometer, and this provides view angle information to answer the question: how close to the ideal angle is the device being held? Ideally, the device is held vertically in front of the user's face, as if he were talking face-to-face with the remote participants. For comfort, though, we tolerate some small angle from the vertical, for example 15 degrees from vertical (with respect to Earth’s gravity). In a coordinate system where x and y are in the plane of the device, and z is the coordinate towards the viewer, we use the magnitude of the z component to detect the view angle. The x and y components are useful to ensure the display is not rotated. If front-facing depth cameras become available on mobile platforms, we can improve on this technique by sensing the relative angle between the mobile device and the person’s face regardless of the position of the person (sitting up, lying down, etc.)

The second part of our solution is to fix the angle at which the device is held. We chose the approach of giving subtle visual feedback that has the end effect of causing the user to change how he is holding the device. There are many ways to provide this feedback, and our goal is to do so in a way that is not intrusive but naturally causes the user to fix the angle of the phone. Our current method is to reduce the contrast or color saturation of the incoming video in a manner proportional to the wrong view angle. The participant naturally adjusts the device to a good angle because doing so restores the contrast or color. Since there are times when a bad angle is intentional or cannot be helped (e.g. when the user is lying down or walking while conferencing), it is important to have an easy way to disable the feedback, although we do not yet do this in our prototype.

Better Sizing and Stable Positions

To solve the problem of postage-stamp-sized people in mobile multi-party conferences, we use the mobile device as a “window” into a larger virtual space and can exploit accelerometer, compass, gyroscope, and touch sensor data to navigate the space. Consider participant “Mary” in a multi-party conference. Her phone or tablet receives video streams over the network, each of which provides video to display one or more remote participants. To provide the illusion that the remote participants are arranged in front of and facing Mary, as they would be if sitting across a physical table, we choose or calculate their arrangement and then use Mary’s device as a window onto that arrangement. Assuming we have Alice, Bob and Carol as remote participants, we can arrange them in that order from left to right in a virtual space. If Mary holds her mobile display up and pans it from left to right or rotates it from left to right around a vertical axis, she will see Alice, then Bob and then Carol as if they were in front of her but seen through the mobile display. We could also tilt the phone slightly to one side or another (as in

Figure 2) to move between participants. (We have made both of these techniques available in our prototype.) Slides or other shared artifacts can similarly be positioned in the virtual space. Note that the degree of panning, turning or tilting required to move through the virtual space is adjustable; a small movement can lead to a larger change of view. This is good, because requiring a large movement can result in bad view angles.

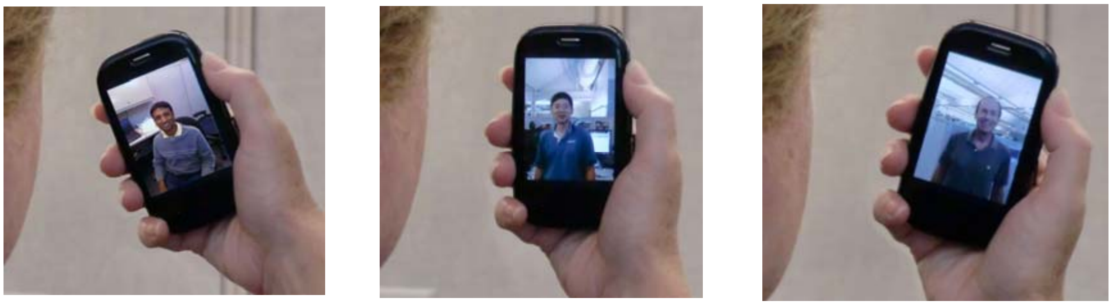

Figure 2: using the tilt option for view change app on Palm device

Using our solution also gives us an important advantage in the face of constrained networks and devices, since we can limit the number of simultaneous video streams that must be received and decoded by the device to just the visible participants and perhaps their neighbors. We require all the audio streams, but can be more selective about the video streams, which require more bandwidth.

Panning, turning and tilting use accelerometer, gyroscope and compass data, but we can use other kinds of sensor input as well. With a touch interface, we can scroll the display left or right to move to different remote users or slides or other shared artifacts. We could instead use a rear-facing camera to detect device motion (as the scene viewed by the camera changes) to augment or replace the gyroscopic or accelerometer inputs, or we could use the front-facing camera to detect the user's head position in the camera's field of view and use changes in head position to augment or replace the gyroscopic or accelerometer inputs.

Without enough display space to provide high-quality images of all the remote participants at once, the local user may not have enough information about which way to turn to view the current active speaker. Our current solution to this problem uses spatial audio cues through head phones (the voices of people positioned further to the left come proportionally more from the left ear phone, and vice versa). In the future, we could also provide a small strip of thumbnail images at the top of the display (Figure 3), arranged in order of the remote participants, with the currently viewed speaker highlighted (and perhaps the active speaker differently highlighted) so the local user knows which way to pan or scroll the display, or so he can just touch a picture to bring up the image of a particular remote participant. We do not automatically switch to the stream with the active speaker, since it proves disconcerting to users to lose control over whose face they choose to view.

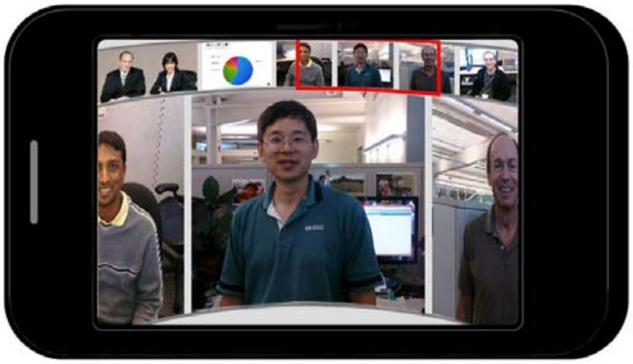

Figure 3: mock-up of a possible multi-party UI

Evidence the solution works

As of this draft, the Palm phones we are using do not yet have front-facing cameras, gyroscopes or compasses, so we have been testing our ideas in advance of these expected features. To demo bad view angle detection and feedback, we have implemented an application on the Palm Pre Plus that proportionally grays out (reduces color saturation) of video as the angle of the phone is tilted away from the vertical. Through this test we have determined that the precision of the accelerometer data is sufficient to provide responsive, smooth feedback. Informal testing with over ten users shows that the feedback is intuitive, since users immediately adjusted the position of the phone in their hands to restore color to the video.

We have implemented a combination of demos to explore viewing a virtual space of participants through the mobile device. Using the touch interface, we can scroll left or right through images of participants arranged in a virtual space. Using the accelerometer, moving the phone quickly to the right or left moves from one picture to the next in the indicated direction. Also, tilting the phone to the right or left scrolls across the images to the left or right until the user returns the phone to a neutral position. In informal user testing, the tilt interface seems to be as easy a way to scroll through images as the touch interface, but it must be lightly applied so that it does not contribute to our other problem: bad view angles. We believe that panning or rotating the phone or using the touch interface will probably beat the tilt interface in the end, but this must be verified through user testing.

Competitive approaches

As far as we know, nobody else offers the features we describe in this paper. For addressing bad view angles, our most obvious competitor is the industry-standard small mirror window overlaid on the display (e.g. in Apple's FaceTime) in which the user can see what he looks like to the other participants, but we have found through formal user testing [1] that consumers generally prefer not to see themselves all the time in a mirror window. They prefer such feedback to be visible only when something is wrong. This is even more important on small displays where a larger percentage of the display is occluded by the mirror window. Another approach for improving view angle, which also helps to avoid tired arms, is to use a phone stand such as the amusingly named iPlunge (http://www.worldwidefred.com/iplunge.htm). This requires a horizontal surface like a table, which is not always available, and unless the surface is high enough, the angle is still bad. Most mobile video conferencing solutions such as FaceTime do not yet offer multi-party conferences, but those that do, such as Damaka, are limited to a small number of participants (usually four) who then become postage-stamp-sized people on the display.

Current status and next steps

Our demos show new ways to use mobile device sensors to provide natural, simple solutions for two fundamental problems of small devices: bad viewpoint and small display area. We will take advantage of the new sensors on webOS handsets and tablets to complete the solutions we describe above. At that point we will begin formal user tests. If the results of the tests are encouraging, we will include these features in our port to webOS of our experimental video conferencing application [2].

References

[1] A. Mitchell, M. Baker, C. Wu, R. Samadani, and D. Gelb, "How Do I Look?" *HP Labs Technical Report HPL-2010-175*, HP Laboratories, November 2010.

[2] D. Tanguay, D. Gelb, and H. Baker, "Nizza: A Framework for Developing Real-time Streaming Multimedia Applications." *HP Labs Technical Report HPL-2004-132,* HP Laboratories, August 2004.