April 06, 2018

Automatically emphasizing displayed content in video communication interfaces

Bailiang Zhou

Automatically emphasizing displayed content in video communication interfaces

ABSTRACT

Video calls, video conferences, and similar real-time communications between devices display users and other subject matter to other users participating in the communication in a user interface. This disclosure describes techniques in which display of particular video content can be emphasized in the user interface based on particular characteristics of the video content. The emphasized display can be used to automatically highlight video content at appropriate times, allow easier viewing of video content by other users in a communication or other video capture or display application, etc.

KEYWORDS

- video call
- video chat
- motion detection
- chat window
- facial expression

BACKGROUND

Real-time video communications includes video calls, video chat, and videoconferences between devices, including mobile devices. Faces of users participating in the video communication are typically displayed in views or windows in a user interface based on cameras capturing the faces at the user devices. Other content can be displayed in a view in the user interface, for example, if the user moves the device or camera, or chooses a different camera such that a different scene or other content is captured. In some systems, one of the views from the participating users is sized larger and/or positioned in a central location in the

user interface, based on which of the users is currently talking. However, limited available space of the user interface can lead to other views to be displayed in a small size, especially if several users are participating in the communication. Users thus cannot see details of content displayed in the views, nor notice when views display new content, especially on small screens of mobile devices.

DESCRIPTION

This disclosure describes techniques that allow display of video content in a view (e.g., window) associated with a user in a video communication to be emphasized based on changes of the content being captured and displayed in the view. The techniques include detecting changes in captured and displayed video content, which can indicate that the content is likely to be of greater interest to the participating users and that the content should be highlighted. The rendering of the view and/or the content of interest is then updated in the interface in a particular way to emphasize the content, e.g., by enlarging the display of the content in the user interface.

Detecting qualifying video content characteristics

Described techniques include detecting video characteristics in displayed views that indicate that the content should be emphasized in the user interface. For example, the characteristics can include changes of video content. Such changes include motion and/or pixel changes displayed in the view. For example, it is determined whether a threshold number of pixels have changed in the displayed view, e.g., a threshold percentage of pixels of the view, within a threshold time period. Pixel analysis can be used to determine which pixels have changed to new pixel values in comparison to the previously-displayed pixel values in the view.

If larger than the threshold number of pixels have changed, the content is determined as likely of interest to the user and selected to be emphasized in the user interface.

In some examples, a scene change captured by a camera and displayed in the view can provide a large number of changed pixels. For example, if the user moves a mobile device or switches cameras on the mobile device (e.g., from a front camera to a rear camera), a large number (or all) of the pixels in the view may change. In another example, an object may be moved or introduced in the view, e.g., the user holds up a new object that is captured in the view, or a second person moves into the view. In another example, an existing object in the view may be moved, e.g., a face may move in a particular direction, a facial expression may change on a face, etc.

When users provide permission, object detection techniques can also be used to identify types of objects in the view, which can assist in determining whether a change has occurred in the view and/or whether the change is significant and indicates the content and view should be emphasized. For example, when users provide permission for use of facial detection technology, faces and facial features such as mouths, eyes, etc. can be detected. Since a person's body language and facial expressions can be critical for communication, pixel comparisons for particular facial regions of faces are performed, e.g., to analyze the user's eyes and mouth and facial muscle movement to detect changes in facial expressions and/or changes in moods (e.g., detecting a change from happiness, indicated by smile, to sadness, indicated by a frown or straight mouth). Users are provided with options to decline or modify consent for use of facial detection/recognition techniques, and such techniques are not utilized when users do not provide permission.

Additionally, compression rates of transmitted video data can also be examined to determine if changes are occurring in a displayed view. For example, for video transmission having a constant bit rate, the amount of compression applied increases as more pixel changes occur, to maintain the constant bit rate. This can be used as an indicator of the rate of pixel change.

For example, a number of the last displayed frames of video, e.g., a threshold number of frames, such as frames occurring within the last 3 seconds or 5 seconds, are examined to determine whether to maintain a view in the emphasized state. If previously-changing pixels are displayed for more than this number of seconds without changing, such pixels are considered static pixels. The percentage of changing pixels in this instance may fall under the threshold percentage, and if so, the view is removed from emphasized status, or reduced in priority for such status.

Other characteristics that can indicate that content should be emphasized include types of image features depicted by the content, if users provide consent for detection of types of image features. Users are provided with options to selectively provide permission for detection of such image features. For example, with user permission, types of objects, such as faces, landscape objects (trees, buildings, clouds, etc.), monuments, articles, event signifiers (pumpkins, cakes, etc.), etc., can be detected and can be determined as being designated by users to indicate that the view displaying such types of objects should be emphasized. In another example, particular hand gestures can be detected if a hand has been detected in the view, such as an open palm gesture, a pointing gesture, etc., which can indicate a signal to the system from the user to emphasize the content.

Certain types of content may be designated as eligible to cause the content's view to be emphasized, and/or other types of content may be designated as ineligible. For example, if objects such as floors or ceilings are detected moving within the view, this content can disqualify the view from being emphasized, as this content is likely to be unintentional video capture during camera motion. In another example, content such as an offensive hand gesture or other offensive content can be considered ineligible to cause its view to be emphasized.

Optionally, user input is obtained to confirm that a particular change in pixels is intended by the user, or to otherwise confirm that a view should be emphasized. For example, the user can provide a touch gesture on a touchscreen (e.g., double tap) after switching the active camera from the front camera to the back camera, to indicate that a change in content is intended. This can reduce occurrences of emphasizing unintended changes in the video, such as moving a mobile device around while the camera is unintentionally capturing video.

Emphasizing a view and video content

Described techniques include emphasizing the display of a view once the characteristics of the content of the view have been determined to qualify the view for such emphasis, e.g., the content is determined as likely interesting to users. A view and its content can be rendered differently in a number of ways to emphasize the content.

For example, the size of a view can be increased to provide an emphasized view. The increase in size of the view accordingly increases the size of the displayed content in the view. In some cases, e.g., when an interface displays the views in a fixed interface area, the increase in size of a first view may be simultaneous with a reduction in the size of one or more other views in accordance with the increase in size of the first view.

In some examples, e.g., when participating users in a video call provide permission for automatic camera control, the camera capturing the content can be controlled to automatically zoom to capture the content at a higher camera zoom setting. In cases where a face is being zoomed, facial detection techniques can be used with user permission to cause the view to zoom in on the facial features of the face. Emphasizing a view can include moving the position of the view within an interface area. For example, an emphasized view can be moved to the top of an interface area, or closer to the center of an interface area.

The amount of content change can be used to determine the amount of emphasis for the views. For example, a large amount of content change (large percentage of pixels changing) can cause a correspondingly large size increase of the view. The content change in multiple different views can be compared with each other to determine the resulting view changes. For example, if a first view has a greater number of pixels change than a second view, the first view can be enlarged and the second view not enlarged; or the first view and second view can each be enlarged by an amount proportional to the number of changed pixels in the respective view.

Various types of content changes in a view are analyzed to determine priorities of display emphasis. For example, if two or more views are simultaneously displaying content changes, a view having a content change of a higher priority will be emphasized instead of (or to a greater degree than) a different view having a type of content change of a lower priority. For example, switching cameras on the mobile device (e.g., from a front camera to a rear camera) or an equivalent scene change, can be assigned the highest priority. Other content change types include scene changes caused by moving a camera in space; new objects being moved into the view; existing objects in the view being moved within the view; a person speaking (mouth detected moving); a change in facial expression; static non-facial objects; and

a static face with no expression, which can be assigned priorities, each having a lower priority than the earlier listed type of content change.

Emphasizing views can also be based on priorities of the types of content changes. For example, a content change due to switching from a front camera to a rear camera (highest priority in this example) can cause an increase in view size by 50%, while a content change from a facial expression changing (lower priority in this example) can cause an increase in view size of 25%. A range of different view sizes can be provided based on the type of content change or priority scores, and different views in the interface can be changed to different sizes and displayed simultaneously in the user interface.

Certain types of objects can be designated to have a maximum size change of the view. Different types of objects can have different maximum sizes. For example, a change in facial expression can be associated with a maximum size increase of 20%, while a change in camera scene can be associated with a maximum size increase of 70%.

A change in view size can be limited such that the view does not change too radically in size for a given change in content. For example, the view size can be changed from a smallest size to a middle-range size, instead of directly to a largest size from the smallest size.

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 social network, social actions or activities, profession, content items, 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.

Examples

Figs. 1-4 show examples of a video chat interface in which views have been changed in size and location based on detected content changes.

Fig. 1 illustrates an example video call or chat interface 100 displayed on a device, e.g., by a mobile device display screen. Interface 100 can be displayed by each device participating in the chat. Two user participants are displayed in separate views 102 and 104, whose faces are captured by cameras in real time at each device of the call.

Fig. 2 shows a change in the size of displayed views 102 and 104 in the interface 100 of Fig. 1, based on a change in video content shown in view 102. For example, the user displayed in view 102 in Fig. 1 has switched the capturing camera of her device from the front camera to the rear camera, such that the scene shown in Fig. 2 is captured and displayed in view 102, causing a large change in pixels in the view 102. The user displayed in view 104 of Fig. 1 has changed his facial expression in Fig. 2, causing a much smaller change in number of pixels in view 104 than in view 102. Due to the larger change in pixels and/or a higher priority of the scene change in view 102, the view 102 is highlighted by being increased in size, while the view 104 is reduced in size to accommodate the larger size of view 102. The frames of the

video content in these views can be cropped to fit the new aspect ratios of the views shown in Fig. 2.

Fig. 3 shows another example, where views 302, 304, and 306 are displayed in a video chat interface 300, which show three faces of users participating in the video chat. View 302 has a larger size than views 304 and 306 because the user displayed in view 302 is talking.

Fig. 4 shows the video chat interface 300 after the user shown in view 304 has moved his camera and mobile device to cause another person to appear in the view 304. This causes a greater number of pixels to change in view 304 than in views 302 and 306, which in turn causes view 304 to be displayed having the largest size. The position of view 306 is rearranged in Fig. 4 relative to Fig. 3, from the bottom portion of the interface 300 to the upper portion of the interface 300, to accommodate the larger size of view 304.

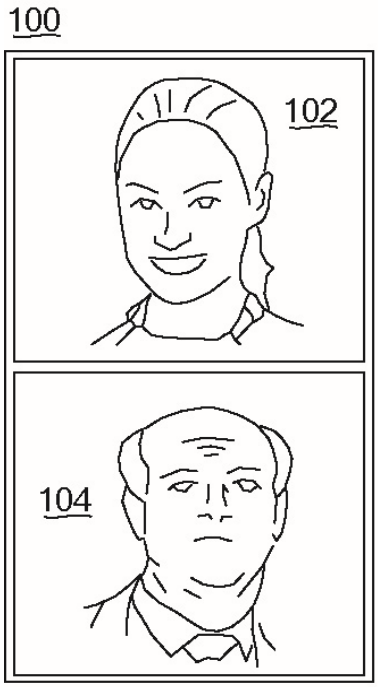

FIG. 1

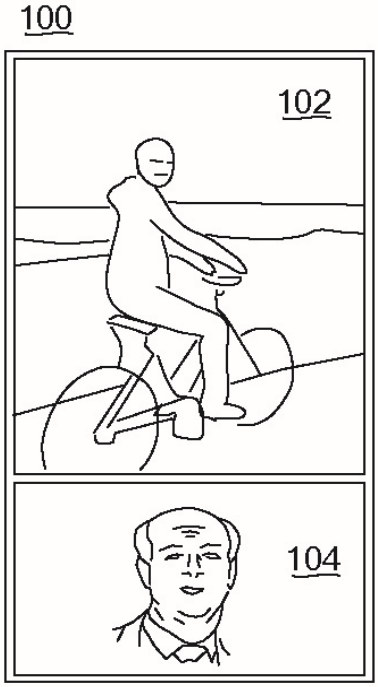

FIG. 2

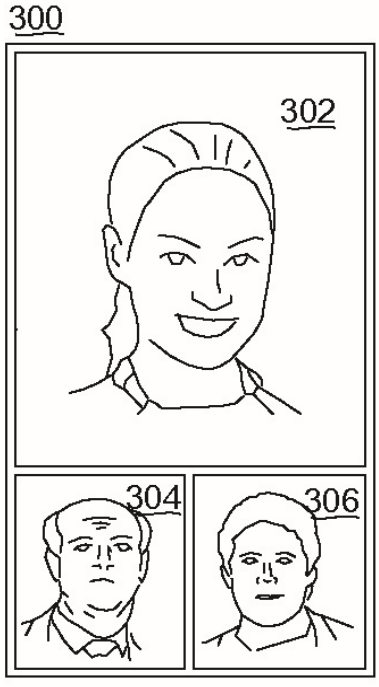

FIG. 3

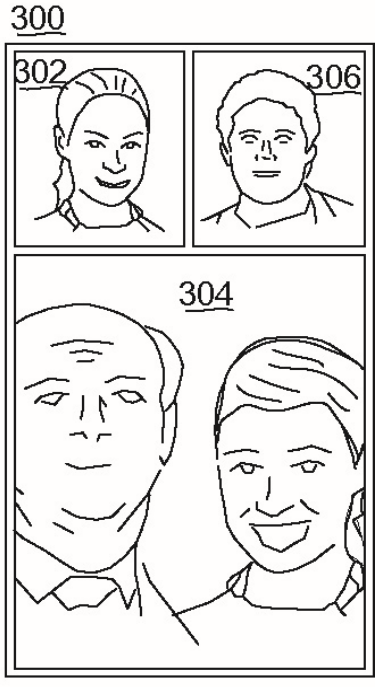

FIG. 4

CONCLUSION

This disclosure describes techniques for emphasizing the display of video content in user interfaces for video communications. The emphasized video content is determined as likely of interest to users based on changes detected in the video content or based on other content characteristics, determined with user permission. These techniques provide user interfaces that help users to notice and view details of new and changed content, including changes of facial expressions. A user interface is made flexible and dynamic in accordance with the video content displayed in that interface.

In video chat sessions, after an initial period, the value of displayed visual information may decrease unless changes in the video information occur. The described techniques present a dynamic interface that changes in presentation in accordance with changing visual information. Described techniques provide a way to provide facial expression capture and display in a more effective way to enhance human communication, and provide a way to maximize the usage of a small-sized screen to display new and expressive video content in video communications.