August 2022

Camera Focus Adjustment Using Depth Estimated via Ultra-wideband (UWB) Handshake

D Shin

Camera Focus Adjustment Using Depth Estimated via Ultra-wideband (UWB) Handshake

ABSTRACT

Many video conferencing applications support portrait mode where the participant is in focus and their surroundings are blurred. Portrait mode based on computer vision techniques can sometimes be unsatisfactory due to difficulties in separating the user and the background. This disclosure describes techniques that use existing ultra-wideband (UWB) hardware on commodity devices to perform crisp and more accurate segmentation. A UWB handshake protocol is utilized to estimate the distance and angle between the camera and another device that the conference participant is wearing. The estimate is used to automatically adjust camera focal length to focus on the conference participant while blurring other objects. The techniques can make high-resolution portrait mode video conferencing affordable for users with commodity devices.

KEYWORDS

- Ultra-wideband (UWB) chip
- Video conferencing
- Depth estimation
- Depth camera
- Portrait mode
- Person detection

BACKGROUND

In many video conferencing applications, users can set the camera to focus on the user with the background blurred. Such an operation provides better aesthetics and enhances privacy. Accurate implementation of the focus and blur operations requires semantic understanding of various objects in the world in order to determine the object on which the camera is to focus at any given time and adjust the focal distance accordingly.

Typically, such operation is realized with the application of computer vision algorithms on images obtained from the camera feed. The output of the algorithms provides adjustments necessary for achieving autofocus via a physical change in the focal length of the camera lens and/or alteration of the focus via digital techniques. However, such an approach can be less than optimal as it can often be difficult to separate the user and the background depending on the properties of the scene, such as textures, light levels, contrast, etc. For instance, a user can blend in with a low-textured background, making it difficult to separate the user and the background using only object segmentation methods that employ computer vision.

Other technologies, such as LIDAR (Light Detection and Ranging), offer greater accuracy in detecting distances between the camera and various objects. However, such technologies are too expensive and/or bulky to be incorporated within commodity mobile devices used for the purposes of video conferencing.

DESCRIPTION

This disclosure describes techniques that use existing hardware on commodity devices to perform crisp and accurate segmentation between human users and other objects within a scene. Specifically, the techniques employ ultra-wideband (UWB) depth sensing principles to estimate the distance between the camera of the video conferencing device and another device, such as a smartphone, earbuds, smartwatch, etc., that the user is wearing or carrying on their person. With user permission, a distance estimate is generated and is used for automatically adjusting the focal length of the device camera to provide maximal focus on the user while blurring the other objects within the video feed.

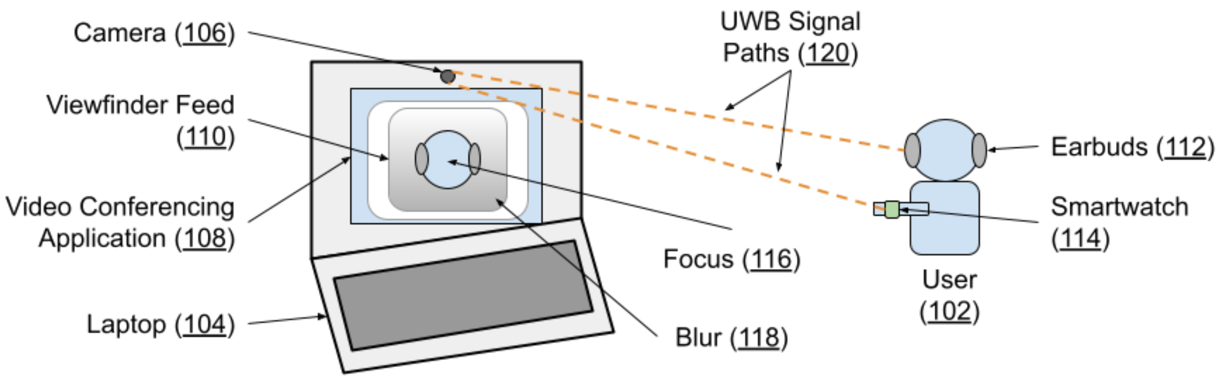

Fig. 1: Adjusting camera focus and blur based on distance measurement via UWB signals

Fig. 1 shows an example of operational implementation of the techniques described in this disclosure. A user (102) wearing earbuds (112) and a smartwatch (114) is using a video conferencing application (108) on a laptop (104). The distance between the laptop and the user is estimated using UWB depth sensing via UWB signal paths between the laptop and the earbuds and/or smartwatch. The estimated distance is used to adjust the focus of the laptop camera (106) to focus (116) on the user while blurring (118) the rest of the scene within the viewfinder (110) of the camera.

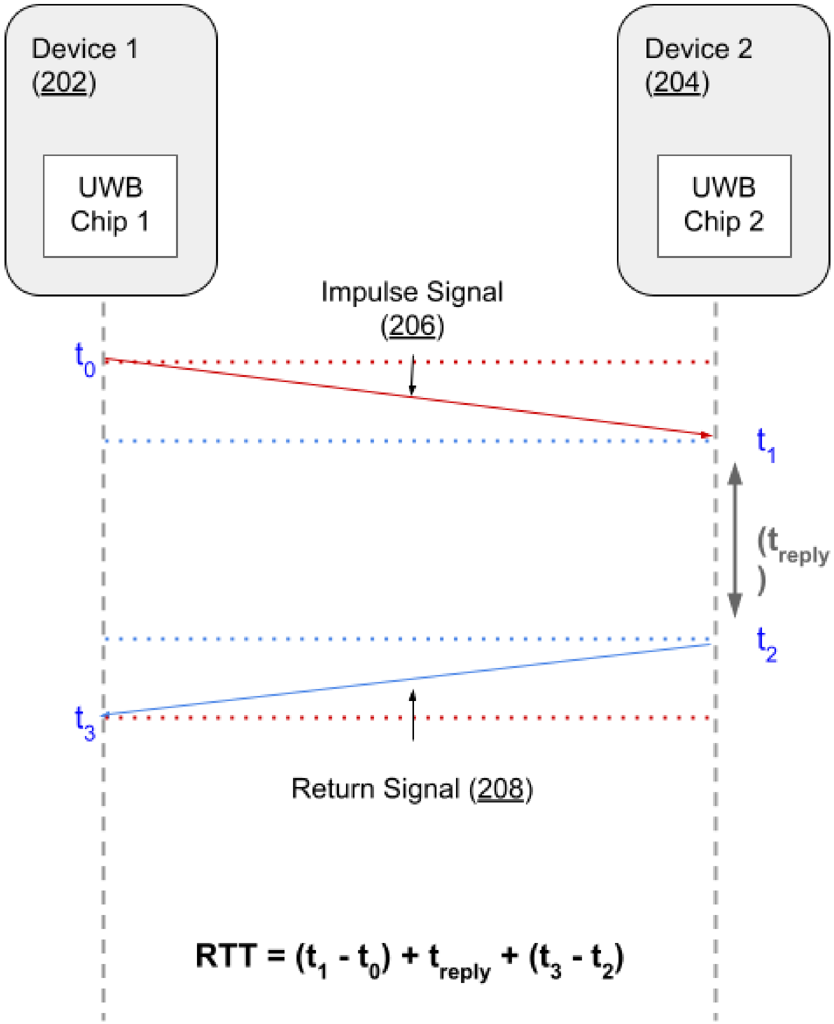

Fig. 2: UWB handshake protocol for estimating distance between two devices

The operation described above is realized via UWB chips in each of the devices to estimate the absolute distance via a handshake protocol based on the round-trip time (RTT) for a UWB impulse signal. As shown in Fig. 2, an impulse signal (206) transmitted from a UWB chip in a device (202) at time t_0 is received by the UWB chip in another device (204) at time t_1. The receiving device requires a fixed amount of time, t_{reply}, for generating a response signal (208), which is transmitted back at time $t_2 = t_1 + t_{reply}$ and received by the UWB chip in the first device at time t_3, where $RTT = t_3 - t_0$ and $(t_1 - t_0) = (t_3 - t_2)$. Therefore, the distance between the two chips, or equivalently, the two devices, can be computed as:

$d = 1/2 * c * (RTT - t_{reply})$, where c is the speed of light.

If the first device that originated the signal contains multiple antennas for receiving the reply signal, then the phases of the signal received at each of the antennas can be compared with each other. Application of beamforming principles can estimate the angle between the two devices based on phase deviations of the waveform at the different antennas of the first device.

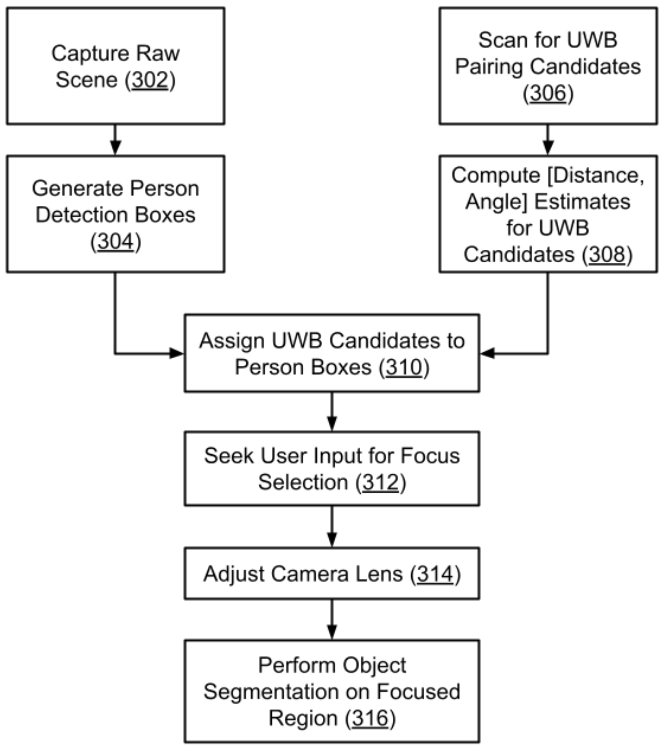

Fig. 3: Method for adjusting camera focus based on UWB depth estimates

Fig. 3 illustrates an example method to adjust camera focus based on depth estimates. Upon starting a video conference on a device, the space in the vicinity of the device is scanned with permission for UWB devices that can be candidates for pairing (306). The estimated distance and angle between the video-conferencing device and each discovered candidate device is computed via the handshake protocol described above and stored in temporary memory for further processing downstream (308).

A comparison between the stored distance and angle estimates with detected locations of human users (304) in the raw image scene captured with the camera (302) of the video conferencing device can be used to segment the scene by performing UWB-to-image target assignment (310). For instance, horizontal locations and (x, y) coordinates of person-detection boxes within the scene that are generated using off-the shelf engines, such as Region-based Convolutional Neural Network (RCNN), can be correlated with the distance and angle estimates to assign users to the locations within the scene. For example, as shown in Fig. 4, if two users (401 and 402) with distance and range measures of $[r_1, \theta_1]$ and $[r_2, \theta_2]$ respectively are present within the scene and $\theta_1 < \theta_2$ then the left person-detection box within the scene corresponds to the first user. A comparison between the distances can provide additional confidence for the assignment with, for instance, $r_1 >> r_2$ corresponding to $dimension_{leftbox} < dimension_{rightbox}$.

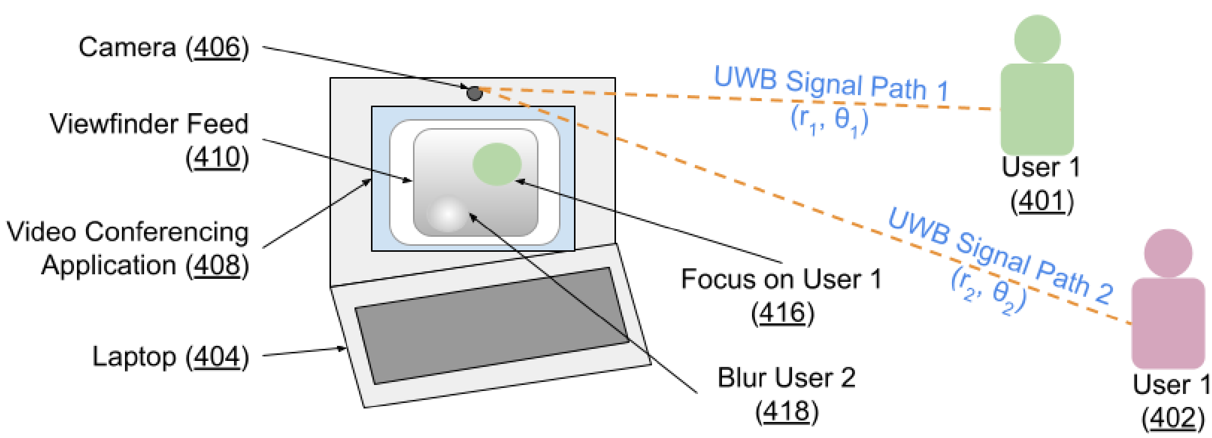

Fig. 4: Focusing on a selected user in scenes containing multiple users

When multiple persons are present within view of the device camera, users participating in the video conference can use the user interface (UI) of the video conferencing application to choose the person(s) of interest on whom the camera should focus (312). The user can indicate the correct person(s) by a suitable UI mechanism, such as touching the region within the camera viewfinder feed containing the person(s). For example, consider the case where Alice, Bob, and Charlie are in the same physical room participating actively in a video conference with remote parties. At some point during the conference, Bob and Charlie are done speaking and move into the background away from the camera while Alice continues to talk. Alice can simply touch the regions within the camera feed where Bob and Charlie are shown to indicate that the camera should be adjusted to blur Bob and Charlie as part of the background and with focus being on Alice.

Based on user input, the range information for the nearest UWB result is queried. For instance, if the user touches the camera viewfinder feed at coordinates (x, y) on the device screen, then the centroid (x_i, y_i) of the person-detection box nearest the user touch input can be determined via the optimization: $argmin_i \| (x, y) - (x_i, y_i) \| \_2^2$. The range information of the person-detection box nearest the user input can be used to control lens parameters and change the focal length accordingly to within the sub-meter accuracy provided by UWB (314). Subsequently, the image obtained with the adjusted focus can be processed via a suitable object segmentation engine to generate a higher-accuracy result (316). In cases where multiple users within the camera feed are located within the same depth of field, a digital blurring effect can be applied by using the UWB angle estimates.

With user permission, the techniques can be incorporated within any video conferencing devices that contain a UWB chip. The target devices for the UWB handshake described above can be any fine ranging (Fira) compatible device, such as a smartphone, earbuds, smartwatch, etc., that includes a UWB chip capable of establishing a back-and-forth UWB ranging channel with the video conferencing device. If users permit, the techniques can be implemented within any video conferencing platform or application. Implementation of the techniques can enhance the accuracy of focusing the camera on the desired users(s) and blurring the other objects, achieving effects akin to expensive LIDAR with commodity hardware at no added cost. As such, the techniques can make high-resolution portrait mode video conferencing affordable for users with commodity devices.

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 context, 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.

CONCLUSION

This disclosure describes techniques that use existing ultra-wideband (UWB) hardware on commodity devices to perform crisp and more accurate segmentation. A UWB handshake protocol is utilized to estimate the distance and angle between the camera and another device that the conference participant is wearing. The estimate is used to automatically adjust camera focal length to focus on the conference participant while blurring other objects. The techniques can make high-resolution portrait mode video conferencing affordable for users with commodity devices.