June 2020

# Interaction Quality Determination Using Device Camera

Herbert Jordan

Konrad Jan Miller

Agoston Weisz

# Interaction Quality Determination Using Device Camera

ABSTRACT

Software such as virtual assistants provide information to users, e.g., as ambient information, e.g., displayed on a device home screen, or in response to user queries. In such contexts, it is difficult to determine whether the user is satisfied with the displayed information or response to the query. This disclosure describes techniques that, with user permission, utilize on-device analysis of images or video obtained from a device camera, to determine user satisfaction.

KEYWORDS

- User feedback
- Gaze detection
- Virtual assistant
- User satisfaction
- Emotion recognition
- Eye tracking
- Smart display

## BACKGROUND

Software such as virtual assistants provide information to users, e.g., as ambient information, e.g., displayed on a device home screen, or in response to user queries. In such contexts, it is difficult to determine whether the user is satisfied with the displayed information or response to the query. Such software or service often includes an active feedback loop for the purposes of evaluation of quality, e.g., to determine whether a user interaction was successful or

not. This feedback can be gathered online (during or immediately after a user interaction) with real users.

For some software and online services, the success of an interaction can be gauged directly. For example, if a user requests the upload of a photo to a cloud service, success is determined by validating the file transfer. However, in certain cases, e.g., knowledge retrieval, the determination of success can be ambiguous. For instance, if a user asks for a piece of information and a response that includes text and a picture is shown on the screen, it is unclear as to whether the provided information is satisfactory to the user or not. In cases where there is no initial user interaction, e.g., the display of ambient information on a home screen of a device in a locked/ standby state, there is no initial query available to judge whether the displayed information is useful to the user or not.

In the case of a virtual assistant that responds to user queries, user interaction that follows after the query response is provided is an uncertain indicator of user satisfaction. For example, the user dismissing a response gives no clue as to whether the user was satisfied or not.

## DESCRIPTION

This disclosure describes techniques that, with user permission, utilize on-device analysis of images or video obtained from a device camera, to determine user satisfaction.

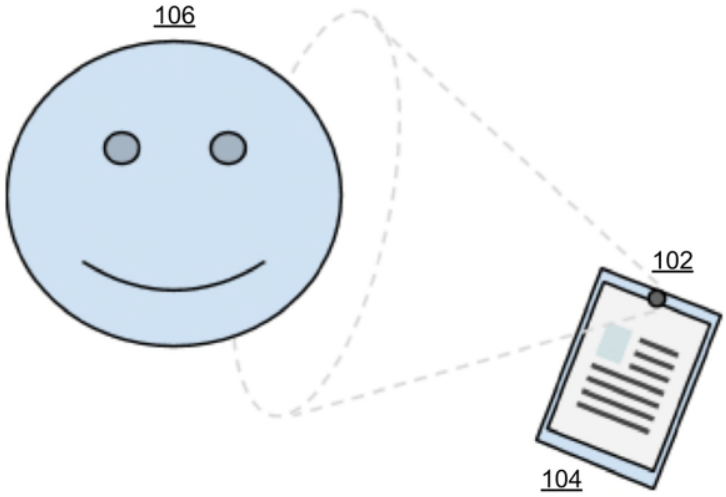


**Fig. 1: Interaction quality determination using device camera**

Fig. 1 illustrates determination of interaction quality using a device camera. Per the techniques, a camera (102) of a mobile device (104) is activated, with specific user permission, to obtain images/ video of the user (106) during or immediately after the user interaction. The obtained images or video are analyzed on-device to determine user satisfaction with the user interaction with a service provided via the device, e.g., interaction with a virtual assistant.

User permission is obtained to utilize images/video from a camera of the user device (e.g., a front-facing camera) that the user is interacting with and to analyze such content on-device to determine the user's satisfaction with the interaction. The user is provided with options to decline permission, in which case the camera is not utilized and no evaluation of the interaction is performed.

The device can be any device with a user-facing camera, e.g., mobile devices (smartphone, tablet, laptop, wearable device, etc.), consumer devices (smart displays, smart appliances, televisions, etc.) or any other type of device. The interaction of a user with the service is analyzed via on-device processing of the camera feed. In particular, with user

permission, one or more of the following determinations can be made to determine user satisfaction:

- The user's attentiveness and attendance, e.g., whether, when, how often, or the duration for which the user consumes information on the device.
- The user's area of focus on the screen, which indicates the section of the screen that is of importance to the user. If permitted, eye-tracking or gaze detection can be used to determine the user's area of focus on the screen.
- The user's reaction to the content on the screen or information provided via other modalities such as audio, e.g., happiness, frustration, etc.
- The number of users that are present in front of the screen.
- Whether the content is shared with other users, e.g., by passing the device to others, by sharing the content over messaging or social media, etc.

Analysis of the camera feed is performed on-device - the images/videos are not transmitted to any other device. The result of the evaluation (e.g., the user's satisfaction level) can be provided to the service or software that the user interacted with, e.g., a virtual assistant. With user permission, the result can be utilized as feedback to improve service quality.

*Example 1: Ambient information displayed on a home screen*

The start-up or home screen of devices or apps is often used to display information, e.g., time, weather, date, or other contextually relevant information, without an explicit user query. Per the techniques, the user viewing the displayed information can be detected, e.g., by eye-tracking and the user turning away from the device after reading can be classified as a successful interaction, supporting the provision of the information. If, however, eye-tracking indicates that the displayed data is never or hardly consumed, it can be considered unsatisfactory and may be

turned off from being displayed on the home screen. With user permission, information permitted by the user, e.g., stocks, breaking news, social media posts, etc., can be rotated in and out of the start-up screen based on user interest, gauged by techniques such as eye-tracking, user attentiveness, etc.

*Example 2: Interactions with a virtual assistant*

Virtual assistant software benefits from lightweight interaction, e.g., the user providing their query in as few steps as possible. If the number of possible queries is relatively few, it is easier to verify the success or failure of a virtual assistant response. For example, if the user types "li" and the virtual assistant displays light controls that the user proceeds to operate, it can be detected that the response of the virtual assistant was correct.

However, if there are a number of ways to complete a query, success is harder to determine. For example, if the user searches for "ca" on their smartphone some possible queries for the user include ***ca***lendar, the current position of their ***ca***r, ***ca***lling someone on their contact list, etc. and corresponding information can be provided on the screen. The user may then simply pause interaction, e.g., not select any of the displayed query-completions, or take a particular action, e.g., read where the next meeting is (from their calendar) and turn the device off. In the absence of further user interaction with the query-completion process (e.g., typing the full word "calendar"), it is difficult to gauge the success of the interaction. Per the techniques of this disclosure, eye-tracking can be utilized to determine the query of interest to the user: if, for example, the user's eyes focused on ***ca***lendar events displayed, the response of the virtual assistant can be deemed correct, even in the absence of further interaction.

*Example 3: Autocompletion of queries with answers*

In cases where the query is not only auto-completed but an available predicted answer is also provided, it is difficult to determine if the auto-completed query is the correct one. For example, if the user types "how old is the re" an example auto-complete response may be "how old is the redwood tree (***2500 years***)" that provides the auto-completion and the answer inline, with an option to view more information by selecting the response. If the user then does not select the auto-completed response, it can be an indication that the response was what the user was looking for, or that the user lost interest or was interrupted. Per the techniques, the camera feed can be analyzed to detect the user's expression to determine if the user was satisfied (or not) with the response. For example, a look of frustration may indicate that the user did not follow through with a tap on the auto-completed response because none of the provided suggestions were good enough. A look of satisfaction, a smile, an eye focus on the served answer, or substantial time spent on the served response can indicate that the user was happy, and hence didn't follow through with tapping on it or executing it.

*Example 4: Presenting information based on the number of users detected*

The interaction with a device can be interpreted based on the number of users detected in front of the screen. For example, in response to a request "let's play a game," user action to not start a single-player game can be interpreted as an unsatisfactory interaction when two or more users are detected and can be used to improve virtual assistant responses.

*Example 5: Content sharing activities*

If the user shares the information displayed on a screen, e.g., a cat video shared in response to a query "show my funny cat videos" with other users, e.g., by passing the device

around or by sharing the video via social media, the interaction may be deemed satisfactory, while if no sharing occurs, it may be an indicator of an unsatisfactory response.

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 queries, user's camera feed, sharing activities), 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. 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

Software such as virtual assistants provide information to users, e.g., as ambient information, e.g., displayed on a device home screen, or in response to user queries. In such contexts, it is difficult to determine whether the user is satisfied with the displayed information or response to the query. This disclosure describes techniques that, with user permission, utilize on-device analysis of images or video obtained from a device camera, to determine user satisfaction.