

Technical Disclosure Commons

Defensive Publications Series

January 2021

Invoking Visual Search With Device Camera Using Intuitive Physical Gestures

Brett Aladdin Barros

Ali Stanfield

Bo Brinkman

Nada Elassal

Paulo Coelho

Follow this and additional works at: https://www.tdcommons.org/dpubs_series

Recommended Citation

Barros, Brett Aladdin; Stanfield, Ali; Brinkman, Bo; Elassal, Nada; and Coelho, Paulo, "Invoking Visual Search With Device Camera Using Intuitive Physical Gestures", Technical Disclosure Commons, (January 11, 2021)

https://www.tdcommons.org/dpubs_series/3951



This work is licensed under a [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/).

This Article is brought to you for free and open access by Technical Disclosure Commons. It has been accepted for inclusion in Defensive Publications Series by an authorized administrator of Technical Disclosure Commons.

Invoking Visual Search With Device Camera Using Intuitive Physical Gestures

ABSTRACT

Invoking a visual search typically requires a user to open an application that supports the functionality before pointing the camera at the space or objects for which visual search is desired. Such an approach for invoking a visual search is time-consuming and cumbersome and limits the types of actions that can be supported by these capabilities. This disclosure describes techniques that enable users to invoke visual searches from their device cameras via intuitive physical gestures. Users can then take further actions on objects within view.

KEYWORDS

- Visual search
- Camera-based search
- Internet of Things (IoT)
- Smart home
- Flicking gesture
- Swipe, hold, and lift
- Sideways pointing

BACKGROUND

People can use the camera of their mobile device, e.g., smartphones, glasses, other wearable devices, etc. to conduct visual searches based on objects in view of the camera. However, invoking such visual searches typically requires users to open an application that supports search functionality before pointing the camera at the space or objects for which visual search is desired. Such a user experience (UX) is akin to needing to open an application explicitly prior to invoking a virtual assistant, instead of the seamless use of voice interaction.

Initiating visual search does not currently use user input via device sensors, such as microphone, accelerometer, camera, etc. Therefore, current user experience of invoking a visual search is time-consuming and cumbersome and limits the types of actions that are supported.

DESCRIPTION

This disclosure describes user interaction techniques that enable a user to invoke visual searches from their device cameras via various input mechanisms that do not require the user to run a specific application prior to the visual search. To that end, input from one or more device sensors, e.g., inertial measurement unit (IMU), is obtained with permission from the users and is analyzed to determine whether the input is indicative of a physical gesture that serves as an intuitive and seamless indicator of the user's desire to initiate a visual search.

For instance, the user can invoke visual search with one or more of the following physical actions:

- Flicking the device such that the end of the flicking action results in the device camera being pointed at the target(s) of the visual search.
- Swiping, holding, and lifting a finger as a single combined gesture on the device screen while the camera view shows the target(s) of the visual search.
- Sideways pointing of the device (akin to a handshake) with the device cameras pointing to the user's left and/or right. With the user's permission, a pre-scan of the user's surroundings (room) can be performed to enable the device to calculate the current position and the direction of pointing.

With user permission, the above physical gestures can be detected by processing input from one or more device sensors with an appropriate training machine learning model or other suitable technique. A holding gesture performed on objects within camera view can enable query

refinement with relevant visual feedback. Alternatively, or in addition, the user can perform relevant actions on the visual search targets. For instance, the user can invoke visual search by pointing at a nearby lamp in order to turn the lamp on or off.

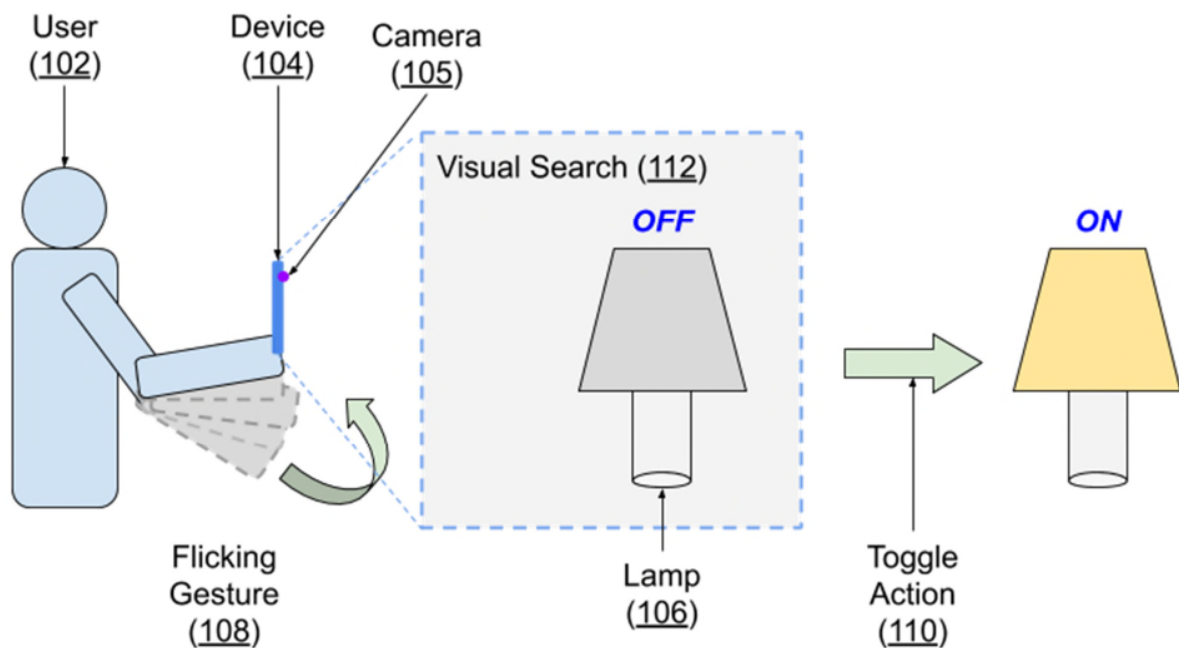


Fig. 1: Flicking gesture to invoke visual search to take an action on an object

Fig. 1 shows an example of operational implementation of the techniques described in this disclosure. As seen in Fig. 1, a user (102) performs a flicking gesture (108) by flicking a device (104) upward such that the device camera (105) points at a lamp (106) near the user. The flicking action, detected with the user's permission, invokes a visual search (112) based on the camera view. The user can then perform a toggle action on the lamp within the visual search field to switch it on from the current turned off state. Alternatively, or in addition, a user interface that enables changing properties such as lamp color, lamp intensity, etc. can be displayed on the device in response to detecting the lamp within the camera field of view. The

detection of the user's physical actions for initiating a visual search is performed locally on the device.

Implementations of the described techniques can significantly enhance the UX of invoking visual search capabilities. Gesture-based invocation provides users an efficient alternative for taking various actions related to physical objects, such as Internet of Things (IoT) objects, smart home devices, etc. Gesture-based initiation of actions can be provided within any device, application, or platform that supports visual search capabilities. Further, the techniques can support actions pertaining to any networked physical objects with appropriate mechanisms, such as an application programming interface (API) that enables sending commands to the object.

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 devices, 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 enable users to invoke visual searches from their device cameras via intuitive physical gestures. Data from device sensors are obtained and analyzed with user permission to detect a physical gesture that serves as an intuitive and seamless indicator of the user's desire to initiate a visual search. Implementations of the described techniques can significantly enhance the UX of invoking visual search capabilities and provide users an efficient alternative for taking various actions related to physical objects.