August 19, 2019

Utilizing Eye-Tracking to Develop Contextual Suggestions

Jeff Palm

Utilizing Eye-Tracking to Develop Contextual Suggestions

Abstract:

This publication describes techniques aimed at developing improved contextual suggestions for predictive text services (*e.g.*, query suggestion, autosuggest, autocomplete) on computing devices utilizing a weighted, multi-item clipboard function. The techniques incorporate the utilization of machine-learned modules that cooperatively analyze user-attention to on-screen entities, such as images or text, and develop corresponding (weighted) contextual suggestions.

Keywords:

Gaze tracking, gaze detection, gaze location, eye tracking, eye movement, pupils, machine-learning, computing device, smartphone, x-y coordinate, weighted clipboard, multi-item clipboard, predictive text service, predictive keyboard, autosuggest, autocomplete, query suggestion

Background:

Computing devices, such as smartphones, computers, and tablets, often allocate a reserved portion of memory (a "clipboard" function) that temporarily stores data so users can copy and paste on-screen entities, such as text and images. Generally, a clipboard function is limited to providing paste options based solely on user-copied items.

For example, a user (James) is typing an academic paper on his computer. In his paper, James desires to quote a paragraph from an image of typed text displayed on the screen of his computer. To do so, James must type the paragraph into his paper while gazing back and forth between the image and his paper. In another example, James is viewing images of automobiles on his smartphone. While looking at the images, James focuses his attention on an image of a Ferrari. Immediately, James opens a new internet browser window and types the letters “Ferrari.”

Instances like these, where a user types text into a computing device as a result of viewing on-screen entities, are frequent. If the clipboard or even predictive text services (*e.g.*, autosuggest, autocomplete) had smart features such that they could determine whether James gave certain recently viewed on-screen entities more attention than others, then the computing device can offer James more-relevant contextual suggestions.

Description:

This publication describes techniques aimed at developing improved contextual suggestions for predictive text services (*e.g.*, autosuggest, autocomplete) through use of a weighted, multi-item clipboard function (*e.g.*, a reserved portion of memory that temporarily stores data).

Figure 1 illustrates an example computing device and elements of the computing device that support the techniques described in this publication.

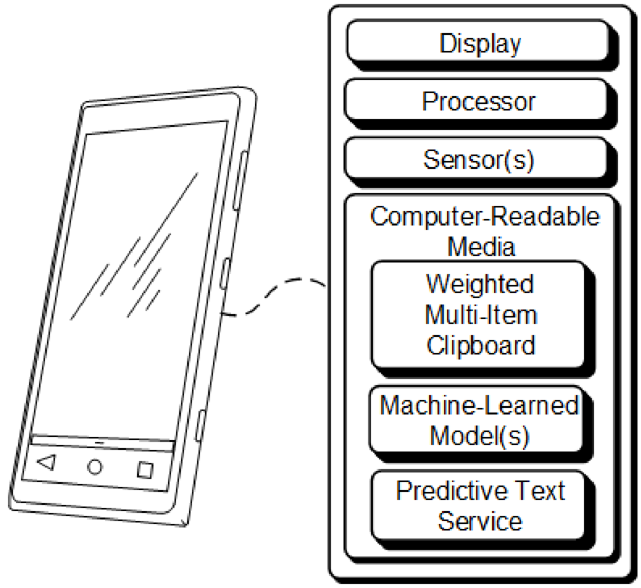

Figure 1

As illustrated, the computing device is a smartphone. However, other computing devices (*e.g.*, computers, tablets, watches) can also support the techniques described herein. The computing device includes a display and a sensor. The sensor (*e.g.*, a front-facing camera) can measure visual input relating to the user. For example, the size of the pupil of the eye, the location of the pupil of the eye, the pitch of the head, the tilt of the head, the yaw of the head, the motion of any eye relative to the head, and the distance from the user to the display. The sensor can measure visual input while the user is interacting with on-screen entities, such as text and images, or navigating through applications on the computing device.

The computing device also includes a processor and a computer-readable medium (CRM). The CRM includes a weighted, multi-item clipboard function, machine-learned modules, and a predictive text service.

The machine-learned modules (*e.g.*, a gaze detection module, a tracking module, a conversion module) are iteratively trained, off-device, by exposure to their respective training scenes, sequences, and/or events. The gaze detection module is trained on labeled images of eyes, the distance of the eyes to the display, and the user's respective gaze direction. The gaze direction module is for providing suggestions relating to the gaze of a user. The tracking module keeps track of the attention of the user relative to portions of the display for providing suggestions regarding the time of user-attention to specific sections of the display. The conversion module is trained to convert regions of a display screen into clipboard items, including copying plain text, translating image text into text, and classifying images as semantic objects. The machine-learned modules are deployed to the CRM after sufficient training.

Once deployed to the CRM, the machine-learned modules can cooperatively operate as follows. Utilizing the sensor, the gaze detection module can determine the user's gaze direction and on-screen entities (*e.g.*, images, text) at which the user is gazing. The tracking module can determine user-attention by estimating the duration of the user gaze on the on-screen entities and the user interaction with the on-screen entities (*e.g.*, zooming in on an on-screen entity, selecting an on-screen entity, highlighting an on-screen entity). The tracking module can, additionally, assign timestamps to the on-screen entities for which user-attention was measured.

Finally, the conversion module can convert the on-screen entities into clipboard objects. To perform the conversion, the conversion module can copy text, translate an image text into text, and/or classify images as semantic objects. Once the conversion is complete, the clipboard objects, along with the timestamps, can be sent to the predictive text service. The predictive text service can apply weights to clipboard objects based on the level of user-attention and the recency of the user-attention. Multiple such clipboard objects can be weighted and provided to the clipboard function to create a weighted, multi-item clipboard. The predictive text service can then provide suggestions (*e.g.*, query suggestions, autosuggest, autocomplete) to the user based on multiple clipboard items on the weighted multi-input clipboard. For example, while typing, the user can access a contextual suggestion from the weighted, multi-item clipboard through either an application-specific menu or a right-click menu. Alternatively or additionally, the predictive text service can offer the contextual suggestion.

An illustration of these machine-learned modules operating cooperatively to implement a weighted, multi-item clipboard is illustrated in Figure 2.

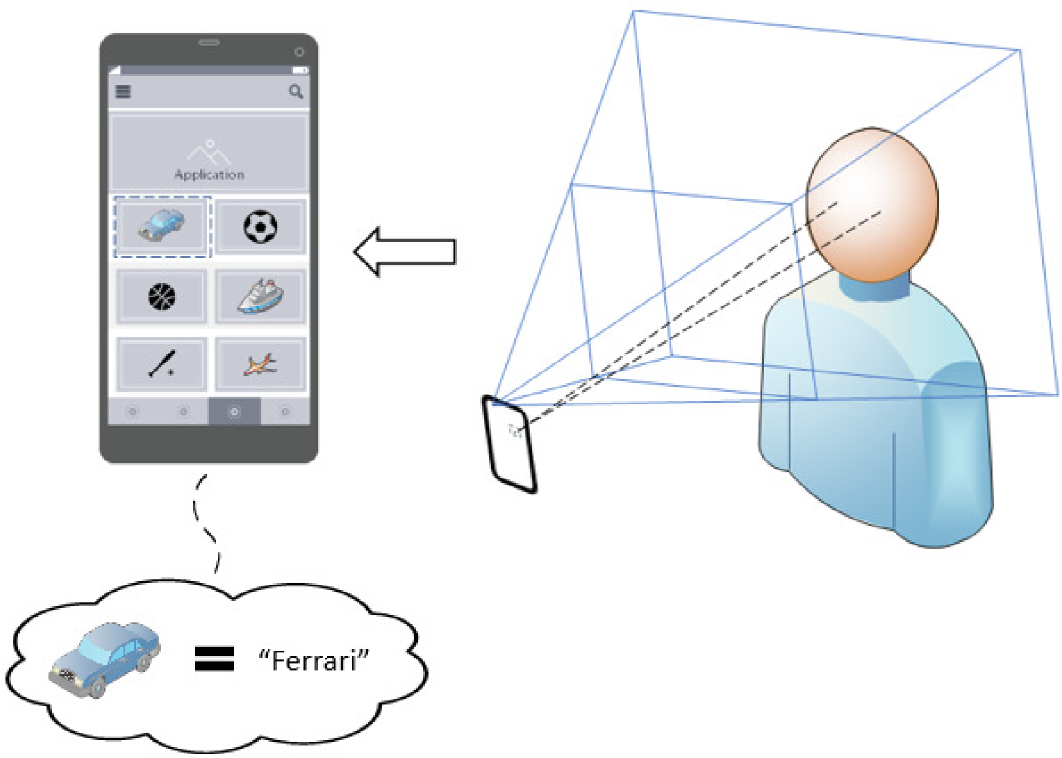

Figure 2

As illustrated, a user (James) focuses his gaze on an image of a Ferrari on the display of a smartphone. James also briefly gazes at an image of a baseball on the display. The gaze detection module, by means of the sensor, detects the on-screen location of James' gazes and the related on-screen entities. After which, the tracking module estimates James' attention to the on-screen entities. Lastly, the conversion module, in consideration of James' attention, converts the images to clipboard objects (*e.g.*, words), and assigns timestamps to the clipboard objects. These clipboard objects are then sent to the predictive text service for weighting. After this process, if James opens a new Internet browser, the top contextual suggestion can be the word “Ferrari,” and a second contextual suggestion can be “baseball.”

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, social activities, profession, 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.

In conclusion, the techniques described herein can develop improved contextual suggestions for predictive text services utilizing a weighted, multi-item clipboard function.

References:

[1] Published Application: US20120094700A1. Expectation assisted text messaging. Filing Date: August 11, 2011.

[2] Published Application: US20140215372A1. Methods, systems and computer readable media for pasting data. Filing Date: April 1, 2014.

[3] Kishi, Reo & Hayashi, Takahiro. Development of an Effective Gaze Writing System with an Interface for Text Copy and Paste. International Journal of Networked and Distributed Computing. 3. 193. 10.2991/ijndc.2015.3.3.7. August 2015.