Int. J. Human-Computer Studies 73 (2015) 91–106

Contents lists available at ScienceDirect

Int. J. Human-Computer Studies

journal homepage: www.elsevier.com/locate/ijhcs

Gaze-based prediction of pen-based virtual interaction tasks ☆

Çağla Çığ *, Tevfik Metin Sezgin

Intelligent User Interfaces Lab, Department of Computer Engineering, Koç University, Istanbul 34450, Turkey

ARTICLE INFO

Article history:
Received 12 June 2014
Received in revised form
19 September 2014
Accepted 20 September 2014
Available online 28 September 2014

Keywords:
Sketch-based interaction
Multimodal interaction
Predictive interfaces
Gaze-based interfaces
Feature selection
Feature representation
Multimodal databases

ABSTRACT

In typical human–computer interaction, users convey their intentions through traditional input devices (e.g. keyboards, mice, joysticks) coupled with standard graphical user interface elements. Recently, pen-based interaction has emerged as a more intuitive alternative to these traditional means. However, existing pen-based systems are limited by the fact that they rely heavily on auxiliary mode switching mechanisms during interaction (e.g. hard or soft modifier keys, buttons, menus). In this paper, we describe how eye gaze movements that naturally occur during pen-based interaction can be used to reduce dependency on explicit mode selection mechanisms in pen-based systems. In particular, we show that a range of virtual manipulation commands, that would otherwise require auxiliary mode switching elements, can be issued with an 88% success rate with the aid of users' natural eye gaze behavior during pen-only interaction.

1. Introduction

Pen-based devices are gaining popularity. Pen-enabled smart phones and tablet computers have penetrated our lives to a great extent due to their mobility, ease of use and affordable prices. However, despite what their name suggests, pen-based devices are not purely pen-based (Plimmer, 2008).

For example, in pen-enabled smart phones, many actions force the user to put the pen aside and switch to multi-finger gestures (e.g. spread/pinch for zoom in/out, and swipe to navigate back/forward). These gestures require the simultaneous use of 2, 3 or even 4 fingers (Fig. 1a). The necessity of switching between pen and multi-touch input goes against the goal of seamless interaction in pen-based devices.

Even the state of the art devices and software specifically built for pen-based interaction lack purely pen-based interaction. For example, graphics tablets preferred mainly by digital artists such as Wacom Cintiq 24HD (Fig. 1b) are often referred to as “heaven on earth” by users. However, even with these high-end models many tasks are still accomplished via on-pen or on-tablet external buttons called “express keys”, “touch rings” and “radial menus”. These buttons allow the user to simulate keystrokes including letters, numbers and modifier keys (e.g. Shift, Alt and Control). To issue a virtual manipulation command (e.g. scroll), the user has to locate the correct button which interrupts the interaction flow, hence causing an overall disappointing experience.

Another example where we lose purely pen-based interaction is with tablet computers. In most pen-based applications, features are hidden in standard context/pop-up menus that are accessed via tapping and/or holding the pen on the tablet screen in various ways (Fig. 1c). In this case, the pen is used to trigger mouse clicks, which fits the traditional GUI/WIMP-based interaction paradigm, rather than that of a purely pen-based interaction (Fig. 1d).

These issues show that existing pen-based systems depend substantially on multi-finger gestures, context/pop-up menus and external buttons which goes against the philosophy of pen-based interfaces as a more intuitive interaction alternative. In this paper, we show that eye gaze movements that naturally accompany pen-based user interaction can be used to infer a user's task-related intentions and goals. The non-intrusive and transparent use of eye gaze information for task prediction brings us closer to the goal of purely pen-based interaction and reduces the reliance on multi-finger gestures, context/pop-up menus and external buttons.

Our approach consists of tracking eye gaze movements of the user during pen-based interaction, fusing the spatio-temporal information collected via gaze and sketch modalities in order to predict the current intention of the user. We use the term “intention” to refer specifically to the intention of the user to issue a virtual manipulation command. Virtual manipulation commands that we can currently predict are: *drag*, *maximize*, *minimize* and *scroll*. Additionally, we can distinguish whether the user intends to issue any of these virtual manipulation commands, or intends to sketch using our special-purpose task class called *free-form drawing*.

☆This paper has been recommended for acceptance by J. LaViola.

* Corresponding author. Tel.: +90 212 338 1540; fax: +90 212 338 1548.
E-mail addresses: ccig@ku.edu.tr (Ç. Çığ), mtsezgin@ku.edu.tr (T. Metin Sezgin).

http://dx.doi.org/10.1016/j.ijhcs.2014.09.005
1071-5819/

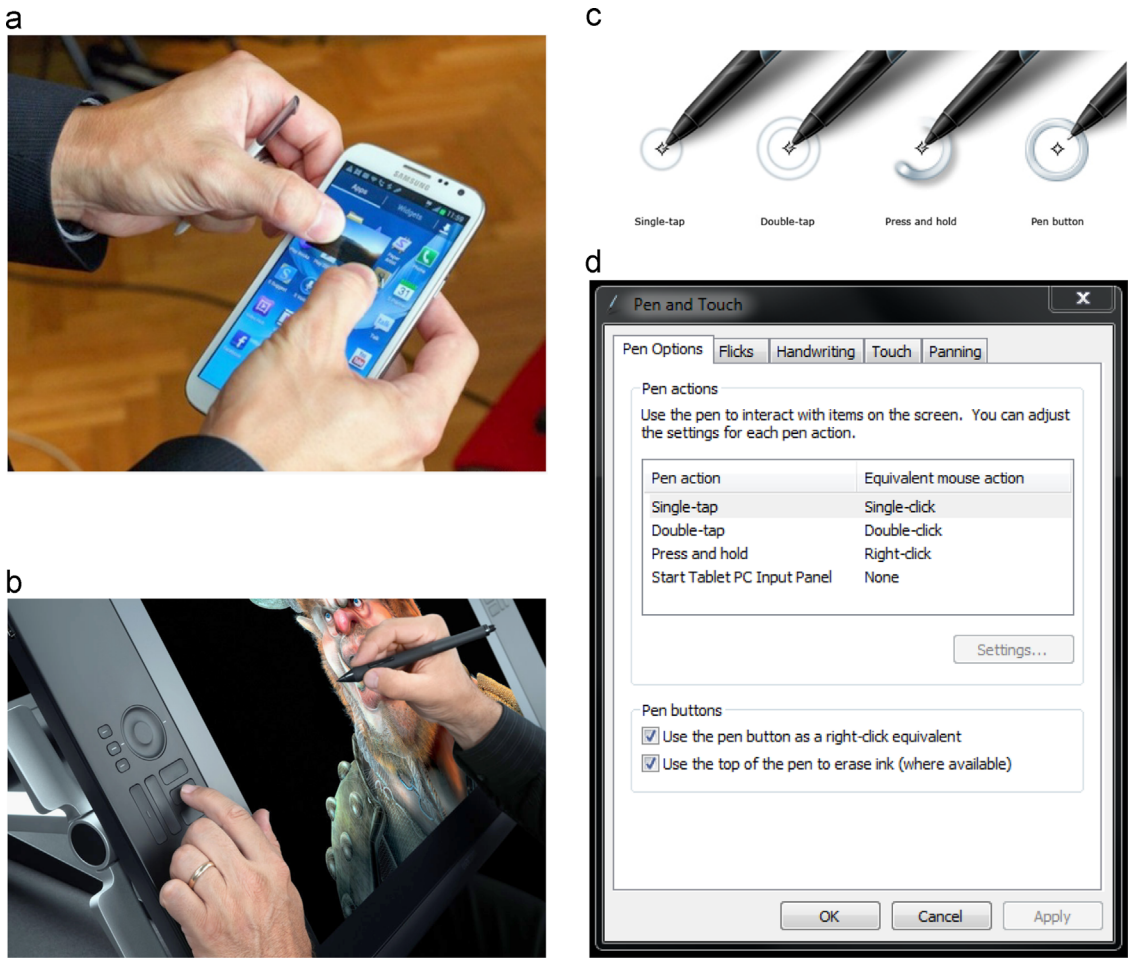

Fig. 1. Some examples to pen-based devices with interaction paradigms that are not purely pen-based. Gaze-based prediction of virtual interaction tasks in pen-based interaction is a step towards systems that require fewer mode changes (Li et al., 2005). (a) Switching between pen and multi-touch input for object manipulation (e.g. image resizing) in pen-based smart phones. (b) On-pen or on-tablet external buttons in pen-based graphics tablets. (c) Various tapping and/or holding techniques to access context/pop-up menus in pen-based tablet computers. (d) The pen is used to emulate a mouse in pen-based tablet computers.

Our overall approach to gaze-based prediction of virtual interaction tasks is depicted in Fig. 2. The left part of the diagram shows how we build our system whereas the right part shows how our system performs predictions. After we build our system, it can be used to infer user's task-related intentions and goals in an event-driven manner where each pen marking triggers prediction.

Briefly, our system is built as follows: Initially we collect sketch and gaze data during a number of pen-based interaction tasks and build a multimodal database. We then extract novel gaze-based features from this database and train a task prediction model using supervised machine learning techniques. These steps are executed only once. Then, our system is ready for prediction. When the user performs a pen action (demarcated by a pen-down and a pen-up event), the synchronized pen trajectory and eye gaze information is used to predict the user's intended virtual task. Predictions are carried out by the previously trained model and the features extracted from the corresponding sketch–gaze data of the user. Detailed description and discussion of our approach can be found in the following sections.

We have three main contributions. First, we present a carefully compiled multimodal dataset that consists of eye gaze and pen input collected from participants completing various virtual interaction tasks. Second, for predicting user intention through gaze, we propose a novel gaze-based feature representation based on human vision, and behavioral studies. Third, we introduce a novel gaze-based task prediction system that uses this feature representation. These features are neither subject- nor interface-specific, and perform better than commonly utilized and well-established sketch recognition feature representations in the literature. We evaluate our system based on several aspects, including the prediction accuracy and scale-invariance. In addition, we run feature selection tests to evaluate the relevance and redundancy of the feature representations. Our prediction system opens the way for more natural user interface paradigms where the role of the computer in supporting interaction is to "interpret user actions and [do] what it deems appropriate" (Nielsen, 1993). It is widely accepted that intelligent mode selection mechanisms that provide low cost access to different interface operations will dominate new user interface paradigms (Negulescu et al., 2010).

Section 2 gives an outline of the state-of-the-art gaze-based interfaces in a categorical manner with relevant examples for each category. Our approach consists of three major parts: data collection, feature extraction and intention prediction. These parts are detailed in Sections 3–5, respectively. Section 6 concludes with a discussion of our work and a summary of future directions.

2. Related work

We have presented a novel gaze-based interface for predicting virtual manipulation commands during pen-based interaction. State-of-the-art gaze-based interfaces fall under two main categories: *command interfaces* and *non-command interfaces*.