Entertainment Computing 5 (2014) 147–156

Contents lists available at ScienceDirect

Entertainment Computing

journal homepage: ees.elsevier.com/entcom

Head movement and facial expressions as game input ☆

Mirja Ilves *, Yulia Gizatdinova, Veikko Surakka, Esko Vankka

Research Group for Emotions, Sociality, and Computing, Tampere Unit for Computer–Human Interaction (TAUCHI), School of Information Sciences, University of Tampere, Kanslerinrinne 1, FIN-33014 Tampere, Finland

ARTICLE INFO

Article history:
Received 9 September 2013
Revised 10 February 2014
Accepted 28 April 2014
Available online 6 May 2014

Keywords:
Camera-based video game
Computer vision
Face detection and tracking
Facial expression classification
Gameplay experience
Emotion

ABSTRACT

This study aimed to develop and test a hands-free video game that utilizes information on the player's real-time face position and facial expressions as intrinsic elements of a gameplay. Special focus was given to investigating the user's subjective experiences in utilizing computer vision input in the game interaction. The player's goal was to steer a drunken character home as quickly as possible by moving their head. Additionally, the player could influence the behavior of game characters by using the facial expressions of frowning and smiling. The participants played the game with computer vision and a conventional joystick and rated the functionality of the control methods and their emotional and game experiences. The results showed that although the functionality of the joystick steering was rated higher than that of the computer vision method, the use of head movements and facial expressions enhanced the experiences of game playing in many ways. The participants rated playing with the computer vision technique as more entertaining, interesting, challenging, immersive, and arousing than doing so with a joystick. The results suggested that a high level of experienced arousal in the case of computer vision-based interaction may be a key factor for better experiences of game playing.

1. Introduction

Enjoyment and other emotion-related factors are central motivators for playing video games. People seek and play video games that are fun and entertaining or elicit other kinds of emotional experiences. Games with different characteristics elicit various emotional responses (e.g., [1]), but control devices for playing can also affect a player's emotions and game experience. Recent evidence shows that new, handheld but more natural controllers (i.e., Wii remote, steering wheel) can lead to higher feelings of spatial presence and game enjoyment than traditional control devices (i.e., joystick, gamepad, keyboard) [2]. However, systems that utilize physical controls have inherent limitations of being unable to detect the presence and identity of players, for example.

More natural, active, and playful gaming has become possible because of advances in computer vision (CV). Through standard video cameras, CV technology provides a low-cost alternative to handheld devices and allows entirely unobtrusive detection of head and body movements or hand gestures, for example, and their use as a game input. Automatic analysis of facial information can help to understand the identities and number of players, as well as their presence and locations. The recognition of facial expressions is also possible with the help of CV. The human face and facial expressions provide a rich source of information about human behavior and emotional state. It can be argued that faces are the main modality in human nonverbal communication, and many expressions can be performed voluntarily; thus, the use of facial expressions could provide a natural method of game control. However, in the past, research in the area of automatic face analysis had focused on the technological aspect, dealing with performance characteristics of different methods such as their speed, accuracy, and robustness [3,4]. Generally, the question of how video games can successfully leverage facial information remains less understood. The literature analysis reveals that although automatic face analysis has started to be utilized in gaming, few studies have attempted to systematically evaluate the usability and user experience aspects of face-responsive games.

This study aimed to investigate the functional and experiential aspects of head movements and facial expressions detected by CV as game input methods. In contrast to earlier studies, we combined active and continuous face tracking with facial expression classification in real time in order to enhance the overall experiences of a gameplay. For this purpose, we designed and implemented the game "Take Drunkard Home," where a player's goal is to steer by head movements a drunken character home as quickly as possible without the latter falling down. Along the way, the player needs to pick up various items and avoid collisions with stationary or

☆ This paper has been recommended for acceptance by Rainer Malaka.

* Corresponding author. Tel.: +358 50 318 5848.
E-mail addresses: mirja.ilves@sis.uta.fi (M. Ilves), yulia.gizatdinova@uta.fi (Y. Gizatdinova), veikko.surakka@uta.fi (V. Surakka), esko.vankka@uta.fi (E. Vankka).

http://dx.doi.org/10.1016/j.entcom.2014.04.005
1875-9521/

moving obstacles. Additionally, the player can influence the behavior of the other game characters by using facial expressions with two affective meanings—positive for smiling and negative for frowning. The game solely relies on automatic face processing and therefore supports better accessibility to video gaming for those players who have difficulties or simply do not want to use physical input devices. We conducted a user study in which the participants played the game using CV technology and a conventional joystick. We recorded the game duration; the number of falls; and the number of picked flowers, beer cans, and hamburgers when the participants played the game. The participants evaluated the functionality of the control methods, as well as their emotional and game experiences.

This paper reviews recent advances in face-responsive video gaming and introduces the game design and CV-based control methods used in this study. Then it presents the results of the empirical evaluation of the game and further discusses issues related to the future development of video games that utilize facial information.

2. Background

Quite recently, the game industry has added new input devices and techniques to traditional controls such as the keyboard, mouse, joystick, and gamepad. The controlling of games is not limited anymore to the use of the hands; new input devices and techniques allow more natural and more playful, physically active gaming. Thus, the motion-sensing capability of Nintendo's Wii[1] remote enables the detection of acceleration and movement in three dimensions. There are also technologies that enable game control without any handheld device. For example, Microsoft Kinect[2] enables players to control the Xbox 360 using body movements and gestures without touching the control device. The Microsoft Kinect sensor consists of an infrared laser projector, combined with two cameras that detect the positions and movements of people in three dimensions. Some game technologies, such as Sony EyeToy[3] in Sony PlayStation2, use CV for the gestural control of games. Because of advances in CV and hardware processing capabilities, movement- and gesture-based interaction has also become possible in conventional computer systems. The CV technology has been improving constantly, and the field has recently demonstrated a number of face analysis methods that have proven to function well even in challenging conditions [3,5,6]. Considering that real-time, accurate, and robust measurement of facial visual information is only a matter of time, it is important to understand how this information can be used in the context of game interaction. At this point, we emphasize the need for early user studies as an integral part of the development of CV-based user interfaces and their successful integration into video games.

From the implementation standpoint, controlling video games or any other graphical user interface is typically based on pointing and activation methods. The pointing method identifies the object of interest, and the activation method allows implying a certain action in a virtual game world. From the design standpoint, video games can be controlled in two ways, implicit and explicit. Implicit control allows for automatic adaptation or adjustment of the game environment and interaction modalities to the player's spontaneous behavior. Explicit control means that the player consciously produces facial expressions, head movements, and body gestures to directly control the game interaction. This type of gaming replaces traditional gaming with physical input devices that are primarily based on the point-and-click concept. Furthermore, the use of faces for game control (implicit or explicit) has two important advantages. First, the human face is highly expressive, with more than 40 muscles that alone or in combinations produce visually detectable changes in facial appearance. Therefore, facial expressions, together with head movements, can potentially provide a diverse, intuitive, and fine-grained means of game control. Second, growing evidence indicates that the use of physical movements of the head and face can enhance the overall experience of game playing.

The proposed video games to date can be roughly divided into three categories, according to the inner mechanisms of how facial information has been utilized for game interaction:

(1) Digital interactive mirror. The avatar explicitly repeats the player's head movements and facial expressions (the top left image of Fig. 1). Real-time animation of sometimes impressively photorealistic avatars has become an increasingly popular area of research (e.g., [7]), partly due to its potential utilization in the film industry. Additional graphical add-ons such as makeup, various head-wearing objects, or emoticons can be drawn on top of the player's face to enhance the experience of presence and role-playing [8].
(2) Viewpoint, directional navigation, and action trigger control. Information on face position and head orientation is directly used to change the player's point of view in first-person games or to steer the avatar in the game environment. Additionally, the detected head gestures and facial expressions are used to imply a certain action in the game world. Conventionally, the navigation of the avatar in a two- or three-dimensional world has been performed with handheld devices such as the keyboard, mouse, joystick, and gamepad. It has been shown that head rotation and movement can substitute for the use of physical devices in navigational tasks and provide a more natural and intuitive means of game control [8,9]. The top right image in Fig. 1 shows the "eating game" [10], which implies the idea of transferring the head movement sideways to a horizontal motion of the "eater" character that is located at the bottom of the game space. The player controls the character's mouth-opening movement by opening his or her own mouth. The bottom image in Fig. 1 shows a top-down, strategy-like "walking game" [10], where a circular movement template is used to move the character from one cell to another in a labyrinth, by means of head gestures. The player can also produce facial expressions to pick up different items.
(3) Affective control. Facial expressions are naturally utilized to bring affective information to the gameplay and, depending on the game design, implicitly or explicitly execute emotion-related or emotion-guided activities. This idea closely relates to the concept of affective gaming, meaning that the player's emotional state influences the game's difficulty level or aesthetics, for example [11]. Previously, in order to assess the user's affective state, information on the player's heart rate, skin conductance, and respiration had been detected and further utilized in manipulating the gameplay [12]. In CV-based games, the player's spontaneous facial expressions and body gestures are detected fully unobtrusively without the attached physiological sensors and used to adapt the game to the supposed affective state of the player.

Our research focus belongs to the last two categories on augmenting video games with information on the player's real-time head position and facial expressions. We concentrate our literature search on those studies involving empirical verification and explain how the proposed game designs influence user experiences.

[1] http://www.nintendo.com/wii.
[2] http://www.microsoft.com/en-us/kinectforwindows/.
[3] http://us.playstation.com/ps2/accessories/eyetoy-usb-camera-ps2.html.