Monitoring of Facial Characteristic Movement for the Control of Mobile Devices: Issues and Recommendations

Andreas Komninos, Peter Barrie

Mobile and Ubiquitous Computing Research Group
Dept. of Computer Science
Glasgow Caledonian University
70 Cowcaddens Road
Glasgow G4 0BJ
{andreas.komninos, peter.barrie}@gcal.ac.uk

Abstract. This paper discusses the possibility of monitoring the movement of facial characteristics such as the eye and nose, for the control of mobile devices. The areas of concern behind this research are highlighted and the paper describes an experiment, during which observations relating to the movement of the user's facial characteristics were made. The paper concludes with suggestions on how these observations may be exploited.

1. Introduction

Eye tracking has been the focus of study by many researchers, not only as a method of observing user interaction and behaviour with systems or other activities, but also as a method of computer control. This area of application is especially useful to disabled users who suffer from limited dexterity or do not have use of their hands, as well as in situations where the system operator's hands might be tied up elsewhere, such as in the case of a fighter pilot.

Computer control through eye tracking or web-cams has been available to disabled users and computer game players for many years, helping to enhance the usability of computers in a variety of situations. With the advent of 3G mobile phones that incorporate user-facing cameras, we began to wonder whether the tracking of user head movements or facial feature movements might open up new possibilities in creating more natural human computer interfaces in the future.

2. Related work

Controlling computers through the user's facial characteristics or gestures is not a novel idea. Several researchers have focussed on this area with profound success. A variety of techniques, such as eye tracking, infrared hotspot tracking, blink detection etc have been employed with considerable success and some are being used in commercial products readily available today. Most eye-tracking or feature tracking implementations can be classified into three categories: template based methods, appearance based methods and feature based methods.

In the template based methods[1-4], a generic eye model, based on the eye shape, is designed first. Template matching is then used to search the image for the eyes. The appearance based methods [5-7] detect eyes based on their photometric appearance. These methods usually work on the basis of a large amount of training data, representing the eyes of different subjects. Several parameters such as different face orientations, and different illumination conditions are considered in the collection of suitable data. These data are then used to train a classifier such as a neural network. Feature based methods [8-12] explore the characteristics (such as edge and intensity of iris, the color distributions of the sclera and the flesh) of the eyes to identify some distinctive features around the eyes.

Most of the techniques are a variation on the same theme: identify an area of interest in an image, such as the iris, the eyelid or an infrared marker, and measure its movement through a sequence of captured images. In such techniques, it is assumed that the image capture device and the subject remain still throughout the capture and analysis process. In projects where cameras are attached to glasses on the user's head, the cameras also remain relatively fixed in relation to the user's eyes.

3. Issues in facial feature tracking for mobile devices

Upon examining the existing techniques, we identified several problems which might restrict their applicability to mobile devices. Firstly, since mobile device position relies on the user's grasp, they are inherently unstable. Even when the user is static, the hand and arm cannot provide sufficient stability that will ensure the user can indeed be considered to be still, in relation to the camera. This problem might be particularly aggravated by a list of physiological factors and conditions such as age.

The second problem is the poor resolution (VGA, at most) in front facing cameras on 3G devices. This means that accurately tracking small facial features such as the iris is very difficult. This problem is aggravated by the distance at which the device is naturally held. Further to this, given the small screens featured in today's mobile devices, considerably less movement is required for a facial characteristic such as the eye in order to point towards a particular area on the screen.

Adding to these problems, another factor was the actual position at which a mobile device is held. This position varies according to the user's posture (standing, sitting, slouching), vision problems and lighting conditions. It can be generally observed that the user will typically look down on the device, thus the iris would be partially obscured by the lower eyelid. The upper eyelid will also be lowered, with the eyelashes further reducing the visible area of the eye and further shading the visible area considerably.

The final problem we anticipated was the effect of the surrounding environment to the visibility of the user's face. Mobile devices are used both indoors and outdoors, if the dominant light source in the environment, whether that is the sun or a ceiling light, is behind the user's face, this will result in a very dark image of the user. In such circumstances, it might be impossible to distinguish and differentiate between facial characteristics and detection of a passive light reflective feature (the face) might not be a viable option. In such cases, an active light emitting source such as an LED would be the only solution.

4. Experiment description

To investigate the extent of these problems, we decided to conduct an experiment based on our predicted use of facial tracking on devices such as mobile phones. Our assumption is that such devices might benefit from the following modes of input (figure 1): a) Five degree of freedom cursor control (up, down, left, right, click) and b) Accurate translation of facial feature movement into screen coordinates plus click.

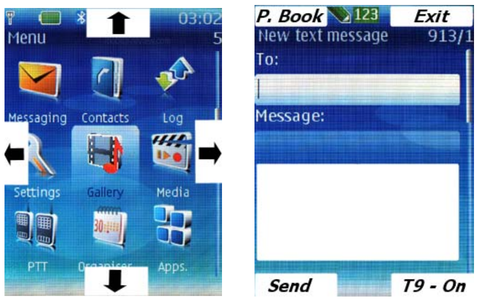

Fig. 1. A mock-up of cursor control is shown on the left. The mock-up on the right shows four commands at the corners of the screen. A user might "blink to click" any of the commands, or "blink to move" to the text input fields. The command areas might be context-sensitive, with different commands displayed, according to the user's current control focus.

We decided to emulate the control of a mobile phone using the eyes and the nose as a pointing device. Specifically, we asked 14 subjects to look at the four corners of the phone's screen, whilst being videoed by the front facing camera of a mobile phone (E1000). This was first done with the subjects holding the phone at a distance and posture which they felt natural, and repeated with the subjects holding the phone at a such distance that their face filled the viewfinder screen (the top and bottom of their face was at the top and bottom edges of the image). Subsequently, we carried out the same experiment, asking the subjects this time to imagine "pointing their nose" to the four corners of the screen, rather than just looking at it. Again, videos of the users were taken at a natural device holding distance, and at a closer distance, similar to the previous method.

5. Findings

Because we had to consider the user's natural movement, we decided that an appropriate solution would be to examine the eye movement in relation to the corners of the eye. This would provide a stable and relative measure for movement, independently of the user's head position.

Detecting the precise eye movement at a natural distance proved to be difficult, even though we found it to be directly observable in all but 3 of the participants. We attempted to measure the distance between the edge of the iris to the inner corner of the best-illuminated eye, however, we realised that because of the low resolution of the camera, this would be impossible. The estimated distances between eye corner and iris were between 3-7 pixels and given the antialiasing and compression employed in the video, this made any attempt to measure inaccurate.

We observed that eye movement could be perceived by the following two characteristics: firstly, the amount of visible sclera (white of the eye) between the iris and inner or outer eye corner. This was particularly useful in observing lateral movement. We also observed that looking at the bottom of the screen resulted in further and notable dropping of the upper eyelid. Further to this fact, we observed in all but two participants, that a detectable and significant movement of the entire head followed the movement of the eyes around the corners of the screen. We also observed that lighting conditions prevented us from perceiving any eye movement of two subjects and hairstyle (fringe) and dark coloured eye-shadow also made eyelid movement detection impossible with two female subjects.

Regarding eye movement at a closer distance, it was much more easily detectable. However, again, the distance in pixels between the iris edge and the eye's inner corner did not exceed 9 pixels, which made any measurement inaccurate. As in the previous observation, head movement was also detected in this instance, although this time it was accentuated slightly. Again eyelid movement was highly useful in determining whether the subject was looking at the upper or lower corners of the screen.

In these experiments, we detected slight head movement which followed the completion of the movement of the subjects' eyes almost immediately. We hypothesised that it would be possible to measure the movement of the head by examining the position of the facial characteristic which is closer to the camera, and whose movement would thus appear to be the greatest in the captured video. This facial characteristic is naturally the tip of the nose. We measured the distance in pixels between nasal positions when asking our subjects to "point" to the corners of the screen with their nose. We discovered that a vertical difference of 9.5 pixels and lateral difference of 8.3 pixels on average in the position of the centroid of the nasal tip could be noted for all of our subjects, which is significant, considering the dimensions of the captured video (176x144). We repeated our measurement for the same tasks performed this time with the device at a closer distance, where we found greater differences of about 15.1 pixels laterally and 18.7 pixels vertically. This was anticipated, given the device was much closer to the user.

6. Discussion

It appears from our observations that examining the overall head or fixed facial feature movement, is a more promising alternative than eye tracking. However, 10 out of 14 of our subjects reported at the end of the experiments that they would be conscious of using their head movements as a method to control their device in public, as they felt it would look “awkward”. We believe that there is still some possibility for combining eye movement with cues from the apparent slight head movement we encountered, in order to create a more natural and acceptable method of control for mobile users. We are in the process of testing a method which relies on the discovery of the position of the eye on the face, splitting the eye image vertically in half across the iris and determining the amount of visible sclera by examining the mean greyscale intensity of the image. We hypothesize that we may be able to establish a direct correlation of the mean greyscale intensity with the amount of visible sclera (and thus iris movement).

References

[1] M. Nixon, Eye spacing measurement for facial recognition, in: Proc. of the Society of Photo-Optical Instrument Engineers, 1985.

[2] A. Yuille, P. Hallinan, D. Cohen, Feature extraction from faces using deformable templates, Int. J.Comput. Vis. 8 (2) (1992) 99–111.

[3] X. Xie, R. Sudhakar, H. Zhuang, On improving eye feature extraction using deformable templates, Pattern Recognit. 27 (1994) 791–799.

[4] K.M. Lam, H. Yan, Locating and extracting the eye in human face images, Pattern Recognit. 29 (1996) 771–779.

[5] A. Pentland, B. Moghaddam, T. Starner, View-based and modular eigenspaces for face recognition, in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR_94), Seattle, WA,1994.

[6] W. min Huang and R. Mariani, Face detection and precise eyes location, in: Proc. Int. Conf. on Pattern Recognition (ICPR_00), 2000.

[7] J. Huang, H. Wechsler, Eye detection using optimal wavelet packets and radial basis functions (rbfs), Int. J. Pattern Recognit. Artif. Intell. 13 (7) (1999) 1009–1025.

[8] S. Kawato, J. Ohya, Real-time detection of nodding and head-shaking by directly detecting and tracking the between-eyes, in: Proc. 4th IEEE Int. Conf. on Automatic Face and Gesture Recognition, 2000, pp. 40–45.

[9] G.C. Feng, P.C. Yuen, Variance projection function and its application to eye detection for human face recognition, Int. J. Comput. Vis. 19 (1998) 899–906.

[10] G.C. Feng, P.C. Yuen, Multi-cues eye detection on gray intensity image, Pattern Recognit. 34 (2001) 1033–1046.

[11] Y. Tian, T. Kanade, J.F. Cohn, Dual-state parametric eye tracking, in: Proc. 4th IEEE Int. Conf. on Automatic Face and Gesture Recognition, 2000.

[12] B.D. Lucas, T. Kanade, An iterative image registration technique with an application to stereo vision, in: International Joint Conference on Artificial Intelligence.