Face and Gaze Tracking as Input Methods for Gaming Design

Florin Nanu[1]; Stefan Petrescu[1]; Peter Corcoran[2,3], *Fellow IEEE;* Petronel Bigioi[2,3], *Senior Member IEEE*
DigitalOptics Corp. (Romania) [1]*; DigitalOptics Corp. (Ireland)* [2]*; National University of Ireland, Galway, Ireland* [3]

***Abstract*—Real-time face detection combined with eye-gaze tracking and face analysis can provide a new means of user input in a gaming environment. Game designers can use facial information in various ways in designing the user interface (UI) and to provide smarter modes of interaction with players.**

I. INTRODUCTION

Gaming is one of the fastest developing IT industries in recent years. Providing new means of man-machine interaction can change the way gaming is perceived and enhance the underlying user experience. With the proliferation of portable entertainment devices in various form factors, it becomes more important to monitor the user to determine their engagement and reaction to the gameplay.

Here we present solutions to enable real-time human face detection and tracking combined with real-time eye tracking. Additionally, these solutions can be complemented with techniques for the determination of human emotions to provide feedback into the gaming user interface and gameplay logic [1]. Portions of the underlying work are already deployed in a number of commercial devices in either software and/or hardware accelerated versions.

II. THE NEED

The presence of a user-facing camera in modern portable devices (mobile phones, gaming portable devices, etc.) is not a new concept and has been installed by various manufacturers for a variety of communication use-cases. Other scenarios are where the user captures self pictures and/or movies. The presence of this user-facing camera opens up a number of additional possibilities and usage scenarios as described in this section.

A. Power Management

An important aspect for battery-operated single user devices, such as smartphones, gaming consoles and laptops is optimization of power usage. When the user is not directly viewing the monitor, then the device should implement progressive levels of power management – beginning with a dimming of the display backlight and culminating with a complete shut-down of the device. Assuming the eye gaze can be detected successfully and the capability of the display to have the backlight controlled on specific areas, the operating system could also implement selective region of interest ROI highlighting, while keeping the remainder of the screen dimmed.

B. Auto Stereoscopic 3D Displays

3D displays are beginning to appear in single-user consumer devices. However, one of the main barriers to success seems to be the need for special glasses, an inconvenient requirement for many users. One solution to this problem is a single user auto stereoscopic display where the 3D effect is optimized for a certain distance to the display (e.g. 35 or 40 cm depending on the display size). Such displays rely on the use of a specialized coating on the display known as a *parallax barrier*. This is a coating placed in front of a display, to allow it to show a stereoscopic image. For an LCD, it consists of a layer of material with a series of precision slits, allowing each eye to see a different set of pixels, thereby creating a sense of depth through parallax. The viewer must be positioned centrally, and at a specific distance in order to correctly experience the 3D effect.

Active research is ongoing into displays where the point of parallax for the display can be adapted using software control. A key requirement for such an approach to function properly is the capability to accurately measure real-time face and eye movement, location and tracking data at frames rates of 30 fps or more. Without such real-time face and eye data, it is not possible to control the display in a real-time feedback loop.

C. Automatic Authentication and Profile Switching

Using face detection in combination with face recognition we can continuously authenticate users. This enables a range of new user-specific services and applications and also opens the door to micropayment based business models. For certain shared devices such as gaming consoles, the same information can be used for automatic user profile switching to allow personal customizations and settings to be automatically applied when a known user is detected.

Device security is also enhanced if a device requires authentication from a *master user* to activate new services or enable certain device features.

D. Enhanced User Interface

Knowing, with a high degree of precision, where the eyes of the player are focused and having the ability to track facial features and emotions, game designers can provide enhanced character representation (e.g. animated avatars), as well as enhanced content presentation (e.g. the changing perspective of the gameplay window). User emotions can also be used as feedback to the gaming engine to dynamically adjust difficulty levels and/or decide the content presented to the user.

E. Facial Gesture Control

Various applications and games may accept user input based on facial event monitoring. In this respect, the presented technology can detect with a high degree of precision, various eye related events (single eye blink, double eyes blink, other blink combinations and eye(s) shut) that can also be extended to the mouth area (e.g. open/close mouth detector). Those events can be associated with frequent actions in a game that would otherwise require keyboard interaction (e.g. one eye shut in a shooting game can immediately bring up the scope mode).

III. Implementation

A. Algorithm Flow

Figure 1 illustrates the algorithm flow of a real-time face detector used to restrict the search for the eyes. To speed-up the algorithm, the initial face detection is performed on a low-resolution image (typically WQVGA resolution) with a minimum face size of 14x14 pixels. Once the face is detected, the principle face coordinates are up-sampled to match the image resolution used for a precise face re-detection step (typically at WVGA resolution). At this step the minimum face size is 64x64 pixels and is selected so that the eye size is large enough to ensure proper detection and tracking.

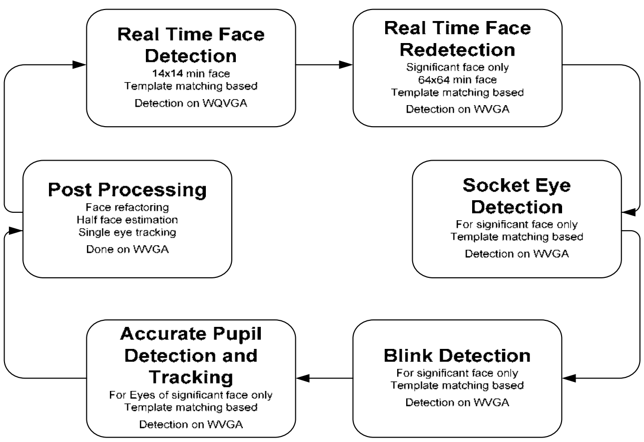

Figure 1: Eye Detection and Tracking Algorithm Flow

The main steps of eye and pupil detection and tracking are shown in Figures 2 and 3. Note that eye-state, particularly eye-blink is also determined and recorded [2]. The last step is to predict the position of the face in the next frame and eventually refine and provide the adjusted current face rectangle. During this prediction step a decision is also made whether a half face re-detection should be employed or not [3]. The whole cycle will be repeated for the next frame.

B. Algorithm Performance

Measurements have been performed on a Beagle board, equipped with a TI OMAP3530 DCBB72 having a 720MHz ARM Cortex-A8 CPU with NEON™ extensions, 16/16 KB L1 cache, 256 KB L2 cache. These preliminary experiments on the Beagle board do not include the first step. The locking for one face is performed directly on large resolution image (e.g. WVGA) and takes up to 22ms for the first detection. Once the face is detected, the tracking takes only 11 ms per frame, achieving real-time 30 frames per second effective processing. The algorithm uses up to 8MB of RAM and takes about 515KB of non-volatile flash storage including code, templates and data.

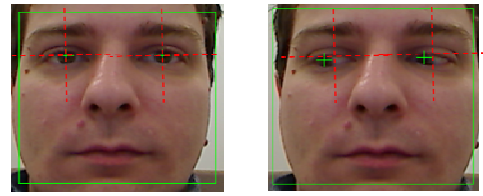

Figure 2: Eye Center Tracking (red) vs Pupil Center Tracking (green)

Figure 3: Pupil Tracking Examples

The algorithm performance is summarized in Table1.

Table 1: Eyeball Detection and Tracking Characteristics

Parameter	Value/Explanation
Input Image Size	800x480
Number of faces	1 face
Minimum face size	32x32
Yaw angle	[-25,+25] degrees with 2˚ precision measured between 30-45 cm to subject
Pitch angle	[-20,+30] degrees with 2˚ d precision measured between 30-60 cm to subject
Roll angle	[-30,+30] degrees with 2˚ precision measured between 30-60 cm to subject; *Note: with processing power expense, it can work on full trigonometric circle*
Horizontal translation distance	[-27, +27] cm, with 2mm accuracy @ 30-60cm subject distance and 72˚ FOV
Vertical translation distance	[-24, +24] cm, with 2mm accuracy @ 30-60cm subject distance and 72˚ FOV

IV. Summary and Conclusions

The current technology is able to provide the following facial parameters at 30 frames per second:

- Eye socket centers and pupil eye centers (with sub-pixel accuracy)
- Eyes stabilized precision face rectangle
- Distance to the subject (based on f35mm equiv focal length of the optical subsystem)
- Gaze angle as combination between pupil and socket eye detections
- Stabilized face roll angle (1-2˚ accuracy)
- Out-of-plane face angles (Yaw/Pitch)

References

[1] Bacivarov, I.; Corcoran, P.M.; , "Facial expression modeling using component AAM models — Gaming applications," *Games Innovations Conference, 2009. ICE-GIC 2009. International IEEE Consumer Electronics Society's* , vol., no., pp.1-16, 25-28 Aug. 2009

[2] Bacivarov, I.; Ionita, M.; Corcoran, P.; , "Statistical models of appearance for eye tracking and eye-blink detection and measurement," *Consumer Electronics, IEEE Transactions on* , vol.54, no.3, pp.1312-1320, August 2008

[3] Wei Chen; Tongfeng Sun; Xiaodong Yang; Li Wang; , "Face detection based on half face-template," *Electronic Measurement & Instruments, 2009. ICEMI '09. 9th International Conference on* , vol., no., pp.4-54-4-58, 16-19 Aug. 2009