Multimedia Browser Controlled by Head Movements

Lukasz Kosikowski, Piotr Dalka, Andrzej Czyzewski
Gdansk University of Technology, Multimedia Systems Department *

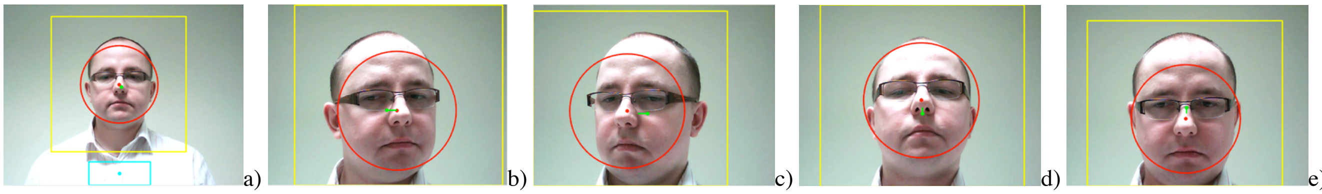

Figure 1: Sample results of head region (yellow rectangle) and head shape detection (red ellipse) without head gesture (a) and for gestures recognized by the application: turning the head left (b), turning the head right (c) turning the head up (d) turning the head down (e)

1 Introduction

A contactless multimedia content browser for personal computers is presented, where the user browses data using movements of his/her head only.

The presented solution supports browsing static images, videos and music clips can be browsed subsequently and zoomed on demand. Video clips can be viewed and paused. Additionally, a user may fast-forward or rewind the content. The same functionality applies to listening audio files.

Multimedia files are arranged in a multi-level, hierarchical structure. A user navigates through the structure and displays an element by moving the head up, down, left and right. Keeping the head in a tilted position for the longer time is also recognized. An action executed in the system depends on the type of the content a user is viewing (e.g. moving the head to the right selects the next picture or allows for fast-forwarding audio files). The content for the multimedia browser is chosen and organized with a separate configuration application that was also developed within the framework of the project.

The application is especially suitable to standalone, multimedia terminals where users may get acquainted with a company or a store offer in the fast and convenient way. The application may also be used by disabled people.

2 Interface description

Multimedia Browser is a standalone application developed within the Windows Presentation Foundation technology, which is one of the newest Microsoft technologies. Net framework 3.5 as a base of WPF allows for getting full advantage of graphics hardware acceleration using DirectX library. A rich user interface has been created. Many visual effects and transformation has been implemented including transparency animations, reflection effects, scale animations, and 3D path animations. On the computer's screen complex animations and high resolution multimedia content are presented simultaneously.

The Multimedia Browser has been designed using Model View ViewModel (MVVM) design pattern suited especially for WPF technology. The most important aspect of WPF that makes MVVM a convenient pattern to use is the data binding infrastructure. By binding properties of a view to a ViewModel, we have got a loose coupling between the two entities. It was not also nessesary to write code in a ViewModel that directly updates a view. The view Layer has been written in XAML language (eXtensible App Markup Language) whereas logic and model have been written in C# language.

The system requires only one hardware component - a standard web camera that captures images of the user face. The core of the system is formed by a head movement analyzing algorithm that finds a user face and tracks head movements in the real time. Face detection is based on a cascade of boosted classifiers working with Haar-like features [Viola and Jones 2001]. Head (face) movements in vertical and horizontal directions are detected by comparing the current face position to the mean face position over past video frames [Dalka and Czyzewski 2009]. Head movements are traced with a Finite State Machine (Fig. 2). State transitions are triggered by various spatial and temporal conditions. Whenever a state of the machine changes, an event is sent to the GUI application supposed to react accordingly. These events include tilting head up, down, left or right, keeping the head in a tilted position for a longer time (only horizontal direction) and returning to the normal position. Small head movements are sufficient to operate the multimedia browser. Additionally, a user presence in a camera field of view is reported. The system is immune to the presence of many faces in a video stream; only one face is tracked. The system is able to distinguish head movements resulting from a tilt of the head in vertical or horizontal direction from head movements caused by whole body motion (e.g. changing position in from of a camera or passing by in the camera field of view). No user adaption or calibration is required to work with the engineered system.

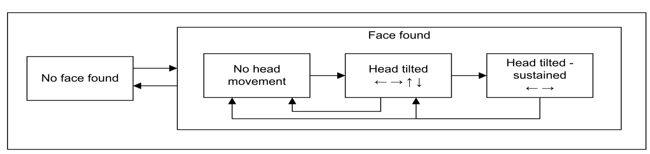

Figure 2: Overview of the head movement detection algorithm states (with possible head tilt direction marked with arrows) and available transitions

References

VIOLA P., JONES M., 2001, Rapid Object Detection using a Boosted Cascade of Simple Features, *IEEE CVPR*.

DALKA P., CZYZEWSKI A., 2009, LipMouse - a novel multimodal human-computer interaction interface, *SIGGRAPH*.

* e-mail: {kosiq, dalken, , andcz}@sound.eti.pg.gda.pl

SIGGRAPH 2010, Los Angeles, California, July 25 – 29, 2010.
ISBN 978-1-4503-0210-4/10/0007

Research funded within the project No. POIG.01.03.01-22-017/08, entitled "Elaboration of a series of multimodal interfaces and their implementation to educational, medical, security and industrial applications". The project is subsidized by the European regional development fund of the and by the Polish State budget".