KINterestTV

Can we measure in a non-invasive way, the interest that a user has in front of his television displaying its content ?

Project leaders: Leroy Julien (UMONS), Mancas Matéi (UMONS)
Participants: Rocca François (UMONS), Zajega François (UMONS)

Project Objectives

The main goal of this project is to analyze the users behavior in front of a television by using 3D cameras such as "Kinect" **[http://www.xbox.com/en-US/kinect]** and automatically understand their implicit behavior and interests for the media. This research is part of the EU FP7 project LinkedTv [http://www.linkedtv.eu].

The eNTERFACE project will be divided into two main axes:

- Development and integration of existing technologies, obtained in the framework of the research project LinkedTv. The objective is to develop a global software integrating:
    a. Face detection and tracking
    b. Face recognition
    c. Head direction (looking towards the TV or not) and depending on the relative precision, extracts some interests areas on the TV screen
    d. People counting and their biometric signature (adults, kids, sex, height)
    e. Extraction of “expressive” features (excitement, relative positioning of people …)
    f. Couch detection

- Data analysis. Beyond producing a feature extraction tool, we want to exploit these technologies to detect the interest that may bring the user to the media currently playing. A module of analysis of viewer behavior (face direction, physical activity, specific gestures) changes compared to his behavior history will be built to exhibit viewer interest towards the TV content. The final goal of this module is to detect and to measure how much of a media has attracted the interest of a user in order to improve dynamically his user profile.

Background information

LinkedTV is an integrated and practical approach towards experiencing Networked Media in the Future Internet! Networked Media will be a central element of the Next Generation Internet. Online multimedia content is rapidly increasing in scale and ubiquity, yet today it remains largely still unstructured and unconnected from related media of other forms or from other sources. Our vision of future Television Linked To The Web (LinkedTV) is of a ubiquitously online cloud of Networked Audio-Visual Content decoupled from place, device or source. Accessing audio-visual programming will be “TV” regardless whether it is seen on a TV set, smartphone, tablet or personal computing device, regardless of whether it is coming from a traditional or new media broadcaster, a Web video portal or a user-sourced media platform. More informations can be found on the LinkedTV website : http://www.linkedtv.eu

In this workshop, we will focus our work on the concept of reactive tracking for implicit understanding of the user behavior applied to the next TV generation. Behavioural tracking technologies generally regroup all means to observe, analyse and process non-verbal behaviour of a user. This is a very large topic spanning from speech (audio) analysis and gestures understanding to emotions reading, mouse clicks or navigation history. The possibility of automatically understanding human behaviour has already vastly been explored. In [Vinciarelli09], one may find a very good survey on social signal processing and behaviour tracking for non-verbal communication. In this section, we will focus on audiovisual features and the behavioural tracking which falls under the scope of computer vision technologies. Moreover, we focus on TV or home related experiences applications where the state of the art is much less deployed.

We work with the Microsoft Kinect sensor, that is a cheap and effective device which also gained popularity within a large public with its explicit gesture analysis for video games. Also other 3D camera-based explicit interfaces developed for TVs and interactive adds [INTEL12] are more and more popular and lead the public to be more and more aware about these technologies for home and TV-based applications. This trend already pushed some TV manufacturers like Samsung to propose new cameras directly embedded into the TVs [SAM12]. Even if those efforts mainly intend to provide explicit control on TV interfaces, the same systems can also be used in implicit interfaces and behavioural tracking. Moreover some implicit data like the sex or age are already inferred from the cameras.

Within the work on implicit interfaces, we can find the work of Microsoft research on the Kinect to extract face emotions and provide them to avatars where the face emotions are extracted in real time (right image) and then simulated on the left image avatar [MIC_AVA12]. Interfaces showing data use real-time context as social networks or related songs to provide more context-aware information [MIC_AMB11].

Detailed technical description

Technical description

The eNTERFACE project will be divided into two parts.The first objective is to develop a software integrating face detection and tracking, face recognition, face direction, detect the number of viewers and their biometric signature and “expressive” features extraction. We will use 2D and 3D video processing mainly based on OpenCV and Point Cloud Libraries. Subparts of the system have already been developed and they only need to be integrated into a global software. The first part of this workshop will be to define a flexible and efficient software architecture, that can be extended by adding new analysis modules, while responding to the needs of computer vision algorithms.

The second objective is about data analysis. We will use the results from the first objective to try to infer how much a media has attracted the interest of a user in order to improve his profile. During preliminary testing, in the case of interaction with a second screen as a tablet, we observed that it was possible to identify the portions of media that attracted the attention of the user. For this, we developed a 3D head tracker based on point clouds analysis and synchronized with the media currently playing. It is thus possible to identify the focus of the user and link them to the media. Other planned developments would analyze both areas of interest but also the trajectories of the gaze. The assembly can be correlated with a scene analysis and context-bound. For example, it would therefore be possible to determine whether the gaze passes from one person to the tv and vice versa. For this second goal, we will have to develop different scenarios to analyze and code. For this, we use software like ELAN [ELAN].

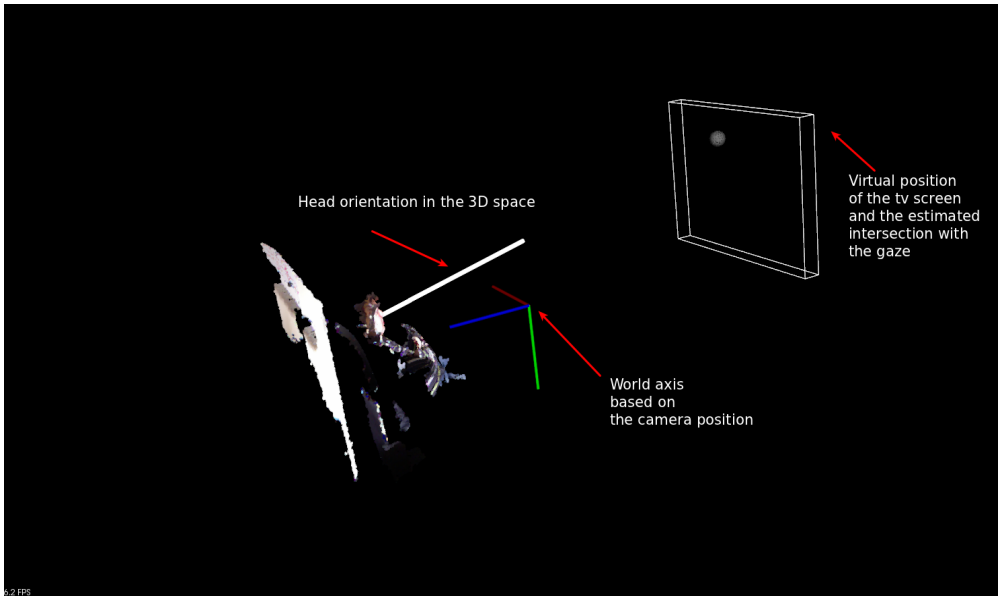

Figure 1: The 3D face point cloud orientation (white vector) and its intersection with the virtual screen.

Figure 1 shows the current state of the technology. A head tracker is applied to a 3D point cloud obtained with Kinect. We estimate the orientation of the face and the probable intersection with a virtual TV display. Figure 2 shows the kind of experimental setup which will be achieved in eNTERFACE.

Figure 2: Experimental setup

Resources needed

No particular resources are needed, apart from a quiet, isolated room and a TV screen for experiments. We will bring the Kinects with us.

Project management

Two subteams will be established working on each of the two developpement axes. The integration team will work the whole months to get a final software (and also integrating the analysis module of the second team). The second team will mainly focus on the research part of the project and turn the low level features extracted into higher level cues about people information. The second team can have data anytime as a first version of the face tracking is already done, they do not need to wait for the first team.

Work plan and implementation schedule

	Development Team	Research Team
1st week	Software architecture	Build scenarios based on the head pose estimation
2de week	Integration of tracking modules	Database and annotation for ground truth analysis
3rd week	Integration of new biometric modules	Feature extraction from face and body gesture to detect interest
4th week	Integration from the Research Team	Trials

Benefits of the research

The benefits of this project will be to analyze the users behavior in front of a television by face detection and body activity. We will also analyse who is in front of the TV (age, sex and name) based on face recognition. In a second step we will automatically understand their interests through previous features analysis.

Profile of team:

Leaders

Leroy Julien : he holds an Electrical Engineering degree (Ir.) from the University of Mons, Engineering Faculty (since June 2010). His master thesis was conducted in partnership with Infrabel, manager of Belgian rail network, on the automatic detection of defects in the catenary. His main research interests are : social signal processing, modeling of human behavior, point clouds processing and 3D animation. His PhD thesis focuses on modelisation of proxemic behavior based on computer vision techniques. One of his goals is to develop a new methodology to measure and model proxemics behaviors that is : accurate, ecological and less time consuming. This work is conducted in partnership with psychologists and psychiatrists.
Mancas Matei : he holds an ESIGETEL Audiovisual Systems and Networks engineering degree (Ir.), and a Orsay Univ. D.E.A. degree (MSc.) in Information Processing. He also holds a PhD in applied sciences from the FPMs on computational attention since 2007. His research deals with signal saliency and understanding (http://tcts.fpms.ac.be/attention).

Staff

Rocca François : he holds an Electrical Engineering degree from the FPMs since June 2011. He did his master's thesis in the field of emotional speech analysis, and more especially on laughter frequencies estimation. He is currently pursuing a PhD thesis on facial animation by markerless motion capture at UMONS.

Zajéga François : he is a visual artist mainly interested in video analysis, visualisation and interfaces. He has studied infography in Saint-Luc, Bruxelles. He has been a fulltime contributor to the numediart Reseach Program on Digital Art Technologies since 2010.

Other researchers needed

- A computer scientist with good programing skills (C/C++) with optional interest to social signal processing and computational attention. He/she would work on system integration and help on social signal processing and computational attention algorithms.
- A psychologist to set-up pertinent scenarios details and the most interesting psychological tests.
- Of course any interested people are welcome both from engineering and humanities.

References:

ALDOMA, Aitor. 3D Face Detection and Pose Estimation in PCL. 2012.

S.Papadopoulos, E. Schinas, V. Mezaris, R. Troncy and I.Kompatsiaris, Social Event Detection at MediaEval 2012: Challenges, Dataset and Evaluation,*In Proceedings of the MediaEval Benchmarking Initiative for Multimedia Evaluation (MediaEval 2012) held on 4 & 5 October 2012 in Pisa, Italy.*

[Vinciarelli09] A.Vinciarelli, M.Pantic and H.Bourlard Social Signal Processing: Survey of an Emerging Domain Image and Vision Computing Journal, Vol. 27, No. 12, pp. 1743-1759 (2009).

[INTEL12] Intel and SoftKinetic develop interfaces for interactive adds: http://techcrunch.com/2012/01/30/softkinetic-and-intel-partner-for-minorityreport-style-ads/

[SAM12] Samsung integrates cameras directly in their TVs: http://www.sananews.net/english/2012/01/samsung-will-includeintegrated-camera-and-microphone-in-smart-tv-2-0/

[MIC_AVA12] From real humans to avatars in real-time. http://www.microsoft.com/presspass/features/2012/jan12/01-03Future.mspx

[MIC_AMB11] Ambient intelligence at Microsoft: http://www.engadget.com/2011/05/03/microsofts-home-of-the-future-lullsteens-to-sleep-with-tweets/

[Ballendat10] Ballendat, T., Marquardt, N., and Greenberg, S. Proxemic Interaction: Designing for a Proximity and Orientation- Aware Environment. Proc. of ITS'10, ACM (2010).

[Greenberg11] Greenberg, S., Marquardt, N., et al. Proxemic interactions:the new ubicomp? interactions 18, ACM (2011), 42–50.

[Elan] http://tla.mpi.nl/tools/tla-tools/elan/