REALISATION OF AN AUDIO & VIDEO LABORATORY FOR PRECISE OBJECT LOCALISATION AND TRACKING

Robert Manthey[1]*, Hussein Hussein*[2]*, René Erler*[1]*, Richard Siegel*[1]*, and Danny Kowerko*[1]

[1]*Department of Computer Science, Chemnitz University of Technology, Chemnitz, Germany*
[2]*Department of Literary Studies, Free University of Berlin, D-14195 Berlin, Germany*
firstname.lastname@informatik.tu-chemnitz.de, hussein@zedat.fu-berlin.de

Abstract: This paper presents the realisation of an audiovisual laboratory for detection, localisation, classification and tracking of objects in an indoor environment using visual as well as audio information. The laboratory is property of the junior professorship Media Computing at the Chemnitz University of Technology. It was funded by Federal Ministry of Education and Research by the program of Entrepreneurial Regions InnoProfile-Transfer.

Visual information is retrieved by 10 Intenta S2000 optical embedded smart stereo sensors. Each capture a video signal of HD resolution, corresponding depth information and basic metainformations like sitting or standing of detected persons. For audio processing, a total of 64 microphones and 16 loudspeakers (Genelec & Tannoy) are used. Three microphone arraies are constructed using 16 $\times$ Nowsonic Calibration microphones, 16 $\times$ MXL 840 microphones, and 24$\times$ Justin JM-714 microphones. The active tracking area measures (L $\times$ W $\times$ H = 4 $\times$ 3.5 $\times$ 3.5 meter) within a room (L $\times$ W $\times$ H = 7.2 $\times$ 6 $\times$ 4 meter) and comprises a total of 82 sensors. The camera sensors and microphones can be mounted in different positions, directions and heights. The loudspeakers can be set up freely within the tracking area using standard monitor stands. Within a separate air-conditioned server room, all audio signals are pre-amplified and AD/DA converted using industry standard rack-mounted audio hardware, such as RME & Focusrite interfaces and a Rosendahl masterclock. A server cluster and workstations with high-end Nvidia P6000 graphics cards provide the raw processing power for all processing tasks within the scope of use cases for the laboratory.

The visual sensors be used to trigger an identification process of objects and persons as well as a tracking operation to follow them in the field of view. Based on this, they can analyse the behaviour of persons and start interpretation of the meaning of the movements of their extremities. With the cooperation of different sensors occlusions and ambiguities can be solved to improve the quality of detections, identifications and behaviour interpretations.

Software which is available for users in the laboratory comprises commercial products, such as Steinberg Cubase 8.5, as well as self-developed solutions, such as an audio & video localisation and annotation tools. This will be complemented by documenting acoustic sources (e.g. the speakers) and detectors (microphones) into a Blender-animated 3D laboratory using the video sensor data. Interested individuals will also have the opportunity to make use of the laboratory's extra peripheral equipment, e.g. a Yamaha DGX-660 Keyboard, a Roland A-88 MIDI Controller-Keyboard, further types mobile wireless of microphones, HTC Wive VR glasses and wireless Internet of Things (IoT) smart home sensor probes from for motion capturing.