INTERACTIVE MULTIMODAL INFORMATION MANAGEMENT

IM2 FINAL EVENT - INTERACTIVE POSTER SESSION

1 Idiap — "An interactive game exploiting state-of-the-art speech technologies"

Authors:
Blaise Potard
Petr Motlicek

Presenter:
Blaise Potard

We present an interactive computer game exploiting state-of-the-art speech processing technologies (Speech recognition and Speech synthesis) for Human-machine interaction.
This game simulates the TV-based quiz game "Who wants to be a millionaire?", shown in many countries. The game exploits core speech technologies developed at Idiap and is extended to support an open-architecture distributed system, enabling it to be played on different platforms, and in several languages.

2 Idiap — "Automatic Staging of Text with Emotions"

Authors:
Lakshmi Saheer
Milos Cernak

Presenter:
Lakshmi Saheer
Milos Cernak

Current day text-to-speech technologies are mature enough to be acceptable in quality for the users. There is still a large gap between a synthesised speech and a real human speech due to lack of expressions and emotions. Geneemo is a technology for automatic addition of emotions and expressions to any audio. The audio can be generated by any text-to-speech technology. The aim of the project is to make human computer interactions as natural as possible with expressive speech. This also opens up a portfolio of applications replacing real human voices. We propose a online tool that will allow people to enter text, tag it with emotions and expressions and generate the audio with the corresponding effects.

3 Idiap — "Very Low Bit Rate Speech Encoding Using Incremental Speech Recognition And Synthesis"

Authors:
Milos Cernak
Phil Garner
Petr Motlicek

Presenter:
Milos Cernak

We present a framework of a HMM-based recognition/synthesis speech coder with phonetic vocoder, focusing on real-time aspects of the encoding process. While speech recognition fires results with last frame processed, speech synthesis requires long context for re-synthesis of the speech. This introduces a long delay, estimated minimally as two times the duration of the encoded speech. To overcome this problem, we investigate in this work to use of incremental speech recognition and synthesis. The presentation will be accompanied by the interactive demonstration of the proposed English speech encoder with an average syllable-based communication delay.

4 Uni Fribourg — "EmotiBoard - Live Multimodal Emotion Recogntion for Augmented Remote Interaction "

Authors:
Fabien Ringeval
Andreas Sonderegger
Juergen Sauer
Denis Lalanne

Presenter:
Fabien Ringeval

Previous research in the domain of work psychology has indicated that mood, emotion and team members empathy may influence team processes and the outcomes of teamwork, such as performance, cohesion and satisfaction. This suggests that emotion awareness of each team members is important for efficient and satisfactory teamwork. Since research in the domain of affective computing has made significant progress in automatically detecting emotional states of humans, we developed a tool, the EmotiBoard, for providing automatic affective feedback in teamwork during video-conference. This tool aims to reduce the emotion awareness's gap between collaborators, by providing an emotion feedback based on real-time automatic emotion recognizers. A demo of the EmotiBoard will be presented: 2 persons will talk to each other remotely and an adaptive emotional feedback will be displayed according to the emotions that would be detected by the system. Different visualization techniques of the emotional feedback will be proposed to the public, with the possibility to rate the one they prefer.

5 Uni Fribourg — "Two handed mid-air gestural HCI: Point + Command "

Authors:
Matthias Schwaller
Denis Lalanne

Presenter:
Matthias Schwaller

This demo proposes to interact with the computer using two-handed mid-air gestures, maximizing accuracy and minimizing effort. The main idea driving the design of these gestures is that one hand is used for pointing, and the other hand for four standard commands: selection, drag & drop, rotation and zoom. Two gesture vocabularies are proposed in this demo. The demo further presents a novel evaluation methodology and the application developed to evaluate the four commands first separately and then together. At the end of the demonstration, the results of a user evaluation will be presented to the participants.

6 ETHZ — "Real-Time Facial Expression Recognition on a Mobile Device"

Authors:
Joël Smely
Angela Yao
Luc Van Gool

Presenter:
Joël Smely

Nowadays it is quite possible to think about futuristic computers/roboters, which are interacting with humans all day through. One step into this realm is the detection of human facial emotions. This demo presents a user-independent approach for a real-time, frame-by-frame classification of facial expressions. The system is build out of a classification forest learning relations between facial image patches from the entire set of faces. The randomized forest can also be updated by modifying his trees post-training, thus adapting the system to the current environment. The approach was tested on a publicly available database and achieved encouraging results comparable to the state of art in addition beiing a real-time system. The real-time functionality was tested on mobile smart phones running the open-source operating system Android on them, and will be shown during the demo.

7 ETHZ — "Real-time Face Analysis using Random Forests"

Authors:
Gabriele Fanelli
Matthias Dantone
Juergen Gall
Andrea Fossati
Luc Van Gool

Presenter:
Matthias Dantone

We present a random forest-based framework for real time face analysis. Our demo consists of four different tasks: face detection, head pose estimation, fiducial point detection and facial attribute classification. We first detect the face using a state-of-the face detection algorithm. After resizing the face to a common size a random regression forest estimates the head pose and the location of the 10 facial features. For head pose estimation and fiducial point detection we use a voting approach, where each patch extracted from the image can directly cast a vote for the head pose or each of the facial features. Our system proves capable of handling large rotations, partial occlusions, and the noisy and blurry images in the wild. For the attribute classification we first normalize the face and then we extract features around the fiducial points. With a linear SVM and the extracted features we are capable of classifying 25 facial attributes (e.g. gender, sunglasses, ...). Finally, the face detection, head pose estimation and facial feature detection work on each frame independently and achieve real time performance without resorting to parallel computations on a GPU.

8 Idiap — "Content Linking for Conversations and Multimedia "

Authors:
Maryam Habibi
Nikolaos Pappas
Chidansh Bhatt
Andrei Popescu-Belis

Presenter:
Maryam Habibi
Nikolaos Pappas

This presentation covers two related research strands. The first part will describe a method for content extraction from conversations, which is used in real time to formulate queries and suggest related content to the conversation participants. The second part generalizes this concept to multimedia recommendation: for each segment of an audio-visual recording, such as a lecture, the most relevant segments from other lectures are proposed, based on a variety of criteria.
The large screen will be used to show: (a) a demo of each system, i.e. one with real-time ASR and related content, and the other one with recommendation-based navigation through lecture segments; (b) a poster explaining each system and the research behind it; (c) a standalone demo video for each system.
The first author is a PhD student supported by IM2.IP2; the second one a PhD student on a related EU project (inEvent); the third one a postdoc supported by the AROLES SNF project for tech transfer from IM2.

9 Idiap — "Mediaparl"

Authors:
Imseng David
Nanchen Alexandre

Presenter:
Imseng David
Nanchen Alexandre

Mediaparl is enabling the Canton of Valais's Members of Parliament, and the general public, to find the content of any parliamentary motion, resolution, or debate, on video – a pioneering tool which is also attracting the interest of the federal government in Bern. We present the interactive tool that was developed at Idiap research institute

10 Uni Geneva — "Quantized Ranking for Permutation-based Indexing"

Authors:
Hisham Mohamed
Stéphane Marchand-Maillet

Presenter:
Hisham Mohamed

Similarity search, translating into the nearest neighbor search problem, finds many applications for information retrieval and visualization, machine learning and data mining. The context of Big Data imposes to find approximate solutions. Permutation-based indexing is one of the most recent techniques for approximate similarity search. We propose a quantized representation of the permutation lists with its related data structure for effective retrieval. We detail and illustrate our novel indexing strategy and demonstrate its scaling capabilities on very large datasets of high-dimensional items.

11 Uni Geneva — "Detecting aesthetic highlights in movies based on physiological and behavioral coupling of spectators: a pilot study "

Authors:
Theodoros Kostoulas
Guillaume Chanel
Sophie Jarlier
Patrizia Lombardo
Thierry Pun

Presenter:
Theodoros Kostoulas

The annotation of multimedia content with discriminating tags is crucial for multimedia information retrieval. Further, the exploration of the temporal dimension of emotions represented in and elicited by moments of aesthetic importance, throws light on the nature and the responses to art. Within this work we focus on the development of a movie highlights' detection system based on physiological and behavioral coupling among spectators in an ecological situation, since they are watching a movie in a theater. Further, we investigate how this system can be helpful in film studies in assessing the aesthetic value of movie scenes. The primary goal, in the first phase of our work, was to collect and analyze physiological and behavioral data from spectators watching a movie in a controlled environment, i.e. a classroom. The physiological data consist of galvanic skin response, skin temperature, blood flow, breathing pattern and facial thermal responses. The behavioral data consist of acceleration measurements and audio-video collected utilizing conventional and thermal camera. The objective of the first phase was to validate the experimental setup of the collection campaign and to uncover possible flaws. In the second phase of our work, the collection of physiological and behavioral data takes place in a movie theater. The participants of this workshop will have the chance to get an insight of the demonstrable outcome of the aforementioned two phases of the project.

12 EPFL — "Augment Meetings with Opportunistic Search"

Authors:
Nan Li
Pierre Dillenbourg

Presenter:
Nan Li

In meetings, people sometimes happen to come across information in the meeting environment or in the discussion. This kind of accidentally encountered information might be helpful to the ongoing meeting. We question if computers can be designed to generate such information systematically so that its chance of being encountered by the users as well as the usefulness increases. In the past years, we have been investigating how to design computer systems to present spontaneous, up-to-date and context-specific search suggestions in meetings, as well as how people react upon these suggestions. Our results will be presented in this presentation.

13 EPFL — "Gaze perception methologies for Human-Human and Human-Robot interactions"

Authors:
Samira Sheikhi
Kenneth Funes
J.M. Odobez

Presenter:
Samira Sheikhi
Kenneth Funes

Gaze estimation plays a central role for a large set of applications. In many situations remote cameras are preferred over wearable sensors to minimize user cooperation. Depending on the distance to the camera, the eyes might be visible and we could use this information for estimating the gaze direction. In contrast, when the distance is large, other methodology is needed for gaze estimation. We will then present two main strategies for these cases:

A) When the eyes are visible to the system, like in scenarios where the person is close to the camera. We will present methods which use the eyes' information to infer the actual gaze direction. To this end we rely on recent consumer RGB-D cameras, which help to address some of the main challenges of gaze estimation.

B) In cases where the eyes are not completely visible, like where a robot is interacting with a number of people, we would rely on the head pose and its dynamics to have an estimate of the eye gaze.

For both cases we would make use of the 3D positions of the objects in the scene to map the gaze and recognize what the humans are actually looking at. In addition contextual information like robot dialog context can be added to make a better recognition of the visual focus of attention.

These methods enable addressing a large set of applications in human-human and human-machine interaction

14 EPFL — "Multimodal detection of transparent household containers"

Authors:
Magrelli Silvia
Cretton Gaetan
Shirvani Boroujeni Mina
Norris Basilio
Billard Aude

Presenter:
Magrelli Silvia
Cretton Gaetan

Transparent objects are ubiquitous in interior environment; however their detection and tracking is far from the performances of detectors limited to opaque objects. We are interested in tracking transparent objects for human behavior analysis as well as robotic grasping tasks. Using a combination of multispectral camera (color and near infrared) and a depth camera, we explored several approaches based on recent state-of-the-art publications.

We will demonstrate the detection of transparent containers on a tabletop setting fit for robotic grasping.

15 EPFL — "Swiss Cheese for visual search"

Authors:
Ivan Ivanov
Martin Rerabek
Touradj Ebrahimi

Presenter:
Martin Rerabek

Within the frame of IM2.IP1, MMSPG at EPFL has developed Cheese, an advanced image management platform for online use and mobile devices. Beside standard features such as image upload, tagging and keyword based search, it offers the user visual similarity based search, object based tagging and semi-automatic tag propagation. For improved interoperability between different image repositories and applications, the platform supports the export and import of image files with embedded metadata in JPSearch - Part 4 compliant format.

16 EPFL — "Real-time facial expression recognition"

Authors:
Anil Yüce
Hua Gao
Jean-Philippe Thiran
Presenter:
Anil Yüce
Hua Gao

In this presentation, we will show our recent work on facial expression recognition. The system built uses the variations of Local Curvature Gabor Binary Patterns in a GentleSVM framework to detect 17 AUs or 6 basic expressions. The comparative nature of the features proposed allows tracking the relative intensity of the facial expressions and we obtain very high accuracy expression detection performance due to their representation power. Finally, an implementation of the system will be demonstrated in a scenario where facial expressions are recognized in real time.

17 Idiap — "Cross-domain personality inference: From Social Media to Small Group Meetings"

Authors:
Oya Aran
Daniel Gatica-Perez

Presenter:
Oya Aran

We investigate the use of YouTube content as a domain to learn social behavior of individuals, particularly personality. Our aim is to transfer the knowledge that can be extracted from conversational videos in video blogging sites to small group settings to infer the extraversion trait with nonverbal cues. We use a YouTube dataset containing personality impression scores of 442 people as the source domain and a small-group meeting dataset from a total of 102 people as target domain. Our results show that for the extraversion trait our transfer learning method is able to achieve higher accuracy than using only the data recorded in small group settings.

18 Idiap — "Sensing Collaborative Groups in Competitive Environments"

Authors:
Dairazalia Sanchez-Cortes
Trinh-Minh-Tri Do
Daniel Gatica-Perez

Presenter:
Dairazalia Sanchez-Cortes

We plan to capture individual behavioral cues of entrepreneurs in a competitive environment, by sensing their interactions during ICC 2013 using smartphones. The contextual information captured with the devices will be used to model interaction and communication patterns while involved in a competitive activity. The demo will capture our general approach and initial findings.