Issue N° 60 / February 2012

Editor : Valérie Devanthéry
E-mail : info@im2.ch

INTERACTIVE MULTIMODAL INFORMATION MANAGEMENT

IM2 Newsletter

Contents

Event

HRI 2012

Boston, Massachusetts, USA
March 5-8, 2012

http://hri2012.org

Cover Story

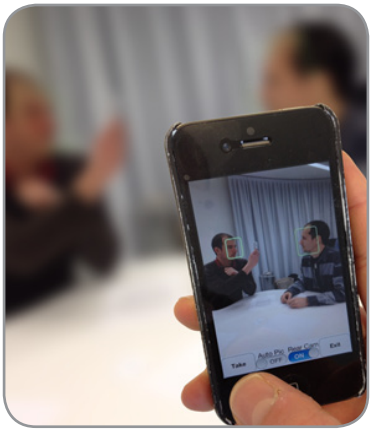
60 Seconds Minutes

A SMARTPHONE-BASED COMIC STRIP GENERATOR OF FACE-TO-FACE GROUP MEETINGS

At EPFL, a team of PhD students under the supervision of Dr. Frédéric Kaplan and Prof. Pierre Dillenbourg are working on a joint mini project in the framework of IM2/IP2 to develop a smartphone-based comic strip generator of face-to-face group meetings, based on visual and acoustic related features. During the meeting several multimodal corpora are captured via smartphones, automatically time stamped, elaborated and centralized in a database. After the meeting the secretary can make use of a web-based wizard to generate a meeting minute in the form of comic-strip.

An iPhone application is distributed to the participants prior to the meeting. Once in the meeting, one of the participants starts a new session from her iPhone. The others use the application to log in to the meeting. Thanks to the geo-location assistance they can easily find the originator of the session. A Login Server provides the backend functionality for localizing the meetings and supporting multiple meetings in parallel.

After this step the application becomes a content producer of audio and visual elements. The application continuously streams audio related low level features and photos to the feature server.



The user can enable/disable the automatic picture taking, can manually take pictures (which are marked with a priority flag) and can choose which camera (front, rear) to use (on iPhone 4 and 4S only).

The feature server processes all these multimodal corpora in order to segment faces, recognize documents, extract annotations, detect untrained textual patterns and segment the meeting timeline into meaningful episodes based on acoustic related features. All these elements are stored into a database.

After the meeting a web application equipped with a wide range of controllers will show all the stored features in a temporally coherent manner and contextual to the originators.

To be continued on page 2

www.im2.ch

IM2, c/o IDIAP Research Institute, Centre du Parc,
Rue Marconi 19, P.O. Box 592, 1920 Martigny
info@im2.ch - www.im2.ch

IM2 is the Swiss National Centre of Competence in Reseach (NCCR) on Interactive Multimodal Information Management, lead by the Idiap Research Institute in Martigny, Switzerland. The National Centers of Competence in Research are managed by the Swiss National Science Fundation on behalf of the Federal Authorities.

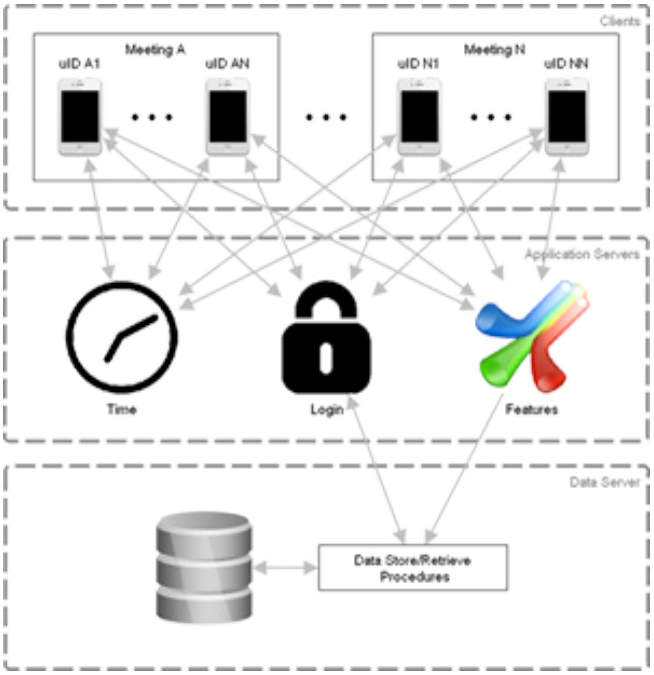
Editor : Valérie Devanthéry
E-mail : info@im2.ch

60 Seconds Minutes (continued)

The meeting speech segmentation into topics determines the spatial arrangement of the comic strip and for each comic frame provides recommendations, in the comic language, concerning the participants' subjectivity. The meeting secretary will make use of all these means to arrange the final comic strip of the meeting.

Andrea Mazzei focuses his efforts on the extraction and classification of all the visual features. Flaviu Roman works with distributed software engineering including iPhone clients and PC servers and provides the platform for content generation and logging capabilities. He also investigates methods for the prosody-based unsupervised segmentation of the meeting speech into meaningful episodes. Himanshu Verma is the leading developer of the data storage and retrieval infrastructure and the web-based wizard for the comics strip generation.

Andrea Mazzei
andrea.mazzei@epfl.ch



Completed Thesis, Mohamamd Soleymani, UNIGE

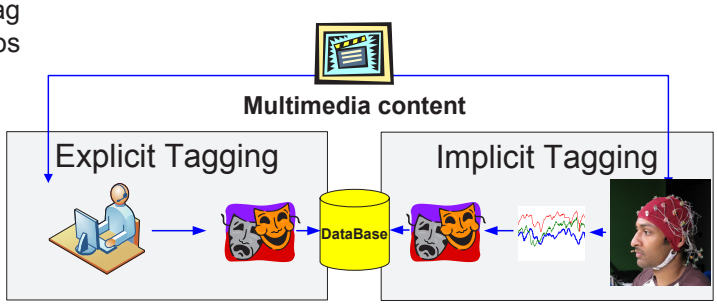
IMPLICIT AND AUTOMATED EMOTIONAL TAGGING OF VIDEOS

On November 4th 2011, Mohamamd Soleymani successfully defended his PhD thesis, entitled "Implicit and automated emotional tagging of videos", at the Doctoral program in computer science at the University of Geneva under the supervision of Prof. Thierry Pun. His work was part of the research contributed by Idiap to IM2. IP1.

Emotions play an important role in viewers' content selection and use. The main aim of this study is to detect and estimate affective characteristics of videos based on the content and viewers' response. These emotional characterizations can be used to tag the content. Implicit or automated tagging of videos using emotions help recommendation and retrieval systems to improve their performance. The analysis and evaluations directions in this thesis are twofold: first, methodology and results of emotion recognition methods employed to detect emotion in response to videos are presented.

Second, methodology and results of emotional understanding of multimedia using content analysis are provided. In conclusion, promising results have been obtained in emotional tagging of videos. However, emotional understanding of multimedia is a challenging task and with the current state of the art methods a universal solution to detect and tag all different content which suits all the users is not possible.

Mohamamd Soleymani
Mohammad.Soleymani@unige.ch



IM2, c/o Idiap Research Institute, Centre du Parc,
Rue Marconi 19, P.O. Box 592, 1920 Martigny,
info@idiap.ch - www.idiap.ch

INTERACTIVE MULTIMODAL INFORMATION MANAGEMENT

IM2 Newsletter

Issue N° 60 / February 2012

Editor : Valérie Devanthéry
E-mail : info@im2.ch

Sree Hari Krishnan Parthasarathi, IDIAP

PRIVACY-SENSITIVE AUDIO FEATURES FOR CONVERSATIONAL SPEECH PROCESSING

Sree Hari Krishnan Parthasarathi (Idiap Research Institute) successfully defended his PhD thesis at EPFL on November 4, 2011. His dissertation is entitled "Privacy-Sensitive Audio Features for Conversational Speech Processing".

The members of his doctoral jury were Jean-Marc Vesin (EPFL, president of the jury), Dan Ellis (Columbia University), Simon King (University of Edinburgh), Jean-Phillipe Thiran (EPFL), Herve Bourlard (Idiap, thesis director) and Daniel Gatica-Perez (Idiap, thesis co-director).

Hari's work was part of the research contributed by Idiap to IM2.IP1.

The work described in this thesis takes place in the context of capturing real-life audio for the analysis of spontaneous social interactions. Towards this goal, we wish to capture conversational and ambient sounds using portable audio recorders. Analysis of conversations can then proceed by modeling the speaker turns and durations produced by speaker diarization. However, a key factor against the ubiquitous capture of real-life audio is privacy. Particularly, recording and storing raw audio would breach the privacy of people whose consent has not been explicitly obtained.

In this thesis, we study audio features instead - for recording and storage - that can respect privacy by minimizing the amount of linguistic information, while achieving state-of-the-art performance in conversational speech processing tasks. Indeed, the main contributions of this thesis are the achievement of state-of-the-art performances in speech/nonspeech detection and speaker diarization tasks using such features, which we refer to, as privacy-sensitive. Besides this, we provide a comprehensive analysis of these features for the two tasks in a variety of conditions, such as indoor (predominantly) and outdoor audio. To objectively evaluate the notion of privacy, we propose the use of human and automatic speech recognition tests, with higher accuracy in either being interpreted as yielding lower privacy.

Our studies showed that the proposed approaches yield performances comparable to state-of-the-art features on the two tasks while preserving privacy

Dr Krishnan Parthasarathi is currently a postdoc at ICSI, Berkeley *(sparta@icsi.berkekely.edu)*.

Daniel Gatica-Perez
gatica@idiap.ch

Completed Thesis, Anindya Roy, IDIAP

BOOSTING LOCALIZED FEATURES FOR SPEAKER AND SPEECH RECOGNITION

Anindya Roy obtained his PhD at EPFL on Nov 11 2011 after the public defence of his thesis: "Boosting Localized Features for Speaker and Speech Recognition".

In his PhD thesis, Mr. Anindya Roy proposed to use local features, selected by a boosting algorithm, for the tasks of speaker and speech recognition.

He investigated more particularly:

1. the application of local binary features from computer vision to speech processing,
2. the robustness of these features to audio noise, and
3. their low complexity compared to traditional holistic features.

Additionally to this work, Mr. Roy proposed an extension of these local binary features to audio-visual processing, as well as a novel feature set for face detection.

His work was part of the research contributed by Idiap to IM2.IP1.

Sébastien Marcel
Sebastien.Marcel@idiap.ch

INTERACTIVE MULTIMODAL INFORMATION MANAGEMENT

Issue N° 60 / February 2012

IM2 Newsletter

Editor : Valérie Devanthéry
E-mail : info@im2.ch

News

KeyLemon: Track your face evolution with LemonDay

Use face recognition to log into your session

KeyLemon launches its new version with its new plugin LemonDay.

Each time your face is recognized, an updated photo will automatically be stored in your computer. LemonDay allows you to track your face evolution by saving automatically one image day after day. Collect photos of yourself, create a movie of your evolution and share it with your friends in the most simple and effective way.

KeyLemon works as a password manager to log on to your personal Windows account and for popular internet sites. When you connect to a website (Facebook, Twitter and / or LinkedIn), KeyLemon automatically logs you into your account by using your face.

A short movie of the LemonDay is available *http://www.keylemon.com/LemonDay*

Valérie Devanthéry
valerie.devanthery@idiap.ch

Pomelo Sàrl selected by PME Magazine

Published in November 2011

Pomelo Sàrl, young IM2 start-up from LASA at EPFL and active in shopper behavior studies, has been selected by PME Magazine for its innovative ideas and technology, to figure among the «150 new entrepreneurs who make the Suisse Romande» in its special issue on that topic, published in November 2011.

The issue praises pomelo's work using its eye-tracking technology – developed during the IM2 project – that has created solutions for improving the marketing process in the retail domain, and has already extended to the european market.

Basilio Noris
basilio.noris@epfl.ch

Selected publications

Evaluation of meeting support technology.
S. Tucker, and A.Popescu-Belis
In Multimodal Signal Processing, Human Interactions in Meetings, Cambridge University Press, 2011.

Towards a Descriptive Depth Index for 3D Content: Measuring Perspective Depth Cues
L.Goldmann, T.Ebrahimi, P.Lebreton and A.Raake
In Proceedings of the 6th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM), January 2012.

Geotag Propagation in Social Networks Based on User Trust Model
I.Ivanov, P.Vajda, J.-S.Lee, L.Goldmann and T.Ebrahimi
In Multimedia Tools and Applications (Springer), Special Issue on Social Mining and Search, Vol. 56, Nr. 1, pp. 155-177, January 2012

VlogSense:Conversational Behavior and Social Attention in YouTube?
J.-I. Biel and D. Gatica- Perez
In ACM Transactions on Multimedia Computing, Communications, and Applications, Special Issue on Social Media, Vol. 7S, No. 1, Oct. 2011.

The Kaldi Speech Recognition Toolkit.
D.Povey, A.Ghoshal, G. Boulianne, L.Burget, O.Glembek, N.Goel, M.Hannemann, P.Motlicek, Y.Qian, P.Schwarz, J.Silovsky, G.Stemmer, and K.Vesely
In Proceedings of IEEE 2011 Workshop on Automatic Speech Recognition and Understanding, Hilton Waikoloa Village, Big Island, Hawaii, US, December 2011.

Towards Certification of 3D Video Quality Assessment
A.Perkis, J.You, L.Xing, T.Ebrahimi, F.De Simone, M.Rerabek, P.Nasiopoulos, Z.Mai, M.Pourazad, K.Brunnstrom, K.Wang and B.Andren
In Proceedings of the 6th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM), January 2012.

Fast and flexible Kullback-Leibler divergence based acoustic modeling for non-native speech recognition.
D. Imseng, R. Rasipuram, and M.Magimai.-Doss,
In Proc. of IEEE Automatic Speech Recognition and Understanding Workshop (ASRU'11), Hawaii, December 14, 2011.

User requirements for meeting support technology.
D.Lalanne, and A.Popescu-Belis
In Multimodal Signal Processing: Human Interactions in Meetings, p.210-221, Cambridge University Press, 2011.

Meeting browsers and meeting assistants.
S. Whittaker, S. Tucker, D. Lalanne
In Multimodal Signal Processing, Human Interactions in Meetings, Cambridge University Press, 2011.

IM2, c/o Idiap Research Institute, Centre du Parc,
Rue Marconi 19, P.O. Box 592, 1920 Martigny,
info@idiap.ch - www.idiap.ch