**IDIAP**

| # | Title | Authors | Abstract |
|---|-------|---------|----------|
| 1 | Iterative Relevance Feedback with Adaptive Exploration/Exploitation Trade-off | *Nicolae Suditu and Francois Fleuret* | Content-based image retrieval systems have to cope with two different regimes: understanding broadly the categories of interest to the user, and refining the search in this or these categories to converge to specific images among them. Here, in contrast with other types of retrieval systems, these two regimes are of great importance since the search initialization is hardly optimal (i.e. the page-zero problem) and the relevance feedback must tolerate the semantic gap of the image's visual features. We present a new approach that encompasses these two regimes, and infers from the user actions a seamless transition between them. Starting from a query-free approach meant to solve the page-zero problem, we propose an adaptive exploration/exploitation trade-off that transforms the original framework into a versatile retrieval framework with full searching capabilities. Our approach is compared to the state-of-the-art it extends by conducting user evaluations on a collection of 60,000 images from the ImageNet database. |
| 2 | Microphone Array Beampattern Characterization for Hands-free Speech Applications | *Mohammad J. Taghizadeh, Philip N. Garner, Hervé Bourlard* | Spatial filtering is the fundamental characteristic of microphone array based signal acquisition, which plays an important role in applications such as speech enhancement and distant speech recognition. In the array processing literature, this property is formulated upon beam-pattern steering and it is characterized for narrowband signals. This paper proposes to characterize the microphone array broadband beam-pattern based on the average output of a steered beamformer for a broadband spectrum. Relying on this characterization, we derive the directivity beam-pattern of delayand- sum and superdirective beamformers for a linear as well as a circular microphone array. We further investigate how the broadband beam-pattern is linked to speech recognition feature extraction; hence, it can be used to evaluate distant speech recognition performance. The proposed theory is demonstrated with experiments on real data recordings. |
| 3 | Comparing different acoustic modeling techniques for multilingual boosting | *David Imseng, John Dines, Petr Motlicek, Philip N. Garner, Hervé Bourlard* | In this paper, we explore how different acoustic modeling techniques can benefit from data in languages other than the target language. We propose an algorithm to perform decision tree state clustering for the recently proposed Kullback-Leibler divergence based hidden Markov models (KL-HMM) and compare it to subspace Gaussian mixture modeling (SGMM). KL-HMM can exploit multilingual information in the form of universal phoneme posterior features and SGMM benefits from a universal background model that can be trained on multilingual data. Taking the Greek SpeechDat(II) data as an example, we show that KL-HMM performs best for small amounts of target language data. |

| 4 | Structured Sparse Coding for Microphone Array Location Calibration | *Afsaneh Asaei, Bhiksha Raj, Hervé Bourlard, Volkan Cevher* | We address the problem of microphone location calibration where the sensor positions have a sparse spatial approximation on a discretized grid. We characterize the microphone signals as a sparse vector represented over a codebook of multi-channel signals where the support of the representation encodes the microphone locations. The codebook is constructed of multi-channel signals obtained by inverse filtering the acoustic channel and projecting the signals onto a array manifold matrix of the hypothesized geometries. This framework requires that the position of a speaker or the track of its movement to be known without any further assumption about the source signal. The sparse position encoding vector is approximated by model-based sparse recovery algorithm exploiting the block-dependency structure underlying the broadband speech spectrum. The experiments conducted on real data recordings demonstrate the effectiveness of the proposed approach and the importance of the joint sparsity models in multi-channel speech processing tasks. |
|---|---|---|---|
| 5 | Acoustic Data-Driven Grapheme-to-Phoneme Conversion using KL-HMM | *Ramya Rasipuram* | At Idiap, we are currently investigating the use of graphemes as subword units for speech recognition in the framework of Kullback-Leibler divergence-based hidden Markov model (KL-HMM). In this poster, we present a novel acoustic data driven grapheme-to-phoneme (G2P) conversion approach based on grapheme-based KL-HMM system.<br>The probabilistic relationship between graphemes and phonemes learned through acoustic data (using KL-HMM) is used along with the orthographic transcription of words to infer the phoneme sequence. One of the main application of the proposed G2P conversion approach is in the area of automatic speech recognition or text-to-speech systems for low linguistic resource (under-resource) domains/languages.<br>We demonstrate the potential of the approach on the development of pronunciation lexicon for under-resourced tasks/domains. |
| 6 | Collective Learning to Rank on Network Data | *Majid Yazdani, Ronan Collobert, Andrei Popescu-belis* | This poster proposes a method for learning to rank on network or relational data, which makes use of the features of the nodes as well as the existing links between them. First, a neighborsaware ranker is trained using a pairwise loss function. Then, collective inference is performed using a sparse iterative ranking algorithm, which propagates the results of rankers over the network. The method is applied to three data sets with papers/citations and webpages/hyperlinks. The results show that the proposed algorithm, using both link structure and node attributes, outperforms several other methods: a content only ranker, a link-only one, an unsupervised random walk method, a relational topic model, and a method based on the weighted number of common neighbors. In addition, the propagation algorithm improves results even when no prior link structure is known, and scales efficiently to large networks. |

## IP1: Integrated Multimodal Processing

| 7 | Mid-air hand gesture human-machine interaction: design, recognition and evaluation | Matthias Schwaller, Denis Lalanne | This poster presents research works on mid-air hand gesture human-machine interaction, which is the topic of a PhD thesis at the university of Fribourg. Research activities include the design, development and evaluation of mid-air hand gestures recognizers with different kind of gestures. Further, the use of various forms of feedback is considered to augment usability and precision, and also to recover from recognition errors. Finally, several evaluations are presented, considering not only the performance of the recognizers, but also usability and effort metrics, so that gestures can be performed during a long time period, avoiding movements tiring or uncomfortable for users. |
|---|---|---|---|

| 8 | Learning a Skeleton Distribution | *Ke Sun* | The representative samples can be pictured as the skeleton of a point cloud. We learn a discrete distribution defined over all samples, so that these skeleton points have large probabilities and the outliers have probabilities close to zero. The basic assumption is that any observation is generated from a nearby skeleton point. The learning objective is to minimize the communication cost from a random sample to its generation source. Experiments show that the learned distribution highlights a compact size of key positions. It is further applied to a denoising task as an indirect method of evaluation. The clustering structures of image datasets are best preserved among several methods investigated. |
| --- | --- | --- | --- |
| 9 | Social Media Enrichment | *Ivan Ivanov, Touradj Ebrahimi* | Social networks and social media sharing websites have become increasingly popular in recent years. They allow users to store, share, search and consume social media very easily. In this poster, we present two approaches on how social media can be enhanced in an efficient way through tag propagation and trust modeling. In many cases, content shared online is tagged manually. One way to make this time-consuming manual tagging process more efficient is to propagate tags from a small set of tagged images to the larger set of untagged images automatically by making use of the similarity between image content (e.g., famous landmarks) and its context (e.g., associated geotags). In such scenario, however, a wrong or a spam tag can damage the integrity and reliability of the automated propagation system. Therefore, for reliable tag propagation, we suggest adopting user trust model based on a social feedback from the users of the social network. The effectiveness of the proposed methods for tag propagation and user trust modeling is demonstrated through a set of experiments on datasets of still images depicting various classes of objects. |

## IP2: Human Centered Design & Evaluation

| 10 | Using Crowdsourcing to Compare Document Recommendation Strategies for Conversations | Maryam Habibi | This paper explores a crowdsourcing approach to the evaluation of a document recommender system intended for use in meetings. The system uses words from the conversation to perform just-in-time document retrieval. We compare several versions of the system, including the use of keywords, retrieval using semantic similarity, and the possibility for user initiative. The system's results are submitted for comparative evaluations to workers recruited via a crowdsourcing platform, Amazon's Mechanical Turk. We introduce a new method, Pearson Correlation Coefficient-Information Entropy (PCC-H), to abstract over the quality of the workers' judgments and produce system-level scores. We measure the workers' reliability by the inter-rater agreement of each of them against the others, and use entropy to weigh the difficulty of each comparison task. The proposed evaluation method is shown to be reliable, and demonstrates that adding user initiative improves the relevance of recommendations. |
| --- | --- | --- | --- |
| 11 | Mood feedback and its influence on teamwork in different interactive work environments | *Andreas Sonderegger, Luisa Bergholz, Denis Lalanne, Fabien Ringeval & Jürgen Sauer* | The EmotiBoard is a vertical interactive wall on which team members can interact in co-presence or remotely and which provides a visual feedback of each team member's emotional state. This poster examines the influence of mood feedback on different outcomes of teamwork in two different collaborative work environments. Employing a 2 x 2 between subjects design, mood feedback (present vs. not present) and work environment (face-to-face vs. distributed teams) were manipulated experimentally. Sixty four participants completed three different tasks, measuring emotions, performance, workload, and team satisfaction. Although almost 40% of the participants reported that they did not consider the mood feedback, the results indicated that the evaluation of the other team members' emotional state was more accurate when the mood feedback was presented. Furthermore, performance was higher in the face-to-face condition compared to the distributed team condition. Similar effects were reported for subjective measures of workload. Implications for the use of emotion feedback in different collaborative work environments are discussed. |

| 12 | The Communication Board: an applicative framework to study user's affective and social behaviors in a context of computer-mediated collaboration | *Fabien Ringeval, Andreas Sonderreger, Basilio Noris, Aude Billard, Juergen Sauer, Denis Lalanne* | The Communication Board is composed of a vertical surface on which users can interact remotely and where an emotional feedback of other participant is displayed on the screen. The EmotiBoard application is an applicative framework that uses IM2 technologies to: (i) Analyse user's affective and social behaviors in the context of remote collaboration and (ii) Elicit factors influencing collaboration, both to facilitate user evaluations and to set up real time technologies supporting computer-mediated collaboration. The poster presents the set up of the EmotiBoard and the work achieved through this framework with two distinct experiences: (i) EmotiBoard2, which consists in estimating how emotion perception and gaze of a person might be influenced by an emotional feedback and (ii) EmotiBoard3, which aims at characterizing both affective and social behaviors from multimodal recordings (audiovisual and physiological) in a dyadic collaboration scenario. |
| --- | --- | --- | --- |
| 13 | Exploring the Design of Context-sensing Display to Support Opportunistic Search in Meetings | *Nan Li, Pierre Dillenbourg* | In meetings, people sometimes happen to come across information e.g. in conversation or meeting documents. This kind of accidentally encountered information may be or not be relevant to the main meeting topic, but stirs people's temporary interest and mediates group conversation. In some cases people even need to search these unexpected terms in the Web. This phenomenon is called opportunistic search. We question if calm technologies can be designed to support opportunistic search in meetings. Assuming information sharing in collocated meetings may largely come from conversation and collaboration context, we propose to design an semi-ambient display that provides just-in-time relevant meeting information to support opportunistic browsing and searching. |
| 14 | 60 Seconds Minutes | *Andrea Mazzei, Himanshu Verma, Flaviu Roman, Frederic Kaplan, Pierre Dillenbourg* | We present a meetings support system which processes participants-generated content to create meeting minutes in the form of comic strips. We created a iOS client application that participants use during the meeting to collect photos of faces, documents, notes or any other relevant element. A server receives all these these visual corpora and extract relevant features from them and segments the meeting timeline based on the intensity of the voice signal of the meeting. Finally a web-based application supports the meeting secretary in preparing the layout for the creation of a comic strip document of the meeting, which is then disseminated to participants. |

### IP3: Social Signal Processing

| 15 | Automatic detection of conflict escalation in multi-party conversation | *Samuel Kim, Sree Harsha Yella and Fabio Valente* | We study conflict escalation in multi-party conversations which refers an increase in the intensity of conflict during conversations. Specifically, we focus on automatic detection of conflict escalation in broadcasted political debates toward a machine mediated conflict management system. We introduce a fully automated feature extraction method equipped with an automatic speaker diarization algorithm rather than manual segmentation. Empirical results to classify the types of conflict escalation, i.e., escalation, de-escalation, and constant, show that we can obtain 45.9% weighted accuracy along with the automatically extracted features, which is significantly higher than a chance level. |

| 16 | Speaker Diarization of Overlapping Speech based on Silence Distribution in Meeting Recordings | *Sree Harsha Yella, Fabio Valente* | Speaker diarization of meetings can be significantly improved by overlap handling. Several previous works have explored the use of different features such as spectral, spatial and energy for overlap detection. This paper proposes a method to estimate probabilities of speech and overlap classes at a segment level which are later incorporated into an HMM/GMM baseline system. The estimation is motivated by the observation that significant portion of overlaps in spontaneous conversations take place where the amount of silence is less, e.g., during speaker changes.<br>Experiments on the AMI corpus reveal that the probability of occurrence of overlap in a segment is inversely proportional to the amount of silence in it. Whenever this information is combined with acoustic information from MFCC features in an HMM/GMM overlap detector, improvements are verified in terms of F-measure. Furthermore the paper investigates the use of exclusion and labelling strategies based on such detector for handling overlap in diarization reporting F-measure improvements from 0.29 to 0.43 in case of exclusion and from 0.15 to 0.22 in case of labelling. Consequently speaker diarization error is reduced by 8\% relative compared to the baseline based solely on acoustic information. |

## IM2 OTHERS

| 17 | Distributed Metric Inverted files | *Hisham Mohamed and Stéphane Marchand-Maillet* | We present parallel strategies for indexing and searching permutation-based indexes for high dimensional data using inverted files. The techniques are parallelized using MapReduce MPI (MRO-MPI). We study performance, efficiency, and effectiveness of our strategies on high dimensional datasets of millions of images. Experimental results show a good performance compared to the sequential version with the same efficiency and effectiveness. |