Scientific Presentations - Abstracts list

Monday, August 31st

Session 1 – Chairman: Jean-Marc Odobez

Title Quality Measures and Stacking Classifiers in Multimodal Biometric Recognition
Speaker Andrzej Drygajlo (EPFL)
Schedule 16:30 – 17:00
Abstract Existing approaches to biometric person recognition with quality measures make a clear distinction between the single-modality applications and the multi-modal scenarios. This study bridges this gap with Q-stack, a stacking-based classifier ensemble, which uses the class-independent signal quality measures and baseline classifier scores in order to improve the accuracy of uni- and multi-modal biometric classification. The seemingly counterintuitive notion of using class-independent quality information for improving class separation is explained. The authors present Q-stack as a generalised framework of classification with quality information, including age as a metadata quality measure. The authors further demonstrate the application of Q-stack on the task of biometric identity verification using face and other modalities, and show that the use of the proposed technique allows a systematic reduction of the error rates below those of the baseline classifiers, in scenarios involving single and multiple classifiers and multi-modalities.

Title Parts-based face verification using local frequency bands
Speaker Chris McCool (IDIAP)
Schedule 17:00 – 17:30
Abstract We extend the Parts-Based approach of face verification by performing a frequency-based decomposition. The Parts-Based approach divides the face into a set of blocks which are then considered to be separate observations, this is a spatial decomposition of the face. This work extends the Parts-Based approach by also dividing the face in the frequency domain and treating each frequency response from an observation separately. This can be expressed as forming a set of sub-images where each sub-image represents the response to a different frequency of, for instance, the Discrete Cosine Transform. Each of these sub-images is treated separately by a Gaussian Mixture Model (GMM) based classifier. The classifiers from each sub-image are then combined using weighted summation with the weights being derived using linear logistic regression. It is shown on the BANCA database that this method improves the performance of the system from an Average Half Total Error Rate of 24.38% to 15.17% when compared to a GMM Parts-Based approach on Protocol P.

Title Dealing with asynchrony in audio-visual speech recognition
Speaker Virginia Estellers (EPFL)
Schedule 17:30 – 18:00
Abstract We present two different approaches to compensate for asynchrony in audio-visual speech recognition: using multimodal classifiers with different degrees of asynchrony or processing the data to align the audio and visual streams for a simple synchronous classifier. First, we propose a new asynchronous model and compare its performance to existing ones. The improvement obtained suggests the necessity of a processing method reducing the effects of stream asynchrony when a synchronous classifier is used. Both the model and the processing step outperform traditional audio-visual speech recognition systems in experiments with the CUAVE database.

Title Nonverbal small-group characterization: classification and mining
Speaker Dinesh Babu Jayagopi (IDIAP)
Schedule 18:00 – 18:30

Abstract Characterizing group conversations using nonverbal behaviour is a key problem in human interaction modelling, with applications related to browsing and retrieval of specific conversations where certain types of behaviours are exhibited. Modelling group interaction is challenging both in social science and in computing, where methods for understanding group conversations from audio and visual nonverbal cues have started to become popular, motivated by the fact that nonverbal behaviours carry a wealth of information about the group members' relationships. In this talk, we discuss two different approaches to model nonverbal aspects of small-group interaction using a classification and a mining framework.

Session 2 – Chairman: François Fleuret

Title Semi-supervised learning for handwriting recognition
Speaker Volkmar Frinken (UniBern)
Schedule 16:30 – 17:00
Abstract The talk will be about Semi-Supervised Learning in Handwriting recognition and will actually not differ that much from the IM2 presentation earlier this year in Lausanne. I will explain the idea of Semi-Supervised Learning in general and Self-Learning in particular and the recognition system (the features and the recognizer). The main focus of the talk will be about different retraining rules, meaning which words are selected for retraining. Experiments with two types of recognition systems are presented (open vocabulary and closed vocabulary recognition) and the effect of the retraining rules of these two systems.

Title The Hub and ezHub API
Speaker Mike Flynn (IDIAP)
Schedule 17:00 – 17:30
Abstract The Hub is a data distribution and storage mechanism, suitable for use in projects where annotation data is needed in real-time, or from the past. The ezHub is a simple, elegant and efficient Java API to produce and consume Hub data, with no need to know about the intricacies of connection, data-formats, error-handling, threading issues, etc.

Title Mining query logs with topic models
Speaker Donn Morrison (UniGE)
Schedule 17:30 – 18:00
Abstract This research details the application of topic models for query log mining. Working in the context of a query-by-example retrieval system, we formalise a User Relevance Model which posits that users make relevance judgements based on the existence (or lack thereof) of concepts in both queries and documents. Latent variable models such as the singular value decomposition (SVD) and non-negative matrix factorisation (NMF), among others, are unsupervised statistical learning methods that decompose a matrix of co-occurrences into a product of component matrices. They facilitate a projection of the observations into a lower dimensional space by representing queries and documents as linear combinations of topics. We demonstrate how these topics can be extracted from query logs and how they can be used to improve retrieval and propagate meta-data.

Title Hough Transform-based Mouth Localization for Audio-Visual Speech Recognition
Speaker Gabriele Fanelli (ETHZ)
Schedule 18:00 – 18:30
Abstract We present a novel method for mouth localization in the context of multimodal speech recognition where audio and visual cues are fused to improve the speech recognition accuracy. While facial feature points like mouth corners or lip contours are commonly used to estimate at least scale, position, and orientation of the mouth, we propose a Hough transform-based method. Instead of relying on a predefined sparse subset of mouth features, it casts probabilistic votes for the mouth centre from several patches in the neighbourhood and accumulates the votes in a Hough image. This makes the localization more robust as it does not rely on the detection of a single feature. In addition, we exploit the different shape properties of eyes and mouth in order to localize the mouth more efficiently. Using the rotation invariant representation of the iris, scale and orientation can be efficiently inferred from the localized eye positions. The superior accuracy of our method and quantitative improvements for audio-visual speech recognition over monomodal approaches are demonstrated on two datasets.