Poster Session - Abstract list

Wednesday, September 3

Summer Institute 2008
IM2 & Affective Sciences
September 1-3, 2008
Riederalp

IM2.AP, Audio Processing			
1	Agglomerative Information Bottleneck for Speaker Diarization of Meetings Data	*Deepu Vijayasenan*	In this work, we investigate the use of agglomerative Information Bottleneck (aIB) clustering for the speaker diarization task of meetings data. In contrary to the state-of-the-art diarization systems that models individual speakers with Gaussian Mixture Models, the proposed algorithm is completely non parametric . Both clustering and model selection issues of non-parametric models are addressed in this work. The proposed algorithm is evaluated on meeting data on the RT06 evaluation data set. The system is able to achieve Diarization Error Rates comparable to state-of-the-art systems at a much lower computational complexity.
2	Confidence Measures for Recognition Error Detection in LVCSR	*Petr Motlicek*	The poster presents some techniques for detection of recognition errors and detection of out-of-vocabulary words in large vocabulary speech recognition systems. These techniques are based on confidence measures and combination of different classifiers. All results are presented on Wall Street Journal task with reduced recognition vocabulary.
3	Hilbert Envelope Based Features for Far-Field Speech Recognition	*Samuel Thomas, Sriram Ganapath*	Automatic speech recognition (ASR) systems, trained on speech signals from close-talking microphones, generally fail in recognizing far-field speech. In this paper, we present a Hilbert Envelope based feature extraction technique to alleviate the artifacts introduced by room reverberations. The proposed technique is based on modeling temporal envelopes of the speech signal in narrow sub-bands using Frequency Domain Linear Prediction (FDLP). ASR experiments on far-field speech using the proposed FDLP features show significant performance improvements when compared to other robust feature extraction techniques (average relative improvement of 43% in word error rate).
4	Hilbert Envelope Based Spectro-Temporal Features For Phoneme Recognition	*Sriram Ganapathy*	This work investigates the use of spectro-temporal features extracted from the Hilbert envelope of the signal for ASR. Sub-band Hilbert envelopes of the speech signal are estimated using Frequency Domain Linear Prediction (FDLP).
5	Exploiting contextual information for speech/non-speech detection	*Sree Hari Krishnan Parthasarathi, Petr Motlicek, Hynek Hermansky*	We present a data-driven approach to weighting the temporal context of signal energy to be used in a simple speech/non-speech detector (SND). The optimal weights are obtained using linear discriminant analysis (LDA). Regularization is performed to handle numerical issues inherent to the usage of correlated features. The discriminant so obtained is interpreted as a filter in the modulation spectral domain. Experimental evaluations on the test data set, in terms of average frame-level error rate over different SNR levels, show that the proposed method yields an absolute performance gain of 10.9 %, 17.5%, 7.9% and 8.3% over ITU's G.729B, ETSI's AMR1, AMR2 and a state-of-the-art multi-layer perceptron based system, respectively. This shows that even a simple feature such as full-band energy, when employed with a large-enough context, shows promise for applications.
IM2.BMI, Brain machine interfaces			
6	Characterizing EEG correlates of exploratory behavior.	*Nicolas Bourdaud, Ricardo Chavarriaga, José del R. Millãn*	In order to study the exploratory behavior using EEG, an N-arm bandit paradigm previously used in other studies which have shown, using fMRI techniques, bilateral activation in the frontal and parietal cortex during exploration, has been used and adapted to EEG recording. Up to our knowledge, no previous study has been done on it using EEG. A model of how subjects make their decision has been built and the label of each trials derived from it. Because of the complexity of the task, the EEG correlates of the exploratory behavior are not necessarily time locked with the action. So the EEG processing methods used should be designed in order to handle signals that can shift in time across trials. Assuming that the non-overlapping regions of the probability

			distributions of the EEG signal over time in the two conditions (exploration, exploitation) correspond to time samples where the characteristic phenomena for each condition takes place, we are able to identify peaks of EEG activity that is not necessarily time-locked with any time clue but that are relevant to each condition. The results show that the bilateral frontal and parietal areas are also the most discriminant in the EEG activity which strongly suggests that the EEG signal conveys also the information about the exploratory behavior. Classification of single trial has also been done and has shown to provide best accuracies in the low frequencies, below 23 Hz. In addition, classification based on combined low frequency bands provides better performances than classification based on single band. This suggests that the information in spread in the frequency domain.
7	Probabilistic human-robot interaction: Navigating with uncertain user interfaces.	*Xavier Perrin*	Our motivation is to develop a driver assistant using a new kind of interaction between the human and the machine. In our proposed semi-autonomous navigation system, the human monitors the suggestions of travel direction made by an autonomous robot. He is then involved in the control loop only when he disagrees with the robot's proposition by providing it with an error signal. In this poster, we recall our experiments about the different types of feedback the robot could use for proposing an action and their influence on the recognition of the error signal from the brain activity (EEG processing). Then, we present the robotic controller and its different components, relying on Bayesian reasoning techniques in order to overcome the uncertainty related to sensory information and human answers. An experiment in simulation shows the robustness of our approach to different user interface's accuracy. A real robot is also used for an indoor experiment, demonstrating the validity of our approach.
8	Recent Progress on Single-Trial Recognition of Error-Related Potentials	*Pierre W. Ferrez,* *José del R. Millãn*	Brain-computer interfaces (BCIs) are prone to errors in the recognition of the subject's intent. In this study, we show the feasibility of simultaneously classifying motor imagery for BCI control and detecting error-related potentials (ErrP) to filter out erroneous commands in a real-time system. We also show the potential benefit of using inverse solutions such as the Cortical Current Density (CCD) inverse model to improve ErrP detection. Two healthy volunteer subjects participated in real-time BCI experiments where they were mentally moving a cursor using motor imagery (imagination of a movement of the left hand for "Left" and of the right foot for "Right") and where the system was simultaneously canceling the movement if the presence of ErrP was detected in a short window following the movement of the cursor. The system was using a 1 second window to classify motor imagery and a 400 ms window just after the movement of the cursor to detect the presence of ErrP. For both subjects, the BCI error rate without the use of ErrP detection to filter out erroneous commands is just above 30%. This error rate drops to 7% on average when ErrP detection is integrated. The average ErrP detection is around 80% for both subjects. Finally, the bit rate is multiplied by 3 with the integration of ErrP detection. Ten subjects participated in further experiments showing that ErrP classification is significantly better with the CCD inverse model than with EEG. Indeed, the average classification rate is 80.8% for EEG and 84.5% for the CCD inverse model. Furthermore, the most relevant solution points of the CCD model for ErrP detection are localized in compact clusters partially localized on the anterior cingulate cortex (ACC) and on the pre-supplementary motor area (pre-SMA).
9	Towards anticipation based Brain-Computer Interfacing (aBCI)	*Garipelli Gangadhar,* *Ricardo Chavarriaga,* *José del R. Millãn*	Anticipation increases the efficiency of a daily task by partial advanced activation of neural substrates involved in it. Single trial recognition of this activation can be exploited for a novel anticipation based Brain-Computer Interfacing (aBCI). The current poster describes the general framework and how electroencephalogram (EEG) correlates of anticipation to relevant stimuli can be used to decode the human intention in real time. We first recorded an anticipation related potentials using classical Contingent Negative Variation (CNV) paradigm using GO and NOGO conditions with 9 subjects and develop techniques for single-trial recognition. Our offline analysis showed classification rates on average of 67% with a increasing trend over sessions, indicating subject's learnability of the anticipation task. Secondly, we developed a technique called Time Aggregation of Classification (TAC), for the fast recognition of these potentials. With TAC, we not only improve the classification accuracy but also achieve fast decisions. Thirdly, we report online experiments with two subjects. The preliminary online experiments show single trial accuracies up to 80% in some sessions for both subjects with an average of 70% and 67%, respectively.
IM2.DMA, Database management and meeting analysis			
10	Objective quality evaluation of color images by perceptual weighting of single-channel metrics.	*Francesca De Simone,* *Touradj Ebrahimi*	We present a new approach for the design of a full reference objective quality metric for the assessment of color pictures. Our goal is to build a multi-channel metric based on the perceptual weighting of single-channel metrics. A psycho-visual experiment is thus designed in order to determine the values of the weighting factors. This metric is expected to provide a new useful tool for the quality assessment of compressed pictures in the framework of codec performance evaluation.

11	The Hub	*Mike Flynn,* *Alexandre Nanchen*	The Hub is a real-time data distribution and storage mechanism, intended to support the processing, recording and playback of annotation data in IM2. It enables multiple "producers" to send data to multiple "consumers", in real time, and have the data recorded for future browsing. A layered set of APIs allow easy access to the data. Recent improvements are explained.
12	Topickr: Flickr Groups and Users Reloaded	*Radu-Andrei Negoescu,* *Daniel Gatica-Perez*	With the increased presence of digital imaging devices there also came an explosion in the amount of multimedia content available online. Users have transformed from passive consumers of media into content creators. Flickr.com is such an example of an online community, with over 2 billion photos (and more recently, videos as well), most of which are publicly available. The user interaction with the system also provides a plethora of metadata associated with this content, and in particular tags. One very important aspect in Flickr is the ability of users to organize in self-managed communities called groups. Although users and groups are conceptually different, in practice they can be represented in the same way: a bag-of-tags, which is amenable for probabilistic topic modeling. We present a topic-based approach to represent Flickr users and groups and demonstrate it with a web application, Topickr, that allows similarity based exploration of Flickr entities using their topic-based representation, learned in an unsupervised manner.
13	What Did You Do Today? Discovering Daily Routines from Large-Scale Mobile Data	*Katayoun Farrahi,* *Daniel Gatica-Perez*	We present a framework built from two Hierarchical Bayesian topic models to discover human location-driven routines from mobile phones. The framework uses location-driven bag representations of people's daily activities obtained from celltower connections. Using 68 000+ hours of real-life human data from the Reality Mining dataset, we successfully discover various types of routines. The first studied model, Latent Dirichlet Allocation (LDA), automatically discovers characteristic routines for all individuals in the study, including “going to work at 10am", “leaving work at night", or “staying home for the entire evening". In contrast, the second methodology with the Author Topic model (ATM) finds routines characteristic of a selected groups of users, such as ``being at home in the mornings and evenings while being out in the afternoon", and ranks users by their probability of conforming to certain daily routines.
14	Multi-Eval: an evaluation framework for multimodal dialogue annotations	*Paula Estrella,* *Andrei Popescu-Belis*	This poster presents the first version of Multi-Eval, a new framework for the evaluation of multimodal dialogue annotations. Multi-Eval offers the possibility to compare two annotations (either proposed by IM2 project or uploaded by users) according to some standard evaluation metrics. In this scenario, the users select two available annotations, choose one or more evaluation metrics and receive in return the score obtained by comparing one annotation against another.
15	An RFID-based infrastructure for recording, archiving, and indexing meetings	*Nicolas Pittet, Denis Lalanne*	This poster describes an RFID-based infrastructure for recording meetings with multiple cameras and microphones, compressing and archiving the recordings automatically, and for indexing at low cost meetings through online annotations.
IM2.HMI, Human-machine interaction			
16	The HephaisTK Multimodal Interfaces Creation Toolkit : Architecture and Scripting Language	*Bruno Dumas,* *Denis Lalanne,* *Rolf Ingold*	This poster presents HephaisTK, a project which targets (a) the development of novel multimodal fusion mechanisms and (b) the creation of an open-source toolkit allowing the rapid creation of multimodal interfaces. This poster hence presents the current version of HephaisTK, an agent-based framework dedicated to the creation of multimodal interfaces and of SMUIML, the language used to script the framework. Finally, it brushes the future plans of this work.
17	Personal Information management through interactive visualization	*Florian Evequoz*	Personal information (PI) overload is an issue everyone has to cope with. While existing PI management approaches mainly rely on searching mechanisms or semantic tagging, we believe that much insight into our PI can be gained by providing visual browsing tools making use of the similarity links between different pieces of PI and some chosen facets that we consider crucial for supporting our biographic and documentary memory: its temporal, social and thematic facets. An early user-requirements survey that we conducted tends to confirm this tendency. Our system, called WotanEye, extracts the metadata related to the social and thematic facets of emails, that we consider representative of the structure of one’s PI, and links them to all the other documents and event records available in one’s personal digital memories. Moreover, visualization techniques are proposed for browsing PI using dynamic query refinement strategies over the chosen facets. WotanEye allows browsing emails, digital documents and agenda items in parallel through various facets (i.e. temporal, social, thematic) used in conjunction. Finally, user evaluation issues are discussed along with our future plans to assess our visual and holistic approach.

17A	Task-based Evaluation of Meeting Browsers: from BET Task Elicitation to User Behavior Analysis	*Andrei Popescu-Belis, Philippe Baudrion*	This poster presents the results of the application of the task-based Browser Evaluation Test (BET) to meeting browsers, that is, interfaces to multimodal databases of meeting recordings.
IM2.MCA, Multimodal context abstraction			
18	Sparsity and high dimensionality in image retrieval	*Donn Morrison, Enikö Szekely*	The goal of this research is to analyze sparsity and high dimensionality for both low- and high-level features in order to improve image retrieval. High-level (semantic) features are captured through relevance feedback ``sessions'' which describe a query in terms of images which can be relevant, irrelevant, or neutral with respect to the underlying query concept (what the user has in mind). By collecting many of these sessions over a period of time, a semantic space can be constructed over the documents. High dimensionality affects retrieval accuracy because of the sparsity of such spaces where distances tend to be equal. We provide theoretical and practical results for both long-term user interaction and dimension reduction.
19	Image Content Annotation: Bottom-up Visual Attention Model as a Link between Content Analysis and Textual Tags	*Marija Uscumlic*	Although the multimedia content analysis techniques have shown to be effective for automatic analysis of multimedia data, it still has a difficulty in handling semantic high-level information of these data automatically. This open issue is called the semantic gap. Nowadays, the success of social networking has inspired new strategies in solving this problem. In particular, the semantic gap can be bridged combining results of content analysis with the information that users provide in social network environments. We focus on extraction of interesting semantic parts in images that are likely to be tagged by the users. We use a bottom-up attention model which has been shown to be successful in predicting the locations of interesting objects in images. The object that a user tags can be considered as a good approximation of an interesting object. This is the link between textual tags and content analysis that we concentrate on. By using the bottom-up attention model in the image segmentation process, relevant parts can be extracted for further processing. Then, the duplicate detection method is applied in order to group images with similar contents. In future work, automatic merging of textual tags is planned to be performed based on the grouping results.
20	Role Recognition in Multiparty Recordings	*Sarah Favre, Hugues Salamin, Alessandro Vinciarelli*	This poster presents two different approaches for the recognition of roles in multiparty recordings. The first approach uses only the interactions between the people as source of information while the second approach combines the lexical choices made by people playing different roles and the Social Networks describing the interactions between the persons. The experiments for the first approach have been performed over several corpora, including broadcast data and meeting recordings, for a total of roughly 90 hours of material. The results are satisfactory for the broadcast data (from 74 to 87 percent of the data time correctly labeled in terms of role), while they still must be improved in the case of the meeting recordings (around 45 percent of the data time correctly labeled). The experiments for the second approach have been performed over a corpus of 138 meeting recordings (over 45 hours of material) and show that around 70 percent of the time is labeled correctly in terms of role. Both approaches outperform significantly chance.
21	World-scale Mining of Objects and Events from Community Photo Collections	*Till Quack, Bastian Leibe, Luc Van Gool*	In this paper, we describe an approach for mining images of objects (such as touristic sights) from community photo collections in an unsupervised fashion. Our approach relies on retrieving geotagged photos from those web-sites using a grid of geospatial tiles. The downloaded photos are clustered into potentially interesting entities through a processing pipeline of several modalities, including visual, textual and spatial proximity. The resulting clusters are analyzed and are automatically classified into objects and events. Using mining techniques, we then find text labels for these clusters, which are used to again assign each cluster to a corresponding Wikipedia article in a fully unsupervised manner. A final verification step uses the contents (including images) from the selected Wikipedia article to verify the cluster-article assignment. We demonstrate this approach on several urban areas, densely covering an area of over 700 square kilometers and mining over 200,000 photos, making it probably the largest experiment of its kind to date.

IM2.MPR, Multimodal Processing and recognition			
22	Using Haar LBP features for Fast Illumination Invariant Face Detection	*Anindya Roy, Sébastien Marcel*	Face Detection is the first step in many visual processing systems like Face Recognition, Emotion Recognition, Lip Reading etc., and hence is of paramount importance in today's world. In this work, we have endeavored to combine Haar features with Local Binary Patterns, two widely used and successful concepts in Vision tasks, to design a Face Detection system which has the advantages of both. We have obtained preliminary results showing the superiority of our method over usual Haar features proposed by Viola and Jones.
23	Associating Audio-Visual Activity Cues in a Dominance Estimation Framework	*Hayley Hung*	We address the problem of both estimating the dominant person in a meeting from a single audio source and identifying them visually in a multi-camera setting. We use a speaker diarization algorithm to perform speaker segmentation and clustering, representing when they spoke. Using a greedy ordered audio-visual association algorithm, we investigate using the speaker clusters to find the corresponding person in one of the video channels. The difficulty of the problem is that firstly the speaker diarization output is noisy (e.g. for participants who speak little) and often produces an unequal number of clusters to true participants. Secondly, personal visual activity from natural upper torso motion, which can include highly deformable pose changes and perspective distortion, is computed through computationally efficient coarse features. Our results using almost 2 hours of audio-visual data from 4-participant meetings show a strong correlation between the estimated speaker diarization and visual activity features, enabling the identification of the most dominant person as a pair of audio-visual channels.
24	Classifying multivariate time series with static Bayesian networks: a phase space perspective	*Jonas Richiardi*	Using probabilistic models for time series offers advantages over steady-state signal analysis techniques if the signals under consideration are not deterministic, but stochastic, as is often the case in pattern recognition of real-world signals. We can obtain models describing the amount of spread in the random variables, meaning that we can take into account the uncertainty on the realisation of the random variables (features) due to intra-class variability. Based on dynamical/chaotic systems theory we show empirically that computing derivatives in time of the multivariate signal yields a higher-dimensional representation that allows for better classification than using the base signals alone. We also show that using derivative coordinates for the phase space has several advantages over the more often used delay coordinates, including better adequation to multivariate time series.
24 A	Object of Interest Detection using Object and Gaze Direction Detection	*Basilio Noris, François Fleuret, Aude Billard*	This work addresses the problem of identifying the object of interest in a speaker/listener scenario. This is done using Color and Shape – based object detection to identify a number of objects appearing in the environment and by applying Gaze detection to identify which of those objects are in the user focus of attention.
IM2.VP, Visual/Video Processing			
25	Vision-supported speech understanding	*Beat Pfister, Gabriele Fanelli, Beat Fasel, Jean-Philippe Thiran*	Speech-based dialog systems can drastically be impaired in acoustically adverse conditions. In particular other voices and non-stationary noise may cause problems. The robustness of such systems can be improved by means of visual information along two axes: On the one hand optical user tracking allows e.g. to distinguish between the user's and other voices. On the other hand information of the acoustic channel may be improved by visual information such as the lips movement and thus improve speech recognition. The aim of this project is to demonstrate the benefit for speech recognition and dialog processing from the smart use of the vision channel. As demonstration scenario we use an publicly accessible information kiosk where users can participate in a voice imitation contest.
26	Minimum distance between pattern transformation manifolds: Algorithm and Applications	*Effrosyni Kokiopoulou, Pascal Frossard*	Transformation invariance is an important property in pattern recognition, where different observations of the same object typically receive the same label. This paper focuses on a transformation invariant distance measure that represents the minimum distance between the transformation manifolds spanned by patterns of interest. Since these manifolds are typically nonlinear, the computation of the manifold distance becomes a non-convex optimization problem. We propose to represent a pattern of interest as a linear combination of a few geometric functions extracted from a structured and redundant basis. Transforming the pattern results in the transformation of its constituent parts. We show that when the transformation is restricted to a synthesis of translations, rotations and isotropic scalings, such a pattern representation results in a closed-form expression of the manifold equation with respect to the transformation parameters. The manifold distance computation can be then formulated as a minimization problem, whose objective function is expressed as the difference of convex functions (DC). This interesting property

			permits to solve optimally the optimization problem with DC programming solvers that are globally convergent. We present experimental evidence which shows that our method is able to find the globally optimal solution, outperforming existing methods that yield suboptimal solutions.
27	Recognition of Handwritten Historical Documents: HMM-Adaptation vs. Writer Specifc Training	*Emanuel Indermühle, Marcus Liwicki, Horst Bunke*	In this poster we compare two strategies to train a recognizer for handwritten manuscript by Swiss authors. We show that on small training sets, a writer independent training followed by an HMM-adaptation to writer specific data performs better than a writer specific training only.
28	The Spherical Approach To Visual Attention	*Iva Bogdanova, Alexandre Bur, Heinz Hügli*	The conventional algorithms for visual attention are not suited to omnidirectional images which present radial distortions. A computation approach that processes images in spherical space and produces attention maps with a homogeneous response has been proposed. This work investigates how the spherical approach applies to real scenes and particularly to different omnidirectional sensors. A comparison illustrates the capacity of the spherical approach to provide saliency maps with homogeneous response on the sphere and therefore shows its advantages for detecting spots of attention in omnidirectional scenes. Reported experiments refer to omnidirectional images obtained from a multi-camera omnidirectional sensor as well as a parabolic and a hyperbolic catadioptric image sensor.
29	Volterra Series Expansion for Analyzing MLP based Posterior Features in Speech Recognition	*Joel Pinto*	We use Volterra series expansion to analyze a multilayer perceptron based posterior feature extraction system used in speech recognition. This system comprises of a multi-input, multi-output, linear time-invariant system, followed by a multilayered perceptron. By analyzing the identified first order Volterra kernels (linear approximation of the non-linear system), we obtain insights into the spectro-temporal patterns learned by the system in order to discriminate among phonemes
30	Action Snippets: How many frames does human action recognition require?	*Konrad Schindler, Luc Van Gool*	Visual recognition of human actions in video clips has been an active field of research in recent years. However, most published methods either analyse an entire video and assign it a single action label, or use relatively large look- ahead to classify each frame. Contrary to these strategies, human vision proves that simple actions can be recognized almost instantaneously. In this paper, we present a system for action recognition from very short sequences ("snippets") of 1-10 frames, and systematically evaluate it on standard data sets. It turns out that even local shape and optic flow for a single frame are enough to achieve ≈90% correct recognitions, and snippets of 5-7 frames (0.3-0.5 seconds of video) are enough to achieve a performance similar to the one obtainable with the entire video sequence.
31	Head pose tracking and applications to visual focus of attention modeling	*Stephanie Lefèvre, Sileye Ba, Jean-Marc Odobez*	Human interaction mostly occur through verbal cues, however non verbal cues such as gaze and gestures convey important information. Nowadays, with the ubiquitous presence of cameras, computer vision techniques can be used to analyze human interactions through non-verbal cues. Head pose is an important cue to estimate as it can be used as observations to model head gestures or visual focus of attention (who is looking at whom or what). Two situations can be considered for head pose tracking: high resolution and low resolution head image conditions. In presence of high resolution head images, 3D head models allow to track both the head pose and the facial animations. When high resolution head images are not available, the head pose can be tracked with appearance based head models. In low resolution image case, only head pose information will be available. In both head image resolution situations, head pose together with contextual cues, such as the ongoing conversational events or the time that has elapsed since the last slide change, can be used to estimate people's visual focus of attention.
32	Face Detection using Ferns	*Venkatesh Bala Subburaman, Sébastien Marcel*	In this work, we have explored ferns (a set of simple binary features) with Semi-Naive Bayesian classifier for frontal face detection task. The binary feature used in this work is the sign of pixel intensity difference. From a large set of binary feature a subset is selected based on conditional mutual information criteria. Preliminary experiments with a single stage Semi-Naïve Bayesian classifier show good performance under varying illumination conditions when compared to haar-like feature with adaboost learning. With no preprocessing required on the image, training time of less than a few minutes, and with good performance makes it interesting for further exploration.

33	Semi-Supervised Learning for Handwriting Recognition	*Volkmar Frinken*	This poster displays the motivation the idea and our progress in Semi-Supervised Learning for Handwriting Recognition. To reduce the amount of human work in creating a large labeled set for training a recognizer, a small labeled set and a large unlabeled set will be used in the training process. Our experiments show significant increase compare to not using the unlabeled set, however, the accuracy of the recognizer trained on the large labeled set is not reached.
34	Modeling Human Perception of Static Expressions by Discrete Choice Models	*Matteo Sorci*	Facial expressions are probably the most visual method to convey emotions and one of the most powerful means to relate to each other. In this work we investigate and show the use of Discrete Choice Models in modeling static facial expressions. The objective of this work is focused on the behavioral modeling of the observer perception of human expressions. In particular we want to investigate if differences in the observers correspond to differences in their perception of expressions. In order to do that we have developed a facial expression recognition survey on the web aiming at collect data from a population of real human observers, from all around the world, doing different jobs, having different cultural backgrounds, ages and gender, belonging to different ethnic groups, doing the survey from different places (work, home, on travel, etc.). Based on the collected data and on the representation of the expressions by means of different sets of features, we have developed and estimated three different discrete choice models. The proposed models are then compared and validated showing interesting and encouraging results.