Spontaneous vs. Posed Facial Behavior: Automatic Analysis of Brow Actions

Michel F. Valstar[1], Maja Pantic[1,2], Zara Ambadar[3] and Jeffrey F. Cohn[3]
[1] Computing Department, Imperial College London, UK
[2] Faculty of EEMCS, University of Twente, The Netherlands
[3] Psychology and Psychiatry, University of Pittsburgh, USA
{michel.valstar,m.pantic}@imperial.ac.uk, {ambadar,jeffcohn}@pitt.edu,

ABSTRACT
Past research on automatic facial expression analysis has focused mostly on the recognition of prototypic expressions of discrete emotions rather than on the analysis of dynamic changes over time, although the importance of temporal dynamics of facial expressions for interpretation of the observed facial behavior has been acknowledged for over 20 years. For instance, it has been shown that the temporal dynamics of spontaneous and volitional smiles are fundamentally different from each other. In this work, we argue that the same holds for the temporal dynamics of brow actions and show that velocity, duration, and order of occurrence of brow actions are highly relevant parameters for distinguishing posed from spontaneous brow actions. The proposed system for discrimination between volitional and spontaneous brow actions is based on automatic detection of Action Units (AUs) and their temporal segments (onset, apex, offset) produced by movements of the eyebrows. For each temporal segment of an activated AU, we compute a number of mid-level feature parameters including the maximal intensity, duration, and order of occurrence. We use Gentle Boost to select the most important of these parameters. The selected parameters are used further to train Relevance Vector Machines to determine per temporal segment of an activated AU whether the action was displayed spontaneously or volitionally. Finally, a probabilistic decision function determines the class (spontaneous or posed) for the entire brow action. When tested on 189 samples taken from three different sets of spontaneous and volitional facial data, we attain a 90.7% correct recognition rate.

Categories and Subject Descriptors
I.2.10 **[Vision and Scene Understanding]:** motion, modeling and recovery of physical attributes

General Terms
Algorithms, Experimentation, Human Factors

Keywords
Automatic facial expression analysis, temporal dynamics

1. INTRODUCTION
Because of the practical importance of the topic for affective, perceptual, and ambient interfaces of the future and theoretical interest of cognitive scientists [31], [30] machine analysis of facial expressions attracted the interest of many researchers in computer vision and AI. Most of the facial expressions analyzers developed so far target human facial affect analysis and attempt to recognize a small set of prototypic emotional facial expressions like happiness and anger [31], [39]. Exceptions from this overall state of the art in the field of machine analysis of human facial affect include few tentative efforts to detect attitudinal and non-basic affective states like attentiveness [16], fatigue [22], and pain [4] from face video. Few works on user-profiled interpretation of facial expressions have been also reported [32], [17].

To facilitate detection of subtle facial signals like a frown or a wink and to make facial expression information available for use in applications like multimodal-command interfaces, several research groups begun research on machine analysis of facial muscle actions (atomic facial signals, AUs, [13]). A number of promising prototype systems have been proposed recently that can recognize 15 to 27 different AUs (from a total of 44 AUs) in either (near-) frontal view or profile view face image sequences [39], [29]. Most of these works emphasize statistical and ensemble learning and are either feature-based (i.e., use geometric features like facial points or shapes of facial components, e.g., as is the case with the system proposed here) or appearance-based (i.e., use texture of the facial skin including wrinkles, bulges, and furrows). It has been reported that appearance-based methods usually outperform those based on geometric features [3]. Recent studies have shown that this claim does not always hold [29].

In addition, most of the past work on automatic facial expression analysis (either in terms of discrete emotions like surprise and fear or in terms of discrete AUs) is aimed at the analysis of posed (i.e., volitionally displayed) facial expression data. Only recently few works have been reported on machine analysis of spontaneous facial expression data (e.g. [16], [7], [4]).

The body of research in cognitive sciences, which suggests that temporal dynamics of human facial behavior (e.g. the timing and duration of facial actions) is a critical factor for interpretation of the observed behavior, is large and growing [6], [28], [1]. Facial expression temporal dynamics are essential for categorization of complex psychological states like various types of pain and mood [44]. They are also the key parameter in differentiation between posed and spontaneous facial expressions [23], [15]. For instance, it has been shown that spontaneous smiles, in contrast to posed

Figure 1. Facial muscle actions produced by brow movements.

smiles (like a polite smile), are slow in onset, can have multiple AU12 apexes (multiple rises of the mouth corners), and are accompanied by other AUs that appear either simultaneously with AU12 or follow AU12 within 1s [8]. In spite of these findings in basic research, the vast majority of the past work in the field does not take dynamics of facial expressions into account when analyzing shown facial behavior. Some of the past work in the field has used aspects of temporal dynamics of facial expression such as the speed of a facial point displacement or the persistence of facial parameters over time. However, this was mainly done either in order to increase the performance of facial expression analyzers (e.g. [21], [46]) or in order to report on the intensity of (a component of) the shown facial expression (e.g. [46], [26]), but not in order to analyze explicitly properties of facial expression temporal dynamics. Exceptions from this overall state of the art in the field include two studies on automatic segmentation of AU activation into temporal segments (neutral, onset, apex, offset) in frontal- [41] and profile-view [29] face videos.

This paper reports on our method for automatic discrimination between posed and spontaneous facial expressions using temporal dynamics of brow actions. We focus on brow actions because they are frequent in the repertoire of human facial behavior. The brows are often lowered in displays of affective states like anger and fear and are often raised in displays of surprise [13]. Brow lowering is also frequently present in displays of psychological and cognitive states like pain [44], fatigue [42], concentration and puzzlement [9]. Brow raising is also typical for various social signaling actions where they may communicate greetings [25], interest [18], and (dis-)agreement [9], may provide emphasis for speech acts [18], or may contribute to the regulation of turn-taking (like in query) in social interactions [18]. The method proposed here is based on AUs of the Facial Action Coding System (FACS) [13]. FACS is the best known and the most commonly used system that describes facial activity in terms of visually observable facial muscle actions (AUs). Using FACS, any anatomically possible facial expression can be described in terms of one or more of in total 44 AUs defined by the FACS. The brow actions we focus on correspond to the following AUs (Fig. 1): AU1 (inner brow lift), AU2 (outer brow lift), and AU4 (brow lowering).

To capture brow action dynamics, we track 8 characteristic facial points in frontal face video and compute their displacements. The points in question are (Fig. 2): the inner and the outer corners of the eyebrows, the inner corners of the eyes, and the outer corners of the nostrils. To track the points, we use one of the two off-shelf trackers, the FACE-III head tracker of Xiao et al. [45] and the Patras-Pantic Particle Filtering tracking scheme [34]. Using the tracking data, we first detect the presence (i.e., activation) of AU1, AU2 and AU4. For each activated AU, we determine the temporal segments (neutral, onset, apex, and offset). To detect the activated AUs and their temporal segments, we use a recently proposed AU detector, which combines Gentle Boost ensemble learning and Support Vector Machines [41]. We compute further a set of mid-level feature parameters for every temporal segment of each activated AU. These include the segment duration, the mean and the maximum displacements of the four brow points in x- and y-directions, the maximum velocity in x- and y- directions, and the asymmetry in the displacement of the brow points. We also compute the 2nd order polynomial functional representation of the displacements of the brow points and the order in which the AUs have been displayed. We use Gentle Boost to learn the most informative parameters for distinguishing between spontaneous and volitional AUs and use these to train a separate Relevance Vector Machine (RVM) for each temporal segment of each of the three brow-action-related AUs (i.e., in total 9 GentleRVMs). The outcomes of these 9 GentleRVMSs are then combined and a probabilistic decision function determines the class (spontaneous or posed) for the entire brow action. When trained and tested on a set containing 60 samples of volitional facial displays from the MMI Facial Expression database [33], 59 samples of volitional facial displays from the Cohn-Kanade Facial Expression database [24], and 70 samples of spontaneous facial displays from the DS118 dataset [37], the proposed method attained a 90.7% correct recognition rate when determining the class (spontaneous or posed) of an input facial expression.

The outline of the paper is as follows. Section 2 presents the utilized facial point tracking schemes. Section 3 explains the methodology used to detect AUs and their temporal segments. Section 4 described the parameters used in further processing. The classification methodology we use to distinguish between posed and spontaneous brow actions is explained in Section 5. Section 6 describes the datasets we used in our validation study, which is discussed in Section 7. Section 8 concludes the paper.

2. FACIAL POINT TRACKING

Since we are interested in brow actions, that is, in the motion patterns of the eyebrows, we track four characteristic points of the brows (P1-P4 depicted in Fig. 2) to capture these motion patterns. To enable intra-registration, which removes rigid head movements within the input video, and inter-registration, which reduces inter-person variations in the head shape, we track also the nostrils and the inner corners of the eyes (P5-P8 depicted in Fig. 2). These 8 points are (semi-)automatically initialized in the first frame of an

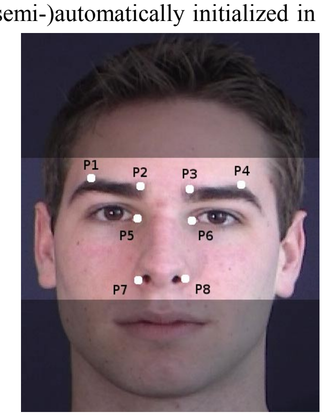

Figure 2. Facial characteristic points used in this work.

input face image sequence using a facial point detector based on Gabor-feature-based boosted classifiers [43]. The positions of the automatically detected points were inspected by a human and, if necessary, corrected. However, due to the high accuracy of the employed facial point detector (Fig. 3), the manual correction was not often needed (in 87% of cases, the correction was not needed).

To track the facial fiducial points, we use one of the two off-the-shelf trackers: the FACE-III head tracker of Xiao et al. [45] that uses the standard Lucas-Kanade optical flow algorithm [27] and the Patras-Pantic Particle Filtering tracking scheme [34]. The reason for using two different trackers is purely practical – while the co-authors of this work affiliated with the University of Pittsburgh were used to work with the FACE-III head tracker, the co-authors affiliated with the Imperial College were accustomed to work with Patras-Pantic Particle Filtering point tracker. Independently of the tracking scheme used to track the facial points for the purposes of this study, the aim was a 'best effort' result, meaning that some of the utilized face image sequences have been tracked more than once using slightly different initialization parameters to remove any errors the trackers made. We did so because the aim of the present study was to investigate whether spontaneously displayed brow actions can be automatically distinguished from volitionally displayed brow actions based on the temporal dynamics of these actions, the aim was not to measure the performance of a fully automated system that accomplishes the classification in question. Fig. 4 depicts typical results of the Patras-Pantic Particle Filtering with Factorized Likelihoods (PFFL) tracking scheme [34], applied for tracking color/intensity-based templates of facial points (for details of the utilized observation models, see [35]). Note that Fig. 4 depicts smoothed tracking results, that is, results after the noise in the output of the PFFL tracker (occurring due to the particle filtering nature of the tracker) is filtered out. To attain this, we apply a temporal filter that removes spurious peaks. The smoothed output of the tracker for a frame at time t is computed by taking the mean of the tracker outputs in the range $[t-2, t+2]$.

To handle possible head rotations and variations in scale of the observed face, we register each frame of the input image sequence with the first frame. If the PFFL tracker was applied for tracking 8 characteristic facial points illustrated in Fig. 2, we carry out the registration process based on 3 referential points: the nasal spine point (say N, calculated as the midpoint between the outer corners of the nostrils P7 and P8, see Fig. 2) and the inner corners of the eyes (P5 and P6, see Fig. 2). We remove variations occurring due to rigid head movements with respect to the camera-lens plane by applying an affine transformation $T1$ to every tracked facial point in each frame. $T1$ is calculated by comparing the locations of the referential points in the current frame with those in the first frame. Inter-person variations in size and location of the facial points are minimized by applying an affine transformation $T2$ to every tracked facial point in each frame. $T2$ is obtained by comparing the locations of the referential points of a given subject in the first frame with the corresponding points in a selected expressionless 'standard' face (the choice of the subject to be used as this 'standard' face does not influence the process). Thus, after tracking any of 8 facial points in an image sequence containing k frames, we obtain a set of coordinates $<p_1 \dots p_k>$ corresponding to the locations of the pertinent point p in each of k frames. Then, the registered coordinates p_i' are obtained as:

$$p_i' = T2(T1(p_i)), \text{ where } i = [1 : k]$$

Figure 3. First-effort results of the utilized face point detector obtained for samples from (left to right): the Cohn-Kanade database, the MMI database, and a cell phone camera.

If the FACE-III head tracker was used to track 8 characteristic facial points, the registration process is carried out by means of a cylindrical head model [45], which is used to estimate the 6 degrees of freedom of head motion: horizontal and vertical position, distance to the camera (i.e., scale), pitch, yaw, and roll. A cylindrical head model is fitted to the initial face region, and the face image is cropped and projected onto the cylinder as the template of head appearance. For any given subsequent frame, the face template is projected onto the image plane assuming that the pose has remained unchanged from the previous frame. Then, the difference between the projected image and the current frame is computed, providing the correction of the estimated head pose. As the templates are updated as the head motions are recovered, the errors of motion recovery accumulate over time. To handle this problem, images of certain reference poses are prepared and when the estimated head pose is close to that in the reference, the head image is re-registered with the relevant reference image. The re-registration also enables the system to recover head pose when the head reappears after occlusion, like when the head moves for a moment out of the camera's view. This procedure recovers head translation, scaling, and rotation for each frame. The parameters are used to stabilize the face image to a frontal view.

3. FACIAL ACTION DETECTION

To recognize AU1, AU2, and AU4 (Fig. 1) occurring alone or in combination in an input face image sequence, we employ an adapted version of the system that detects AUs and their temporal segments (neutral, onset, apex, offset) using a combination of Gentle Boost learning and Support Vector Machines (SVM) [41].

First, for all characteristic facial points Pi depicted in Fig. 2, where $i = [1 : 8]$, we compute two features (the displacement of Pi in y- and x-direction) for every frame t:

$$f_1(Pi) = Pi,y,t - Pi,y,1 \quad (1)$$

$$f_2(Pi) = Pi,x,t - Pi,x,1 \quad (2)$$

Then, for all pairs of points Pi and Pj, where $i \neq j$ and $i,j = [1 : 8]$, we compute in each frame the following features (the distances between the points and the increase/decrease of those distances in correspondence to the first frame; $\|\cdot\|$ is the L_2 norm):

$$f_3(Pi, Pj) = \|Pi - Pj\| \quad (3)$$

$$f_4(Pi, Pj) = f_3(Pi, Pj) - \|Pi,1 - Pj,1\| \quad (4)$$

Finally, we compute the first time derivative df/dt of all features defined above, resulting in a set of 144 features per frame.

We use Gentle Boost [20] to select the most informative features for every class $c \in C$, where $C = \{AU1, AU2, AU4\}$. An advantage of feature selection by Gentle Boost is that features are

Figure 4. Tracking results of the PFFL point tracker for a sample from the Cohn-Kanade Facial Expression database.

selected depending on the features that have been already selected. In feature selection by Gentle Boost, each feature is treated as a weak classifier. Gentle Boost selects the best of those classifiers and then boosts the weights using the training examples to weight the errors more. The next feature is selected as the one that gives the best performance on the errors of the previously selected features. At each step, it can be shown that the chosen feature is uncorrelated with the output of the previously selected features. As shown in [5], when SVMs are trained using the features selected by a boosting algorithm, they perform better.

To detect 3 different brow-action-related AUs occurring alone or in combination in the current frame of the input image sequence (i.e., to classify the current frame into one or more of the $c \in C$, where C = {AU1, AU2, AU4}), we use 3 separate SVMs to perform binary decision tasks using one-versus-all partitioning of data resulting from the feature selection stage. To determine AU predictions for the entire input sequence based on the outputs of 3 SVMs per frame, we compute an adaptive threshold per AU in the following way. Suppose the SVM determined that a video V has m frames in which a certain AU is active. Let N_p be a vector containing for every video in which the pertinent AU is displayed the number of frames for which the SVM estimated that the AU in question is present. Similarly, let N_n be the vector containing per video in which the pertinent AU is not present the number of frames for which the SVM falsely estimated that the AU in question is present. Let $m_p = \min (N_p)$ be the length of the shortest video segment belonging to N_p and let $0 < m_{np} = \max (N_n) < m_p$ be the length of the longest segment belonging to the subset of N_n. The threshold Θ that is used to decide whether the test sample V contains the AU under investigation (i.e., $m > \Theta$) is defined as:

$$\Theta = ½ (m_p + m_{np})$$

To encode temporal segments of the AUs found to be activated in the input image sequence, we proceed as follows. An AU can be either in (i) the onset phase, where the muscles are contracting and the appearance of the face changes as the facial action grows stronger, or in (ii) the apex phase, where the facial action is at its apex and there are no more changes in facial appearance due to this particular facial action, or in (iii) the offset phase, where the muscles are relaxing and the face returns to its neutral appearance, or in (iv) the neutral phase, where there are no signs of activation of this particular facial action. Often the order of these phases is neutral-onset-apex-offset-neutral, but other combinations such as multiple-apex facial actions are also possible. Note that facial actions having multiple apexes are characteristic for spontaneous facial expressions [8].

As every facial action can be divided into these four temporal segments, we consider the problem to be a four-valued multi-class classification problem. We use a one-versus-one approach to multi-class SVMs (mc-SVMs). In this approach, for each AU and every pair of temporal segments we train a separate sub-classifier specialized in the discrimination between the two temporal segments. This results in $\sum_i i = 6$ sub-classifiers that need to be trained ($i = [1 : C - 1]$, C = {neutral, onset, apex, offset}). For each frame t of an input image sequence, every sub-classifier returns a prediction of the class $c \in C$, and a majority vote is cast to determine the final output c_t of the mc-SVM for the current frame t. To train the sub-classifiers, we apply the following procedure using the same set of features that was used for AU detection (see equations (1)–(4) above). For each classifier separating classes c_i, $c_j \in C$ we apply the Gentle Boost, resulting in a set of selected features $G_{i,j}$. We use $G_{i,j}$ to train the sub-classifier specialized in discriminating between the two temporal segments in question (c_i, $c_j \in C$).

4. MID-LEVEL FEATURE PARAMETERS

Our choice of mid-level feature parameters to be used for automatic discernment between spontaneous and volitional brow actions is largely influenced by a number of studies in psychology on spontaneous (produced in a reflex-like manner) and volitional (deliberately produced) smiles.

Intensity: The early study of Ekman and Friesen on felt and false smiles suggested that intensity of contractions of zygomatic major (i.e., the muscle running around the lips that is responsible for lip corner retraction in smiles) is important in distinguishing between felt and 'phony' smiles [12]. Cohn and Schmidt further confirmed this finding, reporting that posed smiles are larger in amplitude than spontaneous smiles [8], [38].

Duration: Ekman and Friesen found that spontaneous smiles last usually between ⅔ and 4 seconds [12]. Hess and Kleck extended these findings and found that deliberate smiles are shorter in total duration than spontaneous smiles [23]. Cohn and Schmidt further confirmed these findings [8], [38]. Fogel et al. [19] found that even a complex family of different smiles can be defined based on the differences in duration and amplitude of the relevant smiles.

Trajectory: Ekman and Friesen reported that in deliberate expressions the onset is often abrupt, the apex held too long, and the offset is either abrupt and/or appears irregular rather than smooth [12]. Hess and Kleck confirmed these observations and reported that in comparison to deliberate smiles, spontaneous smiles are slower in onset and offset time [23]. Cohn and Schmidt reported that spontaneous smiles are slow in onset [8], [38] and they also found that they may have multiple apexes [8].

Symmetry: Ekman et al. [14] reported that spontaneous smiles are more symmetrical than those made deliberately. In one study they found that smiles in response to watching an amusing film were in 96% of cases symmetrical. This finding has been later confirmed by different researchers [8], [38].

Co-occurrences: Evidence that in spontaneous smile the activity of zygomatic major is accompanied by the activity of orbicularis oculi (i.e., the muscle orbiting the eye which when contracted produces crow-feet wrinkles around the outer corner of the eye) dates from the 19th century and Duchenne de Boulogne [10], in whose honour the smiles incorporating orbicularis oculi are called Duchenne's smiles. Duchenne's proposal has been revisited many times by researchers like Charles Darwin, Paul Ekman, and many others. Recent studies by Cohn and Schmidt extend these findings and suggest that the activity of zygomatic major (AU12) may be accompanied not only by the activity of orbicularis oculi (AU6) but also by activity of other AUs that appear either simultaneously with AU12 or follow AU12 within 1s [8].

As suggested by Ekman [11], the findings about spontaneous and posed smiles may be extendable to a wider set of facial actions. In this paper we present our research to check Ekman's proposal for the case of brow actions. To our best knowledge, this is the first study that checks Ekman's proposal, at least as far as brow actions are concerned.

For every temporal segment (neutral, onset, apex, offset) of each activated AU, we calculate a number of mid-level feature parameters based on the displacements of facial fiducial points *Pi*, where $i = [1 : 4]$. We consider only the displacements of points P1–P4 (Fig. 2) in a further processing because we are interested only in brow actions. For a temporal segment d consisting of n frames, we will have signals $s_{i,t} = \{f_1(Pi), f_2(Pi)\}$, where $t = [1 : n]$ is a frame of the temporal segment d and $f_1(Pi)$ and $f_2(Pi)$ are the displacements of a facial point *Pi*, $i = [1 : 4]$ in y- and x-direction, as defined in equations (1) and (2) above. Thus, for the entire temporal segment d, we will have a set of signals $S_i = \{s_{i,1}, \ldots, s_{i,n}\}$ for each facial point *Pi*, $i = [1 : 4]$, resulting in a total of $N_d = 4{*}n$ signals. For each S_i, and for both the y- and x-directions, we first compute the maximum and the mean point displacement (relating to the **intensity** of brow actions), and the maximum and the mean velocity of these displacements (relating to the **speed** and the **trajectory** of AU activations). We do so as follows.

$$mp_1(S_i) = \max(S_i) \quad (5)$$
$$mp_2(S_i) = \mathrm{mean}(S_i) \quad (6)$$
$$mp_3(S_i) = \max(s_{i,t} - s_{i,t-1}) \quad (7)$$
$$mp_4(S_i) = (\textstyle\sum_t s_{i,t} - s_{i,t-1}) / (n - 1) \quad (8)$$

Next, for the pairs of the brow points $\{P1, P4\}$ and $\{P2, P3\}$, we compute a measure of **symmetry**. We say that a given brow action is symmetric in x-direction if the relevant brow points move an equal distance towards each other or away from each other. Otherwise, the brow action in question is asymmetric. We say that a given brow action is symmetric in y-direction if the relevant brow points traverse an equal distance either upwards or downwards. Otherwise, the brow action in question is asymmetric. As the measure of asymmetry in y-direction we use mid-level feature parameter mp_5 defined by equation (9) and as the measure of asymmetry in x-direction we use mid-level feature parameter mp_6 defined by equation (10), where P' and P'' represent a pair of the brow points (i.e., $\{P1, P4\}$ or $\{P2, P3\}$), $t = [1 : n]$, and $f_1(P)$ and $f_2(P)$ are the displacements of a facial point P in y- and x-directions, as defined in equations (1) and (2) above.

$$mp_5(P', P'') = \textstyle\sum_t [\mathrm{abs}(f_1(P') - f_1(P''))] \quad (9)$$
$$mp_6(P', P'') = \textstyle\sum_t [\mathrm{abs}(f_2(P') + f_2(P''))] \quad (10)$$

Next, we describe the overall displacement of a facial fiducial point *Pi*, $i = [1 : 4]$, within the temporal segment d as a function of time $g(t) = at^2 + bt + c$, where $t = [1 : n]$ is a frame of the temporal segment d and $g(t) = s_{i,t}$. Thus, for each S_i, and for both the y- and x-directions, we define the following mid-level feature parameters (relating the **trajectory** of AU activations):

$$mp_7(S_i) = a \quad (11)$$
$$mp_8(S_i) = b \quad (12)$$
$$mp_9(S_i) = c \quad (13)$$

Finally, the total **duration** of a temporal segment d (i.e., n) and the **occurrence order** o of the temporal segment within the entire image sequence ($o = 1$ if d was the first temporal segment of the first activated AU, otherwise $o > 1$) are used as mid-level feature parameters as well.

$$mp_{10}(d) = n \quad (14)$$
$$mp_{11}(d) = o \quad (15)$$

Thus, for each temporal segment of an activated AU, we calculate in total 62 features.

5. SPONTANEOUS OR POSED

The tracking schemes that we utilize to track facial characteristic points in an input face image sequence include a particle-filtering-based and an optical-flow-based tracking algorithm. Although we have aimed to achieve 'best effort' results, meaning that some of the input face image sequences have been tracked more than once using slightly different initialization parameters to remove any errors that the tracker has made in earlier trials, noisy output from the tracking algorithms should be expected. In the case of the PFFL point tracker the noise in the output (occurring due to the particle filtering nature of the tracker) is filtered out by applying a temporal filter that removes spurious peaks. However, since the smoothed output of the tracker for a frame at time t is computed by taking the mean of the tracker outputs in the range $[t\text{-}2, t\text{+}2]$, some noise remains present in the output. For the FACE-III head tracker, the noise in the facial point tracking data occurs mostly due to inaccuracies in registration as well as due to the applied re-registration process (see section 2). This noise in point tracking data is propagated further, resulting in noisy feature-based and mid-level parametric representations of the input data. Eventually, this will influence the predictions of whether a given temporal segment belongs to a spontaneous or to a deliberate brow action. In order to deal with the imperfect data and generate predictions about whether a given temporal segment belongs to a certain class (spontaneous or posed) so that the confidence measure associated with it varies in accordance with the accuracy of the input data, we employ Relevance Vector Machines (RVMs).

A RVM classifier is a probabilistic sparse kernel model identical in functional form to a SVM classifier [40]. In their simplest form, RVMs attempt to find a hyperplane defined as a weighted combination of a few relevance vectors that separate data samples of two different classes. In RVM, a Bayesian approach is adopted for learning, where a prior is introduced over the model weights, governed by a set of hyperparameters, one for each weight. The most probable values of these hyperparameters are iteratively estimated from the data. Sparsity is achieved because the posterior distributions of many of the weights are sharply peaked around 0. Unlike the SVM, the nonzero weights of RVM are not associated with examples close to the decision boundary, but rather appear to represent prototypical examples of classes. These examples are called relevance vectors and, in our case, they can be thought of as representative displays of a brow action. The main advantage of RVM is that while it is capable of a generalization performance comparable to that of an equivalent SVM, it uses substantially fewer kernel functions. Furthermore, predictions in RVM are probabilistic, in contrast to the deterministic decisions provided by SVM. In their original form, RVMs are suitable for solving two-class classification problems. Since for each temporal segment of a brow action we want to determine whether it has

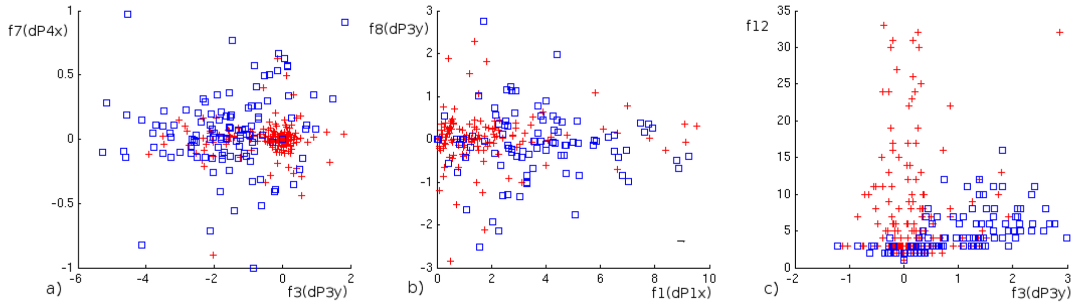

Figure 5. Scatter plots of pairs of the most informative mid-level feature parameters selected by the Gentle Boost for a) the onset temporal segments of AU1, b) the apex temporal segments of AU1, and c) the offset temporal segments of AU1. Crosses denote spontaneous facial data while squares denote data of deliberately displayed facial actions.

been displayed spontaneously or deliberately, our problem is a set of L two-class classification problems. Hence, we use L RVMs, each of which predicts whether a given temporal segment of a specific brow action has been displayed spontaneously or not. In our case $L = 9$ since we have three AUs related to brow actions (i.e., AU1, AU2, AU4), each of which can be in one of the three temporal phases (i.e., onset, apex, offset). We use the Gentle Boost to define a set of mid-level feature parameters defined by equations (5)–(15) that are the most informative for distinction between the 9 classes. Then, we train the RVMs to perform binary decision tasks using one-versus-all partitioning of data resulting from the feature selection stage. As a kernel, we use standard Gaussian radial basis function. For each fold of the cross validation test procedure (section 7), the kernel width has been optimized independently of the test data.

Thus, for an input image sequence I in which m temporal segments d_t of brow actions have been identified by the process defined in section 3, we will have m predictions c_t (one for each d_t), each of which is associated with a confidence measure $p_t \geq 0$, where $c_t \in \{-1, 1\}$ (e.g., -1 for *posed* and 1 for *spontaneous*) and $t = [1 : m]$. Then, the class $C \in \{-1, 1\}$ for the entire brow action shown in the input video I is predicted with a confidence measure $P \geq 0$ in the following way.

$$C = \mathrm{sign}(\gamma),\ P = \mathrm{abs}(\gamma),\ \text{where } \gamma = \textstyle\sum_t p_t c_t,\ t = [1 : m] \qquad (16)$$

6. THE UTILIZED DATASET

In our study, we used a set containing 60 samples of volitional facial displays from the MMI Facial Expression database [33], 63 samples of volitional facial displays from the Cohn-Kanade Facial Expression database [24], and 139 samples of spontaneous facial displays from the DS118 dataset [37]. When selecting this data, we took care that samples of various brow actions were picked from the three databases with the same frequency.

The MMI Facial Expression database contains over 4000 videos and over 600 static images depicting facial displays of single AU activation, multiple AU activations, and six basic emotions. Subjects were 52 adults being 19 to 62 years old; 48% female, 81% being Caucasian, 14% Asian and 5% African. All facial displays were made on command and the recordings were made under constant lighting conditions from frontal, profile, or dual view orientation (the latter using a mirror). Two FACS experts AU-coded the database. When in doubt, decisions were made by consensus. The database contains a large amount of displays of single AU activation recorded for each of the 44 AUs defined in FACS. This means that we can learn to recognize every AU independent of other AUs.

The Cohn-Kanade Facial Expression database has been developed for research in recognition of the six basic emotions and their corresponding AUs. The database contains over 2000 videos of facial displays produced by 210 adults being 18 to 50 years old, 69% female, 81% Caucasian, 13% African and 6% from other ethnic groups. From this database, 480 grayscale videos have been made publicly available. It is currently the most commonly used database for studies on automatic facial expression analysis. All facial displays were made on command and the recordings were made under constant lighting conditions. Certified FACS coders provided AU event coding for all videos. In the publicly available version of this database the expressions are shown until the beginning of the offset phase. In this study, however, we have used the full recordings of facial expression displays (i.e., showing the neutral → onset → apex → offset → neutral temporal pattern of an emotional expression).

The DS118 dataset has been collected to study facial expression in patients with heart disease [37]. Subjects were 85 men and women with a history of transient myocardial ischemia who were interviewed on two occasions at a 4-monthinterval. They averaged 59 years of age (std = 8.24) and were predominantly Caucasian. Spontaneous facial expressions were video-recorded during a clinical interview that elicited AUs related to disgust, contempt, and other negative emotions as well as smiles. The brow actions displayed in the data are often very subtle. Due to confidentiality issues, this FACS-coded dataset is not publicly available.

7. EXPERIMENTAL RESULTS

To evaluate the proposed method for automatic discrimination between spontaneous and deliberate brow actions, we used 119 of samples of volitional brow actions and 70 samples of spontaneous brow actions for which our system for automatic recognition of AUs and their temporal segments [41] generated correct results (as indicated by the ground truth). Although the aim of this study is not to evaluate the performance of a fully automated system for spontaneous and posed brow action detection, but to investigate whether posed brow actions can be automatically distinguished from spontaneous brow actions based on the temporal dynamics of these actions, we would like to make few remarks about the AU recognition results attained by our AU detector. The version of the AU detector that we used in this study has been trained using only samples of volitional facial displays from the MMI database. Hence, it is not surprising that for deliberate facial displays, it achieved high recognition rates. More specifically, from 123 initially used samples of volitional brow actions, 119 have been correctly AU-coded by our system, resulting in 96.7% correct recognition rate, a slightly higher rate than the one reported in the original study for the Cohn-Kanade database [41]. Second, it is also not surprising that for spontaneous facial actions, an AU detector trained on deliberate facial actions achieves low recognition rates. From 139 initially used samples of spontaneous brow actions, 70 have been correctly AU-coded by our system, resulting in 50.4% correct recognition rate. Although this is not a very good result, it is promising, especially when one takes into account that the system was not trained on samples of spontaneous brow actions and that current studies on automatic AU-coding of spontaneous facial data reported correct recognition rates ranging from 26% [4] to 76% [7] for brow actions.

Table 1. Mid-level feature parameters that the Gentle Boost selected as the most informative for determining whether a detected temporal segment *d* of an activated AU has been displayed spontaneously or not. The 1st column lists the 9 relevant classes, the 2nd column lists the total number of mid-level parameters selected by the Gentle Boost for the relevant class, and the 3rd column lists the most informative of the selected parameters defined by equations (5)-(15).

Class	#	Most informative mid-level parameters
AU1-on	7	mp_3 ($S_{3,y}$), mp_7 ($S_{4,x}$), mp_1 ($S_{1,x}$)
AU1-ap	6	mp_1 ($S_{1,x}$), mp_8 ($S_{3,y}$), mp_3 ($S_{2,y}$)
AU1-off	11	mp_3 ($S_{3,y}$), mp_{11} (d), mp_1 ($S_{1,x}$)
AU2-on	3	mp_1 ($S_{1,x}$), mp_2 ($S_{4,x}$), mp_9 ($S_{3,x}$)
AU2-ap	2	mp_2 ($S_{4,x}$), mp_{11} (d)
AU2-off	13	mp_1 ($S_{1,x}$), mp_3 ($S_{3,y}$), mp_9 ($S_{3,x}$)
AU4-on	1	mp_1 ($S_{3,y}$)
AU4-ap	14	mp_1 ($S_{1,x}$), mp_2 ($S_{3,y}$), mp_{11} (d)
AU4-off	14	mp_1 ($S_{3,y}$), mp_{11} (d), mp_1 ($S_{4,x}$)

To evaluate the proposed method for automatic discrimination between spontaneous and deliberate brow actions using the 189 data samples as explained above, we performed a leave-one-subject out cross validation, using in every fold all samples of one subject as the test data and all other samples as the training data. For each of the 9 two-class classification problems (i.e., whether onset, apex, and offset of AU1, AU2, and AU4, is spontaneously displayed or not), Table 1 lists the mid-level feature parameters that the Gentle Boost selected as the three most informative for distinction between the 9 classes. The most informative mid-level feature parameters for classification of onset, apex, and offset temporal segments of AU1 are depicted in Fig. 5. As can be seen from Table 1, Ekman's proposal that the findings about posed and spontaneous smiles may be extendable to other facial actions [11], proved to be correct in the case of brow actions. In brief, the properties of the temporal dynamics of brow actions that help in distinguishing spontaneous from deliberate facial expressions are:
(1) maximal displacement of the relevant facial points (relating to the **intensity** of shown AU),
(2) maximal velocity of this displacement (relating to the **speed** and the **trajectory** of AU activation), and
(3) the occurrence order of the action within the image sequence (relating to **co-occurrences** of AUs).

The symmetry of facial actions, reported to be an important parameter for distinguishing spontaneous from posed smiles did not appear to be important in the case of brow action. This finding is consistent with neuron-physiological evidence which supports the assumption that brow actions produced on command are usually bilateral (shown on both sides of the face) and, in turn, more symmetrical than lower face movements (e.g. smiles), which when produced on command tend to be asymmetrical [36], [11]. More specifically, instructions to deliberately display unilateral brow actions (e.g. AU2) typically yield bilateral movements or no movement at all [36].

Table 2 summarizes the results of evaluating the performance of 9 GentleRVMs using the leave-one-subject-out cross validation. These results can be read as follows: the higher the recognition rate for a certain temporal segment, the more informative the segment is for distinguishing between spontaneous and deliberate brow actions. Hence, onset and offset phases are very important in distinguishing voluntary from involuntary brow actions while the apex phase does not contribute much (or not at all in the case of AU2) to the process. This finding reconfirms and reinforces the observation that temporal dynamics of facial actions play a crucial role in distinguishing spontaneous from deliberate facial displays given that the apex phase of a facial action can be regarded as static in the sense that facial points characterizing the relevant action (i.e., P1–P4 illustrated in Fig. 2 in the case of brow actions) do not change their facial location during this phase.

Table 2. Correct classification rates attained by 9 RVMs using the mid-level feature parameters that the Gentle Boost selected as the most informative for the classification problem in hand, i.e., determining whether a detected temporal segment of an activated AU has been displayed spontaneously or not.

AU	onset	apex	offset
1	0.797	0.703	0.843
2	0.833	0.532	0.863
4	0.751	0.694	0.789

Table 3 summarizes the final classification results achieved by the probabilistic decision function defined by equation (16). These results clearly indicated that the proposed method is effective for distinguishing voluntary from involuntary brow actions. It is interesting to note that the achieved correct classification rate is much higher (90.8%) for the event coding where the event is the entire brow action displayed in the input video than for the event coding where the event is one of the temporal segments of one of the displayed brow actions (75.6%). This reconfirms once again the finding that temporal aspects of facial displays are important in distinguishing voluntary from involuntary facial expressions. More specifically, the observed difference in the achieved correct classification rates can be read as follows. Even brief observations of facial behavior are sufficient to make rather accurate judgments on deliberateness of the shown facial signals. When observations are longer, cues at a higher semantic level become available, revealing temporal relations between isolated cues extracted from fleeting glimpses of behavior and leading to greater prediction accuracy. This reminds (partially) of observations made in the study on thin slices of behavior [2], which suggests that human judgments of very short observations of behavior may be less accurate but that for observations being 30s long or longer, the judgments are accurate and do not become significantly more accurate as the length of the observations increases. We could not confirm or reject the latter assumption for the case of judgments made by a computer system rather than by a human observer, since the samples of facial behavior used in our study were short, lasting in total 5s to 35s.

Table 3. Final classification results achieved by the probabilistic decision function defined by equation (16) for the entire brow actions shown in an input face image sequence. For the purposes of computing the recall and precision, the spontaneous class was considered the target class.

Classification rate:	0.908	Total Spontaneous:	70
Recall:	0.829	Correct Spontaneous:	58
Precision:	0.921	Total Deliberate:	114
		Correct Deliberate:	109

We also measured the effect of incorporating confidence measures p_t into the final decision function defined by equation (16). To do so, we evaluated the performance of the proposed method using a redefined final decision function such that the third term in (16) is defined as $\gamma = \sum_t c_t$, $t = [1 : m]$. The attained correct classification rate was 87%, a decrease of almost 4% in correspondence to the correct classification rate realized when using the originally proposed final decision function. This clearly shows the benefit of using a probabilistic decision function rather than a deterministic final decision function.

8. CONCLUSSIONS

In this paper we proposed a (semi-)automated system for distinguishing posed from spontaneous brow actions. To our best knowledge, this is the first attempt reported in the literature to automatically determine whether an observed facial action has been displayed deliberately or spontaneously. Conform the research findings in psychology on spontaneous and posed smiles, we built our system around characteristics of the temporal dynamics of brow actions and employed parameters like speed, intensity, duration, and the occurrence order of brow actions to classify brow actions present in an input face image sequences as either spontaneous or deliberate facial actions. We attained a 90.7% correct classification rate when testing the proposed system on 189 samples taken from three different sets of spontaneous and volitional facial data. In effect, we confirmed Ekman's proposal that the findings about spontaneous and posed smiles may be extendable to a wider set of facial actions. From our study, it becomes clear that characteristics of brow actions that help in distinguishing spontaneous from deliberate facial displays include intensity, speed, trajectory, and order of occurrence of relevant AU activations. The symmetry of facial actions, reported to be an important parameter for distinguishing spontaneous from posed smiles did not appear to be important in the case of brow action. This is consistent with the neuron-physiological evidence that brow actions produced on command are usually bilateral. In addition, in contrast to the apex phase, the onset and offset phases of brow actions seem to be informative for the classification problem in hand. Given that the apex phase lacks temporal changes due to relative spatial stability of relevant facial points, we argue that this finding reconfirms and reinforces the observation that temporal dynamics of facial actions play a crucial role in distinguishing spontaneous from deliberate facial displays. Finally, the achieved results show a significant increase in the correct classification rate when judgments are made over longer rather than shorter observations. We hypothesise that when observations are longer, cues at a higher semantic level become available, revealing additional temporal relations between isolated cues that can be extracted from short observations of behavior and leading to greater prediction accuracy. In our future research, we will try to extend the findings of the presented study to a wider set of facial actions.

9. ACKNOWLEDGMENTS

The authors would like to thank Paul Ekman of the University of California San Francisco for providing the DS 118 dataset. The work of Zara Ambadar and Jeffrey Cohn has been partially supported by the NIMH grant R01 MH51435.

10. REFERENCES

[1] Ambadar, Z., Schooler, J. and Cohn, J.F. Deciphering the enigmatic face: The importance of facial dynamics in interpreting subtle facial expressions, *Psychological Science, 16*, 5 (May 2005), 403-410.

[2] Ambady, N. and Rosenthal, R. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin*, *111*, 2 (Feb. 1992), 256-274.

[3] Bartlett, M.S., Hager, J.C., Ekman, P. and Sejnowski, T.J. Measuring facial expressions by computer image analysis. *Psychophysiology, 36*, 2 (Mar. 1999), 253-263.

[4] Bartlett, M.S., Littlewort, G., Frank, M.G., Lainscsek, C., Fasel, I. and Movellan, J. Fully automatic facial action recognition in spontaneous behavior. In *Proc. Conf. Automatic Face & Gesture Recognition*, 223-230, 2006.

[5] M Bartlett, M.S., Littlewort, G., Lainscsek, C., Fasel, I. and Movellan, J. Machine learning methods for fully automatic recognition of facial expressions and actions. In *Proc. Int'l Conf. Systems, Man and Cybernetics*, 592-597, 2004.

[6] Bassili, J.N. Facial motion in the perception of faces and of emotional expression", *J. Experimental Psychology, 4*, 3 (Aug. 1978), 373-379.

[7] Cohn, J.F., Reed, L.I., Ambadar, Z., Xiao, J. and Moriyama, T. Automatic analysis and recognition of brow actions in spontaneous facial behavior. In *Proc. Int'l Conf. Systems, Man & Cybernetics, 1*, 610-616, 2004.

[8] Cohn, J.F. and Schmidt, K.L. The timing of facial motion in posed and spontaneous smiles. *J. Wavelets, Multi-resolution & Information Processing, 2*, 2 (June 2004), 121-132.

[9] Cunningham, D.W., Kleiner, M., Wallraven, C. and Bülthoff, H.H. The components of conversational facial expressions. In *Proc. ACM Int'l Symposium on Applied Perception in Graphics and Visualization*, 143-149, 2004.

[10] Duchenne de Bologne, G.B. *Mechanisme de la Physionomie Humaine*. Jules Renouard Libraire, Paris, France, 1862. (Translation: *The Mechanism of Human Facial Expression*. Cambridge University Press, New York, 1990).

[11] Ekman, P. Darwin, deception, and facial expression. *Annals New York Academy of sciences, 1000*, (Dec. 2003), 205-221.

[12] Ekman, P. and Friesen, W.F. Felt, false and miserable smiles. *J. Nonverbal Behavior, 6*, 4 (June 1982), 238-252.

[13] Ekman, P., Friesen, W.V. and Hager, J.C. *Facial Action Coding System*. A Human Face, Salt Lake City, 2002.

[14] Ekman, P., Hager, J.C. and Friesen, W.V. The symmetry of emotional and deliberate facial actions. *Psychophysiology, 18*, 2 (Mar. 1981), 101-106.

[15] Ekman, P. and Rosenberg, E.L. Eds. *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System.* Oxford University Press, Oxford, UK, 2005.

[16] El Kaliouby, R. and Robinson, P. Real-Time Inference of Complex Mental States from Facial Expressions and Head Gestures. In *Proc. Int'l Conf. Computer Vision & Pattern Recognition, 3*, 154, 2004.
[17] Fasel, B., Monay, F. and Gatica-Perez, D. Latent semantic analysis of facial action codes for automatic facial expression recognition, In *Proc. Int'l Workshop on Multimedia Information Retrieval*, 181-188, 2004.
[18] Flecha-Garcia, M.L. Facial gestures and communication: what induces raising eyebrow movements in Map Task dialogues. *Proc. Theoretical & Applied Linguistics – Postgraduate Conf.*, 2001.
[19] Fogel, A., Hsu, H.C., Shapiro, A.F., Nelson-Goens, G.C. and Secrist, C. Effects of normal and perturbed social play on the duration and amplitude of different types of infant smiles. *Developmental Psychology, 42*, 3 (May 2006), 159-173.
[20] Friedman, J., Hastie, T. and Tibshirani, R. Additive logistic regression: a statistical view of boosting. *The Annals of Statistics, 28*, 2 (Apr. 2000), 337-374.
[21] Gralewski, L., Campbell, N. and Voak, I.P. Using a tensor framework for the analysis of facial dynamics. In *Proc. Int'l Conf. Face & Gesture Recognition,* 217-222, 2006.
[22] Gu, H. and Ji, Q. An automated face reader for fatigue detection. In *Proc. Int'l Conf. Face & Gesture Recognition*, 111-116, 2004.
[23] Hess, U. and Kleck, R.E. Differentiating emotion elicited and deliberate emotional facial expressions. *European J. of Social Psychology, 20*, 5 (Sep./Oct. 1990), 369-385.
[24] Kanade, T., Cohn, J.F. and Tian, Y. Comprehensive database for facial expression analysis. In *Proc. Int'l Conf. Automatic Face & Gesture Recognition*, 46-53, 2000.
[25] Kendon, A. A description of some human greetings. In *Comparative Ecology and Behavior of Primates*, Michael, R. and Crook, J., Eds. Academic Press, New York, USA, 1973, 591-668.
[26] Littlewort, G., Bartlett, M.S., Fasel, I., Susskind, J. and Movellan, J. Dynamics of facial expression extracted automatically from video. *Int'l J. Image & Vision Computing, 24*, 6 (June 2006), 615-625.
[27] Lucas, B. and Kanade, T. An interactive image registration technique with an application in stereo vision. In *Proc. Int'l Joint Conf. on Artificial Intelligence*, 674-679, 1981.
[28] Messinger, D.S., Fogel, A. and Dickson, K.L. A dynamic systems approach to infant facial action. *The psychology of facial expression*. Russell, J.A. and Fernandez-Dols, J.M., Eds. Cambridge University Press, Cambridge, UK, 1997, 205-226.
[29] Pantic, M. and Patras, I. Dynamics of facial expressions – recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Trans. Systems, Man, and Cybernetics, Part B, 36*, 2 (Apr. 2006), 433-449.
[30] Pantic, M., Pentland, A., Nijholt, A. and Huang, T.S. Front-end of human computing: Machine analysis of human behaviour. In *Proc. Int'l Conf. Multimodal Interfaces*, 2006.
[31] Pantic, M. and Rothkrantz, L.J.M. Toward an affect-sensitive multimodal human-computer interaction. *Proceedings of the IEEE, 91*, 9 (Sep. 2003), 1370-1390.
[32] Pantic, M. and Rothkrantz, L.J.M. Case-based reasoning for user-profiled recognition of emotions from face images. In *Proc. Int'l Conf. Multimedia and Expo*, 2004, 391-394.
[33] Pantic, M., Valstar, M.F., Rademaker, R. and Maat, L. Web-based database for facial expression analysis. In *Proc. Int'l Conf. Multimedia and Expo*, 2005. (www.mmifacedb.com)
[34] Patras, I. and Pantic, M. Particle filtering with factorized likelihoods for tracking facial features. In *Proc. Int'l Conf. Automatic Face & Gesture Recognition*, 97-102, 2004.
[35] Patras, I. and Pantic, M. Tracking deformable motion. In *Proc. Int'l Conf. Systems, Man and Cybernetics*, 1066-1071, 2005.
[36] Rinn, W.E. The neuropsychology of facial expressions: A review of the neurological and psychological mechanisms for producing facial expressions. *Psychological Bulletin, 95*, 1 (Jan. 1984), 52-77.
[37] Rosenberg, E.L., Ekman, P. and Blumenthal, J.A. Facial expression and the affective component of cynical hostility in male coronary heart disease patients. *J. on Health Psychology, 17*, 4 (July 1998), 376-380.
[38] Schmidt, K.L., Ambadar, Z., Cohn, J.F. and Reed, I. Movement differences between deliberate and spontaneous facial expressions: Zygomaticus Major action in smiling. *J. Nonverbal Behavior, 30*, 1 (Mar. 2006), 37-52.
[39] Tian, Y.L., Kanade, T. and Cohn, J.F. Facial Expression Analysis. In *Handbook of Face Recognition*. Li, S.Z. and Jain, A.K., Eds. Springer, New York, 2005.
[40] Tipping, M.E. The relevance vector machines. In *Proc. Int'l Conf. Adv. Neural Inf. Processing Systems*, 332-338, 2000.
[41] Valstar, M.F. and Pantic, M. Fully automatic facial action unit detection and temporal analysis. In *Proc. Int'l Conf. Computer Vision and Pattern Recognition, 3*, 149, 2006.
[42] Veldhuizen, I.J.T., Gaillard, A.W.K. and De Vries, J. The influence of mental fatigue on facial EMG activity during a simulated workday. *Biological Psychology, 63*, 1 (Apr. 2003), 59-78.
[43] Vukadinovic, D. and Pantic, M. Fully automatic facial feature point detection using Gabor feature based boosted classifiers. In *Proc. Int'l Conf. Systems Man and Cybernetics*, 1692-1698, 2005.
[44] Williams, A.C. de C. Facial expression of pain: An evolutionary account. *Behavioral and Brain Sciences, 25*, 4 (Aug. 2002), 439-488.
[45] Xiao, J., Moriyama, T., Kanade, T. and Cohn, J.F. Robust full-motion recovery of head by dynamic templates and re-registration techniques. *Int'l J. Imaging Systems and Technology, 13*, 1 (Sep. 2003), 85-94.
[46] Zhang, Y. and Ji, Q. Active and dynamic information fusion for facial expression understanding from image sequence. *IEEE Trans. Pattern Analysis & Machine Intelligence, 27*, 5 (May 2005), 699-714.