Pattern Recognition Letters 66 (2015) 13–21

Contents lists available at ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Estimating smile intensity: A better way ☆

Jeffrey M. Girard [a,*], Jeffrey F. Cohn [a,b], Fernando De la Torre [b]

[a] Department of Psychology, University of Pittsburgh, 4322 Sennott Square, Pittsburgh, PA 15260, USA
[b] The Robotics Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA

ARTICLE INFO

Article history:
Available online 17 October 2014

Keywords:
Nonverbal behavior
Facial expression
Facial Action Coding System
Smile intensity
Support vector machines

ABSTRACT

Both the occurrence and intensity of facial expressions are critical to what the face reveals. While much progress has been made toward the automatic detection of facial expression occurrence, controversy exists about how to estimate expression intensity. The most straight-forward approach is to train multiclass or regression models using intensity ground truth. However, collecting intensity ground truth is even more time consuming and expensive than collecting binary ground truth. As a shortcut, some researchers have proposed using the decision values of binary-trained maximum margin classifiers as a proxy for expression intensity. We provide empirical evidence that this heuristic is flawed in practice as well as in theory. Unfortunately, there are no shortcuts when it comes to estimating smile intensity: researchers must take the time to collect and train on intensity ground truth. However, if they do so, high reliability with expert human coders can be achieved. Intensity-trained multiclass and regression models outperformed binary-trained classifier decision values on smile intensity estimation across multiple databases and methods for feature extraction and dimensionality reduction. Multiclass models even outperformed binary-trained classifiers on smile occurrence detection.

1. Introduction

The face is an important avenue of communication capable of regulating social interaction and providing the careful observer with a wealth of information. Facial expression analysis has informed psychological studies of emotion [1–3], intention [4,5], physical pain [6,7], and psychopathology [8,9], among other topics. It is also central to computer science research on human–computer interaction [10,11] and computer animation [12].

There are two general approaches to classifying facial expression [13]. Message-based approaches seek to identify the meaning of each expression; this often takes the form of classifying expressions into one or more basic emotions such as happiness and anger [14,15]. This approach involves a great deal of interpretation and fails to account for the fact that facial expressions serve a communicative function [4], can be controlled or dissembled [16], and often depend on context for interpretation [17]. Sign-based approaches, on the other hand, describe changes in the face during an expression rather than attempting to capture its meaning. By separating description from interpretation, sign-based approaches achieve more objectivity and comprehensiveness.

The most commonly used sign-based approach for describing facial expression is the Facial Action Coding System (FACS) [18], which decomposes facial expressions into component parts called action units (AU). Action units are anatomically-based and correspond to the contraction of specific facial muscles. AU may occur alone or in combination with others to form complex facial expressions. They may also vary in intensity (i.e., magnitude of muscle contraction). The FACS manual provides coders with detailed descriptions of the shape and appearance changes necessary to identify each AU and its intensity.

Much research using FACS has focused on the occurrence and AU composition of different expressions [19]. For example, smiles that recruit the orbicularis oculi muscle (i.e., AU 6) are more likely to occur during pleasant circumstances [20,21] and smiles that recruit the buccinator muscle (i.e., AU 14) are more likely to occur during active depression [22,9].

A promising subset of research has begun to focus on what can be learned about and from the *intensity* of expressions. This work has shown that expression intensity is linked to both the intensity of emotional experience and the sociality of the context [2,23,24]. For example, Hess et al. [24] found that participants displayed the most facial expression intensity when experiencing strong emotions in the company of friends. Other studies have used the intensity of facial expressions (e.g., in yearbook photos) to predict a number of social and health outcomes years later. For example, smile intensity in a posed photograph has been linked to later life satisfaction, marital

☆ This paper has been recommended for acceptance by G. Sanniti di Baja.

* Corresponding author. Tel.: +1 412 624 8826; fax: +1 412 624 5407.
E-mail address: jmg174@pitt.edu, jeffgirard@gmail.com (J.M. Girard).

http://dx.doi.org/10.1016/j.patrec.2014.10.004
0167-8655/

status (i.e., likelihood of divorce), and even years lived [25–29]. It is likely that research has only begun to scratch the surface of what might be learned from expressions' intensities.

Intensity estimation is also critical to the modeling of an expression's temporal dynamics (i.e., changes in intensity over time). Temporal dynamics is a relatively new area of study, but has already been linked to expression interpretation, person perception, and psychopathology. For example, the speed with which a smile onsets and offsets has been linked to interpretations of the expression's meaning and authenticity [30], as well as to ratings of the smiling person's attractiveness and personality [31]. Expression dynamics have also been found to be behavioral markers of depression, schizophrenia, and obsessive-compulsive disorder [32–34].

Efforts in automatic facial expression analysis have focused primarily on the detection of AU occurrence [3], rather than the estimation of AU intensity. In shape-based approaches to automatic facial expression analysis, intensity dynamics can be measured directly from the displacement of facial landmarks [35,36]. Shape-based approaches, however, are especially vulnerable to registration error [37], which is common in naturalistic settings. Appearance-based approaches are more robust to registration error, but require additional steps to estimate intensity. We address the question of how to estimate intensity from appearance features.

In an early and influential work on this topic, Bartlett et al. [38] applied standard binary expression detection techniques to estimate expressions' peak intensity. This and subsequent work [39,40] encouraged the use of the margins of binary-trained maximum margin classifiers as proxies for facial expression intensity. The assumption underlying this practice is that the classifier's decision value will be positively correlated with the expression's intensity. However, this assumption is theoretically problematic because nothing in the formulation of a maximum margin classifier guarantees such a correlation [41]. Indeed, many factors other than intensity may affect a data point's decision value, such as its typicality in the training set, the presence of other facial actions, and the recording conditions (e.g., illumination, pose, noise). The decision-value-as-intensity heuristic is purely an assumption about the data. The current study tests this assumption empirically, and compares it with the more labor-intensive but theoretically-informed approaches of training multiclass and regression models using intensity ground truth.

1.1. Previous work

Since Bartlett et al. [38], many studies have used classifier decision values to estimate expression intensity [39–48]. However, only a few of them have quantitatively evaluated their performance by comparing their estimations to manual (i.e., "ground truth") coding. Several studies [40,45,46] found that decision value and expression intensity were positively correlated during posed expressions. However, such correlations have typically been lower during spontaneous expressions. In a highly relevant study, Whitehill et al. [45] focused on the estimation of spontaneous smile intensity and found a high correlation between decision value and smile intensity. However, this was in five short video clips and it is unclear how the ground truth intensity coding was obtained.

Recent studies have also used methods other than the decision-value-as-intensity heuristic for intensity estimation, such as regression [46,49–52] and multiclass classifiers [53–55]. These studies have found that the predictions of support vector regression models and multiclass classifiers were highly correlated with expression intensity during both posed and spontaneous expressions. Finally, several studies [56–58] used extracted features to estimate expression intensity directly. For example, Messinger et al. [58] found that mouth radius was highly correlated with spontaneous smile intensity in five video clips.

Very few studies have compared different estimation methods using the same data and performance evaluation methods. Savran et al. [46] found that support vector regression outperformed the decision values of binary support vector machine classifiers on the intensity estimation of posed expressions. Ka Keung and Yangsheng [49] found that support vector regression outperformed cascading neural networks on the intensity estimation of posed expressions, and Dhall and Goecke [50] found that Gaussian process regression outperformed both kernel partial least squares and support vector regression on the intensity estimation of posed expressions. Yang et al. [41] also compared decision values with an intensity-trained model, but used their outputs to rank images by intensity rather than to estimate intensity.

Much of the previous work has been limited in three ways. First, many studies [57,50,49,41] adopted a message-based approach, which is problematic for the reasons described earlier. Second, the majority of this work [57,50,49,46,41] focused on posed expressions, which limits the external validity and generalizability of their findings. Third, most of these studies were limited in terms of the ground truth they compared their estimations to. Some studies [38–40] only coded expressions' peak intensities, while others [53,58,54,45] obtained frame-level ground truth, but only for a handful of subjects. Without a large amount of expert-coded, frame-level ground truth, it is impossible to truly gauge the success of an automatic intensity estimation system.

1.2. The current study

The current study challenges the use of binary classifier decision values for the estimation of expression intensity. Primarily, we hypothesize that intensity-trained (i.e., multiclass and regression) models will outperform binary-trained (i.e., two-class) models for expression intensity estimation. Secondarily, we hypothesize that intensity-trained models will offer a smaller but significant boon to binary expression detection over binary-trained models.

We compared these approaches using multiple methods for feature extraction and dimensionality reduction, using the same data and the same performance evaluation methods. We also improve upon previous work by using a sign-based approach, two large datasets of spontaneous expressions, and expert-coded ground truth. Smiles were chosen for this in-depth analysis because they are the most commonly occurring facial expression [59], are implicated in affective displays and social signaling [60,61], and appear in much of the previous work on both automatic intensity estimation and the psychological exploration of facial expression intensity.

2. Methods

2.1. Participants and data

In order to increase the sample size and explore the generalizability of the findings, data were drawn from two separate datasets. Both datasets recorded and FACS coded participant facial behavior during a non-scripted, spontaneous dyadic interaction. They differ in terms of the context of the interaction, the demographic makeup of the sample, constraints placed upon data collection (e.g., illumination, frontality, and head motion), base rates of smiling, tracking, and inter-observer reliability of manual FACS coding. Because of how its segments were selected, the BP4D database also had more frequent and intense smiles.

2.1.1. BP4D database

FACS coded video was available for 30 adults (50% female, 50% white, mean age 20.7 years) from the Binghamton–Pittsburgh 4D (BP4D) spontaneous facial expression database [62]. Participants were filmed with both a 3D dynamic face capturing system and a 2D

Fig. 1. Smile (AU 12) intensity levels from no contraction (left) to maximum contraction (right).

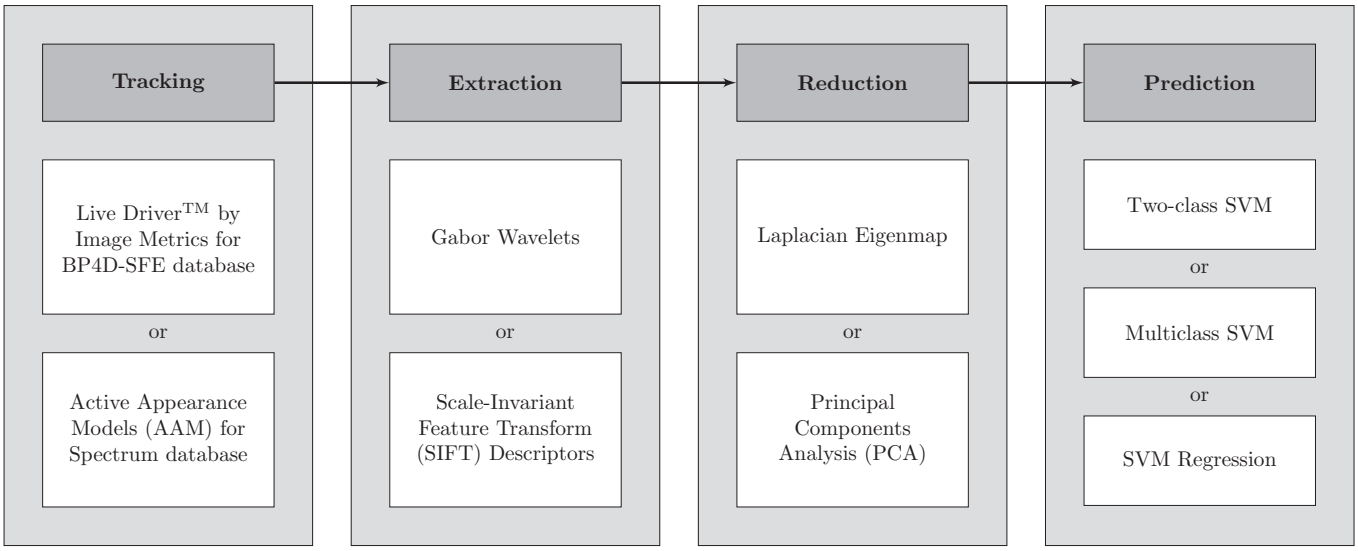

Fig. 2. Techniques used for automatic expression annotation.

frontal camera (520 × 720 pixel resolution) while engaging in eight tasks designed to elicit emotions such as anxiety, surprise, happiness, embarrassment, fear, pain, anger, and disgust. Facial behavior from the 20-s segment with the most frequent and intense facial expressions from each task was coded from the 2D video. The BP4D database is publicly available.

2.1.2. Spectrum database

FACS coded video was available for 33 adults (67.6% female, 88.2% white, mean age 41.6 years) from the Spectrum database [8]. The participants suffered from major depressive disorder [63] and were recorded during clinical interviews to assess symptom severity over the course of treatment [64]. A total of 69 interviews were recorded using four hardware-synchronized analogue cameras. Video from a camera roughly 15° to the participant's right was digitized into 640 × 480 pixel arrays for analysis. Facial behavior during the first three interview questions (about depressed mood, feelings of guilt, and suicidal ideation) was coded; these segments were an average of 100 s long. The Spectrum database is not publicly available due to confidentiality restrictions.

2.2. Manual expression annotation

2.2.1. AU occurrence

For both the BP4D and Spectrum databases, participant facial behavior was manually FACS coded from video by certified coders. Inter-observer agreement – the degree to which coders saw the same AUs in each frame – was quantified using F_1 score [65]. For the BP4D database, 34 commonly occurring AU were coded from onset to offset; inter-observer agreement for AU 12 occurrence was $F_1 = 0.96$. For the Spectrum database, 17 commonly occurring AU were coded from onset to offset, with expression peaks also coded; inter-observer agreement for AU 12 occurrence was $F_1 = 0.71$. For both datasets, onsets and offsets were converted to frame-level occurrence (i.e., present or absent) codes for AU 12.

2.2.2. AU intensity

The manual FACS coding procedures described earlier were used to identify the temporal location of AU 12 events. Separate video clips of each event were generated and coded for intensity by certified coders using custom continuous measurement software. This coding involved assigning each video frame a label of "no smile" or "A" through "E" representing trace through maximum intensity (Fig. 1) as defined by the FACS manual [18]. Inter-observer agreement was quantified using intraclass correlation (ICC) [66]. Ten percent of clips were independently coded by a second certified FACS coder; inter-observer agreement was ICC = 0.92.

2.3. Automatic expression annotation

Smiles were automatically coded for both occurrence and intensity using each combination of the techniques listed in Fig. 2 for tracking, extraction, reduction, and prediction.

2.3.1. Tracking

Facial landmark points indicate the location of important facial components (e.g., eye and lip corners). For the BP4D database, sixty-four facial landmarks were tracked in each video frame using Live Driver™ from Image Metrics [67]. Overall, 4% of video frames were untrackable, mostly due to occlusion or extreme out-of-plane rotation. A global normalizing (i.e., similarity) transformation was applied to the data for each video frame to remove variation due to rigid head motion. Finally, each image was cropped to the area surrounding the detected face and scaled to 128 × 128 pixels.

For the Spectrum database, sixty-six facial landmarks were tracked using active appearance models (AAM) [68]. AAM is a powerful approach that combines the shape and texture variation of an image into a single statistical model. Approximately 3% of video frames were manually annotated for each subject and then used to build the AAMs. The frames then were automatically aligned using a gradient-descent AAM fitting algorithm [69]. Overall, 9% of frames were untrackable, again mostly due to occlusion and rotation. The same normalization

procedures used on the Live Driver landmarks were also used on the AAM landmarks. Additionally, because AAM includes landmark points along the jawline, we were able to remove non-face information from the images using a convex hull algorithm.

2.3.2. Extraction

Two types of appearance features were extracted from the tracked and normalized faces. Following previous work on expression detection [37] and intensity estimation [38,46], Gabor wavelets [70,71] were extracted in localized regions surrounding each facial landmark point. Gabor wavelets are biologically-inspired filters, operating in a similar fashion to simple receptive fields in mammalian visual systems [72]. They have been found to be robust to misalignment and changes in illumination [73]. By applying a filter bank of eight orientations and five scales (i.e., 17, 23, 33, 46, 65 pixels) at each localized region, specific changes in facial texture and orientation (which map onto facial wrinkles, folds, and bulges) were quantified. Scale-invariant feature transform (SIFT) descriptors [74,75] were also extracted in localized regions surrounding each facial landmark point. SIFT descriptors are partially invariant to illumination changes. By applying a geometric descriptor to each facial landmark, changes in facial texture and orientation were quantified.

2.3.3. Reduction

Both types of features exhibited high dimensionality, which makes classification/regression difficult and resource-intensive problems. Two approaches for dimensionality reduction were compared on their ability to yield discriminant features for classification. For each model, only one of these approaches was used. The sample and feature sizes were motivated by the computational limitations imposed by each method.

Laplacian Eigenmap [76] is a nonlinear technique used to find the low dimensional manifold that the original (i.e., high dimensional) feature data lies upon. Following recent work by Mahoor et al. [53], supervised Laplacian Eigenmaps were trained on a randomly selected sample of 2500 frames and used in conjunction with spectral regression [77]. Two manifolds were trained for the data: one using two classes (corresponding to FACS occurrence codes) and another using six classes (corresponding to the FACS intensity codes). The two-class manifolds were combined with the two-class models and the six-class manifolds were combined with the multiclass and regression models (described below). The Gabor and SIFT features were each reduced to 30 dimensions per video frame using this technique.

Principal component analysis (PCA) [78] is a linear technique used to project a feature vector from a high dimensional space into a low dimensional space. Unsupervised PCA was used to find the smallest number of dimensions that accounted for 95% of the variance in a randomly selected sample of 100,000 frames. This technique reduced the Gabor features to 162 dimensions per video frame and reduced the SIFT features to 362 dimensions per video frame.

2.3.4. Prediction

Three techniques for supervised learning were used to predict the occurrence and intensity of smiles using the reduced features. Two-class models were trained on the binary FACS occurrence codes, while multiclass and regression models were trained on the FACS intensity codes. Data from the two databases were not mixed and contributed to separate models.

Following previous work on binary expression detection [79,45], two-class support vector machines (SVM) [80] were used for binary classification. We used a kernel SVM with a radial basis function kernel in all our approaches. SVMs were trained using two classes corresponding to the FACS occurrence codes described earlier. Training sets were created by randomly sampling 10,000 frames with roughly equal representation for each class. The choice of sample size was motivated by the computational limitations imposed by model training during cross-validation. Classifier and kernel parameters (i.e., C and γ, respectively) were optimized using a "grid-search" procedure [81] on a separate validation set. The decision values of the SVM models were fractions corresponding to the distance of each frame's high dimensional feature point from the class-separating hyperplane. These values were used for smile intensity estimation and also discretized using the standard SVM threshold of zero to provide predictions for binary smile detection (i.e., negative values were labeled absence of AU 12 and positive values were labeled presence of AU 12). Some researchers have proposed converting the SVM decision value to a pseudo-probability using a sigmoid function [82], but because the SVM training procedure is not intended to encourage this, it can result in a poor approximation of the posterior probability [83].

Following previous work on expression intensity estimation using multiclass classifiers [53–55], the SVM framework was extended for multiclass classification using the "one-against-one" technique [84]. In this technique, if k is the number of classes, then $k(k-1)/2$ subclassifiers are constructed and each one trains data from two classes; classification is then resolved using a subclassifier voting strategy. Multiclass SVMs were trained using six classes corresponding to the FACS intensity codes described earlier. Training sets were created by randomly sampling 10,000 frames with roughly equal representation for each class. Classifier and kernel parameters (i.e., C and γ, respectively) were optimized using a "grid-search" procedure [81] on a separate validation set. The output values of the multiclass classifiers were integers corresponding to each frame's estimated smile intensity level. These values were used for smile intensity estimation and also discretized to provide predictions for binary smile detection (i.e., values of 0 were labeled absence of AU 12 and values of 1 through 5 were labeled presence of AU 12).

Following previous work on expression intensity estimation using regression [46,49–52], epsilon support vector regression (ε-SVR) [80] was used. As others have noted [46], ε-SVR is appropriate to expression intensity estimation because its ε-insensitive loss function is robust and generates a smooth mapping. ε-SVRs were trained using a metric derived from the FACS intensity codes described earlier. The intensity scores of "A" through "E" were assigned a discrete numerical value from 1 to 5, with "no smile" assigned the value of 0. Although this mapping deviates from the non-metric definition of AU intensity in the FACS manual, wherein the range of some intensity scores is larger than others, it enables us to provide a more efficient computational model that works well in practice. Training sets were created by randomly sampling 10,000 frames with roughly equal representation for each class. Model and kernel parameters (i.e., C and γ, respectively) were optimized using a "grid-search" procedure [81]; the epsilon parameter was left at the default value ($\varepsilon = 0.1$). The output values of the regression models were fractions corresponding to each frame's estimated smile intensity level. This output was used for smile intensity estimation and also discretized using a threshold of 0.5 (so that low numbers rounded down) to provide predictions for binary smile detection.

It is important to note the differences between the three approaches that were tested. In the two-class approach, the five intensity levels of a given AU were collapsed into a single positive class. In the multiclass approach, each of the intensity levels was treated as a mutually-exclusive but unrelated class. Finally, in the regression approach, each intensity level was assigned a discrete numerical value and modeled on a continuous dimension. These differences are clarified by examination of the respective loss functions. The penalty of incorrect estimation in the regression approach is based on the distance between the prediction value y and the ground truth label t, given a buffer area of size ε (Eq. (1)). In contrast, the penalty of misclassification in the two-class approach is based on the classifier's decision value y (Eq. (2)). In this case, the ground truth label is

collapsed into present at any intensity level ($t = 1$) or absent ($t = -1$). As an extension of the two-class approach, similar phenomena occur for the multiclass approach.

$$l_\varepsilon(y) = \begin{cases} 0 & \text{if } |y - t| < \varepsilon \\ |y - t| - \varepsilon & \text{otherwise.} \end{cases} \quad (1)$$

$$l(y) = \begin{cases} 0 & \text{if } (1 - y \cdot t) < 0 \\ 1 - y \cdot t & \text{otherwise} \end{cases} \quad (2)$$

2.3.5. Cross-validation

To prevent model over-fitting, stratified k-fold cross-validation [85] was used. Cross-validation procedures typically involve partitioning the data and iterating through the partitions such that all the data is used but no iteration is trained and tested on the same data. *Stratified* cross-validation procedures ensure that the resultant partitions have roughly equal distributions of the target class (in this case AU 12). This property is desirable because many performance metrics are highly sensitive to class skew [86]. By using the same partitions across methods, the randomness introduced by repeated repartitioning can also be avoided.

Each video segment was assigned to one of five partitions. Segments, rather than participants, were assigned to partitions to allow greater flexibility for stratification. However, this choice allowed independent segments from the same participant to end up in multiple partitions. As such, this procedure was a less conservative control for generalizability. For each iteration of the cross-validation procedure, three partitions were used for training, one partition was used for validation (i.e., optimization), and one partition was used for testing.

2.4. Performance evaluation

The majority of previous work on expression intensity estimation has utilized the Pearson product–moment correlation coefficient (PCC) to measure the correlation between intensity estimations and ground truth coding. PCC is invariant to linear transformations, which is useful when using estimations that differ in scale and location from the ground truth coding (e.g., decision values). However, this same property is problematic when the estimations *are* similar to the ground truth (e.g., multiclass classifier predictions), as it introduces an undesired handicap. For instance, a classifier that always estimates an expression to be two intensity levels stronger than it is will have the same PCC as a classifier that always estimates the expression's intensity level correctly.

For this reason, we performed our analyses using another performance metric that grants more control over its relation to linear transformations: the intraclass correlation coefficient (ICC) [66]. Eq. (3) was used to compare the multiclass SVM and ε-SVR approaches to the manual intensity annotations as their outputs were consistently scaled; it was calculated using between-target mean squares (BMS) and within-target mean squares (WMS). Eq. (4) was used for the decision value estimations because it takes into account differences in scale and location; it was calculated using BMS and residual sum of squares (EMS). For both formulas, k is equal to the number of coding sources being compared; in the current study, there are two: the automatic and manual codes. ICC ranges from −1 to +1, with more positive values representing higher agreement.

$$\mathrm{ICC}(1,1) = \frac{\mathrm{BMS} - \mathrm{WMS}}{\mathrm{BMS} + (k-1)\mathrm{WMS}} \quad (3)$$

$$\mathrm{ICC}(3,1) = \frac{\mathrm{BMS} - \mathrm{EMS}}{\mathrm{BMS} + (k-1)\mathrm{EMS}} \quad (4)$$

The majority of previous work on binary expression detection has utilized receiver operating characteristic (ROC)analysis. When certain assumptions are met, the area under the curve (AUC) is equal to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance [87]. The fact that AUC captures information about the entire distribution of decision points is a benefit of the measure, as it removes the subjectivity of threshold selection. However, in the case of automatic expression annotation, a threshold *must* be chosen in order to create predictions that can be compared with ground truth coding. In light of this issue, we performed our analyses using a threshold-specific performance metric: the F_1 score, which is the harmonic mean of precision and recall (Eq. (5)) [65]. F_1 score is computed using true positives (TP), false positives (FP), and false negatives (FN); it ranges from 0 to 1, with higher values representing higher agreement between coders.

$$F_1 = \frac{2 \times \mathrm{TP}}{2 \times \mathrm{TP} + \mathrm{FN} + \mathrm{FP}} \quad (5)$$

Table 1
General linear model results for smile intensity estimation.

	ICC	F	p
Database			
BP4D	0.765	158.206	**0.00**
Spectrum	0.521		
Extraction			
Gabor	0.602	17.892	**0.00**
SIFT	0.684		
Reduction			
Laplacian	0.639	0.197	0.66
PCA	0.648		
Prediction			
Two-class	0.467[a]	83.360	**0.00**
Multiclass	0.739[b]		
Regression	0.724[b]		
Interaction effects			
Database × extraction		4.627	**0.03**
Database × reduction		3.958	**0.05**
Database × model		8.873	**0.00**
Reduction × model		13.391	**0.00**

Different superscripts indicate significant mean differences in ICC or F_1 score by Tukey HSK test ($p < .05$).

2.5. Data analysis

Main effects and interaction effects among the different methods were analyzed using two univariate general linear models [88] (one for binary smile detection and one for smile intensity estimation). F_1 and ICC were entered as the sole dependent variable in each model, and database, extraction type, reduction type, and classification type were entered as "fixed factor" independent variables. The direction of significant differences was explored using marginal means for all variables except for classification type. In this case, post-hoc Tukey HSD tests [88] were used to explore differences between the three types of classification.

3. Results

3.1. Smile intensity estimation

Across all methods and databases, the average intensity estimation performance was ICC = 0.64. However, performance varied widely between databases and methods, from a low of ICC = 0.23 to a high of ICC = 0.92.

The overall general linear model for smile intensity estimation was significant (Table 1). Main effects of database, extraction method, and supervised learning method were apparent. Intensity estimation performance was significantly higher for the BP4D database than for the Spectrum database, and intensity estimation performance using SIFT features was significantly higher than that using Gabor features. Intensity estimation performance using multiclass and regression models was significantly higher than that using the two-class approach.

Table 2
General linear model results for binary smile detection.

	F_1 score	F	p
Database			
BP4D	0.772	440.209	**0.00**
Spectrum	0.504		
Extraction			
Gabor	0.618	9.740	**0.00**
SIFT	0.658		
Reduction			
Laplacian	0.642	0.501	0.48
PCA	0.633		
Prediction			
Two-class	0.616[a]	4.175	**0.02**
Multiclass	0.661[b]		
Regression	0.636		
Interaction effects			
Reduction × model		5.753	**0.00**

Different superscripts indicate significant mean differences in ICC or F_1 score by Tukey HSK test ($p < .05$).

There was no significant difference in performance between Laplacian Eigenmap and PCA for reduction.

These main effects were qualified by four significant interaction effects. First, the difference between SIFT features and Gabor features was greater in the Spectrum database than in the BP4D database. Second, while Laplacian Eigenmap performed better in the Spectrum database, PCA performed better in the BP4D database. Third, while multiclass models performed better in the Spectrum database, regression models performed better in the BP4D database. Fourth, PCA reduction yielded higher intensity estimation performance when combined with two-class models, but lower performance when combined with multiclass and regression models.

3.2. Binary smile detection

Across all methods and databases, the average binary detection performance was $F_1 = 0.64$. However, performance varied between databases and methods, from a low of $F_1 = 0.40$ to a high of $F_1 = 0.81$.

The overall general linear model for binary smile detection was significant (Table 2). Main effects of database, extraction method, and supervised learning method were apparent. Detection performance on the BP4D database was significantly higher than that on the Spectrum database, and detection performance using SIFT features was significantly higher than that using Gabor features. Detection performance was significantly higher using multiclass models than using two-class models.

These main effects were qualified by a significant interaction effect between reduction method and supervised learning method. PCA reduction yielded higher detection performance when combined with two-class models, but lower detection performance when combined with multiclass and regression models. There was no main effect of dimensionality reduction method and no other interactions were significant.

3.3. Distribution of output values

The decision values of the best-performing two-class model for each database are presented in Fig. 4 as box plots [89]. The boxes represent the first and third quartiles of each smile intensity level, while the line within each box represents the median. The lines extending from each box represent data within 1.5 times the inter-quartile range of the lower and upper quartiles. The regression values of the best-performing regression model for each database are similarly presented in Fig. 5.

Examination of Fig. 4 reveals a slight right-leaning tendency, indicating that more positive SVM decision values are on average more likely to be higher intensity. However, there is substantial overlap between the distributions and a great deal of "clumping" between intensity levels; the distributions for levels 2 through 4 (i.e., "B" through "D") are very similar for the BP4D dataset, while the distributions for levels 3 through 5 (i.e., "C" through "E") are very similar for the Spectrum dataset. Finally, the observed range of values spans −3.6 to 5.5 for BP4D and −8.4 to 8.1 for Spectrum.

Examination of Fig. 5 reveals a stepped and right-leaning pattern, indicating that more positive regression values are on average more likely to be higher intensity. There is some overlapping between the distributions, although the inter-quartile ranges for each group are largely distinct. One exception to this is clumping for levels 4 and 5 (i.e., "D" and "E"), especially for the Spectrum dataset. The observed range of values spans −3.1 to 5.7 for BP4D and −1.2 to 4.8 for Spectrum.

4. Discussion

4.1. Smile intensity estimation

Intensity estimation performance varied between databases, feature extraction methods, and supervised learning methods. Performance was higher in the BP4D database than in the Spectrum database. It is not surprising that performance differed between the two databases, given how much they differed in terms of participant demographics, social context, and image quality. Further experimentation will be required to pinpoint exactly what differences between the databases contributed to this drop in performance, but we suspect that illumination conditions, frontality of camera placement, and participant head pose were involved. It is also possible that the participants in the Spectrum database were more difficult to analyze due to their depressive symptoms. Previous research has found that nonverbal behavior (and especially smiling) changes with depression symptomatology (e.g., [90]). There were also differences between databases in terms of social context that likely influenced smiling behavior; Spectrum was recorded during a clinical interview about depression symptoms, while BP4D was recorded during tasks designed to elicit specific and varied emotions. Participants in the Spectrum database smiled less frequently (20.5% of frames) and less intensely (average intensity 1.5) than did participants in the BP4D database (56.4% of frames and average intensity 2.4). The inter-observer reliability for manual smile occurrence coding was also higher in the BP4D database ($F_1 = 0.96$) than in the Spectrum database ($F_1 = 0.71$). These differences may have affected the difficulty of smile intensity estimation.

More surprising was that intensity estimation performance was higher for SIFT features than for Gabor features. This finding is encouraging from a computational load perspective, considering the toolbox implementation of SIFT used in this study [75] was many times faster than our custom implementation of Gabor. However, it is possible that SIFT was particularly well-suited to our form of registration with dense facial landmarking. Although we did not test this hypothesis in the current study, it would have been interesting to compare these two methods of feature extraction in conjunction with a method of registration using sparse landmarking (e.g., holistic face detection or eye tracking). It is also important to note that the difference between SIFT and Gabor features was larger in the Spectrum database than in BP4D.

For dimensionality reduction, intensity estimation performance was not significantly different between Laplacian Eigenmap and PCA. This may be an indication that the features used in this study were linearly separable and that manifold learning was unnecessary. This finding is also encouraging from a computational load perspective, as PCA is a much faster and simpler technique. However, it is important to note that the success of each dimensionality reduction technique depended on the database and on the classification method used. Laplacian Eigenmap was better suited to the Spectrum database,

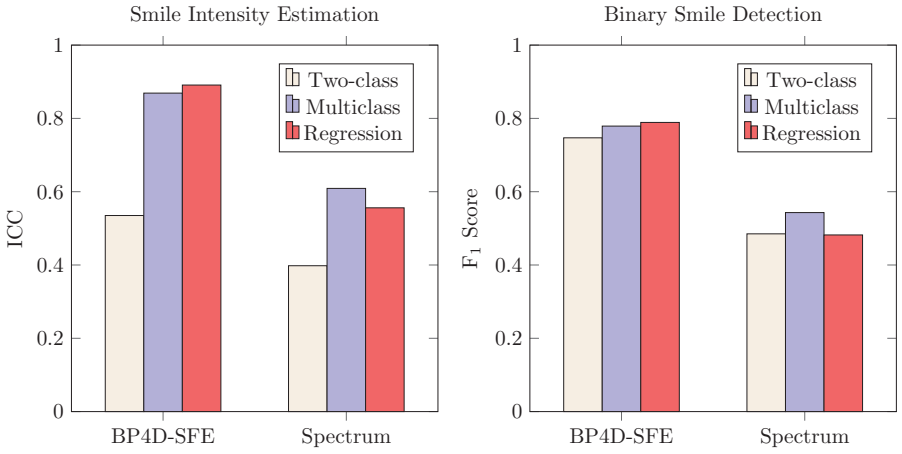

Fig. 3. Average performance for three approaches to supervised learning in two databases.

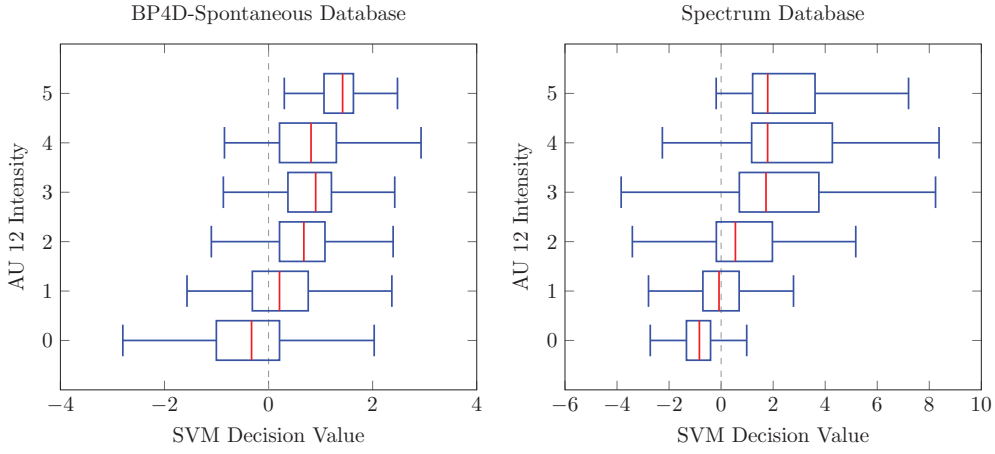

Fig. 4. SVM decision values by AU 12 intensity in two databases.

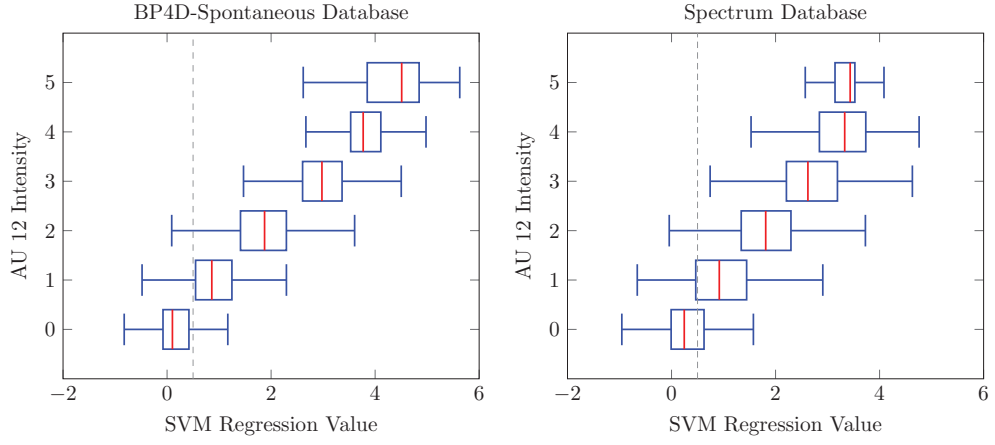

Fig. 5. SVM regression values by AU 12 intensity in two databases.

multiclass models, and regression models; while PCA was better suited to the BP4D database and two-class models.

Most relevant to our main hypothesis are the findings regarding supervised learning method. In line with our hypothesis that the decision-value-as-intensity heuristic is flawed in practice, the intensity-trained multiclass and regression models performed significantly better at intensity estimation than the decision values of two-class models. However, it is important to note that the intensity estimation performance yielded by binary-trained models was not negligible. Consistent with previous reports, decision values showed a low to moderate correlation with smile intensity.

4.2. Binary smile detection

Binary detection performance also varied between databases, feature extraction methods, and supervised learning methods. These differences were very similar to those for expression intensity

estimation. Binary detection performance was higher for the BP4D database than for the Spectrum database, higher for SIFT features than for Gabor features, and no difference between Laplacian Eigenmap and PCA for reduction.

Examination of the decision values for the two-class models (Fig. 4) reveals that there is substantial overlap between the distributions of "no smile" and "A" level smiles. Furthermore, the first quartile of "A" level smiles is negative in both datasets and therefore contributes to substantial misclassification. Examination of the confusion matrices for the multiclass models reveals a similar pattern: detecting "trace" level smiles is difficult.

Surprisingly, detection performance was higher for multiclass models than for two-class models; this difference was modest but statistically significant (Fig. 3). This suggests that the best classifier for binary detection is not necessarily the one trained on binary labels. As far as we know, this is the first study to attempt binary expression detection using an intensity-trained classifier. Although collecting frame-level intensity ground truth is labor-intensive, our findings indicate that this investment is worthwhile for both binary expression detection and expression intensity estimation.

4.3. Conclusions

We provide empirical evidence that the decision-value-as-intensity heuristic is flawed in practice as well as in theory. Unfortunately, there are no shortcuts when it comes to estimating smile intensity: researchers must take the time to collect and train on intensity ground truth. However, if they do so, high reliability with expert human FACS coders can be achieved. Intensity-trained multiclass and regression models outperformed binary-trained classifier decision values on smile intensity estimation across multiple databases and methods for feature extraction and dimensionality reduction. Multiclass models even outperformed binary-trained classifiers on binary smile detection. Examination of the distribution of classifier decision values indicates that there is substantial overlap between smile intensity levels and that low intensity smiles are frequently confused with non-smiles. A much cleaner set of distributions can be achieved by training a regression model explicitly on the intensity levels.

4.4. Limitations and future directions

The primary limitations of the current study were that it focused on a single facial expression and supervised learning framework. Future work should explore the generalizability of these findings by comparing different methods for supervised learning and other facial expressions. Another limitation is the divergence between the number of reduced features yielded by Laplacian Eigenmap and PCA. Future work might standardize the number of features or forego dimensionality reduction entirely (at the cost of computation time or kernel complexity). Finally, future work would benefit from a comparison of additional techniques for facial landmark registration, feature extraction, and dimensionality reduction.

Acknowledgments

The authors wish to thank Nicole Siverling, Laszlo A. Jeni, Wen-Sheng Chu, Dean P. Rosenwald, Shawn Zuratovic, Kayla Mormak, and anonymous reviewers for their generous assistance. Research reported in this publication was supported by the U.S. National Institute of Mental Health under award MH096951 and the U.S. National Science Foundation under award 1205195.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at doi:http://dx.doi.org/10.1016/j.patrec.2014.10.004

References

[1] C. Darwin, The Expression of Emotions in Man and Animals, 3rd ed. Oxford University, New York, 1872.
[2] P. Ekman, W.V. Friesen, S. Ancoli, Facial signs of emotional experience, J. Pers. Soc. Psychol. 396 (1980) 1125–1134.
[3] Z. Zeng, M. Pantic, G.I. Roisman, T.S. Huang, A survey of affect recognition methods: audio, visual, and spontaneous expressions, IEEE Trans. Pattern Anal. Mach. Intell. 31(1) (2009) 39–58.
[4] A.J. Fridlund, The behavioral ecology and sociality of human faces, in: M.S. Clark (Ed.), Review of Personality Social Psychology, Sage Publications, 1992, pp. 90–121.
[5] D. Keltner, Evidence for the distinctness of embarrassment, shame, and guilt: a study of recalled antecedents and facial expressions of emotion, Cogn. Emotion 10(2) (1996) 155–172.
[6] G.C. Littlewort, M.S. Bartlett, K. Lee, Automatic coding of facial expressions displayed during posed and genuine pain, Image Vis. Comput. 27(12) (2009) 1797–1803.
[7] K.M. Prkachin, P.E. Solomon, The structure, reliability and validity of pain expression: evidence from patients with shoulder pain, Pain 139(2) (2008) 267–274.
[8] J.F. Cohn, T.S. Kruez, I. Matthews, Y. Ying, N. Minh Hoai, M.T. Padilla, Z. Feng, F. De la Torre, Detecting depression from facial actions and vocal prosody, in: Proceedings of the IEEE International Conference on Affective Computing and Intelligent Interaction, Amsterdam, 2009, pp. 1–7.
[9] J.M. Girard, J.F. Cohn, M.H. Mahoor, S. Mavadati, D.P. Rosenwald, Social risk and depression: evidence from manual and automatic facial expression analysis, in: IEEE International Conference on Automatic Face & Gesture Recognition, 2013.
[10] R. Cowie, E. Douglas-Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, J.G. Taylor, Emotion recognition in human-computer interaction, IEEE Signal Process. Mag. 18(1) (2001) 32–80.
[11] M. Pantic, L.J.M. Rothkrantz, Toward an affect-sensitive multimodal human-computer interaction, Proc. IEEE 91(9) (2003) 1370–1390.
[12] I.S. Pandzic, R. Forchheimer, The Origins of the MPEG-4 Facial Animation Standard, Wiley Online Library, 2002.
[13] J.F. Cohn, P. Ekman, Measuring facial action by manual coding, facial EMG, and automatic facial image analysis, in: J.A. Harrigan, R. Rosenthal, K.R. Scherer (Eds.), The New Handbook of Nonverbal Behavior Research, Oxford University Press, New York, NY, 2005, pp. 9–64.
[14] P. Ekman, Basic emotions, in: T. Dalagleish, M. Power (Eds.), Handbook of Cognition and Emotion, John Wiley & Sons, UK, 1999, pp. 45–60.
[15] C.E. Izard, The Face of Emotion, Appleton-Century-Crofts, New York, NY, 1971.
[16] P. Ekman, Darwin, deception, and facial expression, Ann. N. Y. Acad. Sci. 1000(1) (2003) 205–221.
[17] L.F. Barrett, B. Mesquita, M. Gendron, Context in emotion perception, Curr. Dir. Psychol. Sci. 20(5) (2011) 286–290.
[18] P. Ekman, W.V. Friesen, J. Hager, Facial Action Coding System: A Technique for the Measurement of Facial Movement, Research Nexus, Salt Lake City, UT, 2002.
[19] P. Ekman, E.L. Rosenberg, What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS), 2nd ed. Oxford University Press, New York, NY, 2005.
[20] P. Ekman, R.J. Davidson, W.V. Friesen, The Duchenne smile: emotional expression and brain physiology: II, J. Pers. Soc. Psychol. 58(2) (1990) 342–353.
[21] M.G. Frank, P. Ekman, W.V. Friesen, Behavioral markers and recognizability of the smile of enjoyment, J. Pers. Soc. Psychol. 64(1) (1993) 83–93.
[22] L.I. Reed, M.A. Sayette, J.F. Cohn, Impact of depression on response to comedy: a dynamic facial coding analysis, J. Abnorm. Psychol. 116(4) (2007) 804–809.
[23] A.J. Fridlund, Sociality of solitary smiling: potentiation by an implicit audience, J. Pers. Soc. Psychol. 60(2) (1991) 12.
[24] U. Hess, R. Banse, A. Kappas, The intensity of facial expression is determined by underlying affective state and social situation, J. Pers. Soc. Psychol. 69(2) (1995) 280–288.
[25] E.L. Abel, M.L. Kruger, Smile intensity in photographs predicts longevity, Psychol. Sci. 21(4) (2010) 542–544.
[26] L. Harker, D. Keltner, Expressions of positive emotion in Women's College yearbook pictures and their relationship to personality and life outcomes across adulthood, J. Pers. Soc. Psychol. 80(1) (2001) 112–124.
[27] M. Hertenstein, C. Hansel, A. Butts, S. Hile, Smile intensity in photographs predicts divorce later in life, Motiv. Emotion 33(2) (2009) 99–105.
[28] C. Oveis, J. Gruber, D. Keltner, J.L. Stamper, W.T. Boyce, Smile intensity and warm touch as thin slices of child and family affective style, Emotion 9(4) (2009) 544–548.
[29] J.P. Seder, S. Oishi, Intensity of smiling in Facebook photos predicts future life satisfaction, Soc. Psychol. Pers. Sci. 3(4) (2012) 407–413.
[30] Z. Ambadar, J.F. Cohn, L.I. Reed, All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous, J. Nonverbal Behav. 33(1) (2009) 17–34.
[31] E. Krumhuber, A.S. Manstead, A. Kappas, Temporal aspects of facial displays in person and expression perception: the effects of smile dynamics, head-tilt, and gender, J. Nonverbal Behav. 31(1) (2007) 39–56.
[32] R. Mergl, M. Vogel, P. Mavrogiorgou, C. Gobel, M. Zaudig, U. Hegerl, G. Juckel, Kinematical analysis of emotionally induced facial expressions in patients with obsessive-compulsive disorder, Psychol. Med. 33(8) (2003) 1453–1462.
[33] R. Mergl, P. Mavrogiorgou, U. Hegerl, G. Juckel, Kinematical analysis of emotionally induced facial expressions: a novel tool to investigate hypomimia in patients suffering from depression, J. Neurol. Neurosurg. Psychiatry 76(1) (2005) 138–140.

[34] G. Juckel, R. Mergl, A. Prassl, P. Mavrogiorgou, H. Witthaus, H.J. Moller, U. Hegerl, Kinematic analysis of facial behaviour in patients with schizophrenia under emotional stimulation by films with "Mr. Bean", Eur. Arch. Psychiatry Clin. Neurosci. 258(3) (2008) 186–191.
[35] M.F. Valstar, M. Pantic, Fully Automatic Facial Action Unit Detection and Temporal Analysis, in: Conference on Computer Vision and Pattern Recognition Workshops, 2006.
[36] M.F. Valstar, M. Pantic, Fully automatic recognition of the temporal phases of facial actions, IEEE Trans. Syst. Man Cybern. 42(1) (2012) 28–43.
[37] S.W. Chew, P. Lucey, S. Lucey, J. Saragih, J.F. Cohn, I. Matthews, S. Sridharan, In the pursuit of effective affective computing: the relationship between features and registration, IEEE Trans. Syst. Man Cybern. 42(4) (2012) 1006–1016.
[38] M.S. Bartlett, G. Littlewort, B. Braathen, T.J. Sejnowski, J.R. Movellan, A prototype for automatic recognition of spontaneous facial actions, in: S. Becker, K. Obermayer (Eds.), Advances in Neural Information Processing Systems, MIT Press, 2003.
[39] M.S. Bartlett, G. Littlewort, M.G. Frank, C. Lainscsek, I.R. Fasel, J.R. Movellan, Automatic recognition of facial actions in spontaneous expressions, J. Multimedia 1(6) (2006) 22–35.
[40] M.S. Bartlett, G. Littlewort, M.G. Frank, C. Lainscsek, I.R. Fasel, J.R. Movellan, Fully automatic facial action recognition in spontaneous behavior, in: IEEE International Conference on Automatic Face & Gesture Recognition, Southampton, 2006, pp. 223–230.
[41] P. Yang, L. Qingshan, D.N. Metaxas, RankBoost with l1 regularization for facial expression recognition and intensity estimation, in: IEEE International Conference on Computer Vision (2009) 1018–1025.
[42] G. Littlewort, M.S. Bartlett, I.R. Fasel, J. Susskind, J.R. Movellan, Dynamics of facial expression extracted automatically from video, Image Vis. Comput. 24(6) (2006) 615–625.
[43] J. Reilly, J. Ghent, J. McDonald, Investigating the dynamics of facial expression, Adv. Vis. Comput. (2006) 334–343.
[44] S. Koelstra, M. Pantic, Non-rigid registration using free-form deformations for recognition of facial actions and their temporal dynamics, in: IEEE International Conference on Automatic Face & Gesture Recognition, 2008, pp. 1–8.
[45] J. Whitehill, G. Littlewort, I.R. Fasel, M.S. Bartlett, J.R. Movellan, Toward practical smile detection, IEEE Trans. Pattern Anal. Mach. Intell. 31(11) (2009) 2106–2111.
[46] A. Savran, B. Sankur, M. Taha Bilge, Regression-based intensity estimation of facial action units, Image Vis. Comput. (2011).
[47] K. Shimada, T. Matsukawa, Y. Noguchi, T. Kurita, Appearance-based smile intensity estimation by cascaded support vector machines, in: ACCV 2010 Workshops, 2011, pp. 277–286.
[48] K. Shimada, Y. Noguchi, T. Kurita, Fast and robust smile intensity estimation by cascaded support vector machines, Int. J. Comput. Theory Eng. 5(1) (2013) 24–30.
[49] L. Ka Keung, X. Yangsheng, Real-time estimation of facial expression intensity, in: IEEE International Conference on Robotics and Automation, Vol. 2, 2003, pp. 2567–2572.
[50] A. Dhall, R. Goecke, Group expression intensity estimation in videos via Gaussian processes, International Conference on Pattern Recognition. (2012) 3525–3528.
[51] S. Kaltwang, O. Rudovic, M. Pantic, Continuous pain intensity estimation from facial expressions, in: G. Bebis, R. Boyle, B. Parvin, D. Koracin, C. Fowlkes, S. Wang, M.H. Choi, S. Mantler, J. Schulze, D. Acevedo, K. Mueller, M. Papka (Eds.), Advances in Visual Computing, Volume 7432 of Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 2012, pp. 368–377.
[52] L.A. Jeni, J.M. Girard, J.F. Cohn, F. De la Torre, Continuous AU intensity estimation using localized, sparse facial feature space, in: Proceedings of the IEEE International Conference on Automated Face & Gesture Recognition Workshops, IEEE, 2013, pp. 1–7.
[53] M.H. Mahoor, S. Cadavid, D.S. Messinger, J.F. Cohn, A framework for automated measurement of the intensity of non-posed Facial Action Units, in: Computer Vision and Pattern Recognition Workshops, Miami, 2009, pp. 74–80.
[54] D.S. Messinger, M.H. Mahoor, S.M. Chow, J.F. Cohn, Automated measurement of facial expression in infant-mother interaction: a pilot study, Infancy 14(3) (2009) 285–305.
[55] S.M. Mavadati, M.H. Mahoor, K. Bartlett, P. Trinh, J.F. Cohn, DISFA: a spontaneous facial action intensity database, IEEE Trans. Affect. Comput. 4 (2013) 151–160.
[56] J.F. Cohn, K.L. Schmidt, The timing of facial motion in posed and spontaneous smiles, Int. J. Wavelets Multiresolution Inform. Process. 2(2) (2004) 57–72.
[57] O. Deniz, M. Castrillon, J. Lorenzo, L. Anton, G. Bueno, Smile detection for user interfaces, in: G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, F. Porikli, J. Peters, J. Klosowski, L. Arns, Y.K. Chun, T.M. Rhyne, L. Monroe (Eds.), Advances in Visual Computing, Volume 5359 of Lecture Notes in Computer Science, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 602–611.
[58] D.S. Messinger, T.D. Cassel, S.I. Acosta, Z. Ambadar, J.F. Cohn, Infant smiling dynamics and perceived positive emotion, J. Nonverbal Behav. 32(3) (2008) 133–155.
[59] J.B. Bavelas, N. Chovil, Faces in dialogue, in: J.A. Russell, J.M. Fernandez-Dols (Eds.), The Psychology of Facial Expression, Cambridge University Press, New York, 1997, pp. 334–346.
[60] U. Hess, S. Blairy, R.E. Kleck, The influence of facial emotion displays, gender, and ethnicity on judgments of dominance and affiliation, J. Nonverbal Behav. 24(4) (2000) 265–283.
[61] U. Hess, R.B. Adams Jr, R.E. Kleck, Who may frown and who should smile? Dominance, affiliation, and the display of happiness and anger, Cognit. Emotion 19(4) (2005) 515–536.
[62] X. Zhang, L. Yin, J.F. Cohn, S. Canavan, M. Reale, A. Horowitz, P. Liu, J.M. Girard, BP4D-Spontaneous: a high-resolution spontaneous 3D dynamic facial expression database, Image Vis. Comput. 32(10) (2014) 692–706. 10.1016/j.imavis.2014.06.002
[63] American Psychiatric Association, Diagnostic and Statistical Manual of Mental Disorders, 4th ed., American Psychiatric Association, Washington, DC, 1994.
[64] M. Hamilton, Development of a rating scale for primary depressive illness, Br. J. Soc. Clin. Psychol. 6(4) (1967) 278–296.
[65] C.J. van Rijsbergen, Information Retrieval, 2nd ed. Butterworth, London, 1979.
[66] P.E. Shrout, J.L. Fleiss, Intraclass correlations: uses in assessing rater reliability, Psychol. Bull. 86(2) (1979) 420.
[67] Image Metrics, Live Driver SDK, 2013, URL: http://image-metrics.com/.
[68] T.F. Cootes, G.J. Edwards, C.J. Taylor, Active appearance models, IEEE Trans. Pattern Anal. Mach. Intell. 23(6) (2001) 681–685.
[69] I. Matthews, S. Baker, Active appearance models revisited, Int. J. Comput. Vis. 60(2) (2004) 135–164.
[70] J.G. Daugman, Complete discrete 2-D Gabor transforms by neural networks for image analysis and compression, IEEE Trans. Acoust. Speech Sig. Process. 36(7) (1988) 1169–1179.
[71] W. Fellenz, J.G. Taylor, N. Tsapatsoulis, S. Kollias, Comparing template-based, feature-based and supervised classification of facial expressions from static images, in: Proceedings of Circuits, Systems, Communications and Computers, 1999.
[72] J.P. Jones, L.A. Palmer, An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex, J. Neurophysiol. 58(6) (1987) 1233–1258.
[73] C. Liu, S. Louis, H. Wechsler, A Gabor feature classifier for face recognition, in: IEEE International Conference on Computer Vision, 2001, pp. 270–275.
[74] D.G. Lowe, Object recognition from local scale-invariant features, in: IEEE International Conference on Computer Vision (1999) 1150–1157.
[75] A. Vedali, B. Fulkerson, VLFeat: An open and portable library of computer vision algorithms, in: Proceedings of the ACM International Conference on Multimedia, 2010, pp. 1469–1472.
[76] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15(6) (2003) 1373–1396.
[77] D. Cai, X. He, W.V. Zhang, J. Han, Regularized locality preserving indexing via spectral regression, in: Conference on Information and Knowledge Management, ACM, 2007, pp. 741–750.
[78] I. Jolliffe, Principal component analysis, in: Encyclopedia of Statistics in Behavioral Science, John Wiley & Sons, Ltd., 2005.
[79] B. Fasel, J. Luettin, Automatic facial expression analysis: a survey, Pattern Recognit. 36(1) (2003) 259–275.
[80] V. Vapnik, The Nature of Statistical Learning Theory, Springer, New York, NY, 1995.
[81] C.W. Hsu, C.C. Chang, C.J. Lin, A practical guide to support vector classification, Technical Report, 2003.
[82] J.C. Platt, Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, in: A.J. Smola, P.J. Bartlett (Eds.), Advances in large margin classifiers, MIT Press, 1999, pp. 61–74.
[83] M. Tipping, Sparse Bayesian learning and the relevance vector machine, J. Mach. Learn. Res. 1 (2001) 211–244.
[84] C.W. Hsu, C.J. Lin, A comparison of methods for multiclass support vector machines, IEEE Trans. Neural Netw. 13(4) (2002) 415–425.
[85] S. Geisser, Predictive Inference, Chapman and Hall, New York, NY, 1993.
[86] L.A. Jeni, J.F. Cohn, F. De la Torre, Facing imbalanced data: recommendations for the use of performance metrics, in: International Conference on Affective Computing and Intelligent Interaction, 2013.
[87] T. Fawcett, An introduction to ROC analysis, Pattern Recognit. Lett. 27(8) (2006) 861–874.
[88] IBM Corp, IBM SPSS Statistics for Windows, Version 21.0, 2012.
[89] M. Frigge, D. Hoaglin, B. Iglewicz, Some implementations of the boxplot, Am. Stat. 43(1) (1989) 50–54.
[90] J.M. Girard, J.F. Cohn, M.H. Mahoor, S.M. Mavadati, Z. Hammal, D.P. Rosenwald, Nonverbal social withdrawal in depression: evidence from manual and automatic analyses, Image Vis. Comput. 32(10) (2014) 641–647. doi:10.1016/j.imavis.2013.12.007