Gender estimation based on smile-dynamics

Antitza Dantcheva, and François Brémond

***Abstract*—Automated gender estimation has numerous applications including video surveillance, human computer-interaction, anonymous customized advertisement and image retrieval. Most commonly, the underlying algorithms analyze the facial appearance for clues of gender. In this work we propose a novel method for gender estimation, namely the use of dynamic features gleaned from smiles and proceed to show that (a) facial dynamics can be used to improve appearance-based gender estimation, (b) that while for adult individuals appearance features are more accurate than dynamic features, for subjects under 18 years old facial dynamics can outperform appearance features. In addition, we fuse proposed dynamics-based approach with state-of-the-art appearance based algorithms, predominantly improving appearance-based gender estimation performance. Results show that smile-dynamics include pertinent and complementary to appearance gender information.**

***Keywords*—soft biometrics, gender estimation, facial dynamics.**

I. Introduction

Human facial analysis has engaged researchers in multiple fields including computer vision, biometrics, forensics, cognitive psychology and medicine. Interest in this topic has been fueled by scientific advances that provide insight into a persons identity, intent, attitude, aesthetics as well as health, solely based on their face images.

Besides establishing an individuals identity, ancillary information may also be gleaned from face images related to personal attributes such as gender, age and ethnicity. Gender and specifically automated gender estimation has been of specific interest for its broad application range, be it in surveillance [67], human computer-interaction, anonymous customized advertisement systems[1] or image retrieval systems [6], leading to numerous commercial applications[2][3][4]. Also, gender has been a prominent soft-biometric trait [21], [23], which can be employed (a) in fusion with other biometric traits to improve the matching accuracy of a biometric system [45], (b) in fusion with other soft biometrics for person authentication [19], [20], or (c) as a filter for search space reduction [22].

Automated gender estimation remains a challenging research area, due to large intra-class variation [52], and also due to challenges concerning illumination, as well as pose, age and ethnicity of a person. Further, facial expressions have a negative affect on the accuracy of automated gender estimation systems. This is why the majority of previous works have extracted and studied appearance-based features under the simplifying assumption of neutral face expressions with reasonably good results.

A. Gender and emotional expression

Deviating from such works, we here introduce the usage of a set of dynamic facial features for gender estimation. Specifically, we focus on extracting dynamic features from a common facial expression, namely the smile, and study how smile-dynamics encrypt gender evidence. The hypothesis is that male and female smile-dynamics differ in parameters such as intensity and duration. This hypothesis is supported in part by a number of cognitive-psychological studies, showing evidence for gender-dimorphism in the human expression [14], [74], [41], [1], [51], [25]. A main observation of such studies has been that females express emotions more frequently than males, and in the context of smile, females tend to smile more often than men in a variety of social contexts [25]. Such observations follow the theorem of men exhibiting restrictive emotionality and thus being unwilling to self-disclose intimate feelings. It is interesting to note, that a gender-based difference in emotional expression is observed as early as in 3 months old, shaped by how caregivers interact to male and female infants [33]; and also observed in toddlers, which appears to be further trained in social interactions [16], [56], [57]. Moreover, females are more accurate expressers of emotion, when posing deliberately and when observed unobtrusively, which is consistent across cultures [11]. The same work assigns happiness and fear as female-gender-stereotypical expressions. On the other hand, faces showing anger are considered more masculine [41], [40], [42], [43], [4], [5], [83] in the context of human gender recognition.

B. Contributions

Motivated from the above, we propose the use of an automated framework for facial dynamics extraction based on signal displacement of facial distances between key facial landmarks. We analyze the properties of 27 such facial distances in smile-video-sequences with emphasis on spontaneous, as well as posed smiles. The proposed dynamic features are fully complementary to appearance based features, and when combined with appearance, can pose an increased difficulty for spoof-attacks. We have adopted the approach from Dibeklioğlu et al. [27], [28], where it has been used for age estimation, as well as spontaneous vs. posed smile detection based on facial dynamics, see also [29], [26].

The use of the framework is instrumental in answering following questions:

- Do facial dynamics provide information about gender in (a) spontaneous smile- and (b) posed smile video sequences?

A. Dantcheva and F. Brémond are with the STARS team of Inria, France. E-mail: Antitza.Dantcheva@inria.fr, Francois.Bremond@inria.fr

[1] articles.latimes.com/2011/aug/21/business/la-fi-facial-recognition-20110821

[2] www.neurotechnology.com/face-biometrics.html

[3] www.visidon.fi/en/Face_Recognition

[4] www.cognitec-systems.de/FaceVACS-VideoScan.20.0.html

- Can facial smile dynamics improve the accuracy of appearance based gender estimation systems?
- Which gender can pose smiles more genuinely?

Related work of a holistic smile-based gender estimation algorithm can be found in Biliński et al. [9].

C. Structure of paper

This work is organized as follows: Section I-D revisits existing works on gender estimation. Section II proceeds to describe the proposed method, elaborating on individual steps (face detection, landmark location, selected features, statistics of dynamic features, feature selection, classification and used appearance features). Section III presents the employed dataset and the subsequent Section IV depicts and discusses related experimental results. Finally Section V concludes the paper.

D. Related work

Gender estimation Existing introductory overviews for algorithms related to gender estimation include the works of Ng et al. [61], Bekios-Calfa et al. [7], Ramanathan et al. [65], Mäkinen and Raisamo [54] and Dantcheva et al. [21]. Based on these works we can conclude that gender estimation remains a challenging task, which is inherently associated with different biometric modalities including fingerprint, face, iris, voice, body shape, gait, signature, DNA, as well as clothing, hair, jewelery and even body temperature. The forensic literature [52] suggests that the skull, and specifically the chin and the jawbone, as well as the pelvis, are the most significant indicators of the gender of a person; in juveniles, these shape-based features have been recorded to provide classification accuracy of $91\% - 99\%$.

Humans are generally quite good at gender recognition from early in life (e.g., [62], [64]), probably reflecting evolutive adaptation. As pointed out by Edelman et al. [30], humans perform facial image based gender classification with an error rate of about 11%, which is commensurate to that of a neural network algorithm performing the same task.

Dynamics have been used in the context of *body-based classification* of gender. Related cues include body sway, waist-hip ratio, and shoulder-hip ratio (see [59]); for example, females have a distinct waist-to-hip ratio and swing their hips more, whereas males have broader shoulders and swing their shoulders more.

Despite these recent successes, automated gender recognition from biometric data remains a challenge and is impacted by other soft biometrics, for example, age and ethnicity; gender dimorphism is accentuated only in adults, and varies across different ethnicities.

Automated Image-based Gender Estimation from Face In gender estimation from face, feature-based approaches extract and analyze a specific set of discriminative facial features (patches) in order to identify the gender of a person. This is a particularly challenging problem, as is implied from the fact that female and male average facial shapes are generally found to be very similar [50].

Another challenge comes to the fore in unconstrained settings with different covariates, such as illumination, expressions and ethnicity. While in more constrained settings, face-based gender estimation has been reported to achieve classification rates of up to 99.3% (see Table I), this performance though significantly decreases in more realistic and unconstrained settings.

The majority of gender classification methods contain two steps preceding face detection, namely feature extraction and pattern classification.

Feature extraction: Notable efforts include the use of SIFT [75], LBP [54], semi-supervised discriminant analysis (SDA) [8] or combinations of different features [36], [79].

Classification: A number of classification methods have been used for gender estimation, and a useful comparative guide of these classification methods can be found in Mäkinen and Raisamo [55]. One interesting conclusion of their work was that image size did not greatly influence the classification rates. This same work also revealed that manual alignment affected the classification rates positively, and that the best classification rates were achieved by SVM.

The area of gender estimation has also received some other contributions such as those that go *beyond using static 2D visible spectrum face-images*. Interesting related work include the work of Han et al. [39], exploring 3D images, Gonzalez–Sosa et al. [35], studying jointly body and face, and Chen and Ross [18], [69], using *near-infrared* (NIR) and thermal images for gender classification.

Expression Recognition Automated expression recognition has received increased attention in the past decade, since it is particularly useful in a variety of applications, such as human computer interaction, surveillance and crowd analytics. The majority of methods aim to classify 7 universal expressions namely neutral, happy, surprised, fearful, angry, sad, and disgusted [82] based on the extracted features used. Classical approaches follow Ekman's *facial action coding system* (FACS) [31], assigning each facial unit to represent movement of a specific facial muscle. In this context, intensity and number of facial units have been studied, as well as of action unit combinations, towards expression recognition. Interesting work can be found in related survey papers [84], [58], [71] and in a related recent expression-recognition-challenge-study [76]. Latest advances involve deep learning [85], [47].

Inspired by cognitive, psychological and neuroscientific findings, facial dynamics have been used previously towards improving face recognition [38], gender estimation [24], age estimation [27], as well as kinship recognition reported in a review article by Hadid et al. [37].

II. Dynamic Feature extraction in Smile-Video-Sequences

Deviating from the above works on gender estimation, we propose to extract dynamic features in smile-video-sequences. The general scheme is shown in Fig. 1. Specifically we focus on signal displacement of facial landmarks, as we aim to study among others the pertinence of different facial landmarks, as well as the pertinence of different statistical properties of facial

TABLE I. OVERVIEW OF FACE-BASED GENDER CLASSIFICATION ALGORITHMS. ABBREVIATIONS USED: PRINCIPAL COMPONENT ANALYSIS (PCA), INDEPENDENT COMPONENT ANALYSIS (ICA), SUPPORT VECTOR MACHINES (SVM), GAUSSIAN PROCESS CLASSIFIERS (GPC), ACTIVE APPEARANCE MODEL (AAM), LOCAL BINARY PATTERN (LBP), ACTIVE SHAPE MODEL (ASM), DISCRETE COSINE TRANSFORM (DCT), SEMI-SUPERVISED DISCRIMINANT ANALYSIS (SDA).

Work	Features	Classifier	Datasets used for evaluation	Performance numbers
Bekios-Calfa et al. (2007) [7]		SVM	UCN (nonpublic), 10,700 images	93.46% ± 1.65%
	PCA	LDA	FERET, 994 images	93.57% ± 1.39%
	LCA	LDA	PAL, 576 images	93.57% ± 1.39%
Xia et al. (2008) [80]	LBP, Gabor	SVM	CAS-PEAL, 10,784 images	93.74%
Mäkinen and Raisamo (2008) [55]	LBP	SVM	FERET, 411 images	86.54%
Baluja and Rowley (2008) [3]	Raw pixels	Adaboost	FERET, 2,409 images	93%
Gao and Ai (2009) [34]	ASM	Adaboost	Private, 1,300 images	92.89%
Toews and Arbel (2009) [75]	SIFT	Bayesian	FERET, 994 images	83.7%
Shan (2010) [72]	LBP	Adaboost	LFW, 7,443 images	94.44%
Guo et al. (2009) [36]	LBP, HOG, BIF	SVM	YGA, 8,000 images	89.28%
Wang et al. (2010) [79]	SIFT, context	Adaboost	FERET, 2,409 images	95.0%
Nazhir et al. (2010) [60]	DCT	KNN	SUMS, 400 images	99.3%
Ross and Chen (2011) [69]	LBP	SVM	CBSR NIR, 3,200 images	93.59%
Cao et al. (2011) [12]	Metrology	SVM	MUCT, 276 images	86.83%
Hu et al. (2011) [44]	Filter banks	SVM	Flickr, 26,700 images	90.1%
Bekios-Calfa et al. (2011) [8]	SDA	PCA	Multi-PIE, 337 images	88.04%
Shan (2012) [73]	Boosted LBP	SVM	LFW, 7,443	94.81%
Ramón-Balmaseda (2012) [66]	LBP	SVM	MORPH, LFW, Images of Groups, 17,814	75.10%
Jia and Cristianini (2015) [46]	Multi-scale LBP	C-Pegasos	Private, 4 million images	96.86%

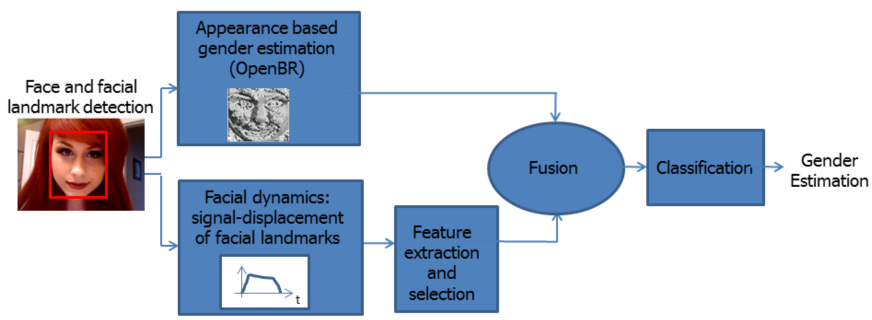


Fig. 1. Proposed framework for automatic gender estimation.

dynamics (e.g. intensity and duration) in the effort of gender estimation.

Towards extraction of such dynamic features, we assume a near frontal pose of the subject and an initial near-neutral expression of the subject (given in the used dataset).

A. Face Detection and Extraction of Facial Landmarks

Firstly we detect the face using the well established Viola and Jones algorithm [78]. We here note that the faces were robustly detected in all video sequences and frames. Within the detected face we identify facial feature points corresponding to points in the regions of the eye brows, eyes, nose and lips (see Fig. 5). Specifically we employ the facial landmark detection algorithm proposed in the work of Asthana et al. [2]. The algorithm is an incremental formulation for the discriminative deformable face alignment framework [81], using a discriminative 3D facial deformable shape model fitted to a 2D image by a cascade of linear regressors. The detector was trained on the $300W$-dataset (a dataset introduced in the context of the 300 faces in-the-wild challenge [70]) and detects 49 facial landmarks (see Fig. 5). For the UvA Nemo-dataset the facial landmarks were detected robustly in all video sequences and frames. We use these points to initialize a sparse optical flow tracking algorithm, based on the Kanade-Lucas-Tomasi (KLT) algorithm [53] in the first frame of each video-sequence. For the here proposed framework we select a subset of facial-points in three different face regions: (a) eye brow region, (b) eye region, (c) mouth region (see Fig. 2) and proceed to extract dynamic features thereof.

B. Extraction of Dynamic Features

We extract dynamic features corresponding to the signal-displacement in facial-distances depicted in Table II. We have selected 27 such facial-distances based on findings on facial

TABLE II. EXTRACTED SIGNAL-DISPLACEMENT-FUNCTIONS CONTRIBUTING TO DYNAMIC FEATURES. ρ DENOTES THE DISTANCE BETWEEN FACIAL LANDMARKS, l_i DENOTES THE i^{th} LANDMARK POINT, AS ILLUSTRATED IN FIG. 2.

Facial Distance	Description	Description by facial landmarks
D_1	Width of right eye	$\rho(\frac{l_{14}+l_{15}}{2}, \frac{l_{17}+l_{18}}{2})$
D_2	Width of left eye	$\rho(\frac{l_8+l_9}{2}, \frac{l_{11}+l_{12}}{2})$
D_3	Length of right eye	$\rho(l_{13}, l_{16})$
D_4	Length of left eye	$\rho(l_7, l_{10})$
D_5	Length of mouth	$\rho(l_{19}, l_{24})$
D_6	Width of mouth	$\rho(l_{22}, l_{25})$
D_7	Center of mouth to left side of upper lip	$\rho(\frac{l_{22}+l_{25}}{2}, l_{21})$
D_8	Center of mouth to right side of upper lip	$\rho(\frac{l_{22}+l_{25}}{2}, l_{23})$
D_9	Center of mouth to left mouth corner	$\rho(\frac{l_{22}+l_{25}}{2}, l_{19})$
D_{10}	Center of mouth to right mouth corner	$\rho(\frac{l_{22}+l_{25}}{2}, l_{24})$
D_{11}	Center of mouth to upper lip	$\rho(\frac{l_{22}+l_{25}}{2}, l_{27})$
D_{12}	Center of mouth to average distance of two mouth corners	$\rho(\frac{l_{22}+l_{25}}{2}, \frac{l_{19}+l_{24}}{2})$
D_{13}	Left side of right eyebrow to nose	$\rho(l_4, l_{26})$
D_{14}	Right side of left eyebrow to nose	$\rho(l_3, l_{26})$
D_{15}	Center of right eyebrow to right side of the right eyebrow	$\rho(l_5, l_6)$
D_{16}	Center of right eyebrow to left side of the right eyebrow	$\rho(l_5, l_4)$
D_{17}	Center of left eyebrow to right side of the left eyebrow	$\rho(l_2, l_3)$
D_{18}	Center of left eyebrow to left side of the left eyebrow	$\rho(l_2, l_1)$
D_{19}	Distance between eyebrows	$\rho(l_3, l_4)$
D_{20}	Left corner of left eye to left mouth corner	$\rho(l_7, l_{19})$
D_{21}	Right corner of left eye to center of mouth	$\rho(l_{10}, \frac{l_{22}+l_{25}}{2})$
D_{22}	Left corner of right eye to center of mouth	$\rho(l_{13}, \frac{l_{22}+l_{25}}{2})$
D_{23}	Right corner of right eye to right mouth corner	$\rho(l_{16}, l_{24})$
D_{24}	Upper side of left eye to right corner of left eyebrow	$\rho(\frac{l_8+l_9}{2}, l_3)$
D_{25}	Upper side of right eye to left corner of right eyebrow	$\rho(\frac{l_{14}+l_{15}}{2}, l_4)$
D_{26}	Upper side of left eye to left corner of left eyebrow	$\rho(\frac{l_8+l_9}{2}, l_1)$
D_{27}	Upper side of right eye to right corner of right eyebrow	$\rho(\frac{l_{14}+l_{15}}{2}, l_6)$

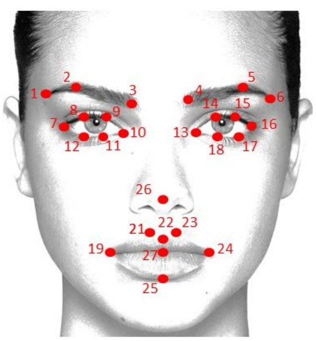


Fig. 2. Subset of landmarks extracted by the Asthana et al. algorithm [2] used in the proposed algorithm.

movements during smile-expressions [68].

1) Temporal smile-segmentation: Generally, the human smile is caused by the contraction of the zygomatic major muscle, which raises the corners of the lips [32], corresponding to "Action Unit Nr. 12" in Ekman's facial action coding system [31]. Temporally segmented, the human smile contains three phases: (a) *onset*: contraction of the zygomatic major muscle and alteration from neutral to expressive state, (b) *apex*: peak period of the expressive state, and (c) *offset*: relaxation of the zygomatic major muscle and change from expressive to neutral state. We here note that there are dozens of smile-classes, differing in appearance and meaning.

The next step in our method is to temporally segment the signal-displacement functions as: (a) onset: duration of monotonous increase, (b) apex: phase between onset and offset, (c) offset: duration of monotonous decrease. Fig. 3 illustrates two examples of signal-displacement in the mouth-region (D_5, mouth length), leading to a smile-curve with differently pronounced onset, apex and offset phases.

We smoothen each of the 27 signal displacement functions by the 4253H-twice smoothing algorithm [77] to flatten minor tracking-flaws.

2) Statistics of Dynamic Features: We proceed to extract statistics from each dynamic function with respect to the particular smile-phases, denoted by the superindices $(^+)$ for onset, $(^a)$ for apex, and $(^-)$ for offset, which we summarize in Table III. We compute the speed as $V(t) = \frac{dD}{dt}$ and the acceleration as $A = \frac{d^2D}{dt^2} = \frac{dV}{dt}$. We denote the number of frames by η, frame rate of the video sequence by ω. Each of the defined 27 signal-displacement-functions are represented by a set of 24 features, resulting in a 648-dimensional feature vector.

C. Feature Selection

We use the Min-Redundancy Max-Relevance (mRMR) algorithm [63] for selecting the permanent dynamic proposed features. mRMR minimizes the redundancy, while selecting

TABLE III. EXTRACTED DYNAMIC FEATURE STATISTICS. η DENOTES THE NUMBER OF FRAMES, D DENOTES THE RESPECTIVE DYNAMIC FEATURE, $V(t) = \frac{dD}{dt}$ DENOTES THE SPEED, $A = \frac{d^2D}{dt^2} = \frac{dV}{dt}$ DENOTES THE ACCELERATION, ω DENOTES THE FRAME RATE OF THE VIDEO SEQUENCE. THE SUPERSCRIPT $+$ DENOTES THE ONSET, SUPERSCRIPT a DENOTES THE APEX, SUPERSCRIPT $-$ DENOTES THE OFFSET.

Feature	**Definition**			
	General	Onset	Apex	Offset
Duration		$\frac{\eta(D^+)}{\omega}$	$\frac{\eta(D^a)}{\omega}$	$\frac{\eta(D^-)}{\omega}$
Duration Ratio		$\frac{\eta(D^+)}{\eta(D)}$		$\frac{\eta(D^-)}{\eta(D)}$
Maximal Amplitude	$max(D)$			
STD of Amplitude	$std(D)$			
Mean Amplitude		$mean(D^+)$	$mean(D^a)$	$mean(D^-)$
Total Amplitude		$\sum(D^+)$		$\sum(\lvert D^-\rvert)$
Net Amplitude	$\sum(D^+) - \sum(\lvert D^-\rvert)$			
Amplitude Ratio		$\frac{\sum(D^+)}{\sum(D^+)+\sum(D^-)}$		$\frac{\sum(D^-)}{\sum(D^+)+\sum(D^-)}$
Maximal Speed		$max(V^+)$		$max(V^-)$
Mean Speed		$mean(V^+)$		$mean(V^-)$
Maximum Acceleration		$max(A^+)$		$max(A^-)$
Mean Acceleration		$mean(A^+)$		$mean(A^-)$
Net Ampl., Duration Ratio	$\frac{(\sum(D^+)-\sum(\lvert D^-\rvert))\omega}{\eta(D)}$			

the most relevant information:

$$\max_{f_j \in F - S_{m-1}} \left[I(f_j, c) - \frac{1}{m-1} \sum_{f_i \in S_{m-1}} I(f_j, f_i) \right], \qquad (1)$$

with I being the mutual information function, c the target class, F the feature set, and S_{m-1} set of $m-1$ features. The mutual information I of a feature f_j and the target class c is computed based on the related probability density functions $p(f_j)$, $p(c)$ and $p(f_j, c)$ as follows

$$I(f_j; c) = \int\int p(f_j, c) \log \frac{p(f_j, c)}{p(f_j)p(c)} df_j dc. \qquad (2)$$

D. Classification

A pattern classifier, trained on labeled data, is used to classify the feature vector into one of two classes: *male* or *female*.

We utilized linear Support Vector Machines (SVM) [15], AdaBoost [6] and Bagged Trees [10] in this work. For SVM the Gaussian RBF kernel is used. The optimum values for C and the kernel parameter γ are obtained by a grid-search of the parameter space based on the training set.

E. Extracted Appearance Features

OpenBR [49] is a publicly available open source software for biometric recognition and evaluation. We utilize the gender estimation algorithm, based on the work of Klare et al. [48]. Specifically, a face image is represented by extracting histograms of local binary pattern (LBP) and scale-invariant feature transform (SIFT) features computed on a dense grid of patches. Subsequently, the histograms from each patch are projected onto a subspace generated using Principal Component Analysis (PCA) in order to obtain a feature vector. Support Vector Machine (SVM) is used for the final gender estimation. The OpenBR gender classification algorithm has been validated on a FERET[5] subset, attaining accuracies of 96.91% and 82.98% for male and female classification, respectively and an overall true classification rate of 90.57% [17], outperforming other algorithms (Neural Network, Support Vector Machine, etc.) on the same dataset [54].

how-old.net is a website (http://how-old.net/) launched by Microsoft for online age and gender recognition. Images can be uploaded and as an output age and gender labels are provided. The underlying algorithm and training dataset are not publicly disclosed.

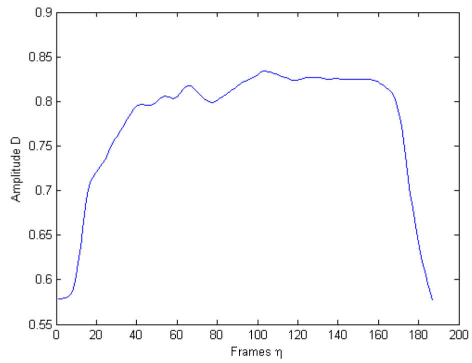
Commercial Off-the-Shelf (COTS) is a commercial face detection and recognition software, which includes a gender classification routine. The underlying algorithm and the training dataset that were used are not publicly disclosed. The system does not provide a mechanism to re-train the algorithm based on an external dataset; instead it is a black box that outputs a label (i.e., male or female) along with a confidence value.

Since the video-sequences of the UvA-NEMO dataset start with the neutral expression of the portrayed subject, the first frame is utilized to extract appearance features.

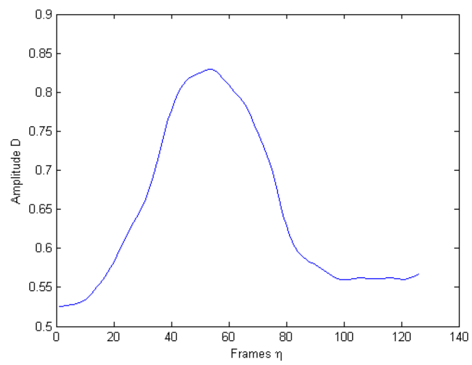
F. Fusion of Dynamic and Appearance Features

We concatenate score-levels obtained from the appearance based-algorithms with features obtained from the feature selection step of the dynamics-framework. We utilize PCA to reduce the dimension and obtain a fused feature vector.

[5]http://www.nist.gov/itl/iad/ig/colorferet.cfm



(a)



(b)

Fig. 3. Signal displacement in mouth-region (D_5, mouth length) for (a) female and (b) male subject from the UvA-NEMO dataset [28]. Example (a) shows the three profound smile phases: *onset* as the monotonic increasing phase, *apex* the (relatively) flat peak phase and *offset* the monotonic decreasing phase, example (b) on the other hand has a less pronounced apex-phase.

III. UvA-NEMO Smile-Dataset

The UvA-NEMO Smile Dataset[6], introduced by Dibeklioğlu et al. [28], consists of multiple video sequences of 400 subjects (185 females, 215 male). The age of the subjects ranges from 8 to 76 years, see Fig. 4 for the age-distribution. For the most of the subjects there are two videos per subject displaying: (a) spontaneous smile and (b) posed smile. To elicit spontaneous smiles, each subject was displayed a short funny video segment. Each video starts and ends with neutral or a near-neutral expression of the subject (see Fig. 5). The pose of the subjects is frontal and the illumination condition is reasonably constant across subjects. The resolution of the videos is 1920 x 1080 pixels at a framerate of 50 frames per second. This dataset has been used for the analysis of smiles for different ages [28] and for smile-based age analysis [27].

We note that the ethnicity of subjects in the UvA-NEMO dataset is predominantly Caucasian, hence the current study does not reflect on covariates such as ethnicity, as well as social and cultural background.

[6]http://www.uva-nemo.org

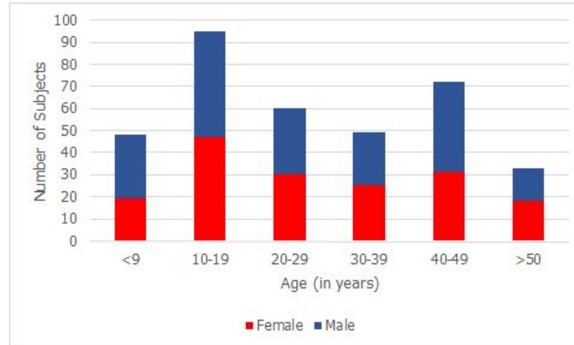
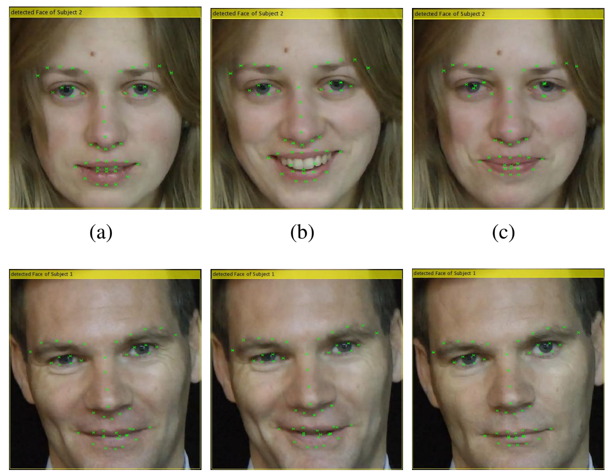


Fig. 4. Age and gender distributions of the subjects in the UvA-Nemo database, part 'spontaneous smile' containing 357 subjects.



(a) (b) (c)

(d) (e) (f)

Fig. 5. Example male and female subjects from the UvA-NEMO dataset expressing spontaneous smiles. Detected face and facial landmarks of (a),(d) the first frame, (b),(e) in a peak-apex-frame, (c),(f) last frame of the video sequence.

A. Effect of Age

The UvA-NEMO dataset consists of images of subjects in the age-range of 8 to 76 years. The ability of dynamics to predict age, and thus the impact of age on a small set of facial dynamics has been previously assessed in the work of Dibeklioğlu et al. [27], where results suggest that facial-dynamics change significantly with age. Consequently we present our results based on age-categories.

IV. Results

In order to evaluate the performance of the proposed gender estimation algorithm, we employ a 15-fold cross-validation scheme. Here, the UvA-NEMO dataset is divided into 15 folds with approximately 24 subjects in each fold. 14 folds are used for training the dynamic gender-estimation algorithm, and the remaining fold is used for testing it. This is repeated 15 times and reported results are the average thereof. Note that the subjects in the training set are not present in the test set.

A. Dynamics versus Appearance

Table IV firstly depicts the discriminative power of the two complementary characteristics individually for spontaneous smiles. As mentioned above, we report age-based gender recognition accuracy. Since training is required for the dynamics based gender estimation (and hence larger amount of subjects per group), we merge age-groups to two main groups: < 20 years and > 19 years and provide the associated results in Table IV. We observe that the *appearance based gender algorithms* perform significantly better for the age category > 19 years and rather poorly in the age category < 20 years. This can be due to age-unbalanced training sets or merely due to poor feature performance for toddlers and adolescents, due to low sexual dimorphism. The related confusion matrices for the age category > 19 years are shown in Table V.

Dynamics based gender estimation: Interestingly, dynamic features (True Gender Classification Rate $TGCR = 59.44\%$) outperform two of the three appearance based features ($TGCR_{OpenBR} = 52.45\%$ and $TGCR_{how-old.net} = 51.05\%$) in the first age-category. While, appearance-based features are more reliable for the age category > 19 years with $TGCR_{OpenBR} = 78.04\%$, $TGCR_{how-old.net} = 93.46\%$, $TGCR_{COTS} = 92.52\%$; dynamics-based features obtain a noticeable accuracy of 67.81%. The latter suggests that facial smile-dynamics carry substantial cues related to gender of the subject. The confusion matrix is rather balanced in the dynamics-based gender estimation (Table V (d)).

We note that *fusion* of appearance and smile-dynamic-based gender estimation either increases the performance of appearance based algorithms (e.g., for OpenBR in both age classes, for *how-old.net* in the younger age-class and for COTS in the older age-class) or does not impact it negatively. Related confusion matrices are shown in Table V.

In our related work [9], we have presented a holistic approach for smile-based gender estimation, that extracts spatio-temporal features based on dense trajectories, represented by a set of descriptors encoded by Fisher Vectors. The associated true gender classification rates account for 86.3% for adolescents, and 91.01% for adults.

B. Spontaneous versus posed smile

We also provide results on the posed-smile subset of the UvA-NEMO dataset presented in Table VI. Interestingly, the associated dynamics-based gender-estimation accuracy resembles strongly the spontaneous-smile-case. The difference in performance origins in the slightly larger posed-smile subset-size, that contributes to larger trainings-sets in the case of dynamics-based gender classification, as well as in the fusion of appearance and dynamics-based features. Nevertheless, the results suggest that dynamics of posed smiles carry significant cues on gender, similarly to spontaneous smiles. The related confusion matrices are shown in Table VII.

This result is in agreement with psychological findings, that show that females are more accurate expressers of emotion, when posing deliberately and when observed unobtrusively [11], hinting that posing a smile carries gender-specific cues.

TABLE VI. POSED SMILE. TRUE GENDER CLASSIFICATION RATES. AGE GIVEN IN YEARS.

Combined Age-Groups	< 20	> 19
Subj. amount	143	225
Dynamics (SVM)	59.44%	66.22%
OpenBR + Dynamics (Bagged Trees)	60.8%	80%
how-old.net	51.05%	93.78%
how-old.net + Dynamics (SVM)	60.8%	92.89%
COTS	76.92%	92%
COTS + Dynamics (Bagged Trees, PCA)	76.92%	92.89%

C. Gender divergence in spontaneous and posed smiles

We seek to answer the question, whether males or females pose smiles more genuinely and whether there is a significant divergence. Towards this, we combine features in all possible sets and compute Euclidean distances between sets in the spontaneous and the associated sets in the posed-smile-case. Fig. 6 illustrates the related results for the most diverging case between males and females. Females have slightly lower distances, suggesting that females pose smiles more realistically; however, the disparity is not significant. This tendency conforms with previous psychological findings [11].

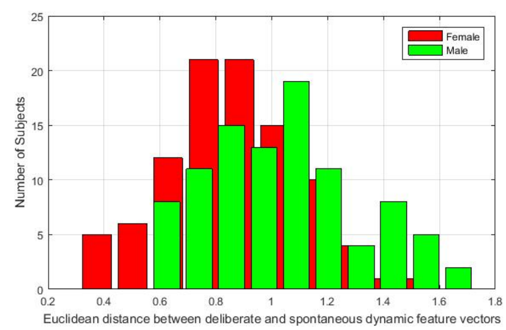


Fig. 6. Distributions of Euclidean distances between posed and spontaneous feature vectors for male and female subjects in the UvA-NEMO dataset.

D. Discriminative Features

We here analyze the individual discriminability of the selected dynamic-features for the 27 distances. Towards this, we estimate gender based on each feature individually. Hence, we train and test an SVM-classifier with each feature individually for the two age groups, < 20 and > 19 years. We report for each age group the most discriminative features respectively (see Table VIII and Table IX). The most striking outcome is that the majority of discriminative features are in the mouth region. It is also interesting to note that while for the younger group D_{10} (*Center of mouth to right mouth corner*) and D_7 (*Center of mouth to left side of upper lip*) and the onset-phase are predominant, for the older group D_5 (*Length of mouth*) and mainly the offset-phase is more profound. This hints that

TABLE IV. SPONTANEOUS SMILE. TRUE GENDER CLASSIFICATION RATES. AGE GIVEN IN YEARS.

Age	< 10	$10-19$	$20-29$	$30-39$	$40-49$	> 49
Subj. amount	48	95	60	49	72	33
OpenBR	58.33%	50.53%	81.67%	75.51%	75%	81.82%
how-old.net	39.58%	56.84%	95%	87.76%	98.61%	87.88%
COTS	77.08%	76.84%	93.33%	89.8%	94.44%	90.91%

Merged Age-Groups	< 20	> 19
Subj. amount	143	214
Dynamics (PCA, SVM)	59.44%	67.81%
OpenBR	52.45%	78.04%
OpenBR + Dynamics (Bagged Trees)	60.1%	78.97%
how-old.net	51.05%	93.46%
how-old.net + Dynamics (Tree)	60.8%	93.46%
COTS	76.92%	92.52%
COTS + Dynamics (Tree)	76.92%	93%

TABLE V. SPONTANEOUS SMILE IN AGE CATEGORY > 19: CONFUSION MATRIX FOR MALES AND FEMALES FOR (A) APPEARANCE FEATURES #1 (OPENBR) (DENOTED AS APP. 1), (B) APPEARANCE FEATURES #2 (*how-old.net*) (DENOTED AS APP. 2) , (C) APPEARANCE FEATURES #3 (COTS) (DENOTED AS APP. 3), (D) DYNAMIC FEATURES (DENOTED AS DYN.), (E) DYNAMIC AND APPEARANCE FEATURES #1 (DENOTED AS DYN. + APP. 1), (F) DYNAMIC AND APPEARANCE FEATURES #2 (DENOTED AS DYN. + APP. 2), (G) DYNAMIC AND APPEARANCE FEATURES #3 (DENOTED AS DYN. + APP. 3).

	(a) App. 1		**(b) App. 2**		**(c) App. 3**		**(d) Dynamics**		**(e) Dyn. + App. 1**		**(f) Dyn. + App. 2**		**(g) Dyn. + App. 3**	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Male	61.8%	38.2%	94.5%	5.5%	99.1%	0.9%	70.0%	30.0%	78.18%	21.82%	94.5%	5.5%	99.1%	0.9%
Female	4.8%	95.2%	7.7%	92.3%	14.4%	85.6%	34.62%	65.38%	20.19%	79.81%	7.7%	92.3%	12.5%	86.54%

TABLE VII. POSED SMILE IN AGE CATEGORY > 19:CONFUSION MATRIX FOR MALES AND FEMALES FOR (A) APPEARANCE FEATURES #1 (OPENBR) (DENOTED AS APP. 1), (B) APPEARANCE FEATURES #2 (*how-old.net*) (DENOTED AS APP. 2) , (C) APPEARANCE FEATURES #3 (COTS) (DENOTED AS APP. 3), (D) DYNAMIC FEATURES (DENOTED AS DYN.), (E) DYNAMIC AND APPEARANCE FEATURES #1 (DENOTED AS DYN. + APP. 1), (F) DYNAMIC AND APPEARANCE FEATURES #2 (DENOTED AS DYN. + APP. 2), (G) DYNAMIC AND APPEARANCE FEATURES #3 (DENOTED AS DYN. + APP. 3).

	(a) App. 1		**(b) App. 2**		**(c) App. 3**		**(d) Dynamics**		**(e) Dyn. + App. 1**		**(f) Dyn. + App. 2**		**(g) Dyn. + App. 3**	
	Male	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
Male	62.34%	37.66%	94.81%	5.19%	98.7%	1.3%	68.1%	31.9%	75%	25%	93.97%	6.03%	99.14%	0.86%
Female	1.52%	98.48%	7.58%	92.42%	15.15%	84.85%	35.78%	64.22%	14.68%	85.32%	8.26%	91.74%	13.76%	86.24%

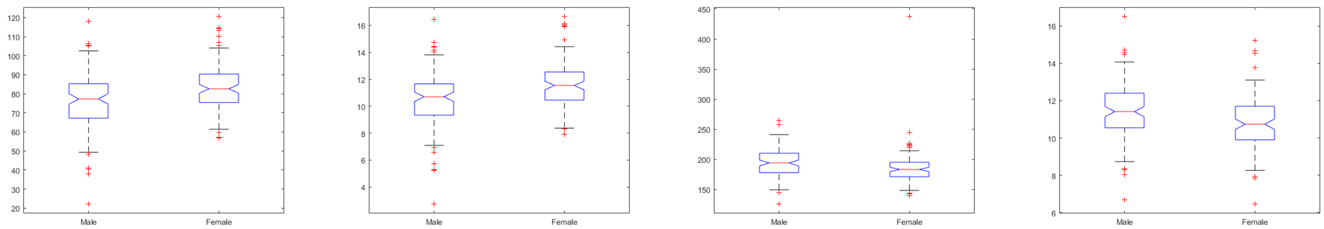
sexual dimorphism can be gleaned from the asymmetrical-onset in adolescents. On a related note, a recent psychological study [13] has found that expressions shown on the left hemi-face (LHF) were rated as more intense, and furthermore that spontaneous expressions start earlier in the LHF. Hence expressions in both hemi-faces are not fully redundant.

Description of most discriminative features In adolescents, females tended to show longer *Duration Ratio – Offset* and longer *Duration – Onset* on the right side of the mouth and higher *Amplitude Ratio – Onset* on the left side of the mouth, than males. In adults, females tended to show higher *Mean Amplitude – Apex* of mouth opening, higher *Maximum Amplitude* on the right side of the mouth, as well as faster *Mean Speed – Offset* on the left side of the mouth, than males. Figure 7 illustrates the boxplots for the five most discriminative features in the age category > 19 years for spontaneous smile.

We here note, that the selected features for the proposed algorithm in previous sections do not correspond to the presented features in this section, since a mutual information function prunes out correlated features in the selection process, which we do not consider here.

V. CONCLUSIONS

In this work we introduced smile-based dynamic facial feature extraction for gender estimation. The proposed dynamics-based gender estimation algorithm predominantly improves the performance of three state-of-the-art appearance-based gender estimation algorithms. We observe that dynamics can outperform appearance-based features for subjects younger than 20 years old; while facial appearance features are more discriminative for older subjects. We show that appearance and dynamics-based features are complementary and the combination thereof beneficial. Our results further suggest that gender



(a) D_{11} Mean Amplitude Apex (b) D_8 Maximum Amplitude (c) D_9 Mean Speed Offset (d) D_5 Mean Acceleration Offset

Fig. 7. Boxplots of most discriminative features in age category > 19 years. Females tended to show longer *Mean Amplitude Apex* of mouth opening, a higher *Maximum Amplitude* on the right side of the mouth, as well as a shorter *Mean Speed Offset* on the left side of the mouth, than males. Further the *Mean Acceleration Offset* of the mouth length is shorter for females than for males.

TABLE VIII. MOST DISCRIMINATE DYNAMIC FEATURES FOR AGE < 20. TGCR...TRUE GENDER CLASSIFICATION RATE.

Distance	Feature	TGCR
D_{10}	Duration Ratio Offset	65.73%
D_7	Amplitude Ratio Onset	63.64%
D_{10}	Duration Onset	62.94%
D_{10}	Total Amplitude Onset	62.24%
D_9	Maximum Amplitude	62.24%
D_{10}	Mean Amplitude Onset	62.24%
D_7	Amplitude Ratio Offset	62.24%
D_{10}	Maximum Amplitude	61.54%

TABLE IX. MOST DISCRIMINATE DYNAMIC FEATURES FOR AGE > 19. TGCR...TRUE GENDER CLASSIFICATION RATE.

Distance	Feature	TGCR
D_{11}	Mean Amplitude Apex	62.15%
D_8	Maximum Amplitude	61.68%
D_9	Mean Speed Offset	61.54%
D_5	Mean Acceleration Offset	60.28%
D_5	Amplitude Ratio Offset	60.28%
D_5	Duration Offset	60.28%
D_5	Maximum Acceleration Offset	60.14%
D_1	Duration Apex	59.81%

is mainly exhibited in dynamics in the mouth-region among the studied facial dynamic-features. Finally, we analyzed the gender-dimorphism of both, spontaneous and posed smiles and observe that both carry substantial cues for gender.

APPENDIX A
ACKNOWLEDGMENT

This work was carried out during the tenure of an ERCIM "Alain Bensoussan" Fellowship Programme. The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement nr. 246016.

REFERENCES

[1] R. B. Adams, U. Hess, and R. E. Kleck. The intersection of gender-related facial appearance and facial displays of emotion. *Emotion Review*, 7(1):5–13, 2015.

[2] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic. Incremental face alignment in the wild. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pages 1859–1866. IEEE, 2014.

[3] S. Baluja and H. A. Rowley. Boosting sex identification performance. *International Journal of Computer Vision (IJCV)*, 71:111–119, 2006.

[4] L. Bayet, O. Pascalis, P. C. Quinn, K. Lee, É. Gentaz, and J. W. Tanaka. Angry facial expressions bias gender categorization in children and adults: behavioral and computational evidence. *Frontiers in psychology*, 6, 2015.

[5] D. V. Becker, D. T. Kenrick, S. L. Neuberg, K. C. Blackwell, and D. M. Smith. The confounded nature of angry men and happy women. *Journal of personality and social psychology*, 92(2):179, 2007.

[6] J. Bekios-Calfa, J. M. Buenaposada, and L. Baumela. Revisiting linear discriminant techniques in gender recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(4):858–864, 2011.

[7] J. Bekios-Calfa, J. M. Buenaposada, and L. Baumela. Revisiting linear discriminant techniques in gender recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 33(4):858–864, 2011.

[8] J. Bekios-Calfa, J. M. Buenaposada, and L. Baumela. Robust gender recognition by exploiting facial attributes dependencies. *Pattern Recognition Letters*, 2013.

[9] P. Bilinski, A. Dantcheva, and F. Bremond. Can a smile reveal your gender? In *International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2016.

[10] L. Breiman. Bagging predictors. *Machine learning*, 24(2):123–140, 1996.

[11] L. R. Brody and J. A. Hall. Gender and emotion in context. *Handbook of emotions*, 3:395–408, 2008.

[12] D. Cao, C. Chen, M. Piccirilli, D. Adjeroh, T. Bourlai, and A. Ross. Can facial metrology predict gender? In *Proceedings of International Joint Conference on Biometrics (IJCB)*, 2011.

[13] E. W. Carr, S. Korb, P. M. Niedenthal, and P. Winkielman. The two sides of spontaneity: Movement onset asymmetries in facial expressions influence social judgments. *Journal of Experimental Social Psychology*, 55:31–36, 2014.

[14] E. Cashdan. Smiles, speech, and body posture: How women and men display sociometric status and power. *Journal of Nonverbal Behavior*, 22(4):209–228, 1998.

[15] C.-C. Chang and C.-J. Lin. Libsvm: A library for support vector machines. *ACM Trans. on Intelligent Systems and Technology*, 2(3):127, 2011.

[16] T. M. Chaplin and A. Aldao. Gender differences in emotion expression in children: a meta-analytic review. *Psychological Bulletin*, 139(4):735, 2013.

[17] C. Chen, A. Dantcheva, and A. Ross. Impact of facial cosmetics on automatic gender and age estimation algorithms. In *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISAPP)*, 2014.

[18] C. Chen and A. Ross. Evaluation of gender classification methods on thermal and near-infrared face images. In *Proceedings of International Joint Conference on Biometrics (IJCB)*, 2011.

[19] A. Dantcheva, J.-L. Dugelay, and P. Elia. Person recognition using a bag of facial soft biometrics (BoFSB). In *IEEE International Workshop on Multimedia Signal Processing (MMSP)*, 2010.

[20] A. Dantcheva, J.-L. Dugelay, and P. Elia. Soft biometric systems: reliability and asymptotic bounds. In *Proceedings of IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2010.

[21] A. Dantcheva, P. Elia, and A. Ross. What else does your biometric data reveal? a survey on soft biometrics. *IEEE Transactions on Information Forensics and Security (TIFS)*, 11(3):441–467, 2016.

[22] A. Dantcheva, A. Singh, P. Elia, and J.-L. Dugelay. Search pruning in video surveillance systems: Efficiency-reliability tradeoff. In *Proceedings of International Conference on Computer Vision Workshops*, 2011.

[23] A. Dantcheva, C. Velardo, A. D'Angelo, and J.-L. Dugelay. Bag of soft biometrics for person identification. New trends and challenges. *Multimedia Tools and Applications (MTAS)*, 51:739–777, 2011.

[24] M. Demirkus, M. Toews, J. J. Clark, and T. Arbel. Gender classification from unconstrained video sequences. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, pages 55–62. IEEE, 2010.

[25] F. M. Deutsch, D. LeBaron, and M. M. Fryer. What is in a smile? *Psychology of Women Quarterly*, 11(3):341–352, 1987.

[26] H. Dibeklioğlu, F. Alnajar, A. Ali Salah, and T. Gevers. Combining facial dynamics with appearance for age estimation. *Image Processing, IEEE Transactions on*, 24(6):1928–1943, 2015.

[27] H. Dibeklioğlu, T. Gevers, A. A. Salah, and R. Valenti. A smile can reveal your age: Enabling facial dynamics in age estimation. In *Proceedings of the 20th ACM international conference on Multimedia*, pages 209–218. ACM, 2012.

[28] H. Dibeklioğlu, A. A. Salah, and T. Gevers. Are you really smiling at me? spontaneous versus posed enjoyment smiles. In *European Conference on Computer Vision (ECCV)*, pages 525–538. Springer, 2012.

[29] H. Dibeklioğlu, A. A. Salah, and Theo Gevers. Recognition of genuine smiles. *Multimedia, IEEE Transactions on*, 17(3):279–294, 2015.

[30] B. Edelman, D. Valentin, and H. Abdi. Sex classification of face areas: how well can a linear neural network predict human performance? *Biological systems*, 4, 1996.

[31] P. Ekman. Facial expression and emotion. *American psychologist*, 48(4):384, 1993.

[32] P. Ekman and W. V. Friesen. Felt, false, and miserable smiles. *Journal of nonverbal behavior*, 6(4):238–252, 1982.

[33] A. Fogel, S. Toda, and M. Kawai. Mother-infant face-to-face interaction in japan and the united states: A laboratory comparison using 3-month-old infants. *Developmental Psychology*, 24(3):398, 1988.

[34] W. Gao and H. Ai. Face gender classification on consumer images in a multiethnic environment. In *Proc. IEEE International Conference on Biometrics*, page 169178, 2009.

[35] E. Gonzalez-Sosa, A. Dantcheva, R. Vera-Rodriguez, J.-L. Dugelay, F. Bremond, and J. Fierrez. Image-based gender estimation from body and face across distances. In *International Conference on Pattern Recognition (ICPR)*, 2016.

[36] G. Guo, C. Dyer, Y. Fu, and T. Huang. Is gender recognition affected by age? In *Proc. International Conference on Computer Vision Workshops*, page 20322039, 2009.

[37] A. Hadid, J.-L. Dugelay, and M. Pietikäinen. On the use of dynamic features in face biometrics: recent advances and challenges. *Signal, Image and Video Processing*, 5(4):495–506, 2011.

[38] A. Hadid and M. Pietikäinen. Combining appearance and motion for face and gender recognition from videos. *Pattern Recognition*, 42(11):2818–2827, 2009.

[39] X. Han, H. Ugail, and I. Palmer. Gender classification based on 3D face geometry features using SVM. In *Proceedings of International Conference on CyberWorlds (CW)*, 2009.

[40] U. Hess, R. Adams Jr, and R. Kleck. Who may frown and who should smile? dominance, affiliation, and the display of happiness and anger. *Cognition & Emotion*, 19(4):515–536, 2005.

[41] U. Hess, R. B. Adams Jr, and R. E. Kleck. Facial appearance, gender, and emotion expression. *Emotion*, 4(4):378, 2004.

[42] U. Hess, R. B. Adams Jr, and R. E. Kleck. When two do the same, it might not mean the same: The perception of emotional expressions shown by men and women. 2007.

[43] U. Hess and P. Thibault. Why the same expression may not mean the same when shown on different faces or seen by different people. *Affective information processing*, pages 145–158, 2009.

[44] S. D. Hu, B. Jou, A. Jaech, and M. Savvides. Fusion of region-based representations for gender identification. In *Proc. International Joint Conference on Biometrics*, 2011.

[45] A. K. Jain, S. C. Dass, and K. Nandakumar. Can soft biometric traits assist user recognition? In *Proceedings of SPIE Defense and Security Symposium*, volume 5404, pages 561–572, 2004.

[46] S. Jia and N. Cristianini. Learning to classify gender from four million images. *Pattern Recognition Letters*, 58:35–41, 2015.

[47] F. Juefei-Xu, E. Verma, P. Goel, A. Cherodian, and M. Savvides. Deepgender: Occlusion and low resolution robust facial gender classification via progressively trained convolutional neural networks with attention. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2016.

[48] B. F. Klare, M. J. Burge, J. C. Klontz, R. W. V. Bruegge, and A. K. Jain. Face recognition performance: Role of demographic information. *Information Forensics and Security, IEEE Transactions on*, 7(6):1789–1801, 2012.

[49] J. C. Klontz, B. F. Klare, S. Klum, A. K. Jain, and M. J. Burge. Open source biometric recognition. In *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, pages 1–8. IEEE, 2013.

[50] J. H. Langlois and L. A. Roggman. Attractive faces are only average. *Psychological science*, 1(2):115–121, 1990.

[51] M.-F. Liébart, C. Fouque-Deruelle, A. Santini, F. Dillier, V. Monnet-Corti, J.-M. Glise, and A. Borghetti. Smile line and periodontium visibility. *Perio*, 1(1):17–25, 2004.

[52] S. R. Loth and M.Y. Iscan. *Sex determination, Encyclopedia of forensic Sciences*, volume 1. Academic Press, San Diego, 2000.

[53] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI*, volume 81, pages 674–679, 1981.

[54] E. Mäkinen and R. Raisamo. Evaluation of gender classification methods with automatically detected and aligned faces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(3):541–547, 2008.

[55] E. Mäkinen and R. Raisamo. Evaluation of gender classification methods with automatically detected and aligned faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 30(3):541–547, 2008.

[56] C. Z. Malatesta, C. Culver, J. R. Tesman, B. Shepard, A. Fogel, M. Reimers, and G. Zivin. The development of emotion expression during the first two years of life. *Monographs of the Society for Research in Child Development*, pages i–136, 1989.

[57] C. Z. Malatesta and J. M. Haviland. Learning display rules: The socialization of emotion expression in infancy. *Child development*, pages 991–1003, 1982.

[58] A. Martinez and S. Du. A model of the perception of facial expressions of emotion by humans: Research overview and perspectives. *The Journal of Machine Learning Research*, 13(1):1589–1608, 2012.

[59] G. Mather and L. Murdoch. Gender discrimination in biological motion displays based on dynamic cues. In *Biological Sciences B*, pages 273–279, 1994.

[60] M. Nazhir, M. Ishaiq, A. Batool, A. Jaffar, and A. M. Mirza. Feature selection for efficient gender classification. In *WSEAS international conference on neural networks, evolutionary computing and Fuzzy systems*, 2010.

[61] C. B. Ng, Y. H. Tay, and B.-M. Goi. Vision-based human gender recognition: A survey. *PRICAI 2012: Trends in Artificial Intelligence. Lecture Notes in Computer Science*, 7458:335–346, 2012.

[62] A. O'Toole, A. Peterson, and K. Deffenbacher. An other-race effect for classifying faces by sex. *Perception*, 25:669–676, 1996.

[63] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1226–1238, 2005.

[64] P. C. Quinn, J. Yahr, A. Kuhn, A. M. Slater, and O. Pascalis. Representation of the gender of human faces by infants: A preference for female. *Perception*, 31(9):1109–1121, 2002.

[65] N. Ramanathan, R. Chellappa, and S. Biswas. Age progression in human faces: A survey. *Visual languages and computing*, 2009.

[66] E. Ramón-Balmaseda, J. Lorenzo-Navarro, and M. Castrillón-Santana. Gender classification in large databases. In *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pages 74–81. Springer, 2012.

[67] D. Reid, S. Samangooei, C. Chen, M. Nixon, and A. Ross. Soft biometrics for surveillance: An overview. In *Handbook of Statistics*, volume 31, 2013.

[68] C. K. Richardson, D. Bowers, R. M. Bauer, K. M. Heilman, and C. M. Leonard. Digitizing the moving face during dynamic displays of emotion. *Neuropsychologia*, 38(7):1028 – 1039, 2000.

[69] A. Ross and C. Chen. Can gender be predicted from near-infrared face images? In *Proceedings of International Conference on Image Analysis and Recognition (ICIAR)*, 2011.

[70] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic. 300 faces in-the-wild challenge: The first facial landmark localization challenge. In *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*, pages 397–403. IEEE, 2013.

[71] G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin. Static and dynamic 3D facial expression recognition: A comprehensive survey. *Image and Vision Computing*, 30(10):683–697, 2012.

[72] C. Shan. Gender classification on real-life faces. *Advanced Concepts for Intelligent Vision Systems*, page 323331, 2010.

[73] C. Shan. Learning local binary patterns for gender classification on real-world face images. *Pattern Recognition Letters*, 33(4):431–437, 2012.

[74] R. W. Simon and L. E. Nath. Gender and emotion in the united states: Do men and women differ in self-reports of feelings and expressive behavior? 1. *American journal of sociology*, 109(5):1137–1176, 2004.

[75] M. Toews and T. Arbel. Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(9):15671581, 2009.

[76] M. F. Valstar, T. Almaev, J. M. Girard, G. McKeown, M. Mehu, L. Yin, M. Pantic, and J. F. Cohn. Fera 2015-second facial expression recognition and analysis challenge. In *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, volume 6, pages 1–8. IEEE, 2015.

[77] P. F. Velleman. Definition and comparison of robust nonlinear data smoothing algorithms. *Journal of the American Statistical Association*, 75(371):609–615, 1980.

[78] P. Viola and M. Jones. Robust real-time face detection. In *Proceedings of IEEE ICCV*, 2001.

[79] J. Wang, J. Li, W. Yau, and E. Sung. Boosting dense sift descriptors and shape contexts of face images for gender recognition. In *Proc. Computer Vision and Pattern Recognition Workshops*, page 96102, 2010.

[80] B. Xia, H. Sun, and B.-L. Lu. Multi-view gender classification based on local gabor binary mapping pattern and support vector machines. In *Proc. International Joint Conference on Neural Networks*, page 33883395, 2008.

[81] X. Xiong and F. de la Torre. Supervised descent method and its applications to face alignment. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 532–539, June 2013.

[82] K. Yu, Z. Wang, L. Zhuo, J. Wang, Z. Chi, and D. Feng. Learning realistic facial expressions from web images. *Pattern Recognition*, 46(8):2144–2155, 2013.

[83] L. A. Zebrowitz and J. M. Montepare. Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, 2(3):1497–1517, 2008.

[84] Z. Zeng, M. Pantic, G. I. Roisman, and T. S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(1):39–58, 2009.

[85] X. Zhao, X. Shi, and S. Zhang. Facial expression recognition via deep learning. *IETE Technical Review*, 32(5):347–355, 2015.

Antitza Dantcheva is a Postdoctoral Fellow at the STARS team, INRIA, France. Previously, she was a Marie Curie fellow at INRIA and a Postdoctoral Fellow at the Michigan State University and the West Virginia University, USA. She received her PhD in Signal and Image Processing in 2011 from Eurecom / Telecom ParisTech in France. She was the recipient of the Best Presentation Award in ICME 2011, the Best Poster Award in ICB 2013 as well as the Tabula Rasa Spoofing Award in ICB 2013. Her research interests are in soft biometrics for security and commercial applications, where she has worked on retrieval of soft biometrics from images, as well as their corresponding analysis.

François Brémond received the PhD degree from INRIA in video understanding in 1997, and he pursued his research work as a post doctorate at the University of Southern California (USC) on the interpretation of videos taken from Unmanned Airborne Vehicle (UAV). In 2007, he received the HDR degree (Habilitation a Diriger des Recherches) from Nice University on Scene Understanding. He created the STARS team on the 1st of January 2012. He is the research director at INRIA Sophia Antipolis, France. He has conducted research work in video understanding since 1993 at Sophia- Antipolis. He is author or co-author of more than 140 scientific papers published in international journals or conferences in video understanding. He is a handling editor for MVA and a reviewer for several international journals (CVIU, IJPRAI, IJHCS, PAMI, AIJ, Eurasip, JASP) and conferences (CVPR, ICCV, AVSS, VS, ICVS). He has (co-)supervised 13 PhD theses. He is an EC INFSO and French ANR Expert for reviewing projects.