Age and Gender Estimation of Unfiltered Faces

Eran Eidinger, Roee Enbar, and Tal Hassner

***Abstract*— This paper concerns the estimation of facial attributes—namely, age and gender—from images of faces acquired in challenging, in the wild conditions. This problem has received far less attention than the related problem of face recognition, and in particular, has not enjoyed the same dramatic improvement in capabilities demonstrated by contemporary face recognition systems. Here, we address this problem by making the following contributions. First, in answer to one of the key problems of age estimation research—absence of data—we offer a unique data set of face images, labeled for age and gender, acquired by smart-phones and other mobile devices, and uploaded without manual filtering to online image repositories. We show the images in our collection to be more challenging than those offered by other face-photo benchmarks. Second, we describe the dropout-support vector machine approach used by our system for face attribute estimation, in order to avoid over-fitting. This method, inspired by the dropout learning techniques now popular with deep belief networks, is applied here for training support vector machines, to the best of our knowledge, for the first time. Finally, we present a robust face alignment technique, which explicitly considers the uncertainties of facial feature detectors. We report extensive tests analyzing both the difficulty levels of contemporary benchmarks as well as the capabilities of our own system. These show our method to outperform state-of-the-art by a wide margin.**

***Index Terms*— Face recognition, identification of persons, support vector machines, neural networks.**

I. Introduction

AT THE most basic level of the languages we speak, how we address a person is largely influenced by who that person is: "sir" or "madam" are used based on the gender of the person being referred to; an older person would often be addressed more formally than a younger one. More generally, languages reserve different words and grammar rules when addressing different people. This phenomenon, at the heart of social interactions, relies on our ability to estimate these individual traits, here, age and gender, at a glance, just from facial appearances. As the roles of computers in our lives grow, and as we interact with them more and more, it is natural to expect computerized systems to be capable of doing the same, with similar accuracy and effortlessness.

Manuscript received March 25, 2014; revised July 1, 2014; accepted September 2, 2014. Date of publication September 22, 2014; date of current version November 12, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Gérard Medioni. *(Corresponding author: Tal Hassner.)*

E. Eidinger and R. Enbar are with Adience, Tel Aviv 6350671, Israel (e-mail: eran@adience.com; roee.e@adience.com).

T. Hassner is with the Department of Mathematics and Computer Science, Open University of Israel, Ra'anana 4353701, Israel (e-mail: hassner@openu.ac.il).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2014.2359646

Yet despite this, and despite the obvious relation to the well-studied problem of face recognition, there has been far less work focused on developing systems for automatic age and gender estimation from face photos. This is at least partially due to the absence of sufficient data: Where face recognition has benefited greatly from high-quality, comprehensive benchmarks such as the Labeled Faces in the Wild (LFW) [1] and the YouTube faces [2] collections, similar data sets are not openly available for age and gender estimation. This is especially perplexing, when considering that estimating facial attributes has been shown in the past to be key to accurate face recognition [3]. This gap in the resources available for the study of the two problems may be traced to the additional challenge of obtaining accurate age labels, compared to subject identities [4]. Regardless, as a consequence, this problem has not enjoyed the same dramatic improvement in capabilities demonstrated for face recognition.

In an effort to close the gap between the capabilities of age and gender estimation systems and face recognition systems (let alone the gap between computer and human capabilities) we take the following steps. First, we offer a public data-set of labeled images, and associated benchmark protocols, designed to reflect the challenges of real-world age and gender estimation tasks. Besides being massive in the number of images and subjects it includes, it is unique in the nature of its images. These were obtained from online image albums captured by smart-phones and automatically uploaded to Flicker before being manually filtered or "curated". It therefore contains many images which would typically be removed by their owners as being faulty or otherwise unsuitable. Compared to other "in the wild" face image collections, most notably LFW or PubFig, it contains images which demonstrate a far wider range of challenging viewing conditions. This is demonstrated in Fig. 1, but also in our experiments, comparing the performances of the same methods on different sets (Section V).

Training a computer vision system to accurately estimate age, runs the risk of over-fitting to the biases of the photos used for the training. This is more true here than in the related problem of face recognition, as represented by, e.g., the LFW benchmark. The reasons are twofold [4]: (i) the additional challenge of preparing suitable data (collecting and labeling photos for accurate age) results in less data being available, and (ii), the multi-class nature of the problem may require a wider variability of examples than the binary "same"/ "not-same" classes used by the LFW.

Our second contribution in this paper addresses this issue by proposing *dropout-SVM* for training linear support vector machine (SVM) classifiers [6]. Our approach follows the "dropout" technique recently proposed for training deep

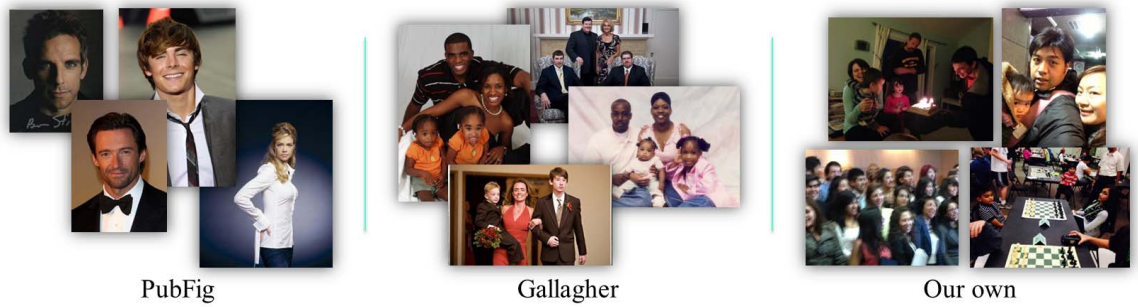

Fig. 1. **Example images from two existing relevant collections and our own, Adience set.** Left: PubFig benchmark [3] images. Despite being considered "in the wild", these images are often clean in terms of viewing conditions, and enjoy participation from the subjects being photographed. Mid: The Gallagher collection [5], provides images with an intentional bias towards groups of people, typically facing the camera and posing for their shots. Right: Images from our collection, automatically uploaded to Flickr, without manual pre-filtering by their owners. Consequently, they include sideways facing subjects, motion blur, poor lighting and more, all of which present additional challenges to automated face analysis systems.

neural networks [7], [8], and shown there to be an extremely powerful means of avoiding over-fitting in these models. Here, we propose using a similar approach when training SVM, in an effort to avoid over-fitting due to the scarcity of available data, rather than the nature of the classification model used.

Bringing these together, we describe a system for age and gender estimation. Handling of facial images is performed by a robust facial alignment technique, our third contribution, which explicitly considers the uncertainty of the facial feature detections used to estimate the aligning transformation. The system employs standard facial features and linear SVM classifiers, but is trained using our proposed dropout-SVM. This system is tested on two variants of our own benchmark, as well as two variants of the Gallagher [5] benchmark. Our tests clearly demonstrate both the elevated challenge of our benchmark as well as the substantial improvement in performance of our proposed method.

To summarize, this paper makes the following contributions.

- **Age/Gender benchmark of unfiltered face photos.** We provide a benchmark designed to reflect more realistic face processing applications than those currently used.
- **Robust face alignment and dropout-SVM.** We describe our own pipeline for inferring facial attributes, which includes a robust face alignment technique as well as the dropout-SVM approach to linear SVM training.
- **Performance evaluation.** We report an extensive evaluation, comparing both alternative benchmarks and their respective levels of difficulty, as well as the capabilities of automatic age and gender estimation systems. We show our own system to outperform others by substantial margins on all the evaluated benchmarks.

Finally, in order to promote reproducibility of our results, our data and code is publicly available from the project webpage: www.openu.ac.il/home/hassner/Adience.

II. Previous Work

Estimating the age of a person appearing in a photo, from that person's facial features, has been studied at length in the past, though far less than the related problem of face recognition. A comprehensive survey of methods and data has previously been offered by [9] and more recently in [10]. Here we provide a cursory overview of this work, referring the reader to those papers for a more in-depth treatment.

Broadly speaking, previous work on this problem can be categorized in terms of how face images are represented and how ages are estimated.

Face representations. Early work on age estimation [16], inspired by studies of aging in biology [17], proposed representing faces by obtaining facial measurements that change with a person's age. These so-called anthropometric models, consider measurements taken from different facial locations, and analyzed using expert-tailored knowledge, represented as manually defined statistical models, to determine age. More recently, [18] presented a similar approach for modeling the age of young people. These methods, based on their need for accurate localization of facial feature detections, may not be suitable for the realistic scenarios considered here.

An alternative approach seeks to model the high dimensional region in some feature space, which characterizes how facial appearances change with age. To this end, previous methods have used subspaces [19] or high dimensional manifolds [14]. Similar to the anthropometric models, however, they assume accurate alignment of faces, and the underlying assumption of the manifold structure of facial appearances across ages, which may not strictly be true.

Here we take a different approach to face image representation. We assume that age appearances and variations can be modeled using example images. Previous related approaches use low-level features to represent these examples. In particular, [4] used local binary patterns (LBP) [20], Gabor features [21] were used by [22], both features were recently used by [23], features inspired by the processing in the visual cortex, the biologically-inspired features (BIF) [24] were used in [25] (as well as others), finally, image patches extracted from different image regions were employed by [26].

Face discrimination. Age estimation, depending on the application domain, can be a regression problem or a multi-label classification task [9]. Here, we focus on age classification, using the terms estimation and classification interchangeably, and review previous work related to this task. Early work on age classification used neural networks to classify face images according to age [27]. Somewhat later, [28] compared a number of different classification schemes, amongst which neural networks were tested as well, and showed that

TABLE I

BENCHMARKS FOR AGE AND GENDER ESTIMATION FROM PHOTOS. WITH THE EXCEPTION OF THE FG-NET AGING AND UIUC-IFP-Y BENCHMARKS, THE TABLE INCLUDES ONLY BENCHMARKS WHICH ARE PRESENTLY AVAILABLE ONLINE TO THE RESEARCH COMMUNITY

Benchmark	Year	♯ **Images**	♯ **Subjects**	♯ **Age groups**	**Gender**	**In the wild**	**Notes**
FG-NET Aging [11]	2002	1,002	82	Accurate ages	Yes	No	This benchmark, though frequently used in the past, to our knowledge, is no longer available for download.
MORPH [12]	2006	55,134	13,000	Accurate ages	Yes	No	Information provided for the academic distribution "Album 2"
UIUC-IFP-Y Internal Aging [13], [14]	2008	8,000	1,600	Accurate ages	Yes	No	Not publicly available, yet may presumably be obtained from its authors.
PubFig [3]	2009	58,797	200	5	Yes	Yes	Celebrity photos from media websites. Biased towards high resolution, clear images with subjects posing for the camera.
Gallagher group photos [5]	2009	5,080	28,231	7	Yes	Yes	Designed for studying group photos, and so contains strong biases towards forward facing, artificially posed faces.
VADANA [15]	2011	2,298	43	4	Yes	Yes	Mostly frontal faces with large number of images per subject
Us, Adience	2014	26,580	2,284	8	Yes	Yes	See Section IV and Table II

their best performance was obtained by using a quadratic regression system, relating the parameters of an active appearance model [29] with numerical age labels. In [30] eleven low-dimensional subspaces, each one representing a different age, are computed. Classification of a face image is performed by considering the likelihood of it belonging to each of these subspaces. Support vector machines were used for classifying age in [14], [31], and more recently [32].

More recently, some have suggested different ways of partitioning the space of face images based on age. One such example is the ordinal subspaces approach of [33], which uses a flat partitioning scheme. Finally, others have proposed hierarchical partitioning models, including [34], which use an "AND-OR" graph partitioning of age progression and [35] which use age classifier hierarchy models.

A. Gender Classification

Gender classification has received considerable attention over the years, both for its potential contribution to face recognition [3], as well as its applications in human computer interaction, soft biometrics [36], [37] and more. For a rigorous survey of the methods developed for this problem over the years, we refer to [38] or, the more recent [39].

Some of the earliest attempts to automatically estimate gender used neural networks [40]. Later, [41] proposed the use of global features for gender classification. The contribution of 3D head structure for gender classification was explored in [42]. Face image intensities were directly classified using SVM in [43] and later again using AdaBoost in [44]. Also using AdaBoost, [45] used local binary patterns (LBP) [20] rather than intensities. Contrary to these, others have considered local image information. These include SIFT features in [46] and Haar wavelets in [47].

The methods listed above were mostly developed and tested using photos obtained under constrained viewing conditions, often using various subsets of the FERET benchmark (see Section II-B). Recently, work has shifted towards more challenging viewing conditions, following a similar development in face recognition research. Real-time performance when classifying gender in real-world images was the emphasis in [48]. Somewhat related to our work here, LBP was used along with SVM classifiers in [49]. In [50] a viewpoint-invariant appearance model was developed and used to represent faces in a gender classification system. Face images were combined with other biometric cues, notably fingerprints, in [51]. Finally, gender classification from video streams was recently proposed in [52].

B. Existing Benchmarks

We survey the benchmarks used by age and gender estimation systems. Various properties of these benchmarks are summarized in Table I. We refer the reader to [9] and [15] for a more throughout survey. We note that this report considers only sets which we know are, or were, publicly available. Other sets listed in [9] may be obtained by direct request, but were excluded from this summary.

Possibly the most well-used benchmark for age estimation has been FG-NET aging set [11]. It consists of about 1,000 images of 82 subjects, labeled for accurate age. These photos were acquired under controlled conditions, and so reflect less challenges than those expected of modern face recognition systems. Not surprisingly, performance on this set has long since saturated, reaching mean average age estimate errors of less than 5% (see [53]).

Another popular benchmark used by many in the past is the MORPH set [12], collected by the Face Aging Group at the University of North Carolina at Wilmington. It is partitioned into several subsets, or albums, of which "Album 2" is available for academic purposes. It contains over 55,000 images of 13,000 individuals. It too, like the FG-NET set, contains images under highly controlled viewing conditions. Over the years, performance on this set has also saturated, with systems demonstrating performances reaching near-perfect scores (e.g., [33]).

The UIUC-IFP-Y Internal Aging [13], [14], extensively used by the SMILE lab at Northeastern University, is not publicly available due to intellectual property limitations, but presumably may be obtained by contacting its authors directly. It offers 8,000 images of 1,600 voluntary Asian

subjects (half male, half female) in outdoor settings. This set too, was produced under lab-controlled conditions, and so unsurprisingly, performance measured by mean average age prediction error on this set has been reported to be near perfect [9].

Recently, following the shift towards face recognition "in the wild" (e.g., the LFW set [1]), benchmarks for age and gender estimation have also been assembled using unconstrained images. The first, originally designed for face recognition, is the Public Figures benchmark (PubFig) [3]. It provides attribute labels for the purpose of improving face recognition systems. It includes images from news and media websites which are typically of high quality, with subjects collaborating with the camera, posing for the shot. Its construction emphasized many images for each individual, and so it includes nearly 60,000 photos of only 200 celebrity faces. Despite providing age and gender information, we are unaware of previous work which used this set for estimating these attributes.

In [5], Gallagher and Chen proposed a benchmark for the study of groups of people, posing for the camera (e.g., family photos). Photos in this collection therefore typically present multiple subjects, in forward facing (towards the camera) poses, each face in relatively low resolution. Thus, the median face included in this set has only 18.5 pixels between eye centers, and 25% of the faces occupy less than 12.5 pixels. The age labels provided in this set make it a convenient choice for studying age estimation, with recently reported results on this set, still very far from reaching the near-perfect performances on the other sets [54].

Finally, the VADANA set was recently proposed in [15]. With 2,298 images of 43 subjects, it is substantially smaller than its recent predecessors, but unlike them, provides multiple images of the same subjects in different ages, allowing for the study of age progression of the same face.

III. Face and Gender Estimation System

A. Overview of Our Approach

We next describe our system for age and gender classification. As a key design choice, we model our system after similar systems successfully applied for face recognition [55]. Specifically, our pipeline consists of detection, alignment, and identification (representation and classification). This choice is motivated by the desire to highlight the specific contributions of the novel aspects of our pipeline – the face alignment of Section III-B and the dropout-SVM proposed in Section III-C – rather than fine-tune elements of a whole system. As a fortunate byproduct, we obtain state-of-the-art results on both our tasks with the same pipeline. Having the same general approach perform well on different problems testifies to its effectiveness.

Detection and alignment. Given a photo, we begin by applying the standard Viola and Jones face detector [56]. Detected faces are then aligned to a single reference coordinate frame using the method proposed in Section III-B.

Representation. Aligned faces are encoded using several popular global image representations. Here we chose the local binary patterns (LBP) of [20], [57], and [58] and the related Four Patch LBP codes (FPLBP) of [59]. These were selected due to their successful application to face recognition problems [60], as well as their efficient computation and representation requirements [61]. Our system is agnostic to the particular image representations used, and so other face descriptors can be used instead or in addition to the ones used here.

Classification. Classification is performed using standard linear SVM [6] trained using the feature vector representations listed above. Here, we examine each descriptor independently, or combine multiple descriptors by concatenating them into single long feature vectors (Section V). Training is performed using our own dropout-SVM scheme described in Section III-C. Classification of gender is performed using a single linear SVM classifier; for the multi-label age classification, we use a one-vs-one linear-SVM arrangement. Our choice of a simple linear-SVM is motivated by the desire to reduce the risk of over-fitting. Simpler classifiers have been shown to work well for these problems in the past (see [62]). Our results show, however, that by training linear-SVM classifiers by dropout-SVM, excellent results may be obtained without apparent over-fitting.

B. Face Alignment With Uncertainty

Past work has acknowledged the major role face alignment plays in the accuracy of face recognition systems. In [55], the "funneling" approach to image alignment was proposed, based on sets of unaligned face image examples, and its impact on recognition performance was demonstrated. An additional significant leap in face recognition performance was obtained using the commercial alignment system used by [63] to resolve in-plane misalignments in the LFW benchmark. More recently, deep learning was used for image alignment in [64]. Here we offer an alternative approach, based on facial feature detection.

Specifically, we employ the robust facial feature detector recently proposed by Zhu and Ramanan [65]. It detects 68 specific facial features, including the corners of the eyes and mouth, the nose and more. By selecting ideal coordinates for each of these points an affine transformation can presumably be obtained and the images aligned. In practice, however, errors in point localizations as well as the variability of face shapes can often result in unstable alignment results.

To address this, we note that some detections are more reliable than others: The corners of the eyes, for examples, are easier to localize than, say, the cheekbones. In order to accurately align faces, these uncertainties should be accounted for. Doing so requires that we know the uncertainty associated with each of the 68 features, but this information can only be estimated once the faces are already aligned. In order to resolve this chicken-and-egg problem, we take an Iterative Re-weighted Least Squares (IRLS) approach [66], [67].

Specifically, we assume reference facial feature points $\{\mathbf{r}_i\}_{i=1..68} = \{(x_i^r, y_i^r)\}_{i=1..68}$ in a frontal-facing face. Given corresponding feature points, $\{\mathbf{q}_{i,j}\}_{i=1..68, j=1..N} = \{(x_{i,j}^q, y_{i,j}^q)\}_{i=1..68, j=1..N}$ (for N training photos), detected on a query photo, we begin by computing an initial, time-0

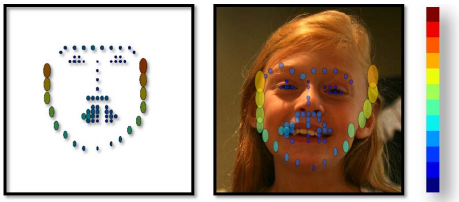

Fig. 2. **Visualization of the uncertainties associated with the 68 facial feature detections.** Left: Uncertainties on the reference coordinate frame. Mid: Uncertainties on a sample face image. Note that ellipses are aligned with the image axes, as we assume variance along the axes is uncorrelated. In both images color codes the amount of uncertainty; color-bar provided on the right.

affine transformation $\mathbf{H}_j^0$, relating facial feature points $\mathbf{q}_{i,j}$, detected in the j'th query photo, with their corresponding reference points, $\mathbf{r}_i$, by using standard least squares. Feature points in all queries are then projected using the standard

$$\mathbf{q}'_{i,j} = \mathbf{H}_j^0 \mathbf{q}_{i,j}.$$

For a given facial feature i, for all images j, we consider the variance along the x axis and the variance along the y axis of these projected points as the uncertainty values associated with this feature. These are used to estimate a new aligning transformation, $\mathbf{H}_j^1$ at time-1, this time, by weighted least squares. Although we tried using the covariance of the projected points, rather than assuming uncorrelated variance along the two axes, the additional computational costs associated with doing so did not provide noticeable improvements in results.

This process can be repeated, obtaining transformations $\mathbf{H}_j^2, \mathbf{H}_j^3$ and so fourth, until convergence [66]. In practice, however, we found that a single iteration was sufficient to provide a considerable improvement in performance (Section V). Fig. 2 demonstrates the uncertainties computed by our method following a single iteration, visualized both in the reference coordinate system and on a sample query photo.

C. Dropout-SVM

A combination of few training samples, high dimensional data, and complex classification models is often a recipe for over-fitting [67]. Here, we use linear-SVM, a simple model, less susceptible to over-fitting. The shortage of data available for the study of age estimation, and the high dimensionality of standard face representations (Section III-A) can nevertheless lead to over-fitting of our classifiers and subsequently, to sub-optimal performance.

Various methods designed to make SVM robust to over-fitting have previously been proposed. Many of these employ regularization terms, added to the objective function [68] or the covariance estimation [69]. Others propose the use of more robust optimization procedures [70]. Naturally, dimensionality reduction techniques may also be used in order to reduce the number of features, and hence the complexity of the learned models compared to the available training instances, thereby improving the models' ability to generalize well. These methods require turning or learning of additional parameters, modified optimizations, and, in the case of dimensionality reduction, loss of potentially valuable information.

By contrast, our approach is inspired by the recent success of *dropout* learning for deep neural networks [7], [8] where over-fitting is a major concern. While training, dropout essentially omits neurons from the network with some probability p_{drop} for each feature, in each sample, to be dropped, usually chosen as $p_{drop} = 50\%$. By doing so, it was claimed that neurons must better adapt to the input data, relying less on other neurons in the network, and so obtain representations which are more distributed and better generalized.

The relation between SVM and neural networks has been noted in the past; SVM can be considered equivalent to a single-layer neural network [71]. Thus, a similar dropout procedure can conceivably be applied to train SVM classifiers: We propose that, rather than omit neurons, training is performed by randomly dropping-out the output of input-layer neurons. For the case of linear-SVM, this is equivalent to assigning the value zero, at random, to training features (elements of the feature vectors used for training). This random selection is applied to each training instance separately; different features are randomly selected and set to zero for different training instances. In Section V we evaluate two variations of this scheme: in one, 50% of the input values are dropped (dropout 0.5), and the other, where each training vector is considered twice, each time with an independently selected random subset of 80% of its values dropped (dropout 0.8).

Dropping out many of the values from the training instances requires that the obtained model be modified accordingly. Specifically, following training with a dropout of rate of p_{drop}, we divide all the coefficients of the resulting linear SVM model by $(1 - p_{drop})$. This compensates for the dropped-out values, and provides a model suitable for test instances which include all their values.

Discussion. Robust classification is often expressed in terms of a classifier's capabilities to generalize beyond a bounded amount of perturbations of the training set [72], [73]. Here, our data is limited both in the number of samples available, and – following the strong alignment discussed above and the invariant descriptors we use – in the natural variation of values for each feature. As a consequence, these perturbations, implicit in the training process, are limited in their capacity to capture the underlying structure of the problem. The process described above can therefore be considered as introducing extreme perturbations to the input, thereby infusing the learning process with a much greater variability of possible values for every feature and increasing the robustness of the learned model. As a consequence, as we show in Section V, it has a profound effect on the performance of our system.

IV. The Adience Benchmark

In order to facilitate the study of age and gender recognition, we provide our own, public data set and benchmark of face photos. Our key design principle is that the data we use should be as true as possible to challenging real-world conditions. As such, it should present all the variations in appearance, noise, pose, lighting and more, that can be expected of images taken without careful preparation or posing. We next describe the database collection process, as well as the testing protocols used with these photos in our tests.

TABLE II
BREAKDOWN OF FACES PER-LABEL IN OUR COLLECTION. T. GND. DENOTES THE TOTAL NUMBER OF PHOTOS IN EACH AGE CATEGORY WITH GENDER LABELS. TOTAL DENOTES ALL PHOTOS IN EACH CATEGORY LABELED FOR AGE (INCLUDING THOSE WITH NO GENDER LABEL)

Complete version (faces in the range of $\pm 45°$ yaw from frontal)									
	0-2	4-6	8-13	15-20	25-32	38-43	48-53	60-	Total
Male	745	928	934	734	2,308	1,294	392	442	8,192
Female	682	1,234	1,360	919	2,589	1,056	433	427	9,411
T. Gnd.	1,427	2,162	2,294	1,653	4,897	2,350	825	869	19,487
Total	2,519	2,163	2,301	1,655	4,950	2,350	830	875	
Front version (faces in the range of $\pm 5°$ yaw from frontal)									
Male	557	691	738	501	1,602	875	273	272	5,824
Female	492	911	956	630	1,692	732	295	309	6,455
T. Gnd.	1,049	1,602	1,694	1,131	3,294	1,607	568	581	13,649
Total	1,843	1,602	1,700	1,132	3,335	1,607	572	585	

A. Database Preparation and Contents

The source for the photos in our set are Flickr.com albums, produced by automatic upload from iPhone 5 or later smartphones. By "opting in", iPhone users can have the photos they take automatically uploaded to their personal Flickr albums, for backup. Flickr users may additionally opt to make these albums publicly available through the Creative Commons (CC) license. We use such photos in our collection.

Image collection consisted of the following steps. Photos downloaded from ∼200 public Flickr albums were processed by first running the Viola and Jones face detector [56], and then detecting facial feature points using a modified version of the code provided by the authors of [65]. Presumably due to the recent "selfie" trend, many faces in these albums appear at different roll angles. To avoid missing these faces, the face detection process was applied to each image, rotated 360° degrees in 5° increments.

Faces for which the facial feature detection step failed were considered too noisy or small and were discarded. The pose estimated using the same code was additionally used to discard faces which were at a greater than ±45° yaw angle from 0 (forward facing). Finally, all images were manually labeled for age, gender and identity using both the images themselves and any available contextual information (image tags and associated text, additional photos in the same album etc.). Though we do not use identity labels here, their contribution to face recognition can be explored in the future.

Table I provides a comparison of the general properties of our collection, compared to other sets. Evidently, our set contains substantially more images than the Gallagher collection, and substantially more subjects than the PubFig set. Fig. 1 compares a few examples of photos from our Adience collection to those in two recent unconstrained sets, and demonstrates the greater variability in terms of pose and viewing conditions of our photos. Breakdown of the number of images included in each class is provided in Table II. We note that not all faces could reliably be labeled for gender (mostly babies) and not all could be labeled for age.

B. Benchmark Protocols

We define test protocols to benchmark the performance of gender and age estimation techniques, using our collection. We use two variations of our data set: the frontal and the complete sets (See Table II). The frontal set includes only roughly frontal facing faces; that is, faces which were determined to be within ±5° yaw angle from a forward facing face. The complete set includes faces with up to ±45° degrees of yaw.

Training and testing is performed using 5-fold cross validation with splits pre-selected to eliminate cases of images from the same Flickr album appearing in both training and testing sets in the same fold. We use the same splits for both the gender and the age classification tasks. Results therefore include both mean classification accuracy (age or gender), including, ± standard error over the five folds.

Gender labeling is a binary classification task. The output is both the accuracy of the classification, as well as an ROC curve. For the multi-class age estimation we report mean classification error across all age groups. Following [74] we provide also the 1-off age error classification rate, where errors of one age group are considered correct classifications.

V. EXPERIMENTS

A. Implementation Details

Our system pipeline is implemented as a combination of C++ and Python code. We produced our own Python implementations for the two descriptors used – LBP and FPLBP. Face detection is performed using standard OpenCV routines, wrapped to consider 360° roll versions of the input image, at 5° increments, and different classifier cascades for improved robustness. Facial feature localization was performed using the code made available by [65], with some optimizations applied for better run-time performance.

The age and gender experiments reported below use our own Adience benchmark, as well as the Gallagher group photos benchmark. Our decision to use the Gallagher set for this purpose follows the conclusion of the BeFIT project, which designated this collection as the most suitable existing age and gender classification benchmark [75]. This, as earlier benchmarks, FG-NET and MORPH-II in particular, have been saturated; results reported on these benchmarks have long since reached near perfect performance, and so slight differences in the reported performances are hard to interpret.

B. Age Classification

We test various components of our system on the Gallagher benchmark and our Adience collection. As the age groups in the two benchmarks are different, direct comparison of the difficulty each data set represents is impossible. We therefore provide these only for gender in Section V-C. Here, we refer to "exact" classification as the mean accuracy, across all age groups, of predicting the true age label. "1-off" implies counting labeling errors, one age group removed from the true label, as correct. Dropout 0.5 denotes dropout-SVM with 50% probability of dropping features; Dropout 0.8 denotes 80% of the input features dropped, randomly and independently, from two copies of the input feature vector.

Gallagher age classification results. We compare our pipeline with what we know to be the state of the art on the Gallagher benchmark, using the testing protocol defined by that benchmark. These include results reported by Shan in [74], and more

TABLE III

AGE CLASSIFICATION ON THE GALLAGHER BENCHMARK. MEAN ACCURACY (± STANDARD ERRORS) OVER THE SEVEN AGE CATEGORIES IN THE GALLAGHER BENCHMARK. BOLDFACE ARE BEST SCORING

	Method	Exact	1-off
1	Shan 2010 [74]	55.9	87.7
2	Alnajar et al. 2012 [54]	59.5	–
	Our system, no alignment		
3	LBP	56.4 ± 0.4	92.7 ± 0.2
4	FPLBP	60.2 ± 0.6	92.0 ± 0.2
5	LBP+FPLBP	57.5 ± 0.5	93.7 ± 0.2
6	LBP+FPLBP+Dropout 0.5	62.7 ± 0.6	94.5 ± 0.1
7	LBP+FPLBP+Dropout 0.8	65.6 ± 0.6	94.0 ± 0.2
	Our system, including alignment		
8	LBP	58.0 ± 0.1	94.1 ± 0.2
9	FPLBP	61.0 ± 0.5	92.2 ± 0.2
10	LBP+FPLBP	59.0 ± 0.4	**95.3** ± **0.2**
11	LBP+FPLBP+PCA 0.5	46.8 ± 0.6	90.1 ± 0.3
12	LBP+FPLBP+PCA 0.8	41.3 ± 0.4	83.9 ± 0.3
13	LBP+FPLBP+Dropout 0.5	64.3 ± 0.6	**95.3** ± **0.3**
14	LBP+FPLBP+Dropout 0.8	**66.6** ± 0.7	94.8 ± 0.3

recently, by Alnajar et al. in [54]. Our method was tested with and without our alignment of Section III-B. Without alignment indicates that faces were aligned using only the annotations available for that collection; namely, the locations of the centers of the eyes. We used these locations to solve for scale and rotation. The results reported for aligned images, refer to the Gallagher set images following our alignment process of Section III-B.

The results in Table III clearly show the substantial improvement of our method compared to previous work, with our 1-off scores reaching near perfect performance on this set. Without alignment and without the dropout-SVM training (Table III, rows 3–5), our performance is comparable to previous work. Dropout-SVM alone provides an additional ∼6% performance boost (rows 6, 7). The addition of a robust alignment step further improves our results by as much as 2.5% (row 10), with our full system, including alignment and dropout-SVM, obtaining the highest results (rows 13, 14).

As previously mentioned, we use dropout-SVM to address potential over-fitting. A well-known alternative is dimensionality reduction. We compare our dropout-SVM with standard PCA, used here to reduce the dimension of the concatenated descriptors to half and to 20% of their size, (corresponding to dropout rates of 50% and 80%). Results are reported in rows 11 and 12. Although PCA was used to reduce dimensionality by only half, its performance falls well below that of dropout. This is likely due to the loss of information inherent in the process. Dropout, by comparison, classifies the entire test descriptors, unchanged.

Adience age classification results. Results on our own benchmark are provided in Table IV. Here, we apply our method following alignment, in the absence of the eye localizations provided in the Gallagher benchmark. Here, dropout-SVM provides a lesser performance gain, though SVM trained with dropout still provides better results than without (row 7). Interestingly, for the 1-off results, dropout learning actually damages performance. We believe this is due to the finer granularity of our age groups and the consequent smaller inter-class variation in appearances. This, along with the dropout-SVM perturbations of the training samples, may be producing classifiers which are over-generalized.

TABLE IV

AGE CLASSIFICATION ON THE ADIENCE BENCHMARK. MEAN ACCURACY (± STANDARD ERRORS) OVER THE EIGHT AGE CATEGORIES IN THE ALIGNED ADIENCE SET. BOLDFACE ARE BEST SCORING

	Method	Exact	1-off
1	LBP	41.4 ± 2.0	78.2 ± 0.9
2	FPLBP	39.8 ± 1.8	74.6 ± 1.0
3	LBP+FPLBP	44.5 ± 2.3	**80.7** ± 1.1
4	LBP+FPLBP+PCA 0.5	38.1 ± 1.4	75.5 ± 0.9
5	LBP+FPLBP+PCA 0.5	32.9 ± 1.6	67.7 ± 1.1
6	LBP+FPLBP+Dropout 0.5	44.5 ± 2.2	80.6 ± 1.0
7	LBP+FPLBP+Dropout 0.8	**45.1** ± 2.6	79.5 ± 1.4

Fig. 3. **Age classification errors made by our full system on the Adience benchmark.** Using LBP + FPLBP, alignment and dropout of 50%. Top row are young people, classified as old. Bottom are old people classified as young.

Fig. 3 provides examples of the age classification errors made by our system (label errors of more than two age groups). It presents young faces mis-labeled as old (top row), and vice versa. Although some examples are extremely challenging, from a computer vision perspective, due to severe blur, occlusions, makeup and more, most if not all ages in these photos can be estimated by a human observer. This testifies to the gap that still remains between human and machine capabilities in performing this task.

C. Gender Estimation

Gallagher gender classification results. We provide average accuracy results on a binary gender classification task, over five-fold cross validation tests. Results are reported in Table V. Fig. 4 (Left) provides ROC curves for the various components of our method, on the aligned Gallagher images (Section III-B). Our results demonstrate the consistent improvement obtained by adding both the alignment (rows 6–8), and then the dropout-SVM training (rows 11,12).

Cross-dataset results. In order to evaluate the relative difficulties of the Gallagher benchmark and our own, we measure the accuracy obtained by using one dataset for training and another for testing. In these tests, train and test sets used the same images defined by the respective test protocols of each benchmark (e.g., tests on an Adience test set were performed using training on a Gallagher training set). We additionally considered the two variants of each collection, where only forward facing faces were included (i.e., faces up to a ±5° yaw angle from directly forward facing). In all cases we used

TABLE V
GENDER ESTIMATION ON THE GALLAGHER BENCHMARK. MEAN ACCURACY (± STANDARD ERRORS)

	Method	Accuracy
	No alignment	
1	LBP	82.4 ± 0.5
2	FPLBP	82.5 ± 0.2
3	LBP+FPLBP	83.1 ± 0.3
4	LBP+FPLBP+Dropout 0.5	85.2 ± 0.3
5	LBP+FPLBP+Dropout 0.8	86.8 ± 0.1
	Including alignment	
6	LBP	84.6 ± 0.3
7	FPLBP	83.6 ± 0.3
8	LBP+FPLBP	86.4 ± 0.2
9	LBP+FPLBP+PCA 0.5	82.5 ± 0.2
10	LBP+FPLBP+PCA 0.8	80.6 ± 0.2
11	LBP+FPLBP+Dropout 0.5	87.5 ± 0.2
12	LBP+FPLBP+Dropout 0.8	**88.6** ± 0.3

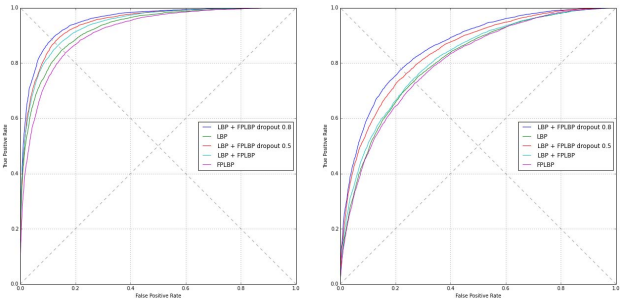

Fig. 4. **ROC curves for gender estimation results.** Left: Results on the Gallagher benchmark; right: results on the Adience benchmark. See text for more details.

the method of Section III-B to align the faces. Results are provided for LBP and FPLBP, as well as the two concatenated and dropout-SVM 0.5 used.

Our results are reported in Table VI. These clearly show the difficulty level of our proposed data, compared to the previous collection. In particular, no matter what training set is used, testing on the full Adience collection produces the lowest scores (shaded third row), with the frontal version of our benchmark coming in second.

An important observation, evident from these results, concerns database bias [76]; that is, how well training examples from one set generalize to test samples in other sets. The columns in Table VI show the variation in performance for different test sets, using each training set in turn. The columns for the full Adience benchmark, show a difference of about 5% between the best and worst performances. This should be compared to the 10% difference between tests sets, when using the Gallagher benchmark for training. These differences imply a greater bias in the Gallagher image collection. This is not surprising considering its original purpose, evaluating group photos, and the manner in which its photos were collected (i.e., using search terms relating to family photos etc.).

Fig. 4 (Right) provides ROC curves demonstrating the contributions of various components of our system, all tested with the aligned version of the Adience benchmark. The curves for LBP, FPLBP, and LBP+FPLBP+Dropout 0.5 represent the same tests used in Table VI. In addition, we provide a few failed gender classification examples in Fig. 5. Here again,

TABLE VI
GENDER ESTIMATION, CROSS-DATASET TRAINING. RESULTS USING THREE VARIANTS OF OUR PIPELINE. COLUMNS REPRESENT THE SOURCE FOR TRAINING AND ROWS THE SOURCE FOR TESTING. WE REPORT AVERAGE ACCURACY IN FIVE-FOLD CROSS VALIDATION TESTS (± STANDARD ERRORS). "GLGR" DENOTES THE GALLAGHER SET; "AD", OUR ADIENCE SET; "FNT" REPRESENTS IMAGES WITH ONLY NEAR-FRONTAL FACES (THE FRONTAL VERSION). NOTE THAT NO MATTER WHAT TRAINING IS USED, THE RESULTS ON THE ADIENCE BENCHMARK (SHADED THIRD ROW) ARE THE LOWEST, TESTIFYING TO ITS ELEVATED DIFFICULTY

Test	LBP Train Source			
	Glgr	Glgr-Fnt	Ad	Ad-Fnt
Glgr	84.6 ± 0.3	84.7 ± 0.2	79.9 ± 0.4	78.5 ± 0.2
Glgr-Fnt	85.4 ± 0.4	86.5 ± 0.2	80.8 ± 0.3	79.9 ± 0.2
Ad	73.3 ± 1.2	73.5 ± 1.4	73.4 ± 0.7	72.9 ± 0.5
Ad-Fnt	74.4 ± 1.2	75.3 ± 0.9	74.6 ± 0.6	75.0 ± 0.3
Test	FPLBP Train Source			
	Glgr	Glgr-Fnt	Ad	Ad-Fnt
Glgr	83.6 ± 0.3	81.2 ± 0.1	79.0 ± 0.3	77.7 ± 0.3
Glgr-Fnt	84.3 ± 0.4	82.9 ± 0.3	79.6 ± 0.4	79.3 ± 0.6
Ad	74.1 ± 0.9	72.0 ± 0.8	72.6 ± 0.9	72.1 ± 1.0
Ad-Fnt	75.5 ± 0.8	74.4 ± 0.8	74.0 ± 0.8	74.1 ± 1.0
Test	LBP+FPLBP+Dropout 0.5 Train Source			
	Glgr	Glgr-Fnt	Ad	Ad-Fnt
Glgr	87.5 ± 0.2	87.1 ± 0.2	83.0 ± 0.3	82.9 ± 0.3
Glgr-Fnt	87.9 ± 0.2	88.4 ± 0.2	84.0 ± 0.2	84.9 ± 0.2
Ad	75.9 ± 1.2	75.8 ± 1.6	76.1 ± 0.9	75.2 ± 1.0
Ad-Fnt	77.4 ± 1.2	77.8 ± 1.3	77.2 ± 0.7	77.0 ± 1.0

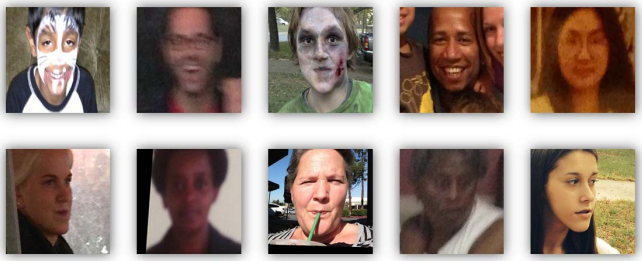

Fig. 5. **Gender classification errors made by our full system on the Adience benchmark.** Using LBP + FPLBP, alignment and dropout of 50%. Top row are males classified as females. Bottom are females classified as males.

these images demonstrate the difficulty of the classification task defined by our benchmark, due to the myriad of confounding factors which affect facial appearances in our images.

VI. CONCLUSIONS

Age, gender and other facial traits represent information important to a wide range of tasks. Despite this, estimating these traits from facial appearances has received less attention than face recognition. Here, we are primarily motivated by the observation that the amount of data available for the study of a computer vision problem, in particular the problems considered here, can have an immense impact on the machine capabilities developed to solve it. In answer to this, we provide two contributions: a new and extensive data set and benchmark for the study of age and gender estimation, and a classification pipeline designed with an emphasis on making the most of what little data is available. In addition, we describe a novel,

robust facial alignment technique, based on iterative estimation of the uncertainties of facial feature localizations. Finally, we provide extensive tests, which demonstrate the improved capabilities of our method, alongside the heightened difficulty level of our new benchmark.

Our tests leave room for future work. This is evident by considering the drop in performance exhibited when using our benchmark compared to previous ones. It is also evident when considering failed results – all of which are be easy for a human to correctly classify, but are still very challenging from a computer vision perspective. To this end, our pipeline can be significantly improved, in the same way face recognition systems have improved in the last few years. One possible direction is considering uncertainties when performing 3D pose normalization as in [77]. In a different line of work, our Adience set is labeled for identity. Though not used here, this information allows the it to be used as a more challenging alternative to the benchmarks used today, consequently providing a new driving force in improving face recognition capabilities beyond their present state.

REFERENCES

[1] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep. TR 07-49, 2007.

[2] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 529–534.

[3] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *Proc. IEEE ICCV*, Oct. 2009, pp. 365–372. [Online]. Available: http://www.cs.columbia.edu/CAVE/databases/pubfig

[4] W.-L. Chao, J.-Z. Liu, and J.-J. Ding, "Facial age estimation based on label-sensitive learning and age-oriented regression," *Pattern Recognit.*, vol. 46, no. 3, pp. 628–641, Mar. 2013.

[5] A. C. Gallagher and T. Chen, "Understanding images of groups of people," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 256–263.

[6] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates, Inc., 2012, pp. 1097–1105.

[8] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. (2012). "Improving neural networks by preventing co-adaptation of feature detectors." [Online]. Available: http://arxiv.org/abs/1207.0580

[9] Y. Fu, G. Guo, and T. S. Huang, "Age synthesis and estimation via faces: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 11, pp. 1955–1976, Nov. 2010.

[10] H. Han, C. Otto, and A. K. Jain, "Age estimation from face images: Human vs. machine performance," in *Proc. Int. Conf. Biometrics*, Jun. 2013, pp. 1–8.

[11] A. Lanitis. (2002). *The FG-NET Aging Database*. [Online]. Available: http://www-prima.inrialpes.fr/FGnet/html/benchmarks.html

[12] K. Ricanek and T. Tesafaye, "MORPH: A longitudinal image database of normal adult age-progression," in *Proc. 7th Int. Conf. Automat. Face Gesture Recognit.*, Apr. 2006, pp. 341–345. [Online]. Available: http://www.faceaginggroup.com/morph

[13] Y. Fu and T. S. Huang, "Human age estimation with regression on discriminative aging manifold," *IEEE Trans. Multimedia*, vol. 10, no. 4, pp. 578–584, Jun. 2008.

[14] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1178–1188, Jul. 2008.

[15] G. Somanath, M. V. Rohith, and C. Kambhamettu, "VADANA: A dense dataset for facial image analysis," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Nov. 2011, pp. 2175–2182.

[16] Y. H. Kwon and N. da Vitoria Lobo, "Age classification from facial images," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1994, pp. 762–767.

[17] L. G. Farkas, *Anthropometry of the Head and Face in Medicine*. New York, NY, USA: Elsevier, 1981.

[18] N. Ramanathan and R. Chellappa, "Modeling age progression in young faces," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. Jun. 2006, pp. 387–394.

[19] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 12, pp. 2234–2240, Dec. 2007.

[20] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[21] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, Apr. 2002.

[22] F. Gao and H. Ai, "Face age classification on consumer images with Gabor feature and fuzzy LDA method," in *Advances in Biometrics*. Berlin, Germany: Springer-Verlag, 2009, pp. 132–141.

[23] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim, "Age estimation using a hierarchical classifier based on global and local facial features," *Pattern Recognit.*, vol. 44, no. 6, pp. 1262–1281, 2011.

[24] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature Neurosci.*, vol. 2, no. 11, pp. 1019–1025, 1999.

[25] G. Guo, G. Mu, Y. Fu, C. Dyer, and T. Huang, "A study on automatic age estimation using a large database," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 1986–1991.

[26] X. Zhuang, X. Zhou, M. Hasegawa-Johnson, and T. Huang, "Face age estimation using patch-based hidden Markov model supervectors," in *Proc. 19th Int. Conf. Pattern Recognit.*, Dec. 2008, pp. 1–4.

[27] T. Kanno, M. Akiba, Y. Teramachi, H. Nagahashi, and A. Takeshi, "Classification of age group based on facial images of young males by using neural networks," *IEEE Trans. Inf. Syst.*, vol. 84, no. 8, pp. 1094–1101, Aug. 2001.

[28] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 1, pp. 621–628, Feb. 2004.

[29] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models-their training and application," *Comput. Vis. Image Understand.*, vol. 61, no. 1, pp. 38–59, Jan. 1995.

[30] K. Ueki, T. Hayashida, and T. Kobayashi, "Subspace-based age-group classification using facial images under various lighting conditions," in *Proc. 7th Int. Conf. Autom. Face Gesture Recognit.*, Apr. 2006, pp. 43–48.

[31] G. Guo, Y. Fu, T. S. Huang, and C. R. Dyer, "Locally adjusted robust regression for human age estimation," in *Proc. IEEE Workshop Appl. Comput. Vis.*, Jan. 2008, pp. 1–6.

[32] D. Cao, Z. Lei, Z. Zhang, J. Feng, and S. Z. Li, "Human age estimation using ranking SVM," in *Biometric Recognition*. Berlin, Germany: Springer-Verlag, 2012, pp. 324–331.

[33] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, "Ordinal hyperplanes ranker with cost sensitivities for age estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 585–592.

[34] J. Suo, S.-C. Zhu, S. Shan, and X. Chen, "A compositional and dynamic model for face aging," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 385–401, Mar. 2010.

[35] P. Thukral, K. Mitra, and R. Chellappa, "A hierarchical approach for human age estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2012, pp. 1529–1532.

[36] G. Mahalingam and C. Kambhamettu, "Can discriminative cues aid face recognition across age?" in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit. Workshops*, Mar. 2011, pp. 206–212.

[37] M. Demirkus, D. Precup, J. J. Clark, and T. Arbel, "Soft biometric trait classification from real-world face videos conditioned on head pose estimation," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 130–137.

[38] E. Makinen and R. Raisamo, "Evaluation of gender classification methods with automatically detected and aligned faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 541–547, Mar. 2008.

[39] D. Reid, S. Samangooei, C. Chen, M. Nixon, and A. Ross, "Soft biometrics for surveillance: An overview," in *Machine Learning: Theory and Applications* (Handbook of Statistics). Amsterdam, The Netherlands: Elsevier, 2013, pp. 327–351.

[40] B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski, "SEXNET: A neural network identifies sex from human faces," in *Proc. Conf. Adv. Neural Inf. Process. Syst.*, 1990, pp. 572–579.

[41] S. Gutta, H. Wechsler, and P. J. Phillips, "Gender and ethnic classification of face images," in *Proc. 3rd Int. Conf. Automat. Face Gesture Recognit.*, Apr. 1998, pp. 194–199.
[42] A. J. O'Toole, T. Vetter, N. F. Troje, and H. H. Bülthoff, "Sex classification is better with three-dimensional head structure than with image intensity information," *Perception*, vol. 26, no. 1, pp. 75–84, 1997.
[43] B. Moghaddam and M.-H. Yang, "Learning gender with support faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 707–711, May 2002.
[44] S. Baluja and H. A. Rowley, "Boosting sex identification performance," *Int. J. Comput. Vis.*, vol. 71, no. 1, pp. 111–119, 2007.
[45] N. Sun, W. Zheng, C. Sun, C. Zou, and L. Zhao, "Gender classification based on boosting local binary pattern," in *Proc. 3rd Int. Conf. Adv. Neural Netw.*, 2006, pp. 194–201.
[46] J.-G. Wang, J. Li, W.-Y. Yau, and E. Sung, "Boosting dense SIFT descriptors and shape contexts of face images for gender recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2010, pp. 96–102.
[47] Z. Yang, M. Li, and H. Ai, "An experimental study on automatic face gender classification," in *Proc. 18th Int. Conf. Pattern Recognit.*, vol. 3. 2006, pp. 1099–1102.
[48] D.-Y. Chen and K.-Y. Lin, "Real-time gender recognition for uncontrolled environment of real-life images," in *Proc. Int. Conf. Comput. Vis. Theory Appl.*, 2010, pp. 357–362.
[49] C. Shan, "Gender classification on real-life faces," in *Advanced Concepts for Intelligent Vision Systems*. Berlin, Germany: Springer-Verlag, 2010, pp. 323–331.
[50] M. Toews and T. Arbel, "Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 9, pp. 1567–1581, Sep. 2009.
[51] X. Li, X. Zhao, Y. Fu, and Y. Liu, "Bimodal gender recognition from face and fingerprint," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2590–2597.
[52] M. Demirkus, M. Toews, J. J. Clark, and T. Arbel, "Gender classification from unconstrained video sequences," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2010, pp. 55–62.
[53] K. Luu, K. Seshadri, M. Savvides, T. D. Bui, and C. Y. Suen, "Contourlet appearance model for facial age estimation," in *Proc. Int. Joint Conf. Biometrics*, Oct. 2011, pp. 1–8.
[54] F. Alnajar, C. Shan, T. Gevers, and J.-M. Geusebroek, "Learning-based encoding with soft assignment for age estimation under unconstrained imaging conditions," *Image Vis. Comput.*, vol. 30, no. 12, pp. 946–953, 2012.
[55] G. B. Huang, V. Jain, and E. Learned-Miller, "Unsupervised joint alignment of complex images," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.
[56] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, May 2004.
[57] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.
[58] T. Ojala, M. Pietikäinen, and T. Mäenpää, "A generalized local binary pattern operator for multiresolution gray scale and rotation invariant texture classification," in *Proc. 2nd ICAPR*, 2001, pp. 397–406 .
[59] L. Wolf, T. Hassner, and Y. Taigman, "Descriptor based methods in the wild," in *Proc. Post-ECCV Faces Real-Life Images Workshop*, 2008.
[60] L. Wolf, T. Hassner, and Y. Taigman, "Effective unconstrained face recognition by combining multiple descriptors and learned background statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 1978–1990, Oct. 2011.
[61] M. Guillaumin, J. Verbeek, and C. Schmid, "Is that you? Metric learning approaches for face identification," in *Proc. 12th ICCV*, Sep./Oct. 2009, pp. 498–505.
[62] J. Bekios-Calfa, J. M. Buenaposada, and L. Baumela, "Revisiting linear discriminant techniques in gender recognition," *Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 4, pp. 858–864, Apr. 2011.
[63] L. Wolf, T. Hassner, and Y. Taigman, "Similarity scores based on background samples," in *Proc. 9th ACCV*, 2009, pp. 88–97.
[64] G. B. Huang, M. A. Mattar, H. Lee, and E. G. Learned-Miller, "Learning to align from scratch," in *Advances in Neural Information Processing Systems 25*. Red Hook, NY, USA: Curran Associates, 2012.
[65] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Proc. Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2879–2886. [Online]. Available: http://www.ics.uci.edu/~xzhu/face/
[66] D. B. Rubin, "Iteratively reweighted least squares," in *Encyclopedia of Statistical Sciences*, vol. 4. Hoboken, NJ, USA: Wiley, 1983, pp. 272–275.
[67] C. M. Bishop *et al.*, *Pattern Recognition and Machine Learning*, vol. 1. New York, NY, USA: Springer-Verlag, 2006.
[68] C. Bhattacharyya, "Robust classification of noisy data using second order cone programming approach," in *Proc. Int. Conf. Intell. Sens. Inf. Process.*, 2004, pp. 433–438.
[69] G. R. G. Lanckriet, L. E. Ghaoui, C. Bhattacharyya, and M. I. Jordan, "A robust minimax approach to classification," *J. Mach. Learn. Res.*, vol. 3, pp. 555–582, Mar. 2003.
[70] H. Xu, C. Caramanis, and S. Mannor, "Robustness and regularization of support vector machines," *J. Mach. Learn. Res.*, vol. 10, pp. 1485–1510, Dec. 2009.
[71] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal margin classifiers," in *Proc. 5th Annu. Workshop Comput. Learn. Theory*, 1992, pp. 144–152.
[72] T. B. Trafalis and R. C. Gilbert, "Robust support vector machines for classification and computational issues," *Optim. Methods Softw.*, vol. 22, no. 1, pp. 187–198, 2007.
[73] S. Sra, S. Nowozin, and S. J. Wright, *Optimization for Machine Learning*. Cambridge, MA, USA: MIT Press, 2012.
[74] C. Shan, "Learning local features for age estimation on real-life faces," in *Proc. 1st ACM Int. Workshop Multimodal Pervasive Video Anal.*, 2010, pp. 23–28.
[75] P. Dago-Casas, D. González-Jiménez, L. L. Yu, and J. L. Alba-Castro, "Single- and cross- database benchmarks for gender classification under unconstrained settings," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Nov. 2011, pp. 2152–2159.
[76] A. Torralba and A. A. Efros, "Unbiased look at dataset bias," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 1521–1528.
[77] T. Hassner, "Viewing real-world faces in 3D," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 3607–3614. [Online]. Available: http://www.openu.ac.il/home/hassner/projects/ViewFaces3D

Eran Eidinger received the B.Sc. (*cum laude*) degree in computer science and electrical engineering, and the M.Sc. (*magna cum laude*) degree in electrical systems engineering from Tel Aviv University, Tel Aviv, Israel, in 2001 and 2004, respectively. His thesis dealt with the cocktail party problem and spectral eigen value decomposition. His fields of research in the industry for the past decade have been around communication systems, equalization, and tracking. In the last four years, he has specialized in machine learning on social and big data, NLP, computer vision, and deep learning.

Roee Enbar received the B.Sc. (*cum laude*) degree in physics and the B.Sc. (*magna cum laude*) degree in electrical engineering from the Ben-Gurion University of the Negev, Beersheba, Israel, in 2012. His research emphasis was signal processing and hyperspectral imagery. He has been with the industry as a practitioner of image and video processing and data analysis. More specifically, online automatic intruder detection in rough visual conditions, and face detection and recognition.

Tal Hassner received the B.A. degree in computer science from the Academic College of Tel Aviv Yaffo, Qiryat Ono, Israel, in 1998, and the M.Sc. and Ph.D. degrees in applied mathematics and computer science from the Weizmann Institute of Science, Rehovot, Israel, in 2002 and 2006, respectively, where he also completed a post-doctoral fellowship. In 2006, he joined the faculty of the Department of Mathematics and Computer Science at the Open University of Israel, Ra'anana, Israel, where he currently holds a Senior Lecturer position (Assistant Professor). He was a recipient of the Best Student Paper Award at the IEEE Shape Modeling International Conference in 2005 and the Best Scoring Method in the LFW Face Recognition Challenge Award at the Faces in Real-Life Images workshop, European Conference on Computer Vision, in 2008. His research interests are in applications of machine learning in computer vision. Recently, his work focused on problems related to face recognition.