ChaLearn Looking at People and Faces of the World: Face Analysis Workshop and Challenge 2016

Sergio Escalera
Universitat de Barcelona

Mercedes Torres Torres
University of Nottingham

Brais Martínez
University of Nottingham

Xavier Baró
Universitat Oberta de Catalunya

Hugo Jair Escalante
INAOE, Mexico

Isabelle Guyon
U. Paris-Saclay / ChaLearn

Georgios Tzimiropoulos
University of Nottingham

Ciprian Corneanu
Universitat de Barcelona

Marc Oliu
Universitat de Barcelona

Mohammad Ali Bagheri
Dalhousie University/ University of Larestan

Michel Valstar
University of Nottingham

Abstract

We present the 2016 ChaLearn Looking at People and Faces of the World Challenge and Workshop, which ran three competitions on the common theme of face analysis from still images. The first one, Looking at People, addressed age estimation, while the second and third competitions, Faces of the World, addressed accessory classification and smile and gender classification, respectively. We present two crowd-sourcing methodologies used to collect manual annotations. A custom-build application was used to collect and label data about the apparent age of people (as opposed to the real age). For the Faces of the World data, the citizen-science Zooniverse platform was used. This paper summarizes the three challenges and the data used, as well as the results achieved by the participants of the competitions. Details of the ChaLearn LAP FotW competitions can be found at `http://gesture.chalearn.org`.

1. Introduction

Automatic face analysis is a research topic that is currently receiving much attention from the Computer Vision and Pattern Recognition communities. Under certain conditions, face recognition, face detection, and facial expression recognition are now considered solved. However, there remain several aspects of face analysis that are still to be considered open problems, including analysis of low resolution images, with poor illumination, large (motion) blur, significant occlusions, and large non-frontal orientation with respect to the camera. These problems can all be attributed to the challenges faced with capturing faces 'in the wild', i.e. under real-world conditions for unconstrained tasks. The applications of face analysis are countless, including such diverse areas as security and video surveillance, human computer/robot interaction, communication, entertainment, and commerce, while having an important social impact in assistive technologies for education and health. However, given the current challenges and the trend of data-driven solutions such as Deep Learning, a major obstacle to advancing the state of the art in this area is a lack of large annotated datasets of *in the wild* data. The 2016 Chalearn Looking at People and Faces of the world challenge and workshop aim to address this issue by organising three competitions on two novel large-scale datasets.

Following previous series on Looking at People (LAP) competitions [14, 13, 9, 12, 4, 10, 42, 15], in 2016, ChaLearn and Faces of the World organized new competitions and workshops on automatic face analysis in still images. Computational methods for face analysis are genuinely important in many applications in their own right, but also often serve as excellent benchmarks for general computer vision and machine learning algorithms. Furthermore,

978-1-5090-1437-8/16 $31.00 © 2016 IEEE
DOI 10.1109/CVPRW.2016.93

facial expressions analysis, age and gender estimation are hot topics in the field of Looking at People that serve as additional cues to determine human behaviour, while accessories worn by people can serve as important cues to context for this analysis. This motivated our choice to organize this new workshop, with a set of three competitions on these topics – age, accessories, smile and gender – to focus the effort of the computer vision community.

These new competitions come as a natural evolution from previous workshops at CVPR2011, CVPR2012, ICPR2012, ICMI2013, ECCV2014, CVPR2015, and ICCV2015. To ensure continuity, we again used our website `http://gesture.chalearn.org` for promotion and challenge entries in the quantitative competition were scored on-line using the Codalab Microsoft-Stanford University platforms `http://codalab.org`.

In the remainder of this paper, we describe in more detail the three face analysis challenges, their relevance in the context of the state of the art, and describe the results achieved by the winners of the challenges.

2. Age estimation challenge

Age estimation is a difficult task which requires the automatic detection and interpretation of facial features. We have designed an application using the Facebook API for the collaborative harvesting and labelling by the community in a gamified fashion.

Age estimation has historically been one of the most challenging problems within the field of facial analysis [37, 17, 22]. It can be very useful for several applications, such as advanced video surveillance, demographic statistics collection, business intelligence and customer profiling, and search optimization in large databases of facial images. Different application scenarios can benefit from learning systems that predict the apparent age, such as medical diagnosis (premature ageing due to environment, sickness, depression, stress, fatigue, etc.), effect of anti-ageing treatment (hormone replacement therapy, topical treatments), or effect of cosmetics, haircuts, accessories and plastic surgery, just to mention a few. Some of the reasons age estimation is still a challenging problem are the uncontrollable nature of the ageing process, the strong specificity to the personal traits of each individual, high variance of observations within the same age range, and the fact that it is very hard to gather complete and sufficient data to train accurate models.

2.1. ChaLearn-AgeGuess Dataset

Due to the nature of the age estimation problem, there is a limited number of publicly available databases providing a substantial number of face images labelled with accurate age information. Table 1 shows a summary of existing databases with their main reference, number of samples, number of subjects, age range, type of age and additional information. This field has experienced a renewed interest from 2006 on, since the availability of large databases such as MORPH-Album 2 [38], which increased by 55$\times$ the amount of real age-annotated data with respect to traditional age databases. Therefore, this database has extensively been used in recent works by applying to it different descriptors and classification schemes. However, all existing datasets are based on real age estimation. In the present challenge, we are running a second round of our previous ICCV 2015 apparent age estimation challenge. For the first stage of the challenge we collected around 2,500 labelled images and got a total of 20,000 votes by many users. For CVPR 2016, we increased the amount of data by adding data from our AgeGuess application. Joining both databases we gathered a total of 4,691 images and nearly 145,000 votes. The challenge received a record participation of about 100 registered teams. One can notice that for this second round of the challenge we augmented the initial data set with 3000 images, allowing participants to develop better feature-learning methods, and at the same time allowing us to perform additional analysis on the results.

Since recognizing the apparent age of a person is highly subjective, for collecting the data we relied on the opinions of many subjects using a new crowd-sourcing data collection and labelling application and the data from AgeGuess platform[1]. We developed a web application in order to collect and label an age estimation dataset online by the community. The application uses Facebook's API to facilitate the access and hence reaching more people with a broader background. It also allows us to easily collect data from the participants, such as gender, nationality and age. We show some panels of the application in Figures 1(a), 1(b) and 1(c).

The web application was developed in a gamified way, i.e. the users or players get points for uploading and labelling images, the closer the age guess was to the apparent age (average labelled age) the more points the player obtained. In order to increase the engagement of the players, we added a global and friends leader board where the users could see their position in the ranking. We asked users to upload images of a single person and we gave them tools to crop the image if necessary. We also asked users to provide the real age for images they uploaded themselves (or a good approximation), allowing more analysis and comparisons between real age and apparent age. Table 2 shows the main characteristics of the previous dataset and the extended one.

Some of the features of the database and its associated challenge are:

- Thousands of faces labelled by many users.
- Images with diverse backgrounds.
- Images taken in non-controlled environments.

[1] AgeGuess: `http://www.ageguess.org/`

Table 1. Age-based Databases and their characteristics.

Database	#Faces	#Subj.	Range	Type of age	Controlled Environment	Balanced age Distribution	Other annotation
FG-NET [26, 25]	1,002	82	0 - 69	Real Age	No	No	68 Facial Landmarks
GROUPS [19]	28,231	28,231	0 - 66+	Age group	No	No	-
PAL [31]	580	580	19 - 93	Age group	No	No	-
FRGC [36]	44,278	568	18 - 70	Real Age	Partially	No	-
MORPH2 [39]	55,134	13,618	16 - 77	Real Age	Yes	No	-
YGA [18]	8,000	1,600	0 - 93	Real Age	No	No	-
FERET[35]	14,126	1,199	-	Real Age	Partially	No	-
Iranian face [5]	3,600	616	2 - 85	Real Age	No	No	Kind of skin and cosmetic points
PIE [43]	41,638	68	-	Real Age	Yes	No	-
WIT-BD [45]	26,222	5,500	3 - 85	Age group	No	No	-
Caucasian Face Database [7]	147	-	20 - 62	Real Age	Yes	No	208 Shape Landmarks
LHI [2]	8,000	8,000	9 - 89	Real Age	Yes	Yes	-
HOIP [16]	306,600	300	15 - 64	Age Group	Yes	No	-
Ni's Web-Collected Database [32]	219,892	-	1 - 80	Real Age	No	No	-
OUI-Adience [8]	26.580	2.284	0 - 60+	Age Group	No	No	Gender
IMDB-WIKI [40]	523.051	20.284	0 - 100	Real Age	No	No	Gender

Table 2. ChaLearn-AgeGuess database characteristics.

Features		ChaLearn	AgeGuess	Total
Images		1506	3185	4691
Users	female	44	1828	1872
	male	110	1143	1253
Votes	female	1753	75136	76889
	male	14897	53117	68004

• Non-labelled face bounding boxes, neither face landmarks, making the estimation problem even harder.

• One of the first datasets in the literature including estimated age as labelled by many users to define the ground truth, with the explicit objective of estimating the age.

• The evaluation metric is weighted by the mean and variance of the labelling by the participants.

• The dataset contains the real age for each image, although this is not used for recognition but only for data analysis purposes. In the same way for all the labellers (the users of the platforms which make estimates of the age of the person in the photo).

We have their nationality, age, and gender, which will allow us to analyse potential correlation between our findings and demographics.

Properties of existing datasets are shown in Table 1.

2.2. Apparent age challenge results

Nearly 100 participants registered for the competition. Finally, at the test step, the teams that submitted their predictions are shown in Table 4. Each prediction is evaluated as $\epsilon = 1 - e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, where x is the prediction, μ and σ are the mean and standard deviation of the human labels. The summary of the methods is shown in Table 3. The summary of the first top ranked methods is shown next.

First place (OrangeLabs).

The winners of the apparent age competition used the VGG-16 Convolutional Neural Network (CNN) trained for face recognition on 2 million faces [34] as the starting point. They fine-tune this CNN on the IMDB-Wiki dataset [40] for the age estimation task using the label distribution encoding [20]. Then authors fine-tune 2 separate CNNs from the obtained CNN on the competition data: the first type CNN is used for age estimation of all ages (this model is trained using label distribution encoding) and the second type of CNN is used for age estimation of children between

Table 3. Summary of the results of top-ranked competitors in the challenge

Rank	Team	Test Error	General Idea	Pre-trained models	Preprocessing	Fusion	Additional data used
1	OrangeLabs	0.2411	Two-phase learning by an ensemble of several CNN models	VGG-16 [33]	face detection / pose estimation / alignment	Score-level fusion	IMDB Wiki Children images from web
2	palm_seu	0.3214	An ensemble of four fine-tuned CNN models	VGG-16 [33]	face detection [44]/ pose estimation [44] / alignment	Score-level fusion	IMDB-WIKI
3	cmp+ETH	0.3361	Ensemble of 8 SO-SVM classifiers learned on the features from the last layer of VGG-16 network	VGG-16 [33]	face detection [29]	Score-level fusion	IMDB-WIKI
4	WYU_CVL	0.3405	Multiple models using grouped deep age networks and random forest regressor	None	face detection / image augmentation	Score-level fusion	WebFace Morph, CACD, FG-Net
5	ITU_SiMiT	0.3668	An ensemble of 3 CNN models originated from VGG-16 and fine-tuned on the challenge data	VGG-16 [33]	face detection / face cropping	Score-level fusion	IMDB-WIKI
6	Bogazici	0.374	A two part model: classification into overlapping age groups and regression among each group	VGG-16 [33]	face detection, intensity averaging	Score-level fusion	None
7	MIPAL_SNU	0.4569	An ensemble of 3 CNNs with different loss functions	ImageNet [24], Pre-trained CNN models with ILSVRC and CACD data	face detection / face cropping	Score-level fusion	ILSVRC 2015, CACD
8	DeepAge	0.4573	Deep Label Distribution Learning (DLDL) framework	VGG-16 [33]	face detection / face cropping	None	1) 53,969 web face images 2) LAP Age Estimation 2015

Table 4. Participating teams in the apparent age-estimation competition, and their performance measured in terms of a Gaussian defined by the mean/standard deviation of the manual annotators.

Position	Team	Test Error
1	OrangeLabs	0.2411
2	palm_seu	0.3214
3	cmp+ETH	0.3361
4	WYU_CVL	0.3405
5	ITU_SiMiT	0.3668
6	Bogazici	0.3740
7	MIPAL_SNU	0.4569
8	DeepAge	0.4573

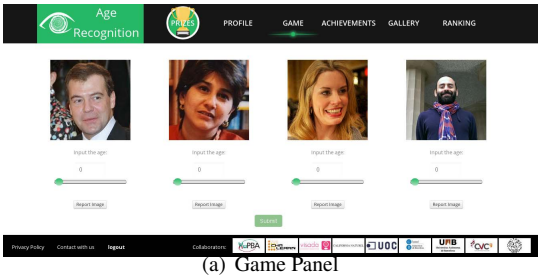
0 and 12 years old (this model is trained using a simple 0/1 classification encoding). During the test phase, if the first type CNN output is above 12, the output is kept as the final one. In the opposite case the output of the second type CNN is kept as the final one. In both cases, the final output is calculated as an expected value of normalized neurons in the output layer. The final age estimation system consists of 11 type one CNN, which are trained by 11-fold cross-validation on the training and validation datasets and 3 second type CNN which are trained using all available children images in training and validation datasets. For both types of CNN, the combination of age estimations is performed on the level of output neurons.

Second place (palm_seu).

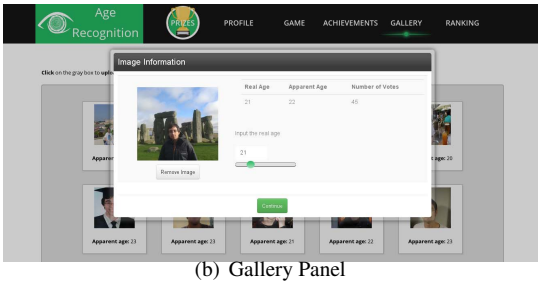
The second ranked team of the apparent age estimation challenge preprocessed the images of the IMDB-WIKI [40] and the competition, including face detection [44], key point detection [44] and face alignment. They divided IMDB-WIKI data set into three parts randomly (each part has about 80,000 pictures) and generated the age distribution of each picture by using a normal distribution function (the assigned age is the mean value and the standard deviation is assumed to be 3). Then they fine-tuned three new deep net models on each IMDB-WIKI data set by using a pretrained vggface model (VGG-16) [34]. The loss function of the net models the is KL-divergence which can be used to measure the difference between the ground truth distribution and the predicted age distribution. This method can be called Deep Label Distribution Learning [46]. Next, the authors fine-tuned these models on the competition data set (including previous [11] and current competition data) by using the same loss function. The authors also fine-tuned a pre-trained vggface model (VGG-16) directly on 2015 and 2016 competition data without pre-training on IMDB-WIKI data set. Therefore there are four nets in total. The final models were employed to extract the last full connected features of four net models, which were used by an ensemble method to generate the final result [46].

Third place (cmp+ETH).

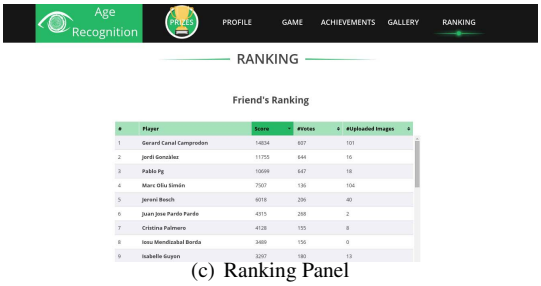
The third ranked participant of the apparent age recognition challenge used Mathias et al. [29] face detector and DEX model [40] for feature extraction. DEX model was learned on the IMDB-WIKI dataset by [40] and fine tuned on the LAP 2015 data. The method does not use any additional data except for the training and validation set of LAP.



(a) Game Panel



(b) Gallery Panel



(c) Ranking Panel

Figure 1. Age Recognition Application. (a) User can see the images of the rest of participants and vote for the apparent age. (b) User can upload images and see their uploads and the opinion of the users regarding the apparent age of people in their photos. (c) User can see the points he/she achieves by uploading and voting photos and the ranking among his/her friends and all the participants of the application.

3. Accessories classification challenge

3.1. Faces of the World Dataset – Accessories

Many works in mid-level face analysis, such as face recognition or facial expression analysis, consider the lower-level face analysis components to be solved. Indeed, problems such as face detection and facial point localisation have seen remarkable progress over the last five years [41]. However, the real reason why people working in mid- or high-level face analysis have come to believe this is because they consistently work with over-simplified datasets, which are often designed to focus on mid-level problems (e.g. expression analysis) and to avoid low-level difficulties. This is not to say these are not realistic scenarios. For example, the SEMAINE database [30] contains recordings of realistic interactions between two people who communicate through a set of teleprompters in an office setting, which has constant and abundant illumination. Participants naturally face the camera while discussing a topic of their choosing. Similarly, the CelebA dataset [27] contains over 200K photographs of celebrities, annotated with 40 attributes. However, CelebA mainly contains frontal face images of celebrities with posed expression (e.g. smile).

The majority of the most popular face databases [41, 38, 35, 27] feature only mild non-frontal views with decent and constant illumination and little to no occlusion. Additionally, very often, these photographs contain only one face to be analysed, which greatly simplifies the analysis.

Another issue with these dataset involves the diversity of the source material. A prime example of this is the MORPH database [38], introduced in Section 2. This dataset, widely used in the fields of face recognition and age estimation [21] is extremely skewed. It contains 156,313 images of men and women from five different ethnic groups (Black, White, Asian, Hispanic and Other). However, out of the total, 115K images are Black subjects, but only 740 Asian subjects. Furthermore, there are only 26,171 images from women, versus the 130,142 images collected from men.

As a consequence, we postulate that analysis made on such skewed datasets cannot be representative enough. The Faces of the World dataset has been created with the aim of overcoming these issues. We have collected over 25,000 publicly-available images from the Internet, with special emphasis in collecting diverse and balanced data. FotW comprises 25% Asian, Black, Hispanic and White subjects, 50% of which are men and 50% are women. Additionally, we made sure to include people from all ages, from newborn babies to elderly people. As a consequence, FotW is a very diverse and technically challenging dataset. A full description of the Faces of the World dataset, including all the additional labels, can be found in [28].

The creation process of the Faces of the World (FotW) dataset, from collection to annotation and post-processing, can be summarised as follows:

1. **Selection:** Over 25,000 images were downloaded from Flickr using Extreme Picture Finder [1]. All of these images are licensed under Creative Commons 4.0, which allows users to use copy and redistribute the material in any medium or format and to remix, transform, and build upon the material. A random selection of the dataset can be seen in Figure 2.

2. **Annotation:** We used Zooniverse [6], a citizen-science platform, to annotate the images collected in the previous step. We created the Faces of the World project within Zooniverse [3], which was organised into three different workflows: Face and gender detection, Accessory classification and Smile classification. Over 500 volunteers participated, providing over 28,000 annotations.

Figure 2. Random selection of images from the Faces of the World dataset

3. **Post-processing:** After the annotation process had finished, faces were extracted from the images using the bounding boxes provided by the Face and Gender Detection workflow. The training (5,651 faces), validation (2,826 faces) and testing (4,086 faces) set of the Faces of the World (FotW) dataset were created. Table 5 shows the annotations corresponding to the Accessories classifications.

Table 5. FotW - Accessories database characteristics.

Accessory	Train	Val	Test
Hat	1151	608	869
Headband	243	109	193
Glasses	1232	614	828
Earrings	770	389	592
Necklace	615	300	559
Tie	151	72	220
Scarf	256	137	256

3.2. Accessories challenge results

Nearly 50 participants registered for the competition. The teams that submitted their prediction are shown in Table 6. Performance was evaluated using the Mean Square Error between participant's predictions and the ground-truth collected through the Zooniverse platform. Table 6 shows the accuracy of each of the participating teams.

Table 6. ChaLearn - Accuracy in accessory classification.

Accessory	SIAT_MMLAB	IVA_NLPR
Hat	0.9469	0.9222
Headband	0.9489	0.9506
Glasses	0.9467	0.9386
Earrings	0.9103	0.8535
Necklace	0.8821	0.8736
Tie	0.9726	0.9606
Scarf	0.9367	0.9401
Average	**0.9349**	0.9199

First place (SIAT_MMLAB).

The winner of the Accessories Classification Challenge used Cascaded Convolutional Neural Networks (CNNs), implemented in Caffe [23]. This method used the celebA dataset [27] as additional training data. Three levels of cascade were trained: first, they fine-tuned pre-trained VGG-Faces [33] with CelebA (with their original 40 attributes), then they fine-tuned the first model with CelebA and only 5 attributes (earrings, hat, glasses, necklace and necktie). Finally, they find-tuned the previous model with the FotW dataset and 7 attributes (earrings, hat, glasses, necklace, necktie, headband and scarf).

4. Smile and Gender classification challenge

4.1. Faces of the World Dataset – Smiles

A different subset of images from the FotW dataset were annotated for gender and expression. This subset is composed of training set (6,171 images), validation set (3,086 images) and test set (8,505 images). The characteristics of this part of the dataset are shown in Table 7.

Table 7. FotW - Smile and Gender database characteristics.

Attribute	Train	Val	Test
Male	2946	1691	4614
Female	3318	1361	3799
Not sure	93	34	92
Smile	2234	1969	4411
No smile	3937	1117	3849

4.2. Smile and Gender challenge results

Nearly 60 participants registered for this competition. The 6 teams that finally submitted predictions are listed in Table 8 together wit the accuracy of each of the participating teams.

Table 8. ChaLearn- Accuracy in Age and Gender classification

Team	Gender	Smile	Mean
SIAT_MMLAB	0.9269	0.8583	0.8926
IVA_NLPR	0.9152	0.8252	0.8702
VISI.CRIM	0.9016	0.8212	0.8614
SMILELAB NEU	0.8999	0.8148	0.8574
Lets Face it!	0.8454	0.8439	0.8446
CMP+ETH	0.7465	0.7189	0.7327
SRI	0.5716	0.5853	0.5784

First place (SIAT_MMLAB) The winner of the Smile and Gender Classification Challenge is again SIAT_MMLAB, again using Cascaded Convolutional Neural Networks (CNNs), implemented in Caffe [23], but the architecture is different from that used in the accessories challenge. Their method again used the celebA [27] as additional training data. Faces are detected and aligned, and then a single cropped face is input to a CNN pipeline for gender detection, and two slightly different face patches are input to the CNN pipeline for smile detection.

The architecture starts with fine-tuning the pre-trained VGG-faces model using CelebA with all 40 attributes. It then creates three separately fine-tuned models, each fine-tuned with only the smile or gender attributes. For smiles two different models with slightly different face patch sizes are created. Finally, the three resulting models are fine-tuned with the Faces of the World data.

5. Conclusion

We presented the 2016 ChaLearn Looking at People and Faces of the World competitions on automatic face analysis. Three competitions were formulated, in apparent age estimation, accessories classification, and gender and smile prediction. Two new datasets were presented to support the three competitions; *ChaLearn-AgeGuess* and *Faces of the World*, totalling almost 30,000 labelled images. Perhaps unsurprisingly, all three competitions were won by teams employing Convolutional Neural Networks.

Acknowledgements

We thank our sponsors: Microsoft Research, University of Barcelona, Amazon, INAOE, Google, NVIDIA corporation, Facebook, and Disney Research. This research has been partially supported by the Spanish projects TIN2013-43478-P, TIN2015-66951-C2-2-R, and European Union Horizon 2020 research and innovation programme under grant agreement No 645378.

References

[1] Extreme Internet Software extreme picture finder. `http://www.exisoftware.com/picture_finder/`. Accessed: 2016-04-15.

[2] LHI image database. Available at `http://www.lotushill.org/LHIFrameEn.html`, 2010.

[3] Zooniverse faces of the world. `https://www.zooniverse.org/projects/pszmt1/faces-of-the-world/`. Accessed: 2016-04-15.

[4] X. Baro, J. Gonzàlez, J. Fabian, M. A. Bautista, M. Oliu, H. J. Escalante, I. Guyon, and S. Escalera. Chalearn looking at people 2015 challenges: action spotting and cultural event recognition. In *ChaLearn LAP Workshop, CVPR*, 2015.

[5] A. Bastanfard, M. Nik, and M. Dehshibi. Iranian face database with age, pose and expression. In *Int. Conf. Machine Vision, 2007*, pages 50–55, Dec 2007.

[6] K. Borne and Z. Team. The zooniverse: A framework for knowledge discovery from citizen science data. In *AGU Fall Meeting Abstracts*, volume 1, page 0650, 2011.

[7] D. M. Burt and D. I. Perrett. Perception of age in adult caucasian male faces: Computer graphic manipulation of shape and colour information. *Royal Society of London. Series B: Biological Sciences*, 259(1355):137–143, 1995.

[8] E. Eidinger, R. Enbar, and T. Hassner. Age and gender estimation of unfiltered faces. *Information Forensics and Security, IEEE Transactions on*, 9(12):2170–2179, Dec 2014.

[9] S. Escalera, X. Baro, J. Gonzàlez, M. Bautista, M. Madadi, M. Reyes, V. Ponce, H. Escalante, J. Shotton, and I. Guyon. Chalearn looking at people challenge 2014: Dataset and results. *ChaLearn LAP Workshop, ECCV*, 2014.

[10] S. Escalera, J. Fabian, P. Pardo, , X. Baró, J. Gonzàlez, H. J. Escalante, D. Misevic, U. Steiner, and I. Guyon. Chalearn looking at people 2015: Apparent age and cultural event recognition datasets and results. In *International Conference in Computer Vision, Looking at People, ICCVW*, 2015.

[11] S. Escalera, J. Fabian, P. Pardo, X. Baro, J. Gonzalez, H. Escalante, and I. Guyon. Chalearn 2015 apparent age and cultural event recognition: datasets and results. In *ICCV, ChaLearn Looking at People workshop*, volume 3, page 4, 2015.

[12] S. Escalera, J. Gonzàlez, X. Baro, P. Pardo, J. Fabian, M. Oliu, H. J. Escalante, I. Huerta, and I. Guyon. Chalearn looking at people 2015 new competitions: Age estimation and cultural event recognition. In *IJCNN*, 2015.

[13] S. Escalera, J. Gonzàlez, X. Baro, M. Reyes, I. Guyon, V. Athitsos, H. Escalante, A. Argyros, C. Sminchisescu, R. Bowden, and S. Sclarof. Chalearn multi-modal gesture recognition 2013: grand challenge and workshop summary. *ICMI*, pages 365–368, 2013.

[14] S. Escalera, J. Gonzàlez, X. Baró, M. Reyes, O. Lopés, I. Guyon, V. Athitsos, and H. J. Escalante. Multi-modal

gesture recognition challenge 2013: Dataset and results. In *ChaLearn Multi-Modal Gesture Recognition, ICMI Workshop*, 2013.

[15] S. Escalera, J. Gonzàlez, X. Baró, and J. Shotton. Special issue on multimodal human pose recovery and behavior analysis. *IEEE Tans. Pattern Analysis and Machine Intelligence*, 2016.

[16] S. J. Foundation. Human and Object Interaction Processing (HOIP) Face Database. Available at `http://www.hoip.jp/`, 2014.

[17] Y. Fu, G. Guo, and T. Huang. Age synthesis and estimation via faces: A survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(11):1955–1976, Nov 2010.

[18] Y. Fu and T. Huang. Human age estimation with regression on discriminative aging manifold. *Multimedia, IEEE Transactions on*, 10(4):578–584, June 2008.

[19] A. Gallagher and T. Chen. Understanding images of groups of people. In *Proc. CVPR*, 2009.

[20] X. Geng, C. Yin, and Z.-H. Zhou. Facial age estimation by learning from label distributions. In *TPAMI*, volume 35, pages 2401–2412. IEEE, 2013.

[21] X. Geng, Z.-H. Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *TPAMI*, 29(12):2234–2240, 2007.

[22] H. Han, C. Otto, and A. K. Jain. Age estimation from face images: Human vs. machine performance. In *ICB'13*, pages 1–8, 2013.

[23] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *Proceedings of the ACM International Conference on Multimedia*, pages 675–678. ACM, 2014.

[24] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C. Burges, L. Bottou, and K. Weinberger, editors, *Advances in Neural Information Processing Systems 25*, pages 1097–1105. Curran Associates, Inc., 2012.

[25] A. Lanitis. FG-NET Aging Data Base, November 2002.

[26] A. Lanitis, C. Taylor, and T. Cootes. Toward automatic simulation of aging effects on face images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(4):442–455, 2002.

[27] Z. Liu, P. Luo, X. Wang, and X. Tang. Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, pages 3676–3684, 2015.

[28] B. Martinez, M. Torres Torres, A. Jackson, A. Bulat, T. Almaev, G. Tzimiropoulos, and M. Valstar. The first faces-of-the-world dataset and challenge: a benchmark for end-to-end unconstrained face analysis.

[29] M. Mathias, R. Benenson, M. Pedersoli, and L. Van Gool. Face detection without bells and whistles. In *Computer Vision–ECCV 2014*, pages 720–735. Springer, 2014.

[30] G. McKeown, M. Valstar, R. Cowie, M. Pantic, and M. Schröder. The semaine database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. *Affective Computing, IEEE Transactions on*, 3(1):5–17, 2012.

[31] M. Minear and D. C. Park. A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers*, 36(4):630–633, 2004.

[32] B. Ni, Z. Song, and S. Yan. Web image mining towards universal age estimator. In *Proceedings of the 17th ACM International Conference on Multimedia*, MM '09, pages 85–94, New York, NY, USA, 2009. ACM.

[33] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. In *BMVC*, 2015.

[34] O. M. Parkhi, A. Vedaldi, and A. Zisserman. Deep face recognition. *Proceedings of the British Machine Vision*, 1(3):6, 2015.

[35] P. Phillips, H. Wechsler, J. Huang, and P. J. Rauss. The {FERET} database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295 – 306, 1998.

[36] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the Face Recognition Grand Challenge. In *CVPR*, pages 947–954. IEEE, 2005.

[37] N. Ramanathan, R. Chellappa, and S. Biswas. Computational methods for modeling facial aging: A survey. *Journal of Visual Languages and Computing*, 20(3):131 – 144, 2009.

[38] K. Ricanek and T. Tesafaye. MORPH: a longitudinal image database of normal adult age-progression. In *Int. Conf. FG*, pages 341–345, 2006.

[39] K. Ricanek and T. Tesafaye. Morph: a longitudinal image database of normal adult age-progression. In *Int. Conf. FG*, pages 341–345, April 2006.

[40] R. Rothe, R. Timofte, and L. Gool. Dex: Deep expectation of apparent age from a single image. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 10–15, 2015.

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

[42] I. G. Sergio Escalera, Vassilis Athitsos. Challenges in multimodal gesture recognition. *Journal on Machine Learning Research*, 2016.

[43] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression (pie) database. In *Int. Conf. FG*, pages 46–51, May 2002.

[44] Y. Sun, X. Wang, and X. Tang. Deep convolutional network cascade for facial point detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3476–3483, 2013.

[45] K. Ueki, T. Hayashida, and T. Kobayashi. Subspace-based age-group classification using facial images under various lighting conditions. In *Int. Conf. FG*, pages 43–48, 2006.

[46] X. Yang, B.-B. Gao, C. Xing, Z.-W. Huo, X.-S. Wei, Y. Zhou, J. Wu, and X. Geng. Deep label distribution learning for apparent age estimation. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, pages 102–108, 2015.