Convolutional Neural Network for Age Classification from Smart-phone based Ocular Images

3 authors:

Ajita Rattani
Wichita State University

119 PUBLICATIONS   **1,408** CITATIONS

SEE PROFILE

Narsi Reddy Sai Donthi Reddy
University of Missouri - Kansas City

15 PUBLICATIONS   **130** CITATIONS

SEE PROFILE

Reza Derakhshani
University of Missouri - Kansas City

97 PUBLICATIONS   **1,170** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Project   Ocular Biometrics in Visible Spectrum View project

Project   Fairness of Biometrics System View project

Convolutional Neural Network for Age Classification from Smart-phone based Ocular Images

Ajita Rattani, Narsi Reddy and Reza Derakhshani

Dept. of Computer Science and Electrical Engineering

University of Missouri- Kansas City, MO, USA

{rattania,derakhshanir@umkc.edu}

Abstract

Automated age classification has drawn significant interest in numerous applications such as marketing, forensics, human-computer interaction, and age simulation. A number of studies have demonstrated that age can be automatically deduced from face images. However, few studies have explored the possibility of computational estimation of age information from other modalities such as fingerprint or ocular region. The main challenge in age classification is that age progression is person-specific which depends on many factors such as genetics, health conditions, life style, and stress level. In this paper, we investigate age classification from ocular images acquired using smart-phones. Age information, though not unique to the individual, can be combined along with ocular recognition system to improve authentication accuracy or invariance to the ageing effect. To this end, we propose a convolutional neural network (CNN) architecture for the task. We evaluate our proposed CNN model on the ocular crops of the recent large-scale Adience benchmark for gender and age classification captured using smart-phones. The obtained results establish a baseline for deep learning approaches for age classification from ocular images captured by smart-phones.

1. Introduction

Age is an important demographic attribute. Humans have an innate ability to reliably estimate age of their peers based on holistic facial features such as skin texture, wrinkles, skin quality, facial hair and chin line [5]. In the context of biometrics, age classification can be employed as a soft-biometric trait in fusion with primary biometric trait to improve the matching accuracy. Further, as recognition difficulty based on biometrics is linked to aging effect [2, 3]. Advances have been made in the form of age-invariant solutions that seek to learn an aging model for age transformation of the input operational image to that of biometric templates [11, 12, 4]. These solutions are usually integrated with existing biometric engines to obtain invariance to the ageing effect.

A number of studies have been conducted on age classification from face images. The proposed approaches utilize geometric information, appearance models and ageing pattern subspaces [5]. Limited studies have been conducted on age estimation from hand [24], fingerprint [13], body [17], and iris [9, 23, 1]. Despite these studies, the ability to automatically deduce age from a biometric sample is far from an accurate and robust solution. One of the main challenges is that age progression varies among individuals and is influenced by factors such as genetics, health, life style, eating habits and stress level [5].

With integration of biometric technologies in mobile phones, biometrics based access control has become convenient alternate to PINs and passwords [18, 21]. In this context, RGB ocular biometrics comprising of scanning regions in the eye and those around it has gained substantial attention. This is because capturing ocular biometrics tantamount to capturing eye image using the regular RGB cameras already available in all mobile devices [7, 19, 6]. However, due to device mobility and operation in an uncontrolled environment, factors such as specular reflection, motion blur, illumination variations and occlusion are introduced in the acquired ocular images. These factors combined with temporal variations can result in high error rates [20].

The aim of this paper is *age classification from ocular images acquired from smart-phones*. This is the first study of its kind as existing studies on age classification have used iris images captured in NIR spectrum instead of mobile RGB captures [9, 23, 1]. This is because of the unavailability of the publicly available mobile ocular biometric databases with annotated age information. We used representation learning and proposed a convolutional neural network (CNN) architecture for this study. We tested our proposed CNN model on the ocular crops of the recent large-scale Adience benchmark for gender and age classifi-

cation [16], captured using smart-phone devices. The proposed age-classification system can be used to enhance the recognition ability and invariance to aging in smart-phone based ocular recognition. Further, it provides privacy benefits and reduces computational cost over scanning full face images for age classification.

In summary, our contributions are: (1) first study utilizing application of CNN for age classification from ocular images acquired by smart-phones, and (2) assessment of the proposed model on the ocular crops of the face images in Adience benchmark.

The rest of this paper is organized as follows: section 2 discuss the existing methods on age classification from ocular images. Section 3 elaborates on the proposed CNN model. Experimental results are discussed in section 4. Conclusions are drawn in section 5.

2. Prior Work on Age Classification from Ocular Images

Sgroi et al. [23] conducted a study to categorize iris images as representing a young or older person. The normalized irises were used to create 630 features based on nine different filter responses tuned to detect spots, thin or thick horizontal and vertical lines, and texture energy in particular regions of the image defined by neighboring rows and columns. For each filter response, 70 features were generated. RandomForest classifier was used for final classification. NIR iris images of 50 subjects with 6 samples per subject between the age of 22 and 25, selected as the *younger* group, and 50 subjects older than the age of 35, selected as the *older* group, taken from the University of Notre Dame dataset were used for experiments. All images in this dataset were acquired by a LG IrisAccess 4000. Reported experimental results showed an accuracy of 64.68%.

Erbilek et al. [9] proposed age classification from iris images into three groups: young, middle and old. Specifically, geometric features from the detected iris (x-coordinate of the centre of the iris, y-coordinate of the centre of the iris, and iris radius), and pupil (x-coordinate of the centre of the pupil, y-coordinate of the centre of the pupil, and pupil radius) were used along with Support Vector Machines, Multi-layer Perceptron, K-nearest Neighbour and Decision Trees. NIR iris images of 50 subjects with 3 samples per subject below the age of 25 , selected as the *young* group, 82 subjects between the age of 25 and 60, selected as the *middle* group, and 33 subjects older than 60, selected as the *old* group, taken from BioSecure Multimodal Database were used for experiments. Experiments were conducted on iris images from 70 subjects acquired using Iris Access EOU3000 from BioSecure Multimodal Database. Authors reported maximum accuracy of 62.94% using Support Vector Machines. The effect of pupil dilation and contraction on the robust extraction of these features was not addressed by authors.

Abaasi and Khan [1] performed statistical experiments to estimate confidence interval for iris-pupil thickness calculated using pupil-iris radius ratio of three different age-groups: children, youth and senior citizen. Significant group differences were observed by applying statistical techniques such as Analysis of Variance (ANOVA). The results of the study on a total of 180 images from CASIA version 4.0 dataset suggested that the proposed methodology can be employed to determine age-group of a person from iris images.

As can be seen, only few studies have been reported on age classification from ocular images. All these studies used geometric or textural information extracted from NIR iris for age classification. Further, experimental investigations were performed on very small sized datasets and the reported accuracies are low, about 63.7%.

3. CNN for Age Classification from Ocular Images

A Convolution Neural Network (CNN) is a type of feed-forward artificial neural network in which the connectivity pattern between its neurons, that have learnable weights and biases, is inspired by the organization of the visual cortex[1]. Each neuron responds to its input by performing a dot product, optionally followed by a non-linearity. The efficacy of CNNs has been very successfully demonstrated for large-scale image recognition [22], pose estimation [26], face recognition [25] to name a few. To our knowledge, this is the first application of CNN for the task of age classification from ocular images captured by smart-phone devices.

An important consideration is over-fitting when using deep CNN due to their large number of model parameters. This could be an issue as datasets for age classification from social media are relatively limited in size. This is because manual labeling of images may be tedious or time consuming process and there may be privacy concerns involved as well in data gathering process from social media [8, 16].

3.1. Network Architecture

Figure 1 shows our proposed six layer CNN model for age classification from ocular images. The network comprises of three convolutional layers and two fully connected layers with small number of neurons. The choice of the simpler model is to avoid overfitting especially for case of small training datasets.

Ocular crops are first converted to grayscale and resized to 32×92 pixels. These resized ocular crops are fed to the network. The first three 2D convolutional layers are then defined as follows:

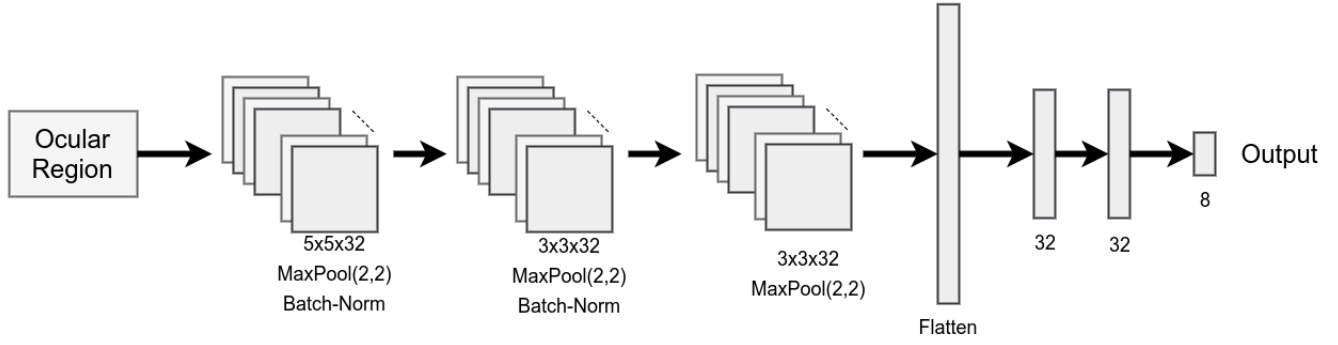[1] http://deeplearning.net/tutorial/lenet.html

Figure 1. The proposed CNN architecture for age classification from ocular images.

Table 1. Detailed architecture of the proposed CNN for age classification from ocular images.

Layer	output shape	# parameters
Input	32x96x1	-
Conv2D	28x92x32	832
Maxpooling	14x46x32	-
BatchNorm	14x46x32	128
Conv2D	12x44x32	9248
Maxpooling	6x22x32	-
BatchNorm	6x22x32	128
Conv2D	4x20x32	9,248
Maxpooling	2x10x32	-
Flatten	640	-
Fully Connected	32	20,512
Fully Connected	32	1,056
Ouput	8	256
Softmax	8	-
Total # Params:		**41,416**

1. First convolutional layer is constructed with 32 filters of size 5×5 to extract features from the input image. The convolutional layer is followed by rectified linear unit (ReLU) with 2×2 stride max pooling.

2. In the second layer, $32 \times 14 \times 46$ features from the previous layer are input to the convolutional layer with 32 3×3 filters. Similar to the first layer, the features from the convolutional layer go through ReLU followed by max-pooling.

3. In the third and the final layer, another 3×3 kernel, 32 filter convolutional layer is fed to the ReLU activation with 2×2 stride max pooling layer.

To lighten the training process and increase the model performance, batch-normalization is adopted after the first and second convolutional layers. The output of the third convolutional layer of shape $32 \times 6 \times 22$ is flattened and provided as input to the following fully connected layers for dense mapping and predicting the age:

1. First and second fully connected layers contain 32 neurons with ReLU as the activation layer.

2. In the final layer, the output from the previous layer of size 32 features are dense mapped to 8 neurons for age-group classification. A soft-max activation layer is used to obtain the probability of the age-group.

The detailed architecture of the CNN along with the number of parameters involved in each layer is given in Table 1. The total number of learnable parameters of the network is $41,416$.

3.2. Network Training and Testing

The proposed model was trained on input data batch size of 64 using Adam optimizer[15] with initial learning rate of 0.001. The weights were initialized using gaussian distribution with variance scaling to the size of the weight matrix. The best weights were chosen using 5-fold cross-validation.

4. Experiments

Our CNN model was implemented using Keras[2], a deep learning library, and Tensorflow 1.1[3], an open source software library for deep learning, as the backend. Training was performed on an Intel $i7$ 7700 3.60GHz processor and Nvidia GTX 1070 GPU.

4.1. The Adience Dataset

We tested the accuracy of our CNN model using the recently released Adience benchmark [16], designed for age and gender classification for face images. The Adience set consist of images automatically uploaded to Flickr

---

Table 2. Breakdown of the ocular crops into eight age-groups.

#	Age-group	# of Images
1.	0-2	1739
2.	4-6	1549
3.	8-13	1631
4.	15-20	1078
5.	25-32	3375
6.	38-43	1869
7.	48-53	597
8.	60-above	622

Table 3. Exact accuracy and 1-off accuracy of proposed CNN model for 5-fold cross-validation and overall performance as mean$\pm$ standard deviation.

Cross Validation	Exact [%]	1-OFF [%]
1	43.70	80.35
2	43.87	79.28
3	49.77	82.01
4	46.77	80.86
5	50.75	82.27
Overall Ocular Crops	**46.97 $\pm$ 2.9**	**80.96 $\pm$ 1.09**
Full Faces in [16]	**49.5 $\pm$ 4.4**	**84.6 $\pm$ 1.7**

from smart-phone devices. As these images have been uploaded without prior manual filtering, they are highly unconstrained, reflecting many of the real-world challenges like extreme variations in head pose, lighting conditions and quality.

For this study, we used $12,460$ in-plane aligned face images with frontal pose from all the 8 age-groups. This is to represent mobile ocular biometric selfie-like use case. Table 2 lists the breakdown of these images into different age categories. To extract the ocular ROI, we used Dlib [14] landmark localization to generate 68 landmarks on the face image. Using these landmark positions, ocular crops comprising of both the eyes, eyebrows and periocular region were extracted from these pre-aligned face images. Figure 2 shows sample ocular regions cropped from the frontal face images of the Adience dataset.

Further, as many of these images were taken in the unconstrained environment with presence of shadows and in poor lighting condition, they were preprocessed using Contrast Limited Adaptive Histogram Equalization (CLAHE) [10] and re-sized to 32×92 pixels.

4.2. Protocol and Results

The performance of the model was tested using exact and 1-off accuracy on 5-fold cross-validation. For each fold, network training and validation was performed on 80% of the images and remaining 20% images were used for testing, with samples from each age-group stratified. Exact accuracy is when the trained model detects the exact age-group of the input image. 1-off accuracy is when the model is off by one adjacent age-group i.e., the subject belongs to the group immediately older or immediately younger than the predicted group. These performance metrics follow the existing work on age classification from face images using Adience dataset [16]. This reported work used CNN for age classification but from full face images using Adience benchmark. The reported exact and 1-off accuracies in [16] are 49.5 ± 4.4 and 84.6 ± 1.7, respectively. These low accuracies for full face images reflect the difficulty involved with the uncontrolled dataset and the challenging nature of the task.

Table 3 present our results on age classification for ocu-

lar region. It can be seen that the proposed model was able to achieve overall exact accuracy of $46.97\% \pm 2.9$ with 1-off accuracy of $80.96\% \pm 1.09$ on test set. Figure 3 show examples of ocular test images that were correctly and incorrectly classified into different age-groups using the proposed CNN. Many of the incorrect classifications were due to degraded images exhibiting extreme conditions such as motion blur and low resolution.

In contrary to existing studies where iris images were categorized into maximum of three age-groups (young, middle and old) [9], we performed age classification into eight groups. Further, the reported accuracies of about 63% for NIR iris are not directly comparable to our results on smart-phone based RGB ocular images. This is because NIR iris images are usually of high quality captured in controlled environment whereas, smart-phone uncontrolled acquisitions are usually of degraded quality. However, our results are quite comparable to the aforementioned age classification study for full face images on Adience dataset [16]. This suggest viability of ocular region in age classification over full face images (see Table 3).

5. Conclusion and Future Work

Limited number of studies have been conducted on age classification from ocular images. Most of these studies utilized texture representations or geometric information for categorization of NIR iris images into young, middle and old age-groups. Despite being evaluated on small datasets captured in controlled conditions, reported accuracies of NIR iris are as low as 63%. Using the recently released Adience benchmark for age and gender classification, in this study we used a CNN for age classification from smartphone based ocular images into eight age-groups. Our proposed CNN is a small network in order to avoid overtraining using the limited training set. Being the first study of its kind, obtained results establish a baseline for deep learning approaches for age classification from ocular images captured by smart-phones.

Our obtained reports are quite comparable to those obtained for full face images using Adience benchmark. This
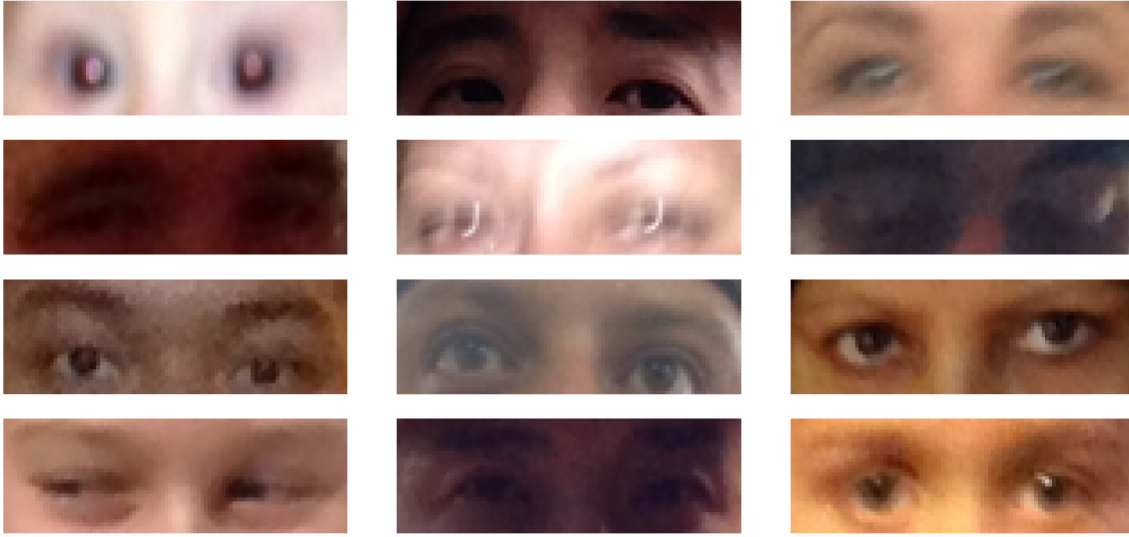
4

Figure 2. Sample ocular regions cropped from the frontal face images in the Adience dataset [14].
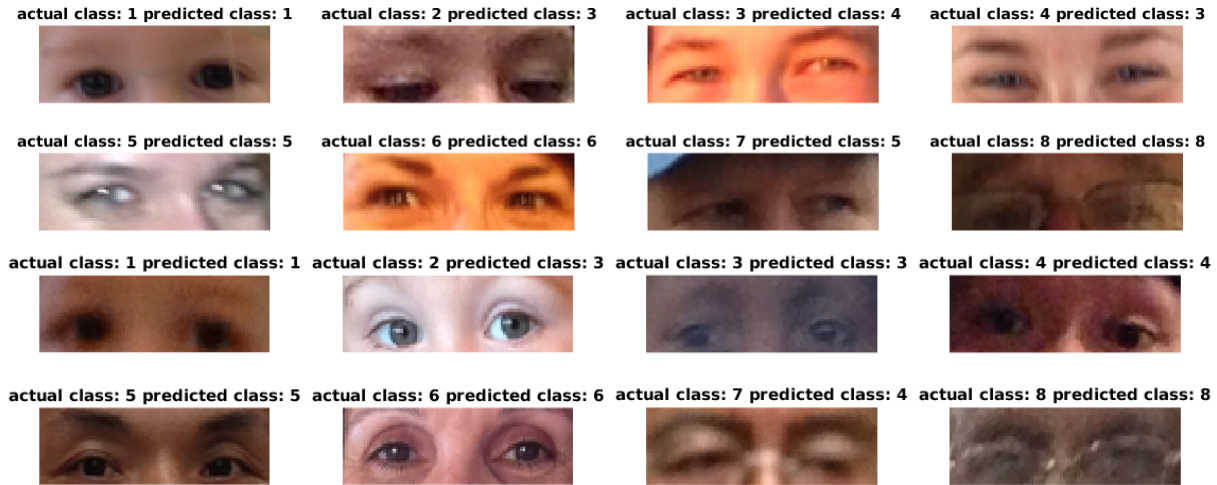

Figure 3. Examples of ocular images correctly (in terms of exact and 1-off accuracies) and incorrectly classified into different age-groups.

suggest efficacy of ocular region in age classification of individuals over full face images. Further, using ocular crops instead of full face images, though more challenging, has applications when full face is not available. For instance, in ocular biometrics pipeline where regions in and around eyes are used for authentication purposes. Using ocular region also has privacy and computational benefits over full face images.

As a part of future work, more elaborate models will be trained on larger datasets. Various pre-trained CNNs will be fine-tuned and combined for age classification from ocular images. Further, gender-specific age classification will be performed and the effect of eye make-up on classification error will be gauged.

6. Acknowledgement

References

[1] A. Abbasi and M. Khan. Iris-pupil thickness based method for determining age group of a person. *International Arab Journal of Information Technology*, 13(6), 2016.

[2] S. Baker, K. W. Bowyer, and P. Flynn. Empirical evidence for correct iris match score degradation with increased time

lapse between gallery and probe matches. In *International Conference on Biometrics*, pages 1170 – 1179, 2009.

[3] K. W. Bowyer, S. E. Baker, A. Hentz, K. Hollingsworth, T. Peters, and P. Flynn. Factors that degrade the match distribution in iris biometrics. *Identity in the Information Society*, 2(3):327–343, 2009.

[4] C. Chen, Y. Chang, K. Ricanek, and Y. Wang. Face age estimation using model selection. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop*, pages 93 – 99, 2010.

[5] A. Dantcheva, P. Elia, and A. Ross. What else does your biometric data reveal? a survey on soft biometrics. *IEEE Transactions on Information Forensics and Security*, pages 1–26, 2015.

[6] A. Das, U. Pal, M. Blumenstein, and M. F. Ballester. Sclera recognition - a survey. In *IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 917–921, Nov 2013.

[7] A. Das, U. Pal, M. A. Ferrer, and M. Blumenstein. SSRBC 2016: Sclera segmentation and recognition benchmarking competition. In *International Conference on Biometrics (ICB)*, pages 1–6, June 2016.

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

[9] M. Erbilek, M. Fairhurst, and M. Abreu. Age prediction from iris biometrics. In *5th International Conference on Imaging for Crime Detection and Prevention*, Stevenage, 2013. The Institution of Engineering and Technology.

[10] P. D. Ferguson, T. Arslan, A. T. Erdogan, and A. Parmley. Evaluation of contrast limited adaptive histogram equalization (clahe) enhancement on a FPGA. In *IEEE International SOC Conference*, pages 119–122, Sept 2008.

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

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

[13] A. k. Saxena and V. K. Chaurasiya. Multi-resolution texture analysis for fingerprint based age-group estimation. *Multimedia Tools and Applications*, pages 1–27, 2017.

[14] E. D. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755 – 1758, 2009.

[15] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[16] G. Levi and T. Hassncer. Age and gender classification using convolutional neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 34 – 42, June 2015.

[17] J. Lu and Y.-P. Tan. Gait-based human age estimation. *IEEE Trans. Inf. Forensics Security*, 5(4):761 – 770, 2010.

[18] M. D. Marsico, M. Nappi, D. Riccio, and H. Wechsler. Mobile iris challenge evaluation (miche)-i, biometric iris dataset and protocols. *Pattern Recognition Letters*, 57:17–23, 2015.

[19] A. Rattani and R. Derakhshani. Ocular biometrics in the visible spectrum: A survey. *Image and Vision Computing*, 59:1 – 16, 2017.

[20] A. Rattani and R. Derakhshani. Online co-training in mobile ocular biometric recognition. In *IEEE International Symposium on Technologies for Homeland Security (HST)*, pages 1–5, April 2017.

[21] A. Rattani, R. Derakhshani, S. K. Saripalle, and V. Gottemukkula. Icip 2016 competition on mobile ocular biometric recognition. In *2016 IEEE International Conference on Image Processing (ICIP)*, pages 320–324, Sept 2016.

[22] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, 115(3):211–252, 2015.

[23] A. Sgroi, K. W. Bowyer, and P. J. Flynn. The prediction of old and young subjects from iris texture. In *International Conference on Biometrics (ICB)*, pages 1 – 5, June 2013.

[24] L. Shamir. Automatic age estimation by hand photos. *Comput. Sci. Lett.*, 3(1):1 – 6, 2011.

[25] Y. Sun, X. Wang, and X. Tang. Deep learning face representation from predicting 10,000 classes. In *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '14, pages 1891 – 1898, Washington, DC, USA, 2014.

[26] A. Toshev and C. Szegedy. Deeppose: Human pose estimation via deep neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1653 – 1660, June 2014.

6