Robust Deep Age Estimation Method Using Artificially Generated Image Set

Jaeyoon Jang*, Seung-Hyuk Jeon*, Jaehong Kim, and Hosub Yoon

Human age estimation is one of the key factors in the field of Human–Robot Interaction/Human–Computer Interaction (HRI/HCI). Owing to the development of deep-learning technologies, age recognition has recently been attempted. In general, however, deep learning techniques require a large-scale database, and for age learning with variations, a conventional database is insufficient. For this reason, we propose an age estimation method using artificially generated data. Image data are artificially generated through 3D information, thus solving the problem of shortage of training data, and helping with the training of the deep-learning technique. Augmentation using 3D has advantages over 2D because it creates new images with more information. We use a deep architecture as a pre-trained model, and improve the estimation capacity using artificially augmented training images. The deep architecture can outperform traditional estimation methods, and the improved method showed increased reliability. We have achieved state-of-the-art performance using the proposed method in the Morph-II dataset and have proven that the proposed method can be used effectively using the Adience dataset.

I. Introduction

Recognizing user information from a single face image is the first step of human–robot and human–computer interaction. For human interaction with robots or computers, the ability to recognize certain human information is required. If a robot does not have information regarding people, they cannot have an appropriate response. Therefore, much research on recognizing various types of information from images is in progress.

Owing to the development of new technology, most of the major information can already be recognized, such as gender, identity, hairstyle, and facial expression. However, some difficult tasks still remain. Age recognition is one of the remaining problems. Although the average face changes according to age and gender, it is not easy to recognize one's age because the variation across individuals is too large.

In general, the age recognition problem can be divided into two types: real biological age estimation and apparent age estimation. The actual biological age is more difficult to recognize than the apparent age because of the influence of various environmental factors during growth. Thus, it is easy to recognize one's approximate age, but it is difficult to recognize one's exact age.

The face of a person changes twice during their lifetime. A baby has a rounded face shape compared to an adult, with a relatively short distance between the eyes and the mouth. After approximately six or seven years of age, an intermediate shape between a baby and an adult occurs. At approximately 17 years old, secondary sexual characteristics appear, and from that time on, the person
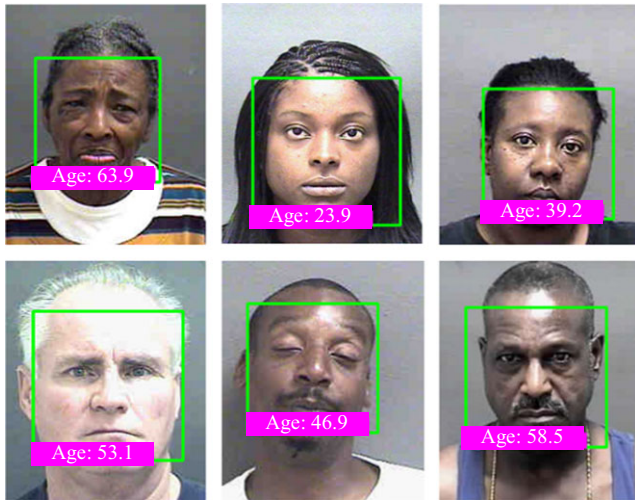
Fig. 1. Age estimation example: real age = {64, 26, 38, 57, 40, 58} (from top-left to bottom-right).

has a face shape similar to that of an adult. The growth of the eyes, nose, and mouth stops, and the shape of the face is relatively long. It is difficult to estimate one's age because of these various changes.

Recently, however, owing to the development of big data and machine learning technology, studies on biological age recognition have become active again. In this study, we generated an artificial face image and made a deep age estimation model using a deep convolutional neural network (CNN). Artificially generated images can be used to mitigate the lack of age data. Thus, we improve the biological age estimation performance using artificially three-dimensionally transformed images (through yaw, pitch, and roll).

The rest of this study is organized as follows: Section II includes a summary of related work from the literature. In Section III, we describe a detailed method regarding artificially generated face images and real age estimation. In Section IV, we describe our experiments and its performance on the MORPH-II dataset. Finally, we compare our proposed method with other state-of-the-art studies. An example is shown in Fig. 1.

II. Related Work

1. Study

In recent decades, numerous studies have been conducted on recognizing human age. A few years ago, most research was focused on feature extraction. They studied how to efficiently extract a feature from an image, how to calculate the age using the extracted feature, and so on. Various human-made features have been applied for age estimation. Some of these focus on the texture or

gradient. The local binary pattern (LBP) [1], [2], Gabor [3], and Sobel [4] filters have been proposed as attempts to extract local texture features.

These studies have generally used a method for extracting landmark points from a face image, generating patches at these positions, and calculating the features of these patches. When using local features, local characteristics can be expressed well, but there is a disadvantage in that this does not reflect the overall morphological characteristics according to age. To compensate for these shortcomings in the previous methods, it has been necessary to use the entire face image, for which it has been difficult to estimate age. In addition, because biological information is more effective in expressing age than texture features, some studies have used biological information [5], [6] for feature extraction.

Some regression methods have improved, and new methods have been developed, such as a support vector regression (SVR) [7], quadratic regression and manifold learning [8], and canonical correlation analysis [9]. These new studies have helped in estimating age by recognizing the inherent problems more reliably.

In addition, some recent studies have used deep CNNs. The methods of age recognition by dividing the face into several parts [10], [11] or recognizing age after identifying the age group as either adult or minor [12] show good performance. In addition, the method of using ordinal regression [13] or the kernel EML [14] show significantly improved results compared with the previous methods. Since this deep learning method requires a large training database, much research on data augmentation is in progress. Some studies created additional data by morphing images to obtain faces in the same position from the faces of several people [15], or by composing hair styles and accessories on face images [16]. In recent years, there have been studies on creating 3D faces from 2D images [17].

2. Dataset

Even as recently as a few years ago, there was not a large amount of data for age recognition. Thus, many studies have over-fitted the training data or cannot significantly improve the performance. To solve the problem of a lack of data, several databases have been produced: FRGC [18], MORPH-II [19], Ni's web-collected database [20], Adience [21], and IMDB-WIKI [22]. However, most of these age databases consist of biological age labels. Because the appearance of a person does not change linearly with age, the age estimation problem using a biological dataset is one of the difficult problems that needs to be solved. For this reason, an apparent age dataset has recently appeared: the

Table 1. Age database.

	Number of subject	Number of image	Age range	Age characteristic
FRGC	568	44,278	18–70	Biological age
MORPH-II	13,618	55,132	16–77	Biological age
Ni's web	N/A	219,892	01–80	Biological age
Adience	2,284	26,580	00–60+	8 age group
IMDB-WIKI	20,284	523,051	00–100	Biological age
ChaLearn 2015	N/A	3,651	N/A	Apparent age
ChaLearn 2016	N/A	5,613	N/A	Apparent age

ChaLearn dataset [23]. Detailed information on this is shown in Table 1.

III. Proposed Method

1. Nonfrontal Face Generation

We generate nonfrontal faces from frontal faces as a face augmentation method to help with the generalization of our age estimation model. The effect of adding a generated nonfrontal face to the training database is described in Section IV. We propose an elaborate face–pose generation method, in which a face in an image is reconstructed from 2D into 3D and then rotated. A face is detected using a state-of-the-art face detector in Dlib [24], as shown in Fig. 2.

This face detector was trained using the histogram of oriented gradients feature and a linear classifier. Facial landmarks are detected using an ensemble of regression trees (ERT) [25]. A two-pixel difference is selected by considering the exponential distance from an estimated landmark when training each node of the ERT. Each cascade regression step of the ERT for facial landmarks is composed of gradient-boosting trees. To build the 3D face model, 68-point annotations in a 300-W database [26], [27] are reconstructed into 3D landmark annotations, based on the structure from the motion method. The 3D landmark annotations allow for the generation of a point distribution model (PDM):

$$X_{\{3n \times 1\}} = M_{\{3n \times 1\}} + \sum_{i=1}^{d} e_i B(:,i)_{\{3n \times d\}} = X(e), \quad (1)$$

where M is the mean shape and B is the basis of the shape variation. The parameter vector e is a shape vector that determines the specific shape. In addition, X contains the 3D positions of n facial landmarks. Contrary to this, we can obtain the projection of any facial landmarks on the PDM of (1):
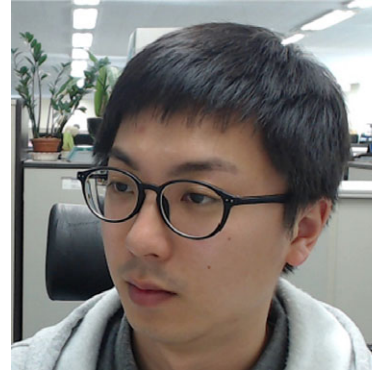

Fig. 2. Face image and 3D landmark points.

$$e_{\{d \times 1\}} = P(Y_{\{3n \times 1\}}) = B^T_{\{d \times 3n\}}(Y - M)_{\{3n \times 1\}}, \quad (2)$$

where Y is any facial landmark transformed by T in (3), and P represents a projection on PDM in (1). We also consider a similarity transformation T that minimizes the Euclidean distance between M and Y:

$$\text{minimize } \|Y - M\|^2,$$
$$\text{where } Y = T(Z),$$
$$T(Z) = s_{\{1 \times 1\}}(I_{\{n \times n\}} \otimes R_{\{3 \times 3\}})Z_{\{3n \times 1\}} + 1_{\{n \times 1\}} \otimes t_{\{3 \times 1\}}, \quad (3)$$

where s, R, and t are the scale, 3D rotation, and 3D translation parameters, respectively. The 3D reconstruction process is described in Algorithm 1:

Algorithm 1. 3D reconstruction process.

Algorithm 1 3D Face Reconstruction	
Input:	M – mean shape in equation (1)
	B – shape basis in equation (1)
	Z – 3D landmarks with x and y-values of detected 2D landmarks and
	z-values of M.
Output:	e – shape parameters in equation (1)
	s, R, t – similarity transformation in equation (3)
1:	while e, s, R, t are converged
2:	Transform Z to Y using equation (3)
3:	Calculate e using equation (2)
4:	e is projected onto equation (1)
5:	X is inverse transformed to X by $1/s$, $-t$.
6:	Replace z-value of Z with z-value of X'
7:	end

A 3D landmark shape is reconstructed through (1) and then transformed using $1/s - t$ [28]. The reconstructed 3D landmarks are a weak perspective transformed by a

rotation around the center of gravity. The texture of the face is divided into a mesh of triangles, and the 3D affine is transformed into rotated meshed triangles. The nonface region is also transformed in terms of the perspective. Gaussian blur is applied to the boundary of a rotated face, because there are cracks between the face region and nonface region. We show examples of rotated faces in Fig. 3.

2. Architecture

To obtain higher performance, we use the VGG-16 architecture from [29] as a pre-trained model. This architecture is one of the famous models in image classification, and many studies have used it as a pre-trained model. It consists of thirteen convolutional layers, five pooling layers, two fully connected layers, and a softmax layer. It uses ReLU as the activation function, 4,096 neurons in a fully connected layer, and 1,000 neurons as the output. To obtain the probability of each age, we change the parameters of the penultimate layer to 100.

3. Expected Value

In the age-regression stage, we use the expected value from [25]. Most previous studies have used single-value regression for age estimation. However, a single-value regression method considers only the target value, and thus it does not fully reflect the probability of the overall range. On the other hand, the expected value can use the probability of the overall range, and can calculate more stochastic ages.

$$E(O) = \sum_{i=0}^{100} pi * yi. \qquad (4)$$

In (4), O is the expected age between 1 and 100, pi is the probability value of each age class from the softmax layer of the CNN architecture, and yi is each class label (age). An overview of this formula is shown in Fig. 4.

IV. Experiments

1. Normalization

To achieve better performance, first, we detect the face region using an MCT-based cascade AdaBoost face detector and find the facial landmark points, including the center of the left eye, center of the right eye, nose tip, and end points of the mouth using an ERT [25]. Then, to achieve an invariant to the rotational change, we make a horizontal face image using the two eye positions. Using a rolling-aligned image and the distance between the two eyes, we set the region of interest as shown in Fig. 5. Finally, the face region is resized to a 224 × 224 image and used for the input of the CNN architecture. At this stage, we use only flip augmentation and do not use a random crop or translation augmentation, because aligned faces are more helpful in extracting the regional characteristics.
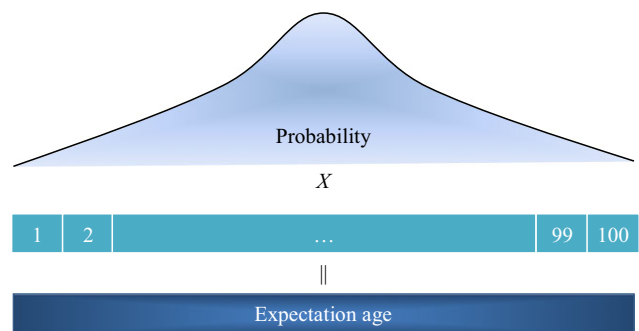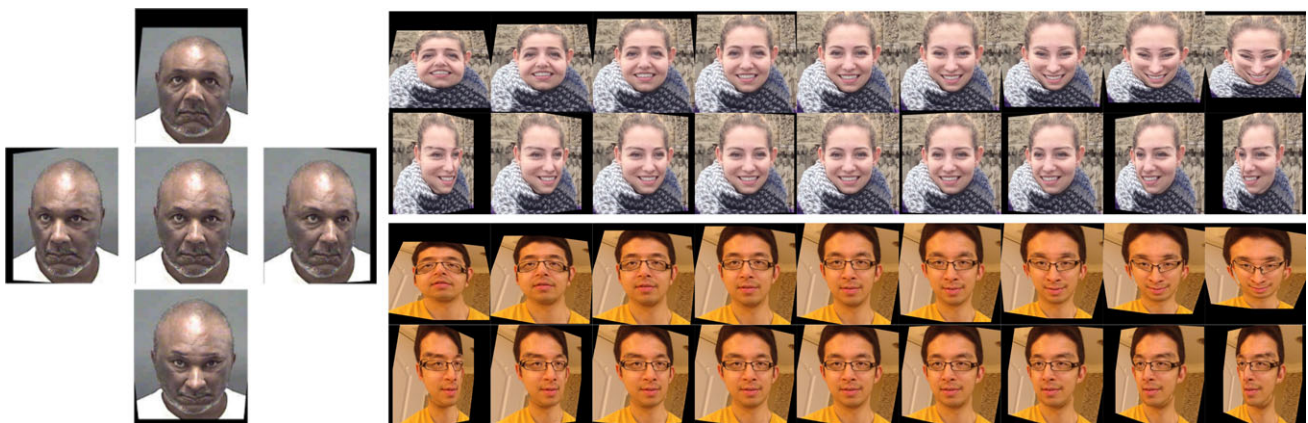
Fig. 4. Calculating the expected age.

Fig. 3. Artificially generated image. Same person showing different characteristics because of the different poses.

Fig. 5. Summary of face image normalization method.

2. Training

The Morph-II database is one of the largest facial age databases, which consists of 55,134 images of over 13,000 people, between 16 and 77 years in age, and including various ancestries, such as Africans, Europeans, Asians, and Hispanics. In our experiments, we randomly divided the dataset into two nonoverlapping sets. The first set is made up of 90% of the whole dataset and is used for training. The remaining 10% is used for testing. After division, we generate artificially transformed face images. We generate four additional faces: $\pm15°$ transformed yaw, and $\pm10°$ transformed pitch. Because the pitch transform causes more substantial differences, it has a narrow boundary. These generated face images have the same characteristics but different shapes, and thus their use has the same effect as using different pictures of the same person. These artificially generated images help with learning the physical traits and help strengthen the generalization tendency. A generated image is shown in Fig. 8. We made approximately 240,000 artificial images using the above method, and can supplement the lack of training data by using these artificially generated images.

Finally, we used artificially generated data to learn through the classification problem. The overall system is shown in Fig. 6.

We applied the same augmentation and training parameters to the Adience database to learn the age model. Since the Adience database uses five-fold cross-validation, five similar models are trained.

3. Age Estimation in Morph-II

In the Morph-II database, the estimation performance is evaluated using the standard mean absolute errors (MAE). The MAE is the average value of the absolute error between the ground-truth value and the estimated value. This is one of the commonly used methods, and thus we evaluate the age estimation performance using this method. We randomly selected 10% of the total dataset for testing, the configuration of which is shown in Fig. 7.

To evaluate our method, we compared it with existing methods in several ways. First, we compared it with a traditional regression method [8], [30]. In [29], the authors extract biologically inspired features from a face image, and then train an SVR using these features. In [30], the authors also trained a linear SVR, but they used a CNN as the feature extractor. They divided the face image into patches, and extracted each local-level feature. Using the local-level features, they aggregated the features through concatenation. Such concatenated features are used for the input of the fully connected layer, and the output is used for the final age feature. Second, we compared some studies on a typical method, such as [13] and [31]. They transformed the traditional regression problem into a binary classification problem or ordinal problem.

The evaluation results are shown in Figs. 8 and 9. Our proposed method shown an MAE of 2.6 on the MORPH-II dataset, and thus outperformed the methods from these
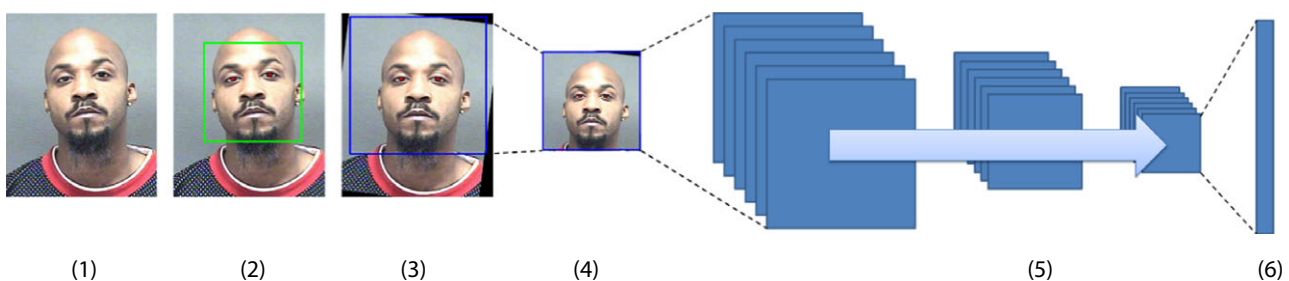

Fig. 6. Age estimation pipeline: (1) original input image, (2) face and landmark detection, (3) alignment using a landmark and set ROI, (4) cropped and resized candidate regions, (5) calculated class probability through the CNN architecture, and (6) calculated expectation age using the class probability and class label.
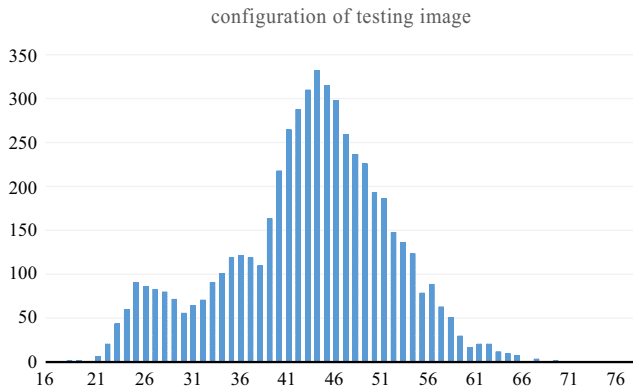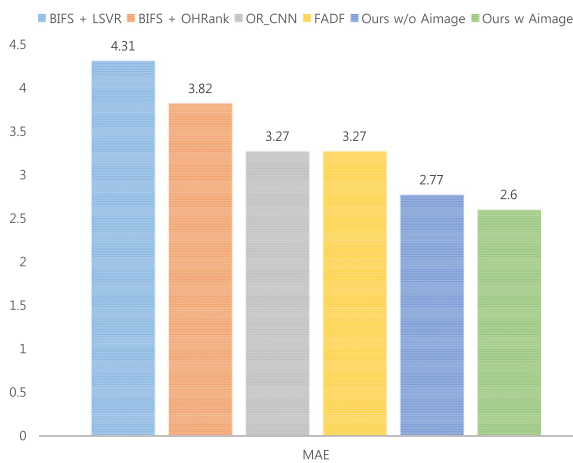
Fig. 7. Configuration of testing image.

Fig. 8. Mean absolute error in Morph-II database. Deeper model without artificially generated data obtain an error of 2.77, and with generated data obtain an error of 2.6. Both outperform traditional methods.
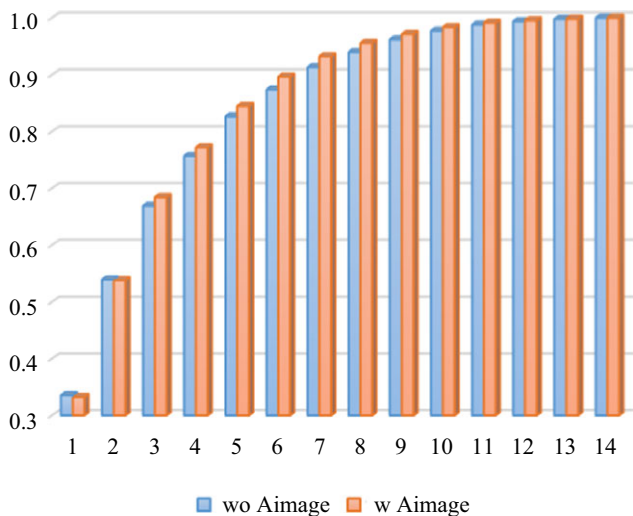

Fig. 9. Experiment results for Morph-II dataset. Training with generated images achieves a better performance.

Table 2. Confusion matrix of testing results in Morph-II database.

	10–19	20–29	30–39	40–49	50–59	60–69	70–
10–19	**1**	3	0	0	0	0	0
20–29	13	**466**	61	5	0	0	0
30–39	2	89	**800**	120	7	0	0
40–49	0	6	497	**2,066**	178	1	0
50–59	0	3	16	393	**679**	10	0
60–69	0	0	0	12	52	**32**	0
70–	0	0	0	0	0	1	**0**

Table 3. Adience dataset benchmark result.

Method	Exact	1-off
01 LBP [21]	41.4 $\pm$ 2.0	78.2 $\pm$ 0.9
02 FPLB P [21]	39.8 $\pm$ 1.8	74.6 $\pm$ 1.0
03 LBP + FPLBP [21]	44.5 $\pm$ 2.3	80.7 $\pm$ 1.1
04 LBP + FPLBP_PCA0.5 [21]	38.1 $\pm$ 1.4	75.5 $\pm$ 0.9
05 LBP + FPLBP + Dropout 0.5 [21]	44.5 $\pm$ 2.2	80.6 $\pm$ 1.0
06 LBP + FPLBP + Dropout 0.8 [21]	45.1 $\pm$ 2.6	79.5 $\pm$ 1.4
07 Over-sample [32]	50.7 $\pm$ 5.1	N/A
08 KNN with BB-FCN [11]	51.1 $\pm$ 5.7	N/A
09 Casecade CNN [33]	52.9 $\pm$ 6.0	88.5 $\pm$ 2.2
10 Ours without 3D augmentation	50.5 $\pm$ 2.6	91.3 $\pm$ 1.3
11 Ours with 3D augmentation	52.6 $\pm$ 1.8	92.4 $\pm$ 1.2

other studies. In addition, as shown in Fig. 9, the performance of the models learned using artificially created images is slightly better on average. These results show that the alternative-pose images of the same person helped to make the learning more robust.

As we can see in Table 2, it was possible to estimate most of the testing data near the actual age. However, as seen with people in their 40s and over, some images were estimated to be younger, which is the result of the disproportionate learning due to the dominance of data from people in their 30s and 40s.

4. Age Estimation in Adience

Since the Morph-II database is well-studied, we used Adience for an additional evaluation of the performance. Four of the five folds were used for learning and the performance was evaluated using the fifth. The final performance was calculated as the average of five results. Our proposed method is compared with several traditional methods [21] and deep CNN-based methods [11], [32],

Table 4. Confusion matrix on the adience benchmark.

Age range in years	0–2	4–6	8–13	15–20	25–32	38–43	48–53	60–
0–2	**0.821**	0.169	0.004	0.001	0.003	0.001	0.002	0.000
4–6	0.263	**0.638**	0.090	0.003	0.004	0.001	0.001	0.000
8–13	0.012	0.181	**0.614**	0.102	0.079	0.007	0.002	0.004
15–20	0.000	0.012	0.117	**0.375**	0.456	0.031	0.007	0.002
25–32	0.000	0.009	0.024	0.125	**0.671**	0.155	0.012	0.004
38–43	0.002	0.001	0.008	0.030	0.462	**0.349**	0.114	0.034
48–53	0.003	0.000	0.007	0.007	0.112	0.381	**0.218**	0.272
60–	0.000	0.000	0.000	0.003	0.047	0.134	0.297	**0.518**

[33]. As can be seen in Table 3, the use of 3D augmentation data can yield better performance. The resulting performance is not state-of-the-art, but it can be relatively good and shows less variance compared to the state of the art. In addition, we obtained state-of-the-art performance in once-off experiments. The confusion matrix in Table 4 confirms that the results are totally stable.

V. Conclusion

This paper proposed a new method to perform age recognition when no additional learning data are available. In addition, unlike the trend in conventional age recognition, we obtained stable performance using a deep architecture. Architectures that have already been proven elsewhere will perform well in age recognition, but our method will yield better results. Our method has achieved state-of-the-art performance using artificially generated images for the Morph-II database. Additionally, we did not obtain the best performance for the Adience database, but we got results comparable to the state of the art using artificially generated images without additional skills.

Finally, the lack of training data can be overcome using artificially generated images, which can provide even more reliable results.

Acknowledgements

References

[1] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, 2002, pp. 971–987.

[2] Z. Yang and H. Ai, "Demographic Classification with Local Binary Patterns," *Int. Conf. Biometrics*, Crystal City, VA, USA, 27–30, Sept. 2007, pp. 464–473.

[3] F. Gao and H. Ai, "Face Age Classification on Consumer Images with Gabor Feature and Fuzzy Lda Method," *Int. Conf. Biometrics*, Alghero, Italy, 2–5, June 2009, pp. 132–141.

[4] J.-D. Txia and C.-L. Huang, "Age Estimation Using AAM and Local Facial Features," *Int. Inform. Hiding Multimedia Signal Process.*, Kyoto, Japan, 12–14, Sept. 2009, pp. 885–888.

[5] G. Guo et al., "Image-Based Human Age Estimation by Manifold Learning and Locally Adjusted Robust Regression," *IEEE Trans. Image Process.*, vol. 17, no. 7, 2008, pp. 1178–1188.

[6] O.M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep Face Recognition," *British Mach. Vis. Conf.*, Swansea, UK, 7–10, Sept. 2015.

[7] G. Guo and G. Mu, "Simultaneous Dimensionality Reduction and Human Age Estimation Via Kernel Partial Least Squares Regression," *IEEE Conf. Comput. Vis. Pattern Recogn.*, Colorado Springs, CO, USA, 20–25 June 2011, pp. 657–664.

[8] Y. Fu and T.S. Huang, "Human Age Estimation with Regression on Discriminative Aging Manifold," *IEEE Trans. Multimedia*, vol. 10, no. 4, 2008, pp. 578–584.

[9] G. Guo and G. Mu, "Joint Estimation of Age, Gender and Ethnicity: CCA vs. PLS," *IEEE Int. Conf. Workshops*

Automatic Face Gesture Recogn., Shanghai, China, 22–26, Apr. 2013, pp. 1–6.

[10] Y. Dong, Y. Liu, and S. Lian, "Automatic Age Estimation Based on Deep Learning Algorithm," *Neurocomput.*, vol. 187, 2016, pp. 4–10.

[11] J. Huang et al., "Age Classification With Deep Learning Face Representation," *Multimedia Tools Applicat.*, vol. 76, no. 19, Oct. 2017, pp. 1–17.

[12] G. Antipov et al., "Apparent Age Estimation From Face Images Combining General and Children-Specialized Deep Learning Models," *Proc. IEEE Conf. Comput. Vis. Pattern Recogn. Workshops*, Las Vegas, NV, USA, 2016, pp. 801–809.

[13] Z. Niu et al., "Ordinal Regression with Multiple Output CNN for Age Estimation," *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, Las Vegas, NV, USA, 27–30, June 2016, pp. 4920–4928.

[14] F. Gurpinar et al., "Kernel ELM and CNN Based Facial Age Estimation," *Proc. IEEE Conf. Comput. Vis. Pattern Recogn. Workshops*, Las Vegas, NV, USA, 2016, pp. 785–791.

[15] J.-J. Lv et al., "Data Augmentation for Face Recognition," *Nerocomput.*, vol. 230, 2017, pp. 184–196.

[16] Y. Nirkin et al., "On Face Segmentation, Face Swapping, and Face Perception," arXiv preprint arXiv:1704.06729, 2017.

[17] D. Kim et al., "Deep 3D Face Identification," arXiv preprint arXiv:1703.10714, 2017.

[18] P.J. Phillips et al., "Overview of the Face Recognition Grand Challenge," In *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recogn.*, San Diego, CA, SUA, 20–25 June 2005, pp. 947–954.

[19] K. Ricanek and T. Tesafaye, "Morph: A Longitudinal Image Database of Normal Adult Age-Progression," In *Int. Conf. Automatic Face Gesture Recogn.*, Southampton, UK, 10–12, Apr. 2006, pp. 341–345.

[20] B. Ni, Z. Song, and S. Yan, "Web Image Mining Towards Universal Age Estimator," In *Proc. ACM Int. Conf. Multimedia*, Beijing, China, 19–24, Oct. 2009, pp. 85–94.

[21] E. Eidinger, R. Enbar, and T. Hassner, "Age and Gender Estimation of Unfiltered Faces, Information Forensics and Security," *IEEE Trans. Inform. Foren. Secur.*, vol. 9, no. 12, Dec. 2014, pp. 2170–2179.

[22] R. Rothe, R. Timofte, and L. Gool, "DEX: Deep Expectation of Apparent Age From a Single Image," *IEEE Int. Conf. Comput. Vis. Workshop*, Santiago, Chile, 7–13, Dec. 2015, pp. 10–15.

[23] X. Baro et al., "Chalearn Looking at People Challenge 2015: Dataset and Results," *CVPR, ChaLearn Looking at People Workshop*, 2015.

[24] D.E. King, "Dlib-ml: A Machine Learning Toolkit," *J. Mach. Learn. Res.* vol. 10, July 2009, pp. 1755–1758.

[25] V. Kazemi and J. Sullivan, "One Millisecond Face Alignment with An Ensemble of Regression Trees," *IEEE Conf. Comput. Vis. Pattern Recogn.*, Columbus, OH, USA, 23–28, June 2014, pp. 1867–1874.

[26] C. Sagonas et al., "300 Faces in-the-Wild Challenge: The First Facial Landmark Localization Challenge," *IEEE Int. Conf. Comput. Vis. Workshops*, Sydney, Australia, 2–8 Dec. 2013, pp. 397–403.

[27] C. Sagonas et al., "A Semi-Automatic Methodology for Facial Landmark Annotation," *IEEE Conf. Comput. Vis. Pattern Recogn. Workshops*, Portland, OR, USA, 23–28, June 2013, pp. 896–903.

[28] J. Saragih and R. Göcke, "Learning AAM Fitting Through Simulation," *Pattern Recogn.*, vol. 42, no. 11, 2009, pp. 2628–2636.

[29] G. Mu et al., "Human Age Estimation Using Bio-Inspired Features," *IEEE Conf. Comput. Vis. Pattern Recogn.*, Miami, FL, USA, 20–25, June 2009, pp. 112–119.

[30] J. Qiu et al., "Hierarchical Aggregation Based Deep Aging Feature for Age Prediction," *Int. Conf. Digital Image Comput.: Tech. Applicat.*, Adelaide, Australia, 23–25, Nov. 2015, pp. 1–5.

[31] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, "Ordinal Hyperplanes Ranker with Cost Sensitivities for Age Estimation," *IEEE, Conf. Comput. Vis. Pattern Recogn.*, Colorado Springs, CO, USA, 20–25, June 2011, pp. 585–592.

[32] L. Gil and T. Hassner, "Age and Gender Classification Using Convolutional Neural Networks," *IEEE Conf. Comput. Vis. Pattern Recogn. Workshops.*, Boston, MA, USA, 7–12, June 2015pp. 34–42.

[33] J.-C. Chen et al., "A Cascaded Convolutional Neural Network for Age Estimation of Unconstrained Faces," *IEEE Int. Conf. Biometrics Theory, Applicat. Syst.*, Niagara Falls, NY, USA, 6–9 Sept. 2016, pp. 1–8.

Jaeyoon Jang received his BE degree in control and measurement engineering from Hanbat National University in 2013. He is currently working towards his PhD in the Department of Computer & Software engineering at the University of Science and Technology, Daejeon, Rep. of Korea. His research interests are HMI and face recognition.

Seung-Hyuk Jeon received his BS degree in control and instrumentation engineering from Seoul National University of Technology, Seoul, Rep. of Korea, in 2011, and his MS degree in computer software engineering from Korea University of Science and Technology, Daejeon, Rep. of Korea, in 2017. Since 2017, he has worked for the SNOW. His main research interests are face alignment, 3D face reconstruction, and face recognition.

Jaehong Kim Received his PhD degree in computer engineering from Kyungpook National University, Daegu, Rep. of Korea, in 2006. He has been a research scientist at ETRI, Daejeon, Rep. of Korea, since 2001. His research interests include socially assistive robotics for the elderly, human–robot interaction, and social HRI frameworks.

Hosub Yoon received his BS and MS degrees in computer science from SoongSil University, Seoul, Rep. of Korea, in 1989 and 1991. He received his PhD degree in image processing from KAIST, Daejeon, in 2003. He joined KIST/SERI in 1991 and transferred to ETRI, Daejeon, Rep. of Korea in 1999. He is a project leader of the HMI (Human Machine Interaction) group in the intelligence robotics division. His major research interests include HRI, AI, image processing, deep learning, audio processing, and pattern recognition.