Multi-Race Age Estimation Based on the Combination of Multiple Classifiers

Kazuya Ueki
NEC Soft, Ltd.
1-18-7 Shinkiba, Koto-ku,
Tokyo 136-8627, Japan
Email: ueki@mxf.nes.nec.co.jp

Masashi Sugiyama
Tokyo Institute of Technology
2-12-1 O-okayama, Meguro-ku,
Tokyo 152-8552, Japan
Email: sugi@cs.titech.ac.jp

Yasuyuki Ihara and Mistuhiro Fujita
NEC Soft, Ltd.
1-18-7 Shinkiba, Koto-ku,
Tokyo 136-8627, Japan

***Abstract*—A considerable amount of research has been conducted on gender and age estimation from facial images over the last few years, and state-of-the-art technology has accomplished a practical accuracy level for a homogeneous race such as Japanese or Korean. However, achieving the same accuracy level *across* multiple races such as Caucasian, African American, and Hispanic is still highly challenging because of the strong diversity of the growth process of each race. Furthermore, difficulty of gathering training samples uniformly over various races and age brackets makes the problem even more challenging. In this paper, we propose a novel age estimation method that can overcome the above problems. Our method combines a recently proposed machine learning technique called *Least-Squares Probabilistic Classifier* (LSPC) with neural networks. Through large-scale real-world age estimation experiments, we demonstrate the usefulness of our proposed method.**

I. Introduction

Automatic age estimation from facial images can be utilized for various purposes such as customer analysis in shops for marketing and digital signage for displaying a customized advertisement for a targeted person who is looking at the screen. For this reason, age estimation has gathered a great deal of attention and has been one of the most frequently researched topics in recent years [2], [3], [4], [6], [11], [13]. Availability of public databases for age estimation [1], [5], [7], [14] will further boost the age estimation research.

State-of-the-art age estimation systems have achieved sufficiently high accuracy if they are applied to a homogeneous race [11], [12], [13], [15]. However, it is still challenging to accurately predict ages for *multiple* races because the growth process is significantly different depending on races. Overcoming this difficulty will allow us to expand the age estimation system to the world-wide market.

Currently customer analysis system for marketing and digital signage system are desired in the world-wide market. In these systems, age groups should be divided into at least 3 categories, such as children, adults, and the elderly, with more than 70–80% of accuracy rates.

This paper proposes a new multi-race age estimation method which consists of neural networks and a recently proposed classifier called *Least-Squares Probabilistic Classifier* (LSPC) [8]. This paper is organized as follows. Section II reviews an outline of LSPC. Section III describes how two types of classifiers, neural networks and LSPCs, are combined. Section IV shows experimental results, and Section V summarizes our contribution and future works.

II. Least-Squares Probabilistic Classifier (LSPC)

In this section, we review LSPC [8], which is a computationally very efficient algorithm for multi-class probabilistic classification.

LSPC models the class-posterior probabilities by the linear combination of kernel functions. The most notable advantage of LSPC is that its global optimal solution can be computed *analytically* without resorting to iterative algorithms. Moreover, the solution of LSPC can be computed separately in a class-wise manner. Therefore, LSPC can be stably computed regardless of the imbalance of sample size in each class. These distinguished features of LSPC are particularly suitable for *multi-race age estimation*: the training database is massive scale, but is imbalanced over attributes such as age, race, and gender.

Furthermore, probabilistic outputs of LSPC (i.e., the *confidence* of prediction) can be effectively utilized for combining LSPC with other classifiers such as neural networks, as described later.

A. Problem Formulation

Let $\boldsymbol{x}$ be an input face feature and $y \in \{1, \ldots, c\}$ be its age-bracket label (say, every 10 years), where c is the number of classes. The goal is to estimate the class-posterior probability $p(y|\boldsymbol{x})$ from training samples $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^n$. The class-posterior probability allows us to classify test sample $\boldsymbol{x}$ to class $\widehat{y}$ with confidence level $p(\widehat{y}|\boldsymbol{x})$:

$$\widehat{y} := \operatorname*{argmax}_y p(y|\boldsymbol{x}).$$

B. Kernel Least-Squares Fitting of Class-Posterior Probability

We model the class-posterior probability $p(y|\boldsymbol{x})$ by the following model:

$$q(y|\boldsymbol{x};\boldsymbol{\alpha}) = \sum_{l=1}^{n} \alpha_l K(\boldsymbol{x}, \boldsymbol{x}_l, y, y_l),$$

where $\{\alpha_l\}_{l=1}^n$ are parameters, and $K(\boldsymbol{x}, \boldsymbol{x}', y, y')$ is a non-negative kernel function.

We determine the parameter $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)^\top$ so that the following squared error J_0 is minimized:

$$\begin{aligned} J_0 &= \frac{1}{2}\int \sum_{y=1}^{c} \left(q(y|\boldsymbol{x};\boldsymbol{\alpha}) - p(y|\boldsymbol{x})\right)^2 p(\boldsymbol{x}) d\boldsymbol{x} \\ &= \frac{1}{2}\int \sum_{y=1}^{c} q(y|\boldsymbol{x};\boldsymbol{\alpha})^2 p(\boldsymbol{x}) d\boldsymbol{x} \\ &\quad - \int \sum_{y=1}^{c} q(y|\boldsymbol{x};\boldsymbol{\alpha}) p(\boldsymbol{x}, y) d\boldsymbol{x} + (\text{constant term}). \end{aligned}$$

Approximating the expectations by sample averages, we obtain the approximation of the first two terms as follows:

$$\widehat{J}(\boldsymbol{\alpha}) = \frac{1}{2}\boldsymbol{\alpha}^\top \widehat{\boldsymbol{H}}\boldsymbol{\alpha} - \widehat{\boldsymbol{h}}^\top \boldsymbol{\alpha},$$

where

$$\begin{aligned} \widehat{\boldsymbol{H}}_{l,l'} &= \frac{1}{n}\sum_{i=1}^{n}\sum_{y=1}^{c} K(\boldsymbol{x}_i, \boldsymbol{x}_l, y, y_l) K(\boldsymbol{x}_i, \boldsymbol{x}_{l'}, y, y_{l'}), \\ \widehat{\boldsymbol{h}}_l &= \frac{1}{n}\sum_{i=1}^{n} K(\boldsymbol{x}_i, \boldsymbol{x}_l, y_i, y_l). \end{aligned}$$

Now the LSPC optimization problem is formulated as

$$\widehat{\boldsymbol{\alpha}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmin}} \left[\frac{1}{2}\boldsymbol{\alpha}^\top \widehat{\boldsymbol{H}}\boldsymbol{\alpha} - \widehat{\boldsymbol{h}}^\top \boldsymbol{\alpha} + \frac{\lambda}{2}\boldsymbol{\alpha}^\top \boldsymbol{\alpha}\right],$$

where the l_2-penalty term $\lambda\boldsymbol{\alpha}^\top\boldsymbol{\alpha}/2$ is included for avoiding overfitting. The solution $\widehat{\boldsymbol{\alpha}}$ can be computed analytically as

$$\widehat{\boldsymbol{\alpha}} = \left(\widehat{\boldsymbol{H}} + \lambda \boldsymbol{I}_n\right)^{-1} \widehat{\boldsymbol{h}}, \tag{1}$$

where $\boldsymbol{I}_n$ denotes the n-dimensional identity matrix. Finally, by rounding up negative outputs to zero and normalizing the output summed up to one, the final LSPC solution is given by

$$\widehat{p}(y|\boldsymbol{x}) = \frac{\max\{0, \sum_{l=1}^{n} \widehat{\alpha}_l K(\boldsymbol{x}, \boldsymbol{x}_l, y, y_l)\}}{\sum_{y'=1}^{c} \max\{0, \sum_{l=1}^{n} \widehat{\alpha}_l K(\boldsymbol{x}, \boldsymbol{x}_l, y', y_l)\}}.$$

In practice, we separate the kernel $K(\boldsymbol{x}, \boldsymbol{x}', y, y')$ for input $\boldsymbol{x}$ and output y, and use the Gaussian kernel for $\boldsymbol{x}$ and the delta kernel for y:

$$K(\boldsymbol{x}, \boldsymbol{x}', y, y') = \exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x}'\|^2}{2\sigma^2}\right) \delta_{y,y'},$$

where σ is the Gaussian width and $\delta_{y,y'}$ is the *Kronecker delta*. The use of the delta kernel for y makes the matrix $\widehat{\boldsymbol{H}}$ block diagonal (given the training samples are sorted according to class labels). Then, the solution (1) can be computed very efficiently in a class-wise manner. We can further reduce the computational complexity and memory usage by limiting the number of kernel bases to a fixed number (say, 2000). The Gaussian width σ and the regularization parameter λ may be optimized by cross-validation.

III. Weighted Combination of Multiple Classifiers

In this section, we propose a novel age estimation method based on LSPC. The flowchart of the proposed method is summarized in Figure 1.

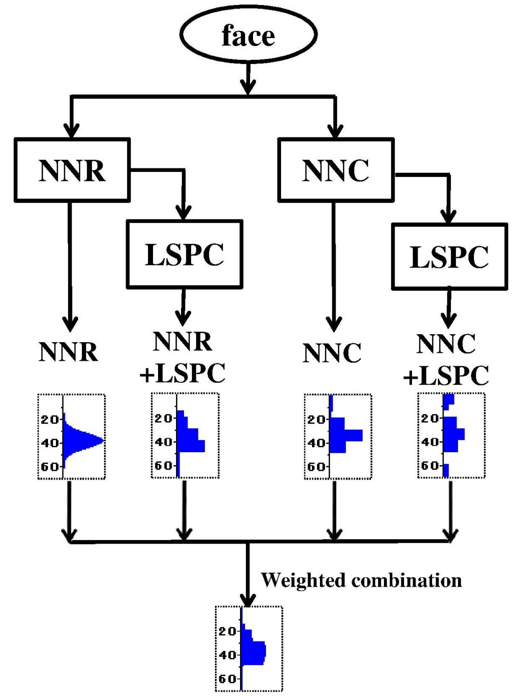

Fig. 1. Flowchart of the proposed method for combining four predictors.

A. Base Predictors

Neural networks were shown to be useful for age estimation [9], [10]. Here, we consider the following four age prediction methods:

- **Neural network regression (NNR):** A neural network is trained as a regressor, which directly estimates a person's real age such as 37 years old.
- **LSPC with NNR (NNR+LSPC):** 100-dimensional feature vectors extracted from the last hidden layer of NNR are used as input to LSPC.
- **Neural network classification (NNC):** A neural network is trained as a classifier, which estimates the confidence of each age class.
- **LSPC with NNR (NNC+LSPC):** 100-dimensional feature vectors extracted from the last hidden layer of NNC are used as input to LSPC.

Combining multiple classifiers is expected to produce more stable estimation performance than a single classifier alone by mutually compensating for the weaknesses of each method. Below, we explain how to combine these four classifiers in a practically useful way.

B. Scoring Methods

First, we explain how to convert the output of each predictor to a score for combining predictors.

1) Scoring for Regression Output: The NNR model outputs a real scalar as an age estimate. We convert this estimate based on the characteristic of human age perception [15] as follows:

Let us define the 'true' age y^* of a subject as the average perceived age y evaluated by those who observed the subject's

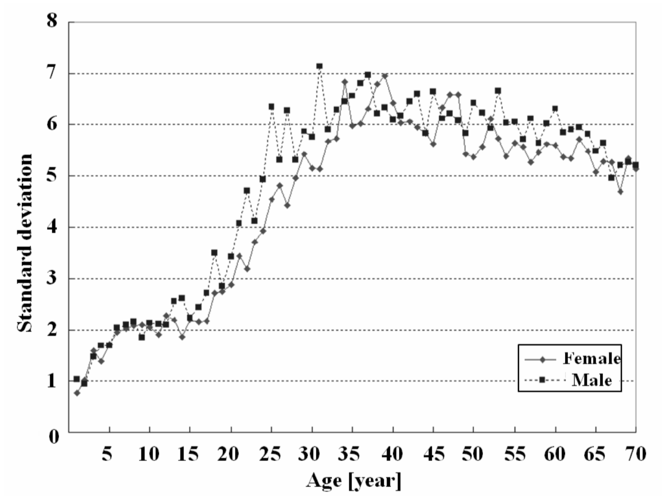

Fig. 2. The relation between subjects' perceived age y^* (horizontal axis) and its standard deviation (vertical axis).

facial images (the value is rounded-off to the nearest integer). Then the standard deviation of perceived age is calculated as a function of y^* and denoted by $w_{\mathrm{age}}(y)$, which is summarized in Figure 2. This graph shows that the perceived age deviation tends to be small in younger age brackets and large in older age groups.

Using this $w_{\mathrm{age}}(y)$, the output of NNR is converted to the score vector as

$$\boldsymbol{S}_{\mathrm{NNR}} = \frac{(s_{\mathrm{NNR}}(0), s_{\mathrm{NNR}}(1), \ldots, s_{\mathrm{NNR}}(69))^\top}{\sum_{y^*=0}^{69} s_{\mathrm{NNR}}(y^*)},$$

where, for $\widetilde{y}$ being the output of NNR,

$$s_{\mathrm{NNR}}(y^*) = \exp\left(-\frac{(y^* - \widetilde{y})^2}{2w_{\mathrm{age}}(\widetilde{y})^2}\right).$$

s_{NNR} is a Gaussian-shaped function which is maximized when y^* agrees with $\widetilde{y}$. The standard deviation of the Gaussian function is set to the standard deviation of perceived age. Thus, When $\widetilde{y}$ is small (younger age), the Gaussian function has a small variance, while when $\widetilde{y}$ is large (older age), the Gaussian function has a large variance.

2) Scoring for Classification Output: The NNC model and LSPC output the class-posterior probability for each class (i.e., age bracket), which can be translated as the confidence level of prediction. We use the confidence level as a score. Let $p_1, p_2, \ldots, p_c$ be the class-posterior probabilities for each class. Then, the score is determined as

$$\boldsymbol{S}_{\mathrm{C}} = \frac{(s_{\mathrm{C}}(0), s_{\mathrm{C}}(1), \ldots, s_{\mathrm{C}}(69))^\top}{\sum_{y^*=0}^{69} s_{\mathrm{C}}(y^*)},$$

where, for i being the age-group class to which the age y^* belongs,

$$s_C(y^*) = p_i.$$

Using this scoring method, we compute the score vectors for NNR+LSPC, NNC, and NNC+LSPC, which are denoted by $\boldsymbol{S}_{\mathrm{NNR+LSPC}}$, $\boldsymbol{S}_{\mathrm{NNC}}$, and $\boldsymbol{S}_{\mathrm{NNC+LSPC}}$, respectively.

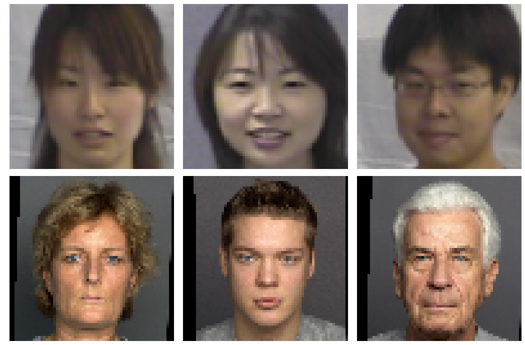

Fig. 3. Examples of face images from our in-house database (top) and FACES database [14] (bottom).

C. Weighted Combination from Output Scores

We blend the score vectors by linear combination:

$$\begin{aligned}\boldsymbol{S} &= (s(0), s(1), \ldots, s(69))^\top \\ &= w_{\mathrm{NNR}}\boldsymbol{S}_{\mathrm{NNR}} + w_{\mathrm{NNR+LSPC}}\boldsymbol{S}_{\mathrm{NNR+LSPC}} \\ &\quad + w_{\mathrm{NNC}}\boldsymbol{S}_{\mathrm{NNC}} + w_{\mathrm{NNC+LSPC}}\boldsymbol{S}_{\mathrm{NNC+LSPC}}.\end{aligned}$$

The weight parameters w_{NNR}, $w_{\mathrm{NNR+LSPC}}$, w_{NNC}, and $w_{\mathrm{NNC+LSPC}}$ are determined so that the prediction error for validation data (which are not used for training predictors) is minimized: weight parameters are selected from $\{0, 0.1, 0.2, \ldots, 1\}$.

Finally, an age estimate obtained from the aggregated predictor is given by

$$\widehat{y} = \underset{y^*}{\mathrm{argmax}}\, s(y^*). \quad (2)$$

IV. Experiments

In this section, we demonstrate the effectiveness of the proposed method in large-scale real-world age estimation. We consider an eight-class age estimation problem: 0–9, 10–14, 15–19, 20–29, 30–39, 40–49, 50–59, and 60–.

A. Face Image Database

Age prediction systems are often used in public places such as shopping malls or train stations. In order to make our experiments realistic, we mainly use in-house facial image samples obtained from video sequences taken by surveillance cameras in addition to the public databases [1], [7], [14] for training predictors. Examples of face images are shown in Figure 3. The recording method, image resolution, face angle, lighting conditions, and the image size are diverse depending on the recording conditions. For example, some subjects were walking naturally, or seated on a stool and keeping their heads still. Our in-house database includes various races such as Caucasians, African Americans, and Asians. We used a face detector for localizing two eye-centers and escaled the image to 64×64 pixels. Then input face features such as the RGB color, x-gradient, and y-gradient are extracted from all the 64×64 pixel values.

For the preparation of test samples, we first selected 800 subjects uniformly over races, genders, and age brackets. The

TABLE I
THE NUMBER OF TRAINING SAMPLES.

(a) male

	younger (–19)	middle (20–49)	older (50–)	total
Caucasian	9,664	58,927	11,345	79,936
Asian	16,484	32,732	19,581	68,797
African American	7,596	42,165	3,258	53,019
Hispanic	7,228	11,927	3,202	22,357
Indian / Middle Eastern	11,720	22,296	3,793	37,809
Total	52,692	168,047	41,179	261,918

(b) female

	younger (–19)	middle (20–49)	older (50–)	total
Caucasian	11,311	42,889	11,655	65,855
Asian	17,719	42,426	18,531	78,676
African American	3,453	17,692	2,086	23,231
Hispanic	5,652	12,829	4,091	22,571
Indian / Middle Eastern	13,509	14,574	2,974	31,057
Total	51,644	130,410	39,337	221,391

TABLE II
THE NUMBER OF TEST SAMPLES.

(a) male

	younger (–19)	middle (20–49)	older (50–)	total
Caucasian	431	1,443	957	2,831
Asian	1,455	1,170	882	3,507
African American	523	1,274	947	2,744
Hispanic	816	1,219	796	2,831
Indian / Middle Eastern	740	1,057	670	2,467
Total	3,965	6,163	4,252	14,380

(b) female

	younger (–19)	middle (20–49)	older (50–)	total
Caucasian	408	1,138	821	2,367
Asian	1,242	1,646	997	3,885
African American	490	1,064	575	2,129
Hispanic	591	1,006	750	2,347
Indian / Middle Eastern	970	1,319	793	3,082
Total	3,701	6,173	3,936	13,810

breakdown of 800 subjects is 10 (persons) $\times$ 8 (age classes) $\times$ 2 (genders) $\times$ 5 (races). We classified the races into 5 categories: Caucasians, Asians, African Americans, Hispanics, and Indian/Middle Eastern.

The number of training and test samples for each gender, age class, and race is summarized in Table I and Table II, respectively Table I shows that the training samples are imbalanced over age classes; especially there are less samples for the 'younger' and 'older' classes than the 'middle' class.

B. Experimental Setup

First, the NNR and NNC models are trained using multi-race face images. To eliminate the imbalance in the number of training samples for each class, we artificially increased the number of samples for minor classes by considering perturbed samples[1] so that all classes have the same number of training samples.

[1]Several perturbation methods are used in our experiments, for instance, Gaussian filtering for blurring, making mirror-reversed images, and randomly changing the contrast and illumination.

TABLE III
SELECTED WEIGHT PARAMETERS FOR FOUR PREDICTORS.

	NNR	NNR+LSPC	NNC	NNC+LSPC
male	0.0	0.7	0.3	0.0
female	0.0	0.6	0.2	0.2

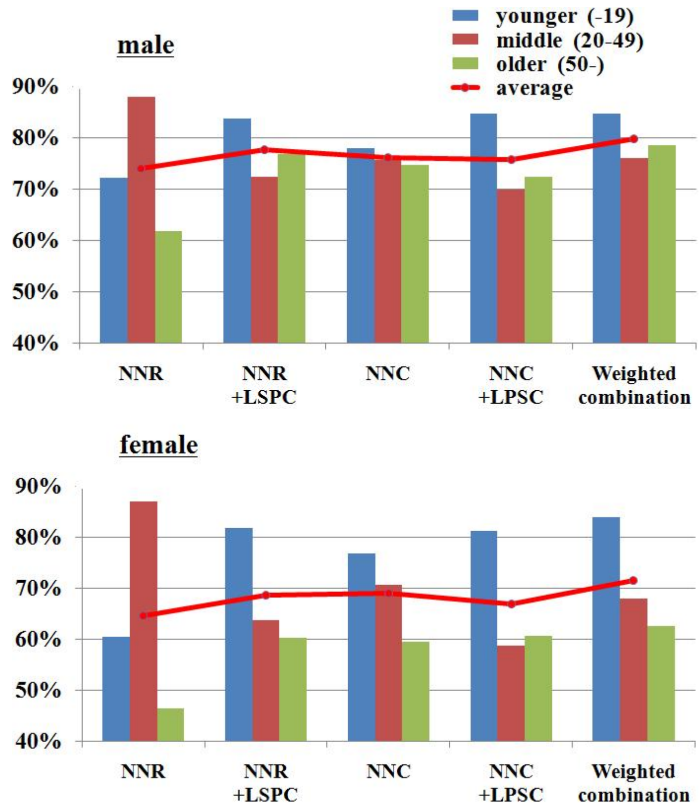

Fig. 4. Comparison of classification rates (upper: male, lower: female). 'Combination' denotes the combined predictor.

In our experiment, age predictors for male and female are separately trained. More precisely, male samples are only used for training age predictors for males, and female samples are only used for training age predictors for females. The test performance is evaluated under the assumption that gender is correctly estimated.

Next, using the hidden layer outputs of the trained NNR and NNC models as input feature vectors, classification models are trained by LSPC (NNR+LSPC and NNC+LSPC). To speed up the computation of LSPC, we use the delta kernel for class labels and limit the number of kernel bases to 2000 for each class, which are randomly chosen from all kernel bases.

C. Experimental Results

Experimental results of 4 different age predictors (NNR, NNR+LSPC, NNC, and NNC+LSPC), in addition to their linear combination are shown in Figure 4. We bind 8 classes into 3 classes (younger, middle, and older age) to make the exposition of experimental results concise.

The NNR method gives accurate age prediction for the 'middle' class, but it performs poorly in the 'younger' and 'older' classes. This tendency can be observed both in male and female cases. This is partially caused by the class imbalance, especially the number of younger and older samples is smaller than the middle age. On the other hand, NNR+LSPC has slightly lower classification accuracy for the 'middle' class, but performs better for the 'younger' and 'older' classes. Overall, NNR+LSPC compares favorably with the plain NNR. This result implies that LSPC is more robust against class imbalance. NNC is stably work well for all classes, while NNC+LSPC performs well for the 'younger' class.

The combined predictor tends to nicely compensate for the weaknesses of each method, and give stably good performance for all classes. Consequently, the overall classification accuracy is improved. Table III describes the optimized weight for each predictor, showing that NNR+LSPC dominates others both in the male and female cases.

Overall, the experimental results show the usefulness of the proposed combined age predictor.

V. Summary and Future Work

The multi-race age estimation is a challenging task, mainly because the growth process of each race is highly diverse and collecting data uniformly over various races and age brackets is practically difficult. In this paper, we proposed to combine neural networks and least-squares probabilistic classifiers for overcoming the above problems. We demonstrated the effectiveness of our proposed method through large-scale real-world age estimation experiments.

In the experiments, we assumed that the genders had been correctly classified and performed age prediction for each gender. We may apply the same idea also to races, i.e., race recognition is first carried out and then age is predicted. This strategy allows us to train age predictors in a race-wise manner. Given the strong diversity of the growth process of each race, this approach would be promising. In our future work, we will investigate this idea.

References

[1] *The FG-NET Aging Database*, http://www.fgnet.rsunit.com/

[2] Y. H. Kwon and N. V. Lobo, *Age classification from facial images*, Computer Vision and Image Understanding, vol.74, no.1, DPP.1–21, 1999.

[3] W. B. Horng, C. P. Lee, and C. W. Chen, *Classification of age groups based on facial features*, Tamkang Journal of Science and Engineering, vol.4, no.3, pp.183–192, 2001.

[4] A. Lanitis, C. J. Taylor, and T. F. Cootes, *Toward automatic simulation of aging effects on face images*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.24, no.4, pp.442–455, 2002.

[5] P. J. Phillips, P. J. Flynn, W. T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. J. Worek, *Overview of the face recognition grand challenge*, Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR2005). pp.947–954, 2005.

[6] X. Geng, Z. Zhou, Y. Zhang, G. Li, and H. Dai, *Learning from facial aging patterns for automatic age estimation*, Proceedings of the 14th ACM International Conference on Multimedia, pp.307–316, 2006.

[7] K. J. Ricanek and T. Tesafaye, *Morph: A longitudinal image database of normal adult age progression*, Proceedings of the IEEE 7th International Conference on Automatic Face and Gesture Recognition (FGR2006), pp.341–345, 2006.

[8] M.Sugiyama, *Superfast-trainable multiclass probabilistic classifier by least-squares posterior fitting*, IEICE Transactions on Information and Systems, vol.E93-D, no.10, pp.2690–2701, 2010.

[9] F. H. C. Tivive and A. Bouzerdoum, *A gender recognition system using shunting inhibitory convolutional neural networks*, Proceedings of the International Joint Conference on Neural Networks (IJCNN2006), pp.5336–5341, 2006.

[10] F. H. C. Tivive and A. Bouzerdoum, *A shunting inhibitory convolutional neural network for gender classification*, Proceedings of the 18th International Conference on Pattern Recognition (ICPR2006), vol.4, pp.421–424, 2006.

[11] Y. Fu, Y. Xu, and T. S. Huang, *Estimating human age by manifold analysis of face pictures and regression on aging features*, Proceedings of the IEEE Multimedia and Expo. pp.1383–1386, 2007.

[12] K. Ueki, M. Miya, T. Ogawa, and T. Kobayashi, *Class distance weighted locality preserving projection for automatic age estimation*, Proceedings of the IEEE international conference on Biometrics: Theory, Applications and Systems (BTAS2008), 2008.

[13] G. Guo, G. Mu, Y. Fu, C. Dyer, and T. Huang, *A study on automatic age estimation using a large database*, International Conference on Computer Vision in Kyoto (ICCV2009), pp.1986–1991, 2009.

[14] N. C. Ebner, M. Riediger, and U. Lindenberger, *FACES - A database of facial expressions in young, middle-aged, and older women and men: Development and validation*, Behavior Research Methods, vol.42, no.1, pp.351–362, 2010.

[15] K. Ueki, M. Sugiyama, and Y. Ihara, *Semi-supervised estimation of perceived age from face images*, Proceedings of International Conference on Computer Vision Theory and Applications (VISAPP2010), pp.319-324, 2010.