Perceived Age Estimation under Lighting Condition Change by Covariate Shift Adaptation

Kazuya Ueki
*VALWAY Technology Center, NEC Soft, Ltd.
Tokyo, Japan
ueki@mxf.nes.nec.co.jp*

Masashi Sugiyama
*Depertment of Computer Science, Tokyo Institute of Technology
Tokyo, Japan
sugi@cs.titech.ac.jp*

Yasuyuki Ihara
*VALWAY Technology Center, NEC Soft, Ltd.
Tokyo, Japan
ihara@mxk.nes.nec.co.jp*

***Abstract*—Over the recent years, a great deal of effort has been made to age estimation from face images. It has been reported that age can be accurately estimated under controlled environment such as frontal faces, no expression, and static lighting conditions. However, it is not straightforward to achieve the same accuracy level in real-world environment because of considerable variations in camera settings, facial poses, and illumination conditions. In this paper, we apply a recently-proposed machine learning technique called *covariate shift adaptation* to alleviating lighting condition change between laboratory and practical environment. Through real-world age estimation experiments, we demonstrate the usefulness of our proposed method.**

***Keywords*-face recognition; age estimation; covariate shift adaptation; lighting condition change; Kullback-Leibler importance estimation procedure; importance-weighted regularized least-squares**

I. Introduction

In recent years, demographic analysis in public places such as shopping malls and stations is attracting a great deal of attention. Such demographic information is useful for various purposes such as designing effective marketing strategies and targeted advertisement based on customers' gender and age. For this reason, a number of approaches have been explored for age estimation from face images [2], [3], [4], and several databases became publicly available recently [1], [5], [6].

The recognition performance of age prediction systems is significantly influenced, e.g., by the type of camera, camera calibration, and lighting variations, and the publicly available databases were mainly collected in semi-controlled environment. For this reason, existing age prediction systems built upon such databases tend to perform poorly in real-world environment.

The situation where training and test data are drawn from different distributions is called *covariate shift* [7], [8], [9]. In this paper, we formulate the problem of age estimation in real-world environment as a supervised learning problem under covariate shift. Within the covariate shift framework, a method called *importance-weighted least-squares* allows us to alleviate the influence of environmental changes, by assigning higher weights to data samples having high test input densities and low training input densities. We demonstrate through real-world experiments that age estimation based on covariate shift adaptation achieves higher accuracy than baseline approaches.

II. Proposed Method

In this section, we describe our proposed method for age prediction.

A. Formulation

Throughout this paper, we perform age estimation based not on subjects' real age, but on their *perceived* age. Thus, the 'true' age of the subject y is defined as the average perceived age evaluated by those who observed the subject's face images (the value is rounded-off to the nearest integer).

Let us consider a regression problem of estimating the age y^* of subject $\boldsymbol{x}$ (face features). We use the following model for regression.

$$f(\boldsymbol{x};\boldsymbol{\alpha}) = \sum_{i=1}^{n_{tr}} \alpha_i K(\boldsymbol{x}, \boldsymbol{x}_i^{tr}), \tag{1}$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{n_{tr}})^\top$ is a model parameter and $K(\boldsymbol{x}, \boldsymbol{x}')$ is a positive definite kernel.

Suppose we are given labeled training data $\{(\boldsymbol{x}_i^{tr}, y_i^{tr})\}_{i=1}^{n_{tr}}$. A standard approach to learning the model parameter $\boldsymbol{\alpha}$ would be *regularized least-squares*.

$$\min_{\boldsymbol{\alpha}} \left[\frac{1}{n_{tr}} \sum_{i=1}^{n_{tr}} (y_i^{tr} - f(\boldsymbol{x}_i^{tr};\boldsymbol{\alpha}))^2 + \lambda \|\boldsymbol{\alpha}\|^2 \right], \tag{2}$$

where $\|\cdot\|$ denotes the Euclidean norm and $\lambda (> 0)$ is the regularization parameter to avoid overfitting.

Below, we explain that merely using regularized least-squares is not appropriate in real-world perceived age prediction and show how to cope with this problem.

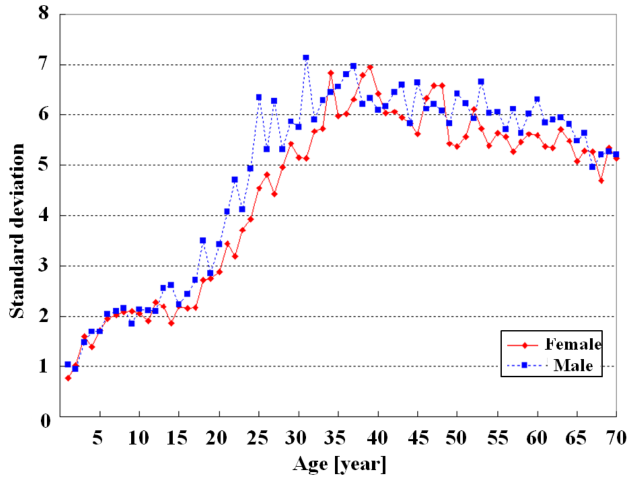

Figure 1. The relation between subjects' perceived age y^* (horizontal axis) and its standard deviation (vertical axis)

B. Incorporating Age Perception Characteristics

Human age perception is known to have heterogeneous characteristics, e.g., it is rare to misregard the age of a 5-year-old child as 15 years old, but the age of a 35-year-old person is often misregarded as 45 years old. In order to quantify this phenomenon, we have carried out a large-scale questionnaire survey: we asked each of 72 volunteers to give age labels y to approximately 1000 face images. Figure 1 depicts the relation between subjects' perceived age y^* and its standard deviation. This shows that the perceived age deviation tends to be small in younger age brackets and large in older age brackets.

In order to match characteristics of our age prediction system to those of human age perception, we weight the goodness-of-fit term in Eq.(2) according to the inverse variance of the perceived age:

$$\min_{\boldsymbol{\alpha}} \left[\frac{1}{n_{tr}} \sum_{i=1}^{n_{tr}} \frac{(y_i^{tr} - f(\boldsymbol{x}_i^{tr}; \boldsymbol{\alpha}))^2}{w_{age}(y_i^{tr})^2} + \lambda \|\boldsymbol{\alpha}\|^2 \right], \tag{3}$$

where $w_{age}(y)$ is the standard deviation of the perceived age (see Figure 1 again).

C. Coping with Lighting Condition Change

When designing age estimation systems, the environment of recording training face images is often different from the test environment in terms of lighting conditions. Typically, training data are recorded indoors such as a studio with appropriate illumination, whereas in real-world environment, lighting conditions have considerable varieties, e.g., strong sunlight might be cast from a side of faces or there is no enough light. In such situations, age estimation accuracy is significantly degraded.

Let $p_{tr}(\boldsymbol{x})$ be the training input density and $p_{te}(\boldsymbol{x})$ be the test input density. When these two densities are different, it would be natural to emphasize the influence of training samples $(\boldsymbol{x}_i^{tr}, y_i^{tr})$ which have high similarity to data in the test environment. Such an adjustment can be systematically carried out by weighting the goodness-of-fit term in Eq.(3) according to the *importance function* $w_{imp}(\boldsymbol{x}) = p_{te}(\boldsymbol{x})/p_{tr}(\boldsymbol{x})$ [7], [8], [9].

$$\min_{\boldsymbol{\alpha}} \left[\frac{1}{n_{tr}} \sum_{i=1}^{n_{tr}} w_{imp}(\boldsymbol{x}_i^{tr}) \frac{(y_i^{tr} - f(\boldsymbol{x}_i^{tr}; \boldsymbol{\alpha}))^2}{w_{age}(y_i^{tr})^2} + \lambda \|\boldsymbol{\alpha}\|^2 \right].$$

The solution can be obtained analytically by

$$\widehat{\boldsymbol{\alpha}} = (\boldsymbol{K}^{tr} \boldsymbol{W}^{tr} \boldsymbol{K}^{tr} + n_{tr} \lambda \boldsymbol{I}_{n_{tr}})^{-1} \boldsymbol{K}^{tr} \boldsymbol{W}^{tr} \boldsymbol{y}^{tr}, \tag{4}$$

where $\boldsymbol{K}^{tr}$ is the kernel matrix whose (i, i')-th element is defined by $K(\boldsymbol{x}_i^{tr}, \boldsymbol{x}_{i'}^{tr})$, $\boldsymbol{W}^{tr}$ is the n_{tr}-dimensional diagonal matrix with (i, i)-th diagonal element defined by $w_{imp}(\boldsymbol{x}_i^{tr})/w_{age}(y_i^{tr})^2$, $\boldsymbol{I}_{n_{tr}}$ is the n_{tr}-dimensional identity matrix, and $\boldsymbol{y}^{tr}$ is the n_{tr}-dimensional vector with i-th element defined by y_i^{tr}.

When the number of training data n_{tr} is large, we may reduce the number of kernel in Eq.(1) so that the inverse matrix in Eq.(4) can be computed with the limited memory; or we may compute the solution numerically by a stochastic gradient-decent method.

D. Kullback-Leibler Importance Estimation Procedure (KLIEP)

In order to compute the solution (4), we need the importance weights $w_{imp}(\boldsymbol{x}_i^{tr}) = p_{te}(\boldsymbol{x}_i^{tr})/p_{tr}(\boldsymbol{x}_i^{tr})$. However, since density estimation is a hard problem, a two-stage approach of first estimating $p_{tr}(\boldsymbol{x})$ and $p_{te}(\boldsymbol{x})$ and then taking their ratio may not be reliable. Here we describe a method called *Kullback-Leibler Importance Estimation Procedure* (KLIEP) [9], which allows us to directly estimate the importance function $w_{imp}(\boldsymbol{x})$ without going through density estimation of $p_{tr}(\boldsymbol{x})$ and $p_{te}(\boldsymbol{x})$.

Let us model $w_{imp}(\boldsymbol{x})$ using the following model:

$$\widehat{w}_{imp}(\boldsymbol{x}) = \sum_{k=1}^{b} \beta_k \exp\left(-\frac{\|\boldsymbol{c}_k - \boldsymbol{x}\|^2}{2\gamma^2} \right), \tag{5}$$

Table I
PSEUDO CODE OF KLIEP. './' INDICATES THE ELEMENT-WISE DIVISION. INEQUALITIES AND THE 'MAX' OPERATION FOR VECTORS ARE APPLIED IN AN ELEMENT-WISE MANNER.

Input: $\{\boldsymbol{x}_i^{tr}\}_{i=1}^{n_{tr}}, \{\boldsymbol{x}_j^{te}\}_{j=1}^{n_{te}}$
Output: $\widehat{w}(\boldsymbol{x})$

Choose $\{\boldsymbol{c}_k\}_{k=1}^{b}$ as a subset of $\{\boldsymbol{x}_j^{te}\}_{j=1}^{n_{te}}$;
$A_{j,k} \leftarrow \exp\left(-\|\boldsymbol{c}_k - \boldsymbol{x}_j^{te}\|^2/(2\gamma^2)\right)$;
$b_k \leftarrow \frac{1}{n_{tr}} \sum_{i=1}^{n_{tr}} \exp\left(-\|\boldsymbol{c}_k - \boldsymbol{x}_i^{tr}\|^2/(2\gamma^2)\right)$;
Initialize $\boldsymbol{\beta}(> \mathbf{0})$ and ε $(0 < \varepsilon \ll 1)$;
Repeat until convergence
  $\boldsymbol{\beta} \leftarrow \varepsilon A^\top (\mathbf{1}./A\boldsymbol{\beta})$;
  $\boldsymbol{\beta} \leftarrow \boldsymbol{\beta} + (1 - \boldsymbol{b}^\top \boldsymbol{\beta})\boldsymbol{b}/(\boldsymbol{b}^\top \boldsymbol{b})$;
  $\boldsymbol{\beta} \leftarrow \max(\mathbf{0}, \boldsymbol{\beta})$;
  $\boldsymbol{\beta} \leftarrow \boldsymbol{\beta}/(\boldsymbol{b}^\top \boldsymbol{\beta})$;
end

Figure 2. Examples of face images under different lighting conditions (left: standard lighting, middle: dark, right: strong light from the side)

where $\{\boldsymbol{c}_k\}_{k=1}^{b}$ is a subset of test input samples $\{\boldsymbol{x}_j^{te}\}_{j=1}^{n_{te}}$. Using the model $\widehat{w}_{imp}(\boldsymbol{x})$, we can estimate the test input density $p_{te}(\boldsymbol{x})$ by

$$\widehat{p}_{te}(\boldsymbol{x}) = \widehat{w}_{imp}(\boldsymbol{x}) p_{tr}(\boldsymbol{x}). \tag{6}$$

We determine the parameters $\boldsymbol{\beta} = \{\beta_k\}_k^b$ in the model (6) so that the Kullback-Leibler divergence from p_{te} to $\widehat{p}_{te}$ is minimized:

$$KL(p_{te} \| \widehat{p}_{te}) = \int p_{te}(\boldsymbol{x}) \log \frac{p_{te}(\boldsymbol{x})}{\widehat{p}_{te}(\boldsymbol{x})} d\boldsymbol{x}$$
$$= \int p_{te}(\boldsymbol{x}) \log \frac{p_{te}(\boldsymbol{x})}{p_{tr}(\boldsymbol{x})} d\boldsymbol{x} - \int p_{te}(\boldsymbol{x}) \log \widehat{w}_{imp}(\boldsymbol{x}) d\boldsymbol{x}.$$

We ignore the first term (which is a constant) and impose $\widehat{w}_{imp}(\boldsymbol{x})$ to be non-negative and normalized. Then we obtain the following convex optimization problem:

$$\max_{\boldsymbol{\beta}} \left[\sum_{j=1}^{n_{te}} \log \left(\sum_{k=1}^{b} \beta_k \exp \left(-\frac{\|\boldsymbol{c}_k - \boldsymbol{x}_j^{te}\|^2}{2\gamma^2} \right) \right) \right],$$
$$\text{s.t.} \begin{cases} \beta_k \geq 0 \quad \text{for } k = 1, \ldots, b, \\ \dfrac{1}{n_{tr}} \displaystyle\sum_{i=1}^{n_{tr}} \left(\sum_{k=1}^{b} \beta_k \exp \left(-\frac{\|\boldsymbol{c}_k - \boldsymbol{x}_i^{tr}\|^2}{2\gamma^2} \right) \right) = 1. \end{cases}$$

A pseudo code of KLIEP is described in Table I. The tuning parameter γ can be optimized based on cross-validation [9].

III. Empirical Evaluation

In this section, we apply the proposed age prediction method to in-house face-age datasets and experimentally evaluate its performance.

We use the face images recorded under 17 different lighting conditions: for instance, average illuminance from above is approximately 1000 lux and 500 lux from the front in the standard lighting condition, 250 lux from above and 125 lux from the front in the dark setting, and 190 lux from left and 750 lux from right in another setting (see Figure 2). Images were recorded as movies with camera at depression angle 15 degrees. The number of subjects is approximately 500 (250 for each gender). We used a face detector for localizing the two eye-centers, and then rescaled the image to 64 $\times$ 64 pixels. The number of face images in each environment is about 2500 (5 face images $\times$ 500 subjects).

As pre-processing, a neural network feature extractor [10] was used to extract 100-dimensional features from 64 $\times$ 64 face images. The kernel regression model (1) with Gaussian kernel was employed for the extracted 100-dimensional data:

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

We constructed the male/female age prediction models only using male/female data, assuming that gender classification had been correctly carried out.

We split the 250 subjects into the *training set* (200 subjects) and the *test set* (50 subjects). The training set was used for training the kernel regression model (1), and the test set was used for evaluating its generalization performance. For the test samples $\{(\boldsymbol{x}_i^{te}, y_i^{te})\}_{i=1}^{n_{te}}$ taken from the test set in the environment with strong light from a side, age-weighted mean square error (WMSE)

$$\text{WMSE} = \frac{1}{n_{te}} \sum_{i=1}^{n_{te}} \frac{(y_i^{te} - f(\boldsymbol{x}_i^{te}; \widehat{\boldsymbol{\alpha}}))^2}{w_{age}(y_i^{te})^2}$$

was calculated as a performance measure. The training set and the test set were shuffled 5 times in such a way that each subject was selected as a test sample once. The final performance was evaluated based on the average WMSE over the 5 trials.

We compared the performance of the proposed method with the two baseline methods:

Baseline method 1: Training samples were taken only from the standard lighting condition and age-weighted regularized least-squares (3) was used for training.

Baseline method 2: Training samples were taken from all 17 different lighting conditions and age-weighted regularized least-squares (3) was used for training.

The importance weights were not used in these baseline methods. The Gaussian width σ and the regularization parameter λ were determined based on 4-fold cross-validation over WMSE, i.e., the training set was further divided into a training part (150 subjects) and a validation part (50 subjects).

In the proposed method, training samples were taken from all 17 different lighting conditions (which is the same as the baseline method 2). The importance weights were estimated by KLIEP using the training samples and additional *unlabeled* test samples; the hyper-parameter γ in KLIEP was determined based on 2-fold cross-validation [9]. We then computed the average importance score over different samples for each lighting condition and used the average importance score for training the regression model. The Gaussian width σ and the regularization parameter λ were determined based on 4-fold importance-weighted cross-validation over WMSE [8].

Table II
THE TEST PERFORMANCE MEASURED BY WMSE.

	Male	Female
Baseline method 1	2.83	6.51
Baseline method 2	2.64	4.40
Proposed method	**2.54**	**3.90**

Table II summarizes the experimental results, showing that, for both female and male data, the baseline method 2 is better than the baseline method 1 and the proposed method is better than the baseline method 2. This illustrates the effectiveness of the proposed method.

IV. SUMMARY AND FUTURE WORKS

Lighting condition change is one of the critical causes of performance degradation in age prediction from face images. In this paper, we proposed to employ a recently-proposed machine learning technique called *covariate shift adaptation* for alleviating the influence of lighting condition change. We demonstrated the effectiveness of our proposed method through real-world perceived age prediction experiments.

In principle, the covariate shift framework allows us to incorporate not only lighting condition change, but also various types of environment change such as face pose variation and camera setting change. In our future work, we will investigate whether the proposed approach is still useful in such challenging scenarios.

REFERENCES

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

[2] 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*, pages 1383–1386, 2007.

[3] 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*, pages 307–316, 2006.

[4] 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 (ICCV 2009)*, pages 1986–1991, 2009.

[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 (CVPR 2005)*, pages 947–954, 2005.

[6] 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 (FGR '06)*, pages 341–345, 2006.

[7] H. Shimodaira. Improving predictive inference under covariate shift by weighting the log-likelihood function. *Journal of Statistical Planning and Inference*, 90(2):227–244, 2000.

[8] M. Sugiyama, M. Krauledat, and K.-R. Müller. Covariate shift adaptation by importance weighted cross validation. *Journal of Machine Learning Research*, 8:985–1005, May 2007.

[9] M. Sugiyama, T. Suzuki, T., S. Nakajima, H. Kashima, P. von Bünau, and M. Kawanabe. Direct importance estimation for covariate shift adaptation. *Annals of the Institute of Statistical Mathematics*, 60(4):699–746, 2008.

[10] 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 (IJCNN '06)*, pages 5336–5341, 2006.