

Research Article

Multifeature Extreme Ordinal Ranking Machine for Facial Age Estimation

Wei Zhao, Han Wang, and Guang-Bin Huang

School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore 639798

Correspondence should be addressed to Wei Zhao; zhao0183@e.ntu.edu.sg

Received 12 May 2015; Accepted 23 August 2015

Academic Editor: Huaguang Zhang

Copyright © 2015 Wei Zhao et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Recently the state-of-the-art facial age estimation methods are almost originated from solving complicated mathematical optimization problems and thus consume huge quantities of time in the training process. To refrain from such algorithm complexity while maintaining a high estimation accuracy, we propose a multifeature extreme ordinal ranking machine (MFEORM) for facial age estimation. Experimental results clearly demonstrate that the proposed approach can sharply reduce the runtime (even up to nearly one hundred times faster) while achieving comparable or better estimation performances than the state-of-the-art approaches. The inner properties of MFEORM are further explored with more advantages.

1. Introduction

With the rapid development of computer vision, pattern recognition, and biometrics, more and more attention has been paid to computer-based human facial age estimation, which will be utilized in the scenarios where an individual's age needs to be obtained without specifically identifying other irrelevant personal information, such as electronic customer information management [1, 2], human-computer interaction (HCI) [3], security surveillance monitoring [4, 5], age-based visual advertisement, and even entertainment.

Unlike other face-oriented problems, the difficulties of computer-based facial age estimation are reflected in the following aspects:

- (1) Difference of aging process: different people have their own living environment, ethnic group, gender, lifestyle, social contact, health condition, and even gene diversity, which all together determine the speed of aging.
- (2) Shape or texture: different forms of aging will emerge in different age levels. For example, from infancy to adolescence, the craniofacial growth (shape growth) is the main change. However, from adult period to

old age, the craniofacial change decreases remarkably and skin transformation (texture change) would be the most prominent change.

- (3) Data insufficiency: we only have a very limited number of aging datasets, especially which can cover all the age range.
- (4) Disturbance: some females tend to show their younger face, so final estimation results will be largely interfered by using cosmetics and accessories.

A lot of facial age estimation approaches have been put forward, some of which are able to obtain rather satisfying performance. Among them, most of the traditional approaches formulate facial age estimation problem based on classification [6–9], regression [4, 10–12], or combination of the two. Suppose we have a dataset of N training samples, $\{(x_i, y_i) \mid i = 1, \dots, N\}$, in which x_i represents the i th face image and y_i represents the corresponding age label. In multiclass classification, every sample will be regarded as a single independent age label for training; as a result, we get a multiclassifier to estimate a person's age. But the trouble is that the age labels have no relationship with each other; that is, each age label is only treated as a separate entity in the training process while, in essence, human's age labels are

sequential. So this kind of multiclassification method may omit some connotative information of the correlation among different age labels, which together compose the fine-ordered age set. For instance, two images with adjacent age labels for the same person will be more similar than those with far-apart labels. In short, multiclass classification cannot take full advantage of the correlation among ordinal age labels. In contrast, the regression method aims to find the best mapping from raw images to the corresponding ages and get a function for age estimation. However, craniofacial and skin changes in different age levels would result in unstable random process in feature space, so the kernels used to assess the similarities among different ages could be drifted. As for the estimation performance, it has been shown in the literature [4, 11, 13] that when different datasets are used for training and testing, the regression method will show better or worse results than the classification-based method. In addition, Guo et al. [4, 14] proposed a hybrid method which combines classification and regression approaches together to make use of both advantages. As a result, the actual performance is further improved to some degree.

In order to overcome the aforementioned defects of classification and regression approaches, ordinal hyperplanes ranker (OHRank) [15] based age estimation has been proposed. As everyone knows, the aging process is diversified for different age levels. As an analogy, the aging process from 22 to 25 would have a different tendency compared to that from 62 to 65. So it is more credible to compare two age labels' relative sequences (smaller or larger) than to compare the differences among labels. In spite of all the above merits, OHRank only utilizes a single-feature set as the feature representation model, so it fails to synthetically include all discriminative information from all available feature sets. More specifically, each feature set has its own advantages over others. For example, the anthropometry models [16] mathematically model the growing of people's head from babyhood to adulthood, so it reveals some information of the face's size and proportion; Active Appearance Models (AAM) [17] can represent both shape and texture information instead of only facial geometry; age manifold [18] learns the age tendency from different face images at every age label. So OHRank cannot integrate these pieces of discriminative information of different feature sets. Later Weng et al. [19] present a multifeature ordinal ranking (MFOR) method to utilize multifeature sets simultaneously, so the feature information's discriminative power is further reinforced. Experiment results demonstrate that MFOR outcompetes other age estimation approaches. However, almost all these approaches including MFOR largely rely on the complicated mathematical optimization solution. Take OHRank and MFOR as an example: these two ordinal ranking-oriented methods have top performance so far but they are all constructed under SVM-based formulation, so the SVM parameters must be computed iteratively by working out complicated and time-consuming optimization problems. As a result, calculation complexity would be a heavy burden for improving efficiency.

Recognizing this point, we propose a multifeature extreme ordinal ranking machine (MFEORM) for facial age estimation. Basically, we divide our approaches into three

stages: (1) representing features using certain feature extraction models, (2) processing the obtained feature sets, and (3) applying certain algorithms to estimate age. On the first stage, we use multiple feature models parallelly to represent our facial image database. For the second stage, since it is more logical and reasonable to distinguish which is older/younger between two facial images than to directly predict the age from images, the more reliable "larger or smaller than" information is used for one binary classification at each age. In this case, the abstract age estimation problem can be downgraded into $K - 1$ binary classification subproblems where K represents the number of total age labels. For the third stage, an ultrafast extreme learning machine (ELM) with kernel function is applied to get a series of classifiers. These classifiers are then integrated together according to a certain rule which will be illustrated in the following. Experiment results explicitly demonstrate that our MFEORM is able to notably reduce the runtime (even up to nearly one hundred times faster) while achieving similar or even better estimation results against state-of-the-art methods.

To sum up, the following contributions are made in this paper:

- (1) Multifeature extreme ordinal ranking machine (MFEORM) for facial age estimation is proposed, which combines the advantages of multifeature space, age's natural characteristics of ordinal information, and extreme learning machine's rapid learning rate while achieving similar or even better performances with much less time compared to state-of-the-art methods. Our approach avoids tediously conducting the iterative computation for mathematical optimization problem and improves efficiency.
- (2) The experiments are conducted comprehensively and thoroughly from the internal and external aspects, respectively, and find out more about MFEORM's particular characteristics and advantages.
- (3) Further properties are explored: (a) the influence of different number of feature models on the final results and (b) the influence of number of dimensions (after PCA dimension reduction) on the final results.

The rest of this paper is organized as follows: firstly, briefs of previous ordinal ranking-oriented age estimation solutions and extreme learning machine will be reviewed in Section 2. Then our proposed method will be detailed in Section 3. After that, experiment results and remarks are reported in Section 4. Finally, Section 5 concludes the paper.

2. Extreme Learning Machine (ELM)

All through the years, conventional learning techniques like support vector machines (SVMs) and neural networks have been suffering from the following: (1) slow training and learning speed, (2) human involvement, and (3) unsatisfactory generalization performance. However, extreme learning machine (ELM) [20–22], which recently draws more and more attention, conquers these drawbacks and gets a satisfying performance to a certain extent. Primarily, ELM is

based on generalized single-hidden layer feedforward neural networks (SLFNs). ELM has the following advantages:

- (1) In ELM, all hidden layer parameters of SLFNs do not need to be tuned and do not rely on training samples. They only need to be randomly generated and reduce human intervention.
- (2) ELM has much faster speed and more superior generalization performance.

Let us start from the structure of SLFNs. Figure 1 shows a typical SLFN's construction. Generally speaking, SLFN can be described as

$$f(\mathbf{x}) = \sum_{i=1}^L \mu_i p_i(\mathbf{x}) = \sum_{i=1}^L \mu_i P(\mathbf{a}_i, b_i, \mathbf{x}), \quad (1)$$

$$\mu_i \in \mathbf{R}^v, \quad \mathbf{x} \in \mathbf{R}^w,$$

where μ_i represents the output weight between the i th neuron and the output node, $p_i(\mathbf{x})$ is the output function $P(\mathbf{a}_i, b_i, \mathbf{x})$ of the i th neuron, \mathbf{a}_i is the weight vector linking to the i th node, and b_i is the bias of the i th node. Particularly, we have

$$p_i(\mathbf{x}) = P(\mathbf{a}_i, b_i, \mathbf{x}) = p(\mathbf{a}_i \cdot \mathbf{x} + b_i) \quad \text{for additive nodes} \quad (2)$$

$$p_i(\mathbf{x}) = P(\mathbf{a}_i, b_i, \mathbf{x}) = p(b_i - \|\mathbf{x} - \mathbf{a}_i\|) \quad \text{for RBF nodes.}$$

Many researchers have built the theoretical foundation [23–25] that SLFN is able to learn N arbitrary distinguishing samples with zero error provided this SLFN has any bounded nonlinear activation function and N hidden neurons at most. More precisely, suppose we have training samples $(\mathbf{x}_i, \mathbf{t}_i)$, where $\mathbf{x}_i \in \mathbf{R}^w$, $\mathbf{t}_i \in \mathbf{R}^v$. Also, we let

$$\mathbf{o}_j = \sum_{i=1}^Q \mu_i p_i(\mathbf{x}_j) = \sum_{i=1}^Q \mu_i P(\mathbf{a}_i, b_i, \mathbf{x}_j), \quad j = 1, \dots, N. \quad (3)$$

Then the abovementioned “zero error” means

$$\sum_{j=1}^Q \|\mathbf{o}_j - \mathbf{t}_j\| = 0. \quad (4)$$

In other words, we can find a combination of μ_i and (\mathbf{a}_i, b_i) such that

$$\sum_{i=1}^Q \mu_i P(\mathbf{a}_i, b_i, \mathbf{x}_j) = \mathbf{t}_j, \quad j = 1, \dots, N. \quad (5)$$

To make it concise, we formulate (5) as

$$\mathbf{U}\boldsymbol{\mu} = \mathbf{T}, \quad (6)$$

where

$$\mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix} = \begin{bmatrix} t_{11} & \cdots & t_{1m} \\ \vdots & \vdots & \vdots \\ t_{N1} & \cdots & t_{Nm} \end{bmatrix},$$

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}(x_1) \\ \vdots \\ \mathbf{u}(x_N) \end{bmatrix} = \begin{bmatrix} P(\mathbf{a}_1, b_1, \mathbf{x}_1) & \cdots & P(\mathbf{a}_Q, b_Q, \mathbf{x}_1) \\ \vdots & \cdots & \vdots \\ P(\mathbf{a}_1, b_1, \mathbf{x}_N) & \cdots & P(\mathbf{a}_Q, b_Q, \mathbf{x}_N) \end{bmatrix} \quad \text{(hidden layer feature mapping),} \quad (7)$$

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1^T \\ \vdots \\ \mu_Q^T \end{bmatrix} \quad \text{(output weight vectors).}$$

In accordance with the above traditional theories of neural network, only when all the hidden layer parameters (e.g., μ_i and (\mathbf{a}_i, b_i)) are adjustable can SLFNs work as the universal approximators. In order to minimize the workload of tuning these parameters, researchers proposed incremental approaches in which parameters of existing hidden neurons remain unchanged while those of newly added neurons will be tuned and then fixed. In all the above methods, the hidden layer parameters need to be adjusted for greater than or equal to once. However, this situation will not happen in

extreme learning machine. For ELM, all the hidden layer parameters do not need to be iteratively tuned; instead, they are randomly selected and independent from training samples. In theory, Huang et al. [26] have proved that, for the purpose of letting SLFNs serve as universal approximators, we can randomly choose the hidden layer parameters and analytically determine the output weight vectors connecting the output layer and hidden layer. In this case, for additive nodes, activation functions can be arbitrary bounded non-constant continuous piecewise functions $g: \mathbf{R} \rightarrow \mathbf{R}$; for RBF

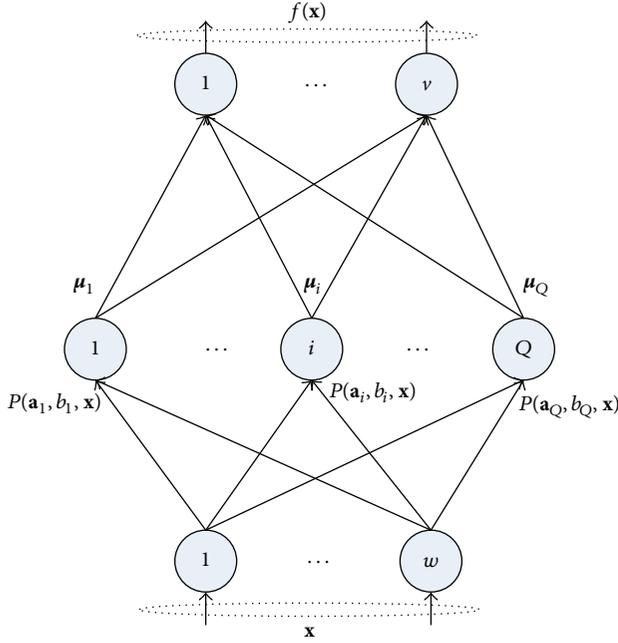


FIGURE 1: A typical example of single-layer feedforward neural network (SLFN).

nodes activation functions can be any arbitrary integrable continuous piecewise functions $g: R \rightarrow R$ and $\int_R g(x)dx \neq 0$. Furthermore, Huang et al. [20] demonstrate that, as long as activation functions are infinitely differentiable, the hidden layer parameters can be randomly selected. After selection, the hidden neuron parameters (\mathbf{a}_i, b_i) will be kept unchanged. Up to now, the sole unknown parameters are output weights μ_i connecting hidden layer and output layer. So the least square method can be used to calculate the output weight vector $\boldsymbol{\mu}$.

In accordance with the theory of feedforward neural network [27], ELM can be mathematically formulated by

$$\begin{aligned} \text{Minimize: } & \|\mathbf{U}\boldsymbol{\mu} - \mathbf{T}\| \text{ (training error),} \\ \text{minimize: } & \|\boldsymbol{\mu}\| \text{ (norm of weights).} \end{aligned} \quad (8)$$

Therefore, training such SLFN is the same as seeking out a least-squares solution of this $\mathbf{U}\boldsymbol{\mu} = \mathbf{T}$ linear system. Suppose that this optimal solution is $\hat{\boldsymbol{\mu}}$, so we have

$$\|\mathbf{U}\hat{\boldsymbol{\mu}} - \mathbf{T}\| = \min_{\boldsymbol{\mu}} \|\mathbf{U}\boldsymbol{\mu} - \mathbf{T}\|. \quad (9)$$

According to the mathematical theory of matrices [28], the smallest norm least-squares solution for the above system is $\hat{\boldsymbol{\mu}} = \mathbf{U}^\dagger \mathbf{T}$, where \mathbf{U}^\dagger represents the Moore-Penrose generalized inverse of \mathbf{U} . In short, ELM can be summarized by the following steps:

- (1) Randomly selecting the hidden layer parameters, that is, the input weight vector \mathbf{a}_i and the hidden layer biases b_i , where $i = 1, \dots, L$ (number of hidden neurons).

- (2) Calculating the hidden layer output, namely, matrix \mathbf{U} .
- (3) Calculating the output weight, namely, vector $\boldsymbol{\mu}$: $\boldsymbol{\mu} = \mathbf{U}^\dagger \mathbf{T}$.

When calculating the Moore-Penrose pseudoinverse \mathbf{U}^\dagger of matrix \mathbf{U} , the orthogonal projection approach can be used with efficiency which can be described as follows:

- (1) When $\mathbf{U}^T \mathbf{U}$ is nonsingular, $\mathbf{U}^\dagger = (\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T$;
- (2) When $\mathbf{U} \mathbf{U}^T$ is nonsingular, $\mathbf{U}^\dagger = \mathbf{U}^T (\mathbf{U} \mathbf{U}^T)^{-1}$.

Furthermore, in order to obtain more stable solution and better generalization performance, Huang et al. [22] added a positive value to the diagonal of $\mathbf{U} \mathbf{U}^T$ or $\mathbf{U}^T \mathbf{U}$ during the calculation of \mathbf{U}^\dagger . In this way, $\mathbf{U}^\dagger = \mathbf{U}^T (\mathbf{I}/C + \mathbf{U} \mathbf{U}^T)^{-1}$ (when $\mathbf{U} \mathbf{U}^T$ is nonsingular) or $\mathbf{U}^\dagger = (\mathbf{I}/C + \mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T$ (when $\mathbf{U}^T \mathbf{U}$ is nonsingular). Consequently, according to $\boldsymbol{\mu} = \mathbf{U}^\dagger \mathbf{T}$ and the output function $f(\mathbf{x}) = \mathbf{u}(\mathbf{x}) \boldsymbol{\mu}$, we have

$$f(\mathbf{x}) = \mathbf{u}(\mathbf{x}) \boldsymbol{\mu} = \mathbf{u}(\mathbf{x}) \mathbf{U}^T \left(\frac{\mathbf{I}}{C} + \mathbf{U} \mathbf{U}^T \right)^{-1} \mathbf{T} \quad (10)$$

with the corresponding $\boldsymbol{\mu} = \mathbf{U}^T (\mathbf{I}/C + \mathbf{U} \mathbf{U}^T)^{-1} \mathbf{T}$ or

$$f(\mathbf{x}) = \mathbf{u}(\mathbf{x}) \boldsymbol{\mu} = \mathbf{u}(\mathbf{x}) \left(\frac{\mathbf{I}}{C} + \mathbf{U}^T \mathbf{U} \right)^{-1} \mathbf{U}^T \mathbf{T} \quad (11)$$

with the corresponding $\boldsymbol{\mu} = (\mathbf{I}/C + \mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T \mathbf{T}$.

On the basis of the abovementioned, the kernel-based ELM is also available, in which kernel matrix can be formulated as

$$\boldsymbol{\Omega}_{\text{ELM}} = \mathbf{U} \mathbf{U}^T, \quad (12)$$

$$\Omega_{\text{ELM},i,j} = u(x_i) \cdot u(x_j) = K(x_i, x_j).$$

So using kernel-based method, output function (10) will become

$$f(\mathbf{x}) = \mathbf{u}(\mathbf{x}) \boldsymbol{\mu} = \begin{bmatrix} K(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}, \mathbf{x}_N) \end{bmatrix}^T \left(\frac{\mathbf{I}}{C} + \boldsymbol{\Omega}_{\text{ELM}} \right)^{-1} \mathbf{T}. \quad (13)$$

Formula (13) indicates that kernel function $K(\mathbf{a}, \mathbf{b})$ (e.g., for RBF kernel, $K(\mathbf{a}, \mathbf{b}) = \exp(-\gamma \|\mathbf{a} - \mathbf{b}\|^2)$) can be directly calculated without knowing specific hidden layer feature mapping $\mathbf{h}(\mathbf{x})$ and the number of hidden nodes L .

3. Proposed Method: Multifeature Extreme Ordinal Ranking Machine (MFEORM)

Basically for the proposed MFEORM, facial age estimation is regarded as an ordinal ranking-oriented issue. More specifically, first the database of the i th feature models can be divided as follows:

$$\begin{aligned} M_k^{i(+)} &= \{(x_j^i, +1) \mid y_j > k\}, \quad 1 \leq k \leq K-1 \\ M_k^{i(-)} &= \{(x_j^i, -1) \mid y_j \leq k\}, \quad 1 \leq k \leq K-1, \end{aligned} \quad (14)$$

where K is the number of total age labels of the database, j is the instance number, and y_j is the age labels. Then the problem can be separated into $K - 1$ classification subproblems. For each subproblem k , the purpose is to find whether the face image is older or younger than k . So each subproblem is equivalent to a basic binary classification problem. In each binary classification problem, it is assigned that $z_k[j] = -1$ if $x_j^i \in M_k^{i(-)}$ and $z_k[j] = +1$ if $x_j^i \in M_k^{i(+)}$. Next the aforementioned rule is applied for all i feature models, which will be combined together as our total training sets. Then RBF-kernel-based ELM can be used to train each of k binary classifiers. After traversing k from 1 to $K - 1$, $K - 1$ binary classifiers are obtained, which imply the ordinal relationships among all age labels. We call these $K - 1$ classifiers the “weak classifiers.” In the end, these “weak classifiers” can be comprehensively integrated to form the final result according to our preference-collection rule.

In age estimation, the most popular performance measurement is the mean absolute error (MAE), which can be described by

$$\text{MAE} = \frac{\sum_{j=1}^N |y_j^* - y_j|}{N}, \quad (15)$$

where y_j^* is the estimated age, y_j is authentic age, and N is the number of test images.

The steps of the proposed MFEORM algorithm are as follows:

- (1) For each target age k ($1 \leq k \leq K - 1$), apply the following steps:
 - (a) For each feature model, separate the raw training data into $M_k^{i(+)}$ and $M_k^{i(-)}$.
 - (b) Integrate all i feature models’ information together as the total training sets.
 - (c) Use the kernel-based extreme learning algorithm to train each binary classifier (the so-called “weak classifiers”) and get a decision function $f_k(\mathbf{x})$ according to (13).
- (2) Construct a preference-collection rule from all subproblems:

$$\text{Estimated age} = 1 + \sum_{k=1}^{K-1} \llbracket f_k(\mathbf{x}) > 0 \rrbracket, \quad (16)$$

where $\llbracket \cdot \rrbracket$ is equal to 1 if the inner part holds and otherwise it is equal to 0. For example, suppose we have trained all the binary classifiers based on our algorithm and now we use one face image \mathbf{P}_1 whose authentic age is 25 to test. So we substitute \mathbf{P}_1 into $f_k(\mathbf{x})$ for all $1 \leq k \leq K - 1$ and get all the $f_k(\mathbf{P}_1)$. Finally we find that $f_k(\mathbf{P}_1) > 0$ holds for $1 \leq k \leq 26$, so the final estimated age is 27.

4. Experiments

4.1. Data Sets. A series of experiments on the popular benchmark aging database FG-NET [29] were conducted. FG-NET has 1,002 grayscale or color facial images of 82 people, which includes comprehensive poses, expressions, and lighting environments, just as Figure 2 shows. All people’s age range is from 0 to 69 in FG-NET and its age level distribution is displayed in Table 1. In order to process uniformly, all facial images in FG-NET were converted to grayscale, aligned and normalized. Finally histogram equalization was conducted in order to decrease the illumination influence.

4.2. Experiment Settings. Basically, three feature models for information extraction were used from FG-NET raw images: Active Appearance Model (AAM) [17], local binary patterns (LBP), and Bioinspired feature (BIF) [30], which would be combined together for a total dataset. Active Appearance Models (AAM) can represent both shape and texture information instead of only facial geometry, which is also popularly selected by other age estimation methods. LBP is also a widely used feature for texture classification in computer vision. BIF was selected because of its high age estimation accuracy. The information extracted from the above three feature models can complement each other; together they were combined as a total dataset. For AAM features, the feature dimension was set to retain 95% of variability. For BIF features, the number of bands was set at 8 (16 scales totally) with 4 orientations each. Also, to reduce the entire feature space, principal component analysis (PCA) was used to reduce the dimension. More specifically, all the three feature models would be reduced to 100 dimensions, respectively. Particularly, the AAM model includes both shape and texture information, so these two subproperties would be reduced to 50 dimensions, respectively (in total 100 dimensions). For extreme learning machine, we used RBF kernel-based ELM. Also, leave-one-person-out (LOPO), a popular test procedure, was utilized for the test strategy, which was suggested in [4, 7, 11, 31–33]. In terms of the accuracy, MFEORM’s experimental results were compared with MFOR [19], OHRank [15], WAS [7], AGES [7], RUN1 [31], RUN2 [33], Multitask Warped Gaussian Process (MTWGP) [11], and SVM. For the time consumed, because the accuracy of other algorithms is lower than MFOR we only need to compare the consumed time results with MFOR. Also, in order to find out the inner properties of MFEORM, different combinations of MFEORM’s inner parameters were tried to analyze its inner characteristics. Furthermore, the number of feature dimensions’ impacts on the final estimation result will be shown and the fact that that the multifeature dataset (three-feature dataset) outcompetes two-feature and single-feature dataset based on the proposed MFEORM will be verified in Section 4.3.

4.3. Experiment Results and Analysis. The biggest advantage of our algorithm is the extreme speed. So firstly the consumed time of MFEORM was tested and compared with MFOR.



(a) 3



(b) 4



(c) 5



(d) 7



(e) 12



(f) 15



(g) 16



(h) 18



(i) 20



(j) 21



(k) 23



(l) 26



(m) 29



(n) 31



(o) 36



(p) 38

FIGURE 2: Facial image examples of one subject in sequential order of different age values in the FG-NET database.

TABLE 1: FG-NET's age level distribution.

Range of age	FG-NET	
	#img.	%
0-9	371	37.03
10-19	339	33.83
20-29	144	14.37
30-39	79	7.88
40-49	46	4.59
50-59	15	1.50
60-69	8	0.80
Total	1002	100

TABLE 2: Consumed time of the compared MFEORM and MFOR.

Method	Runtime(s)	MAE (mean absolute error)
MFEORM [ours]	320.6	4.50
MFOR	27437.0	4.58

Table 2 shows the comparison of MFEORM and MFOR in respect of the runtime and accuracy (MAE). As can be explicitly seen, the MFEORM can shorten the runtime by a significant amount while maintaining high accuracy. Because previous facial age estimation solutions based on multifeature ordinal ranking usually formulate their models as complex constrained-optimization problems with inequality constraints, solving these convex optimization problems needs huge quantities of time. But for our algorithm, the input weights and hidden layer parameters are randomly generated and analytically determined, so the hidden layer parameters need not be tuned like SVM. As a result, the proposed method can be much faster than the previous multifeature ordinal ranking methods.

Table 3 displays the MAE results of different algorithms derived from the FG-NET database, which demonstrates that our algorithm achieves a similar or even better accuracy compared to other popular facial age estimation algorithms.

To take it one step further, next the inner properties of MFEORM are explored. Based on the aforementioned description, there exist two uncertain parameters, regularization coefficient C and kernel parameter γ . For these two parameters, experiments are conducted to explore their influences towards experiment results.

Firstly, different regularization coefficient C and same kernel parameter γ which equals 50 were assigned, respectively. Table 4 shows the tuning results. Finally regularization coefficient C was set at 500.

Next different kernel parameter γ with same regularization coefficient C at 500 was assigned, respectively. Table 5 shows the coarse tuning results. Then γ was fine-tuned and set at 31 with the optimal MAE of **4.50** and runtime of 320.6 seconds.

Remark 1. The estimation accuracy and runtime of our MFEORM are insensitive to its two inner parameters, namely, kernel parameter γ and regularization coefficient C .

TABLE 3: MAEs of compared MFEORM and other popular age estimation algorithms on FG-NET database.

Method	MAE
MFEORM [ours]	4.50
MFOR	4.58
OHRank	4.60
RUN1 [31]	5.78
RUN2 [33]	5.33
AGES [7]	6.82
WAS [7]	8.06
MTWGP [11]	4.83
SVM	7.25

TABLE 4: Different regularization coefficients with fixed kernel parameter in MFEORM.

Regularization coefficient C	MAE	Runtime(s)
10	4.630	310.3
40	4.546	312.3
50	4.546	310.7
100	4.556	311.1
250	4.549	310.0
400	4.545	288.9
500	4.540	310.3
600	4.541	289.7
1000	4.545	309.8
10000	4.549	313.1

TABLE 5: Different kernel parameters with fixed regularization coefficient in MFEORM.

Kernel parameter	MAE	Runtime(s)
10	4.628	297.3
100	4.603	292.2
500	4.692	296.8
1000	4.661	295.4
2500	4.626	297.4
5000	4.695	296.3
10000	4.876	297.2

The influence of different number of feature models on the final results was further explored (regularization coefficient C is set at 500 and kernel parameter γ is 31 uniformly). As Table 6 indicates, conclusion can be made that the more feature models used, the better results. Particularly, in the single-feature test, AAM shows its superior estimation performance in MFEORM, which is better than LBP and BIF features, so we conjecture that this is because AAM gets the feature from both face shape and texture information separately. So next AAM feature may as well be used to analyze the influence of number of dimensions on the final results.

Table 7 displays the tests which were made to explore the influence of number of dimensions (after PCA dimension reduction) on the final results (regularization coefficient C is

TABLE 6: Comparison of different numbers of feature models.

Feature models	MAE	Runtime(s)
Single-feature		
AAM	4.83	247.2
LBP	7.56	269.7
BIF	5.92	245.4
Two-feature		
AAM + BIF	4.74	276.9
LBP + BIF	5.56	271.7
AAM + LBP	4.59	272.7
Multifeature (three-feature)		
AAM + LBP + BIF	4.50	320.6

TABLE 7: Comparison of different dimensions' impacts.

Dimensions (after PCA dimension reduction)	MAE	Runtime(s)
20 (Shape 10 + Texture 10)	5.59	232.5
40 (20 + 20)	4.86	235.7
60 (30 + 30)	4.85	240.2
100 (50 + 50)	4.83	247.2
120 (60 + 60)	4.79	250.3
140 (70 + 70)	4.78	260.2
160 (80 + 80)	4.85	264.4
200 (100 + 100)	4.88	278.0

set at 500 and kernel parameter is 31 uniformly). Generally speaking, the estimation accuracy and runtime of our MFEORM are insensitive to the number of dimensions, which means that there is no need to waste a large amount of time on tuning the inner parameters of MFEORM. However, from a nuanced prospective, when the dimensions are too small, the dataset will suffer from insufficient information from feature models, so the deviation will become larger. In contrast, when the dimensions are oversized, the dataset will include not only main components but also some interference information, which can also lead to a relatively large deviation.

Note that the emphasis and innovation points of this paper and [34] are different. In detail, the target of this paper is to achieve both rapid learning speed and similar or better age estimation performances compared to state-of-the-art methods. As can be seen, for every comparison in the experiment section, the data of both MAE (accuracy) and runtime (speed or efficiency) are listed. However, [34] mainly focuses on the improvement of the ranking rule (preference-collection rule) in the ordinal ranking process so that the estimation performance can be increased. In other words, [34] only pays attention to further improving the estimation performance (accuracy) and does not care about the time or efficiency. Although ELM is also used in [34], that is only for the consideration of saving time; that is to say, if it were not to this end, other classifiers can also be used to replace ELM, such as SVM. Consequently, in [34], the proposed method is suitable for different classifiers.

5. Conclusion

MFEORM combines multifeature space, age's natural characteristics of ordinal information, and extreme learning machine's rapid learning rate, achieving similar or even better performances with much less time compared to state-of-the-art methods. Experimental results demonstrate the fact that the more feature models, the better performance. Further experiments from internal and external aspects have been performed, respectively, so that the following properties and advantages of MFEORM are obtained: the performance of MFEORM including estimation accuracy and runtime is insensitive to its two inner parameters and the number of dimensions (after dimension reduction) so there is no need to consume much time on choosing different inner parameters of MFEORM and the number of dataset dimensions. Note that the proposed method MFEORM has few relationships with the pure ELM ranking problem [35, 36]. Essentially it is a combination of many classification subproblems, though from literal meaning it seems relevant to pure ranking problem and causes misunderstanding.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] Electronic Customer Relationship Management (ECRM), <https://en.wikipedia.org/wiki/ECRM>.
- [2] J. E. Kloepfel, "Step right up, let the computer look at your face and tell you your age," <http://news.illinois.edu/news/08/0923age.html>.
- [3] A. Dix, J. Finlay, G. D. Abowd, and R. Beale, "Human-Computer Interaction," http://svcnogac.nl/wp-content/uploads/2012/04/1.MMI_summary.pdf.
- [4] G. Guo, Y. Fu, C. R. Dyer, and T. S. Huang, "Image-based human age estimation by manifold learning and locally adjusted robust regression," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1178–1188, 2008.
- [5] N. Ramanathan and R. Chellappa, "Face verification across age progression," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3349–3361, 2006.
- [6] C.-C. Wang, Y.-C. Su, C.-T. Hsu, C.-W. Lin, and H. Y. M. Liao, "Bayesian age estimation on face images," in *Proceedings of the IEEE International Conference on Multimedia and Expo (ICME '09)*, pp. 282–285, July 2009.
- [7] X. Geng, Z.-H. Zhou, and K. Smith-Miles, "Automatic age estimation based on facial aging patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2234–2240, 2007.
- [8] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 34, no. 1, pp. 621–628, 2004.
- [9] Z. Yang and H. Ai, "Demographic classification with local binary patterns," in *Proceedings of the International Conference on Advances in Biometrics*, pp. 464–473, 2007.

- [10] Y. Fu and T. S. Huang, "Human age estimation with regression on discriminative aging manifold," *IEEE Transactions on Multimedia*, vol. 10, no. 4, pp. 578–584, 2008.
- [11] Y. Zhang and D.-Y. Yeung, "Multi-taskwarped Gaussian process for personalized age estimation," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 2622–2629, June 2010.
- [12] B. Ni, Z. Song, and S. Yan, "Web image mining towards universal age estimator," in *Proceedings of the 17th ACM International Conference on Multimedia (MM '09)*, pp. 85–94, October 2009.
- [13] B. Xiao, X. Yang, H. Zha, Y. Xu, and T. Huang, "Metric learning for regression problems and human age estimation," in *Advances in Multimedia Information Processing—PCM 2009*, vol. 5879 of *Lecture Notes in Computer Science*, pp. 88–99, Springer, Berlin, Germany, 2009.
- [14] G. Guo, Y. Fu, T. S. Huang, and C. R. Dyer, "Locally adjusted robust regression for human age estimation," in *Proceedings of the IEEE Workshop on Applications of Computer Vision (WACV '08)*, pp. 1–6, IEEE, Copper Mountain, Colo, USA, January 2008.
- [15] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, "Ordinal hyperplanes ranker with cost sensitivities for age estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 585–592, June 2011.
- [16] Y. H. Kwon and N. da Vitoria Lobo, "Age classification from facial images," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '94)*, pp. 762–767, IEEE, Seattle, Wash, USA, June 1994.
- [17] T. F. Cooles, G. J. Edwards, and C. J. Taylor, "Active appearance models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001.
- [18] Y. Fu, Y. Xu, and T. Huang, "Estimating human age by manifold analysis of face pictures and regression on aging features," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, pp. 1383–1386, Beijing, China, July 2007.
- [19] R. Weng, J. Lu, G. Yang, and Y.-P. Tan, "Multi-feature ordinal ranking for facial age estimation," in *Proceedings of the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG '13)*, pp. 1–6, April 2013.
- [20] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1–3, pp. 489–501, 2006.
- [21] G. Huang and C. Siew, "Extreme learning machine with randomly assigned RBF kernels," *International Journal of Information Technology*, vol. 11, no. 1, pp. 16–24, 2005.
- [22] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 42, no. 2, pp. 513–529, 2012.
- [23] S.-C. Huang and Y.-F. Huang, "Bounds on the number of hidden neurons in multilayer perceptrons," *IEEE Transactions on Neural Networks*, vol. 2, no. 1, pp. 47–55, 1991.
- [24] M. A. Sartori and P. J. Antsaklis, "A simple method to derive bounds on the size and to train multilayer neural networks," *IEEE Transactions on Neural Networks*, vol. 2, no. 4, pp. 467–471, 1991.
- [25] G.-B. Huang and H. A. Babri, "Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions," *IEEE Transactions on Neural Networks*, vol. 9, no. 1, pp. 224–229, 1998.
- [26] G.-B. Huang, L. Chen, and C.-K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," *IEEE Transactions on Neural Networks*, vol. 17, no. 4, pp. 879–892, 2006.
- [27] P. L. Bartlett, "The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network," *IEEE Transactions on Information Theory*, vol. 44, no. 2, pp. 525–536, 1998.
- [28] D. Serre, *Matrices: Theory and Application*, Springer, New York, NY, USA, 2002.
- [29] The FG-NET aging Database, <http://www.fgnet.rsunit.com/>.
- [30] G. Guo, G. Mu, Y. Fu, and T. S. Huang, "Human age estimation using bio-inspired features," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 112–119, June 2009.
- [31] S. Yan, H. Wang, T. S. Huang, Q. Yang, and X. Tang, "Ranking with uncertain labels," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, pp. 96–99, IEEE, Beijing, China, July 2007.
- [32] P. Yang, L. Zhong, and D. Metaxas, "Ranking model for facial age estimation," in *Proceedings of the 20th International Conference on Pattern Recognition (ICPR '10)*, pp. 3404–3407, IEEE, Istanbul, Turkey, August 2010.
- [33] S. Yan, H. Wang, X. Tang, and T. S. Huang, "Learning auto-structured regressor from uncertain nonnegative labels," in *Proceedings of the IEEE 11th International Conference on Computer Vision (ICCV '07)*, pp. 1–8, October 2007.
- [34] W. Zhao and H. Wang, "Facial age estimation based on advanced ordinal ranking," *Electronics Letters*, vol. 51, no. 12, pp. 903–904, 2015.
- [35] H. Chen, J. Peng, Y. Zhou, L. Li, and Z. Pan, "Extreme learning machine for ranking: generalization analysis and applications," *Neural Networks*, vol. 53, pp. 119–126, 2014.
- [36] W. Zong and G.-B. Huang, "Learning to rank with extreme learning machine," *Neural Processing Letters*, vol. 39, no. 2, pp. 155–166, 2014.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

