Facial Age Range Estimation with Extreme Learning Machines

Phyo-Kyaw Sai\textsuperscript{a}, Jian-Gang Wang\textsuperscript{b}, Eam-Khwang Teoh\textsuperscript{a}

\textsuperscript{a}Nanyang Technological University, Singapore 639798
\textsuperscript{b}Institute for Infocomm Research, A*STAR, Singapore 138632

Abstract

Face image based age estimation is an approach to classify face images into one of several pre-defined age-groups. It is challenging because facial aging variation is specific to a given individual and is determined by the persons gene and many external factors, such as exposure, weather, gender, living style. Age estimation is a multiclass problem and the number of classes to predict is quite large. There surely is facial aging trend and faces from closed age range have some similar facial aging features. It is difficult to say there are distinct facial aging features for an age. Facial aging features are found to be overlapped among nearby age groups along the aging life and are continuous in nature. In this paper, we emphasised our work on age range estimation with four predefined classes. We applied a fast and efficient machine learning method: Extreme Learning Machines, to solve the age categorization problem. Local Gabor Binary Patterns, Biologically Inspired Feature and Gabor were adopted to represent face image. Age estimation was performed on three different aging datasets and experimental results are reported to demonstrate its effectiveness and robustness.

Keywords: Facial age range estimation, Extreme learning machines, ECOC, Boosting

\textsuperscript{a}\textsuperscript{Email addresses: saip0003@ntu.edu.sg (Phyo-Kyaw Sai), jgwang@i2r.a-star.edu.sg (Jian-Gang Wang), eekteoh@ntu.edu.sg (Eam-Khwang Teoh)
1. Introduction

Age estimation has many applications, e.g. vending machine can prevent the dispensing of alcoholic drinks or cigarette to an underage customer by finding out the estimated age range of the customer using a computer vision system. Facial aging effects display some unique characteristics [1]: the age progression displayed on faces is uncontrollable, individual and time dependent. Such special characteristics of aging variation cannot be captured accurately due to the prolific and diversified information conveyed by the human faces.

1.1. Motivation

A human face conveys much information that we can easily decipher in our day-to-day communication. This includes the identity of a person, as well as gender, look direction, emotion and age. Developing an automated way to estimate the age of a person from the face image is crucial. This is a key motivation for our work. Such capability, when developed successfully, will enable many applications like automated vending and gaming machines and customized digital signages.

The sooner we know the outcome of any environmental setting changes, the faster we move forward in our research work. We can make the decision quickly whether it is desirable to proceed in a certain direction or not. Fast ELM learning phase attracted us to use it in initial investigation. However it was found out to be a good performer as well.

2. Related work

A recent survey on automated age estimation can be found in [2]. Kwon and Lobo [3] first worked on the age classification problem. They referred to cranio-facial research, theatrical makeup, plastic surgery, and perception to find out the features that change with age. A probe gray-scale facial image can be classified into three age groups: babies, young adults, and senior adults. The proposed algorithm is computationally expensive. Adopting the Active Appearance Model (AAM) [4] approach, Lanitis et al. [5] devised a combined shape and intensity model to represent face images. Age is modelled as a function of the vector of the face model parameters. The aging function is defined as linear, quadratic and cubic functions. Later, they [6] reported a quantitative evaluation of the three classifiers (quadratic function,
shortest distance, artificial neural network) using a 400 images database. Geng et al. [1] proposed an aging pattern subspace (AGES) for estimating age from appearance. In order to handle incomplete data such as missing ages in the training sequence, the AGES method models a sequence of individual aging face images by learning a subspace representation. The age of a test face is determined by the projection in the subspace that can best reconstruct the face image. Fu et al. [7] constructed a low-dimensional manifold from a set of age-separated face images and use linear and quadratic regression functions on the low dimensional feature vectors from the respective manifolds to estimate the age of a face. Adopting similarly approach, Guo et al. [8] proposed an age manifold learning scheme for extracting face aging features and design a locally adjusted robust regressor for learning and prediction of human age. Ramanathan et al. [9] proposed a craniofacial growth model that takes into account both psychophysical evidences on how humans perceive age progression in faces and anthropometric evidences on facial growth. The proposed model is used to predict a persons appearance across age and to improve face recognition results across ages. Yan et al. also dealt with the age uncertainty by formulating a semi-definite programming problem [10] or an EM-based algorithm [11]. By boosting Local Binary Pattern (LBP) [12] features, Yang et al. [13] identified a sequence of local features which when combined into a strong classifier performs the task of age classification successfully.

Most of the conventional methods for age estimation are intended for accurate estimation of the actual age. However, it is difficult to accurately estimate an actual age from a face image because facial age progression is specific-dependent. Fortunately, it is not necessary to obtain the precise estimates of the actual age for some applications. Most of the age estimation approaches adopted the regression method to predicate exact age from face image. It is difficult to predicate age by using the limited training samples with discrete age values (sparse, not continuous). More, we use Gabor-based features having high dimension and demand for resource will be very high on big dataset. Therefore, in this paper, we pay attention to the mechanism of human age perception, i.e. we limit the estimation to a few age ranges.

One of the problems in facial age estimation is that a large training database is required to represent the variance of the appearance. The training is a time-consuming process. A fast and efficient machine learning method is expected to overcome it. Extreme Learning Machine (ELM) [14] is a simple and efficient learning algorithm driven from single-hidden layer feedforward
neural networks (SLFNs). The ELM provides several interesting and significant features over traditional popular gradient-based learning. Not like other feedforward network family, training in the ELM will take an extremely short period. Moreover it provides better generalization performance over the gradient-based training methods. This motivates us to develop an ELM based facial age range classifier. In this paper, we proposed a novel facial age range estimation framework in which the ELM is employed to classify age groups.

3. Feature extraction

Similar to face recognition, facial features are extracted from original image to represent the faces. In this framework, as shown in Fig. 1, facial features were extracted from face images manually aligned at the two eye corners. Experiment was carried out to compare the results on features obtained by two extraction methods. Biologically Inspired Features [15] has been verified as a good feature for facial age estimation; we adopted it in this paper. In addition, we compared the BIF with another well-established face feature, Local Gabor Binary Pattern [16]. Both features extraction techniques: Biologically Inspired Features (BIF) and Local Gabor Binary Pattern (LGBP), made use of Gabor features extracted with Gabor filter bank of 5 scales and 8 orientations. All 40 Gabor images generated by Gabor filter bank are divided into processing blocks one after another. Standard deviation or LBP histogram was computed over the block and cascaded into respective

![Figure 1: Local Gabor Binary Pattern processing](image-url)
vector. Since the extracted feature dimension was too high in both Biologically Inspired Features and Local Gabor Binary Pattern, dimension reduction became a necessary step before they were applied to the ELM. Different feature dimension reduction methods like Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), Locality Preserving Projections (LPP) and Orthogonal Locality Preserving Projections (OLPP) had been adopted and the performances were compared in this paper.

4. Extreme learning machine

Extreme Learning Machine (ELM) [14] is a simple and efficient learning algorithm in single hidden layer feedforward neural networks (SLFNs). The ELM has several interesting and significant features over traditional popular gradient-based learning algorithms like feedforward neural networks.

Not like other feedforward network family, training can be done by the ELM in an extremely short period. It can overcome the speed barrier faced by classic learning algorithms. Moreover it provides better generalization performance over the gradient-based training methods. The ELM tends to overcome several issues like local minima, improper learning rate and over-fitting problems straightforward. But traditional classical gradient-based learning algorithms may need to use additional methods like early stopping or weight decay to overcome such issues [17].

The ELM could be taken as the simplest form of learning algorithms for feedforward neural networks. The present form of ELM algorithm is still only valid for single-hidden layer feedforward networks (SLFNs). However, it was proven that SLFNs can approximate any continuous function and implement any classification application [18]. Thus, we could reasonably assume the proposed ELM algorithm can be efficiently used in many applications.

In the ELM, the input weights and hidden layer biases of SLFNs are first assigned with random values[14]. After random assignment of the input weights and the hidden layer biases, SLFNs can be simply regarded as a linear system and the output weights (linking the hidden layer to the output layer) of SLFNs can be analytically obtained by simple generalized inverse operation of the hidden layer output matrices. The structure of ELM is shown in Fig. 2.

For N arbitrary distinct samples \( (x_i - t_i) \), where \( x_i = [x_{i1}, x_{i2}, ..., x_{in}]^T \in \mathbb{R}^n \), \( t_i = [t_{i1}, t_{i2}, ..., t_{im}]^T \in \mathbb{R}^m \) and standard SLFNs with \( N \) hidden
nodes and activation function \( g(x) \) are mathematically modelled as

\[
\sum_{i=1}^{N} \beta_i g(x_j) = \sum_{i=1}^{N} \beta_i g(w_i x_j + b_i) = o_j, \quad j = 1, 2, ..., N
\]  

(1)

where \( w_i = [w_{i1}, w_{i2}, ..., w_{im}]^T \) is the weight vector connecting the \( i^{th} \) hidden node and the input nodes, \( \beta_i = [\beta_{i1}, \beta_{i2}, ..., \beta_{im}]^T \) is the weight vector connecting the hidden node. \( w_i . x_j \) denotes the inner product of \( w_i \) and \( x_j \). The output nodes are chosen linear in the approach.

The standard SLFNs with \( N \) hidden nodes with activation function \( g(x) \) can approximate these \( N \) samples with zeros error means that \( \sum_{j=1}^{N} ||o_j - t_j|| = 0 \), i.e. there exist \( \beta_i, w_i \) and \( b_i \) such that

\[
\sum_{i=1}^{N} \beta_i g(w_i x_j + b_i) = t_j, \quad j = 1, 2, ..., N
\]  

(2)

The above \( N \) equations can be written compactly as

\[
H \beta = T
\]  

(3)

where

\[
H = \begin{pmatrix}
g(w_1 x_1 + b_1) & ... & g(w_N x_1 + b_N) \\
\vdots & \ddots & \vdots \\
g(w_1 x_N + b_1) & ... & g(w_N x_N + b_N)
\end{pmatrix}_{N \times N}, \quad \beta = \begin{pmatrix}
\beta_1^T \\
\vdots \\
\beta_N^T
\end{pmatrix}_{N \times m}
\]
and \( T = \begin{pmatrix} t_{11}^T \\ \vdots \\ t_{N1}^T \end{pmatrix} \). \( H \) is called the hidden layer output matrix of the neutral network; the \( i^{th} \) column of \( H \) is the \( i^{th} \) hidden node output with respect to inputs \( x_1, x_2, \ldots, x_N \). The output weight \( \beta \) can be computed as \( \beta = H^+ T \) where \( H^+ \) is the pseudoinverse of hidden layer output matrix \( H \).

The ELM can be used efficiently in many applications. ELMs are proved to be universal approximators \cite{18} but the performance of sparse high-dimensional applications is still an open interesting aspect in ELMs. The ELM has been applied in many research fields such as biometrics, bioinformatics, Image processing, human action recognition, real-time learning and prediction, security and data privacy etc. In this paper, we applied it to facial age estimation. To the best of our knowledge, this is the first paper which employed the ELM for estimating age range from face image.

In this ELM framework for age estimation, global representation of features was extracted to low dimensional space and classification was performed on the extracted feature space with the ELM classifier to predict the age group. Classification work can be formulated as the following equations.

\[
\begin{pmatrix}
g(w_1.x_1^t) & \ldots & g(w_N.x_1^t) \\
\vdots & \ddots & \vdots \\
g(w_1.x_N^t) & \ldots & g(w_N.x_N^t)
\end{pmatrix}
\begin{pmatrix}
\beta_1 \\
\vdots \\
\beta_N
\end{pmatrix} =
\begin{pmatrix}
l_1 \\
\vdots \\
l_N
\end{pmatrix}
\tag{4}
\]

\[
H \beta = L 
\tag{5}
\]

where \( x_i^t \) is the value of the \( i^{th} \) sample obtained by a specified method; \( w_i \) is the input weight vector connecting the \( i^{th} \) hidden node and the input node; \( \beta_i \) is the weight vector connecting the \( i^{th} \) hidden node and \( l_i \) is the label.

The output weight \( \beta \) can be obtained by equation (7).

\[
\beta = H^+ L 
\tag{6}
\]

where \( H^+ \) is the pseudoinverse of hidden layer output matrix \( H \).

Suppose the image space \( X \) is represented by a set of aligned faces images of \( n \) subjects: \( X = \{x_i\}_{i=1}^n \in \mathbb{R}^{D_1} \). A label set \( L = \{l_i : l_i \in \mathbb{N}\}_{i=1}^n \) associated with \( n \) images provides a total of \( N \) age group labeling. Then Biologically Inspired Feature (BIF) space \( X_b = \{x_i^b\}_{i=1}^n \in \mathbb{R}^{D_3} \) and Local Gabor Binary
Pattern (LGBP) space will be obtained as follows, $X_{gl} = \{x_{gi}^{gl}\}_{i=1}^{n} \in \mathbb{R}^{D_{g}}$.

As mentioned-above, PCA, LDA, LPP and OLPP methods are used to reduce their high dimensional space and to obtain some sense of their global representation. So a total of 8 combination methods will be obtained as $X_{B+PCA}$, $X_{B+LDA}$, $X_{B+LPP}$, $X_{B+OLPP}$, $X_{LG+PCA}$, $X_{LG+LDA}$, $X_{LG+LPP}$ and

(a) (BIF+PCA) : $X_{B+PCA} = \{x_{bi}^{b1}\}_{i=1}^{n} \in \mathbb{R}^{D_5}$
(b) (BIF+LDA) : $X_{B+LDA} = \{x_{bi}^{b2}\}_{i=1}^{n} \in \mathbb{R}^{D_6}$
(c) (BIF+LPP) : $X_{B+LPP} = \{x_{bi}^{b3}\}_{i=1}^{n} \in \mathbb{R}^{D_5}$
(d) (BIF+OLPP) : $X_{B+OLPP} = \{x_{bi}^{b4}\}_{i=1}^{n} \in \mathbb{R}^{D_5}$
(e) (LGBP+PCA) : $X_{LG+PCA} = \{x_{bi}^{b5}\}_{i=1}^{n} \in \mathbb{R}^{D_5}$
(f) (LGBP+LDA) : $X_{LG+LDA} = \{x_{bi}^{b6}\}_{i=1}^{n} \in \mathbb{R}^{D_6}$
(g) (LGBP+LPP) : $X_{LG+LPP} = \{x_{bi}^{b7}\}_{i=1}^{n} \in \mathbb{R}^{D_5}$
(h) (LGBP+OLPP) : $X_{LG+OLPP} = \{x_{bi}^{b8}\}_{i=1}^{n} \in \mathbb{R}^{D_5}$

$X_{LG+OLPP}$ will be fed into the ELM for classification and their respective prediction output $L' = \{l'_i\}_{i=1}^{n}$ will be obtained. The flowchart of the proposed age estimation approach is shown in Fig. 3.

Since the input weights and hidden biases are randomly chosen in the ELM, learning time spent in tuning these parameters is irrelevant. Random assigned computational nodes in the hidden layer are mostly likely to be independent of the training data. It makes the ELM provides better performance on the functional approximation. The ELM appears to be suitable in applications which request fast prediction and response capability[19].

5. Experimental Results

In order to evaluate the proposed ELM facial age range estimation, we collected an aging database from the internet. Motivated by works of web image mining towards universal age estimator [20] and web image mining age estimation framework [21], we tried to collect aging face dataset with Microsoft Programming Application Interface (API) services provided by Microsoft Search Engine Bing [22].

In order to obtain optimized results, trainings followed by testing were done for 10 times and only the best runs were considered for comparison. Since the ELM was trained with random sets of weight inputs and biases, it might not get the optimal setting at first run.
5.1. Database

In order to compare with the existing databases, some benchmark aging databases were also used in our experiments. We have done our experience on three datasets. The images in all datasets are aligned with eye corner points captured manually and cropped to the image size of 65 x 75.

Dataset I

The dataset I consists of aging facial images from two public aging datasets; FG-NET and PAL [23]. Most of the FG-NET age labels are in the younger range and less number of images in old age range. The oldest age of the dataset is 69 years old. Some facial images of the FE-NET database are bad in quality and contain high degree of expressions and poses. The PAL database does not contain face images of children and young teenagers. The starting age in the data base is 18 years old and oldest age is 93 years old. The quality and pose of face in images are good.

The dataset I has four age groups namely child, teen, adult and senior adult. Sample images of dataset I can be seen in Fig. 4. Good images from FG-NET were grouped to from child and teenage group and adult group. Divert age range from PAL dataset were selected to form senior adult and some were added to adult group so that there would be reasonable amount
Table 1: Age distribution of dataset I

<table>
<thead>
<tr>
<th>Group</th>
<th>Age range</th>
<th>Number of images</th>
</tr>
</thead>
<tbody>
<tr>
<td>Child</td>
<td>less than 10</td>
<td>96</td>
</tr>
<tr>
<td>Teen</td>
<td>11 to 19</td>
<td>71</td>
</tr>
<tr>
<td>Adult</td>
<td>20 to 60</td>
<td>116</td>
</tr>
<tr>
<td>Senior Adult</td>
<td>60 and above</td>
<td>100</td>
</tr>
</tbody>
</table>

Figure 4: Sample images of dataset I. A: child; B: teen; C: adult and D: senior adult

of good images in each group. The age distribution of dataset is given in Table 1.

Figure 5: Samples images of dataset II
Database II

The dataset II was created with the images collected over internet by the method which uses Microsoft BING Image Search Application Programming Interface (API)[22]. Querying with different age-labeled text string on API services, age-labeled images were collected and saved with relevant age-labeled information. From collected large pool of image data, image of good quality were selected from certain age and the information was verified manually. This dataset was created with facial images of good quality. Variation on facial pose and expression is small. Some samples of the dataset II are shown in Fig. 5. The age distribution of dataset II is given in Table 2.

<table>
<thead>
<tr>
<th>Group</th>
<th>Age</th>
<th>No. of images</th>
<th>Total Images</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>3, 5, 8, 10, 13</td>
<td>50+50+44+51+40</td>
<td>235</td>
</tr>
<tr>
<td>II</td>
<td>23, 25, 28, 30, 33</td>
<td>49+50+49+57+46</td>
<td>251</td>
</tr>
<tr>
<td>III</td>
<td>43, 45, 48, 50, 53</td>
<td>64+52+39+51+46</td>
<td>252</td>
</tr>
<tr>
<td>IV</td>
<td>63, 65, 68, 70, 73</td>
<td>35+52+46+49+43</td>
<td>225</td>
</tr>
</tbody>
</table>

Database III

Dataset III was created with images from Morph database II [24], a largest collection publicly available aging face database. It has facial images of male and female starting from 16 to 77 years in age. It contains 55,134 images of 13,000 individuals collected over four years. Male African images of age 16, 26, 36 and 46 were selected and used for this dataset with each group containing 250 individual images.

The sample images and the age distribution of dataset III are shown in Fig. 6 and Table 3 respectively.

<table>
<thead>
<tr>
<th>Group</th>
<th>Age</th>
<th>Number of images</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>16</td>
<td>250</td>
</tr>
<tr>
<td>II</td>
<td>26</td>
<td>250</td>
</tr>
<tr>
<td>III</td>
<td>36</td>
<td>250</td>
</tr>
<tr>
<td>IV</td>
<td>46</td>
<td>250</td>
</tr>
</tbody>
</table>
Figure 6: Samples images of dataset III

5.2. Age estimation results

Performance comparison was conducted on 10 fold cross validation where each group of images are equally divided into 10 folds and one fold is always reserved for testing and the other nine folds are used for training.

Figure 7: Overall performance comparison methods on dataset I
Results on Dataset I

Overall performance comparison on all eight approaches was shown in Fig. 7. From the results obtained, we can see that LGBP feature representation method as well as BIF representation method performed quite reasonably well across all dimension reduction techniques. However, it is obvious that LGBP feature extraction method with LDA dimension reduction approach performed better than all other approaches.

![Individual class performance on the best approach](image)

**Figure 8:** Individual class performance of (LGBP+LDA) on dataset I

For individual class performance on (LGBP + LDA) approach shown in Fig. 8, performance on child and senior adult groups was the highest with over 90% accuracy. Adult group maintained a reasonable performance of 80% accuracy. However, teen group performance was unsatisfactory with accuracy of less than 40%. It also reflects the fact that young teens similarity to child and late teens similarity to adult is very high and it is harder to predict them. For adult class, it did not perform that bad since the difference from adult to senior adult is quite high due too skin winkle condition.

Results on Dataset II

In this experience we made sure that equal distribution of images from each age contained in each fold. Overall accuracy comparison of 10 fold cross validation on dataset II was shown in Fig. 9.
In dataset II, LGBP feature representation performed better than BIF feature representation on all dimension reduction techniques. Images of better quality in dataset II might make LGBP representation provide more stable performance. (LGBP + LDA) approach was again top performer among all combinations. Individual class performance of (LGBP + LDA) on dataset II is shown in Fig. 10. According to Fig. 10, there was lowest performance on Class III (40-59). I believe that it is due to its high similarity to its upper and lower classes. By considering the boundary regions: (13 vs 23), (33 vs 43) and (53 vs 63), and their facial similarity, we could assume that the results reflects human perception on age estimation.

**Results on Dataset III**

Dataset III was created from Morph II database containing four age groups with 10-year gap i.e. 16, 26, 36 and 46. Experience was carried out to find out the performance accuracy on face images of 10-year gap ages. The results are shown in Fig. 11.

Again in dataset III, LGBP performed better than BIF and (LGBP + LDA) approach was still the top performer.

As individual class performance, it was the worst performance on age 36 group with average accuracy of 50%. Individual performance on class I and IV is reasonably good but performance on class II and III is unsatisfactory.
Figure 10: Individual class performance of (LGBP+LDA) on dataset II

Figure 11: Overall performance comparison methods on dataset III

This can be seen in Fig. 12.
5.3. Comparison with AdaBoost based approach

In the previous AdaBoost framework[25], most discriminant local features were selected by tree classifiers and the AdaBoost classifier boosts the performance by combining the weak learners. So the features used were just local. In this experiment, we compared the performances between the AdaBoost based approach and the proposed ELM approach. Since BIF and LGBP features are processed on blocks of Gabor output, the square block sizes which gave the best performance on BIF and LGBP was discovered from window size of 4x4 to 10x10. Then their best performance results were computed and compared. From the starting framework (AdaBoost classifier with Gabor feature) to our current work, the performance improvement on dataset (a)I, (b)II and (c)III can be seen in Fig. 13. Significant performance improvement was achieved by using new features as well as adopting new framework. Although overall performance was high, in-between groups still yielded low accuracy.

The CPU time in each process can be found in Table 4. The timing results were obtained on the system with CPU of Intel Core i7-2640M Processor 2.8Ghz (Turbo Boost up to 3.5 GHz), Memory of 8 GB, Windows 7 Professional (64 bit) and MatLab (64 bit) R2010a. The main big improve-
Figure 13: Performance comparison among combination of feature representations and classification algorithms on dataset [a]I, [b]II and [c]III.

6. Conclusion and future work

In this paper, we have developed some ways to improve performance on age group classification from previous to current stage. To increase accuracy, we introduced feature representation which has tolerance toward pose and alignment variation of a few degrees for previous system of AdaBoost classi-

ment through the ELM over AdaBoost is the speed. 3-4 hour-long training process with AdaBoost was completed in seconds by the ELM approach. It is not a problem at all to go through a few runs for obtaining maximum performance. However, conjugate gradient methods [26, 27] move towards the optimum more directly than other methods. We will investigate into conjugate gradient methods for their speed and performance in age estimation.
Table 4: Time used in respective process

<table>
<thead>
<tr>
<th></th>
<th>CPU time</th>
<th>No. of images</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>(Extraction)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIF</td>
<td>48.1287</td>
<td>1000</td>
</tr>
<tr>
<td>LGBP</td>
<td>916.5642</td>
<td>1000</td>
</tr>
<tr>
<td><strong>(Minifold training)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>55.4740</td>
<td>900</td>
</tr>
<tr>
<td>LDA</td>
<td>36.9878</td>
<td>900</td>
</tr>
<tr>
<td>LPP</td>
<td>39.6555</td>
<td>900</td>
</tr>
<tr>
<td>OLPP</td>
<td>42.7911</td>
<td>900</td>
</tr>
<tr>
<td><strong>(Minifold mapping)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PCA</td>
<td>0.0780</td>
<td>100</td>
</tr>
<tr>
<td>LDA</td>
<td>0.0780</td>
<td>100</td>
</tr>
<tr>
<td>LPP</td>
<td>0.3120</td>
<td>100</td>
</tr>
<tr>
<td>OLPP</td>
<td>0.0624</td>
<td>100</td>
</tr>
<tr>
<td><strong>(ELM)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>0.1248</td>
<td>900</td>
</tr>
<tr>
<td>Testing</td>
<td>0.0312</td>
<td>100</td>
</tr>
<tr>
<td><strong>(AdaBoost)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td>12086.000</td>
<td>900</td>
</tr>
<tr>
<td>Testing</td>
<td>0.1404</td>
<td>100</td>
</tr>
</tbody>
</table>

Classifier equipped with ECOC for multi-classification. Performance improvement was achieved depending on dataset by employing BIF feature in previous system. Tree classifiers operating on local features are combined and boosted by AdaBoost methods. So the AdaBoost classifier used only local features which were selected by tree classifiers at each iteration in this method. Even though the ECOC method was used to achieve multi-classification capability, it might jeopardize the performance if the aging process happens to be continuous linear process since ECOC employed code word which is larger than the number of the classes.

Then new age range estimation framework was proposed with Extreme Learning Machine as classifier and output values of dimension reduction methods were fed into the classifier. In this way, the new system had the ability to take global feature representation as inputs. Another advantage over previous system is that the training process is really fast in the ELM. Through experiments, the ELM was found to be not only a good starter
to stand the status quo of the working domain but also a good optimizer which provided reasonable generalization and reliable performance within a short time. Additional performance improvement was obtained through this new framework. Among all approaches, the combination of LGBP and LDA provided the best performance across all datasets on new framework. In this way, we lift initial overall accuracy of 40-50\% on previous system to overall accuracy of 60-80\% on current system. Though we can achieve such a high overall performance for four age group classification, we found out that individual class performance on middle groups was still low. We need a breakthrough in aging feature extraction methods so that computer-based age estimation system can be employed to detect small age range difference.

References


Phyo-Kyaw Sai received B.Eng. degree from School of Electrical and Electronic Engineering (EEE), Nanyang Technological University (NTU), Singapore in 2009. He is a software engineer at HCL Singapore Pte Ltd and pursuing M.Eng. at EEE of NTU, Singapore. His research interest includes Machine learning, computational intelligence and image processing.

Jian-Gang WANG received his Bachelor degree from the Inner Mongolia University in 1985, MEng degree from the Shenyang Institute of Automation, Chinese Academy of Sciences in 1988, and PhD degree from the Nanyang Technological University in 2001. From 1988 to 1997, he was with the Robotics Laboratory at the Chinese Academy of Sciences, where he was appointed Associate Professor in 1995. From 1997 to 1998, he was a Research Assistant at the Department of Manufacturing Engineering and Engineering Management, City University of Hong Kong. He joined the Centre for Signal Processing, Nanyang Technological University as a Research Fellow in January 2001. He is presently a Research Scientist with the Institute for Infocomm Research (I2R). Jian-Gang serves as an Editor of Scholarpedia and as Editor of ISRN Signal Processing Journal, Editor of Communications and Computer Security Journal. He is a Senior Member of the IEEE and a technical committee member of the International
Association of Science and Technology for Development (IASTED). He has served as program committee member for some international conferences and session co-chair for the tenth Asian Conference on Computer Vision (ACCV2010). Dr Wang won the Best Paper Honorable Mention Award 2010 from the Pattern Recognition Journal. His paper is listed as top 25 most cited papers since 2009, extracted from Scopus. His current research interests are in computer vision, biometrics and pattern recognition and has published 2 patents, one book and over 70 papers in these areas.

Eam Khwang TEOH received his BE and ME degrees in Electrical Engineering from the University of Auckland, New Zealand in 1980 and 1982 respectively and the PhD degree in Electrical & Computer Engineering from the University of Newcastle, New South Wales, Australia in 1986. Since June 1985, he has been with the School of Electrical & Electronic Engineering, Nanyang Technological University, Singapore as a Lecturer, Senior Lecture and currently, an Associate Professor in the Division of Control & Instrumentation. Dr Teoh was a Vice Dean in the School of EEE, NTU from June 1996 till May 2005. In 1994, he was voted by the students as NTU Teacher of the Year Award for the School of EEE. In 2012, he received the Nanyang Award for Excellence in Teaching at NTU. He has published more than 250 journal and conference papers. In 2007, one of his papers titled "Lane Detection and Tracking Using B-Snake" by Wang Y., Teoh E.K. and Shen D.G. in Image and Vision Computing, Vol 22, No 4, April 2004, pp 269-280 won the Most Cited Paper Award for the journal Image and Vision Computing. His research interests include computer vision, pattern recognition, biometric recognition systems and medical image processing. He is a member of the IEEE.