Improving Automatic Age Estimation Algorithms using an Efficient Ensemble Technique

Alireza Keshavarz Choobeh

***Abstract*—The age is one of the most important features of a human being, and its estimation is essential in many applications, such as security control and surveillance monitoring. This work presents a framework for facial age estimation based on ensemble of individual age estimators. Both mathematical and experimental proofs show, if the individual age estimators are diverse in error, then to improve the results, we can make the ensemble age estimator using the best selected individual age estimators. We emphasize that although the experiments presented here are performed by neural networks, the proposed framework is readily applicable to any other regressor.**

***Index Terms*—Age estimation, ensemble technique, facial images, neural networks.**

I. Introduction

The problem of predicting the age from a face image has long been a focus of the computer vision and artificial and intelligence communities. Previous efforts on age estimation can be summarized into two types: 1) Age group classification, such as in [1], where a face is classified/labeled according to the age range, and 2) Exact age estimation, such as in [2]–[6], which attempts to guess the age of a human subject.

The first work in facial age group classification was proposed by Kwon and Lobo in [1]. Using measurements of facial features and calculating proportions, they were able to separate face images into three groups: infant, adult, and senior. Lanitis *et al*. [2] are credited with having done the first true work in exact age estimation. In their work, the Active Appearance Models (AAMs) were used for face encoding. Based on aging functions, they used a number of training images in order to learn the relationship between the coded representation of face images and the actual age of subjects in the corresponding images. A good survey of age estimation algorithms can be found in [7].

Age estimation has a large body of applications. Some applications that pass through the author's mind include the following.

- Preventing people to access on computer programs, internet websites and electronic devices that are harmful for specific ages.
- Indexing of face images based on the individual's ages for age-based retrieval of face images from videos.
- Summarizing videos focused on specific ages.
- Improving the accuracy and speed of face recognition systems.

In this paper has been shown that if the estimated ages of the best individual age estimators that are diverse in error are averaged, the results can be improved. The reminder of the paper is organized as follows. Section 2 gives an overview of the system. Aligning shapes using procrustes analysis will be introduced briefly in section 3. The piece-wise linear triangle based warping algorithm are presented in section 4. For the readers' convenience, the artificial neural networks will be introduced in section 5. Section 6 reviews some transforms for generating appropriate texture features. Section 7 gives a mathematical proof for efficiency of the proposed ensemble technique. The experimental results are presented in Section 8, and finally, Section 9 provides concluding remarks.

II. System Overview

The system diagram is shown in Fig. 1. The system inputs are facial shape and gray-level intensities of shape-free face images. The facial shapes are represented by the coordinates of a number of landmarks placed at predefined positions of the face images. Before using the facial shapes, they are aligned by Procrustes analysis to minimize shape variation. Then each image is warped to the mean shape using the piece-wise affine (linear) triangle based warping algorithm so that shape variation within the training set is eliminated and obtaining shape-free faces. The Modified Census Transform and the best 2-D Gabor filters are used for extracting facial texture features from internal facial region. The shape and texture are both vectorized for image representation. Principal Component Analysis (PCA), Artificial Neural Networks (ANNs), and uniform averaging make the other parts of the system. All the system parts will be described in more detail in the next sections.

III. Aligning Shapes Using Procrustes Analysis

There is considerable literature on methods of aligning shapes into a common coordinate frame, the most popular approach being Procrustes analysis [8], [9]. A shape can be aligned to another shape by applying this analysis which yield the minimum distance between the shapes. Procrustes analysis only determines a linear transformation (translation, reflection, orthogonal rotation, and scaling) to reach this goal.

Manuscript received March 12, 2012, revised March 28, 2012.

A. K. Choobeh is with the Young Researchers Club, Buinzahra Branch, Islamic Azad University, Buinzahra, Iran (e-mail: keshavarz_c@yahoo.com).

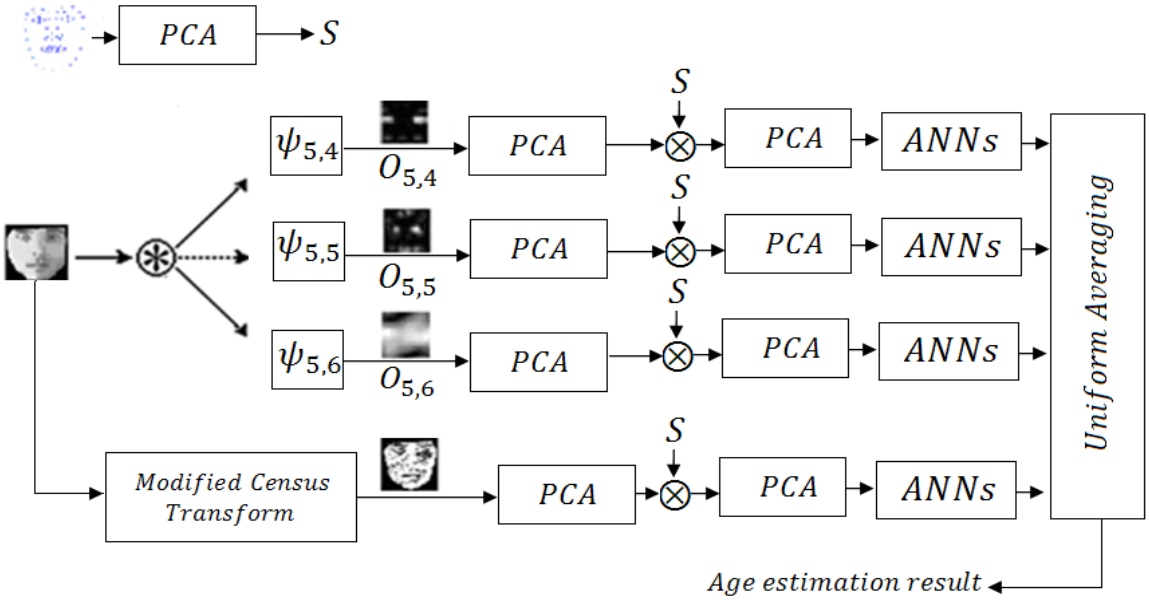

Fig. 1. Block diagram of the ensemble age estimator.

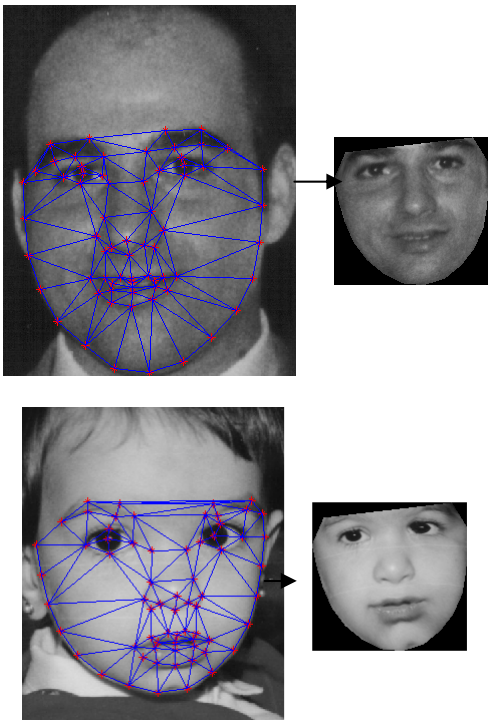

Fig. 2. Warping face images to the mean shape using a piece-wise linear triangle based warping algorithm.

IV. Piece-Wise Linear Triangle Based Warping Algorithm

Face warping to another face (here, mean face) is the process of overlaying the face to another face. This process consists of three steps [10]:

1. Select and match control points (landmarks) in the faces.
2. Determine the triangulation of control points in one of the faces. The triangulation of control points in the other face is automatically obtained from the correspondence between control points in the faces. See Fig. 2.
3. For each pair of corresponding triangles in the faces, using coordinates of three vertex points in two faces as $[(X_1,Y_1),(x_1,y_1)]$, $[(X_2,Y_2),(x_2,y_2)]$, and $[(X_3,Y_3),(x_3,y_3)]$, determine the linear mapping functions $X = f(x,y)$ and $Y = g(x,y)$ that will register the two triangles, where lowercase letters refer to the mean face. (1) Gives the equation of $X = f(x,y)$, using three pair of vertex points.

$$X - f(x,y) = Ax + By + CX + D = 0 \tag{1}$$

where

$$A = \begin{vmatrix} y_1 & X_1 & 1 \\ y_2 & X_2 & 1 \\ y_3 & X_3 & 1 \end{vmatrix} \quad B = -\begin{vmatrix} x_1 & X_1 & 1 \\ x_2 & X_2 & 1 \\ x_3 & X_3 & 1 \end{vmatrix} \quad C = \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix} \quad D = -\begin{vmatrix} x_1 & y_1 & X_1 \\ x_2 & y_2 & X_2 \\ x_3 & y_3 & X_3 \end{vmatrix}$$

The equation of $Y = g(x,y)$ is determined similarly. Given the coordinate of a point in the mean face (x,y), the X-component coordinate of the same point in the warping face can be determined using the linear mapping function of (1). The Y-component of the point is determined similarly. Warping typical face images to the mean shape is shown in Fig. 2.

V. Artificial Neural Networks

Supervised neural network, such as Multilayer Perceptron (MLP) [11] were used as the classifier to classify the parameters of the face based on his/her age. The network is presented with pairs of patterns i.e. the input pattern (the face model parameters) paired with target output (his/her age). Weights are adjusted to decrease the difference between the network's output and the target output. The training set (the set of input/target pattern pairs) is used for training, and is presented to the network many times. After training is stopped, the performance of the network is tested.

The Multilayer Perceptron (MLP) with the RPROP learning algorithm [12] was used in the experiments. The following pseudo-code gives the RPROP learning algorithm.

For all weight and biases{

if $(\frac{\partial E}{\partial w_{ij}}(t-1)\times\frac{\partial E}{\partial w_{ij}}(t)>0)$ then{

$\Delta_{ij}(t)=\text{minimum}\,(\Delta_{ij}(t-1)\times\eta^{+},\Delta_{\max})$

$\Delta w_{ij}(t)=-sign\,(\frac{\partial E}{\partial w_{ij}}(t))\times\Delta_{ij}(t)$

$w_{ij}(t+1)=w_{ij}(t)+\Delta w_{ij}(t)$

}

else if $(\frac{\partial E}{\partial w_{ij}}(t-1)\times\frac{\partial E}{\partial w_{ij}}(t)<0)$ then{

$\Delta_{ij}(t)=\text{maximum}\,(\Delta_{ij}(t-1)\times\eta^{-},\Delta_{\min})$

$w_{ij}(t+1)=w_{ij}(t)-\Delta w_{ij}(t-1)$

$\frac{\partial E}{\partial w_{ij}}(t)=0$

}

else if $(\frac{\partial E}{\partial w_{ij}}(t-1)\times\frac{\partial E}{\partial w_{ij}}(t)=0)$ then{

$\Delta w_{ij}(t)=-sign\,(\frac{\partial E}{\partial w_{ij}}(t))\times\Delta_{ij}(t-1)$

$w_{ij}(t+1)=w_{ij}(t)+\Delta w_{ij}(t)$

}

}

where w_{ij} is the weight from neuron j to neuron i, E is the error function, η^{+} and η^{-} are the increase and decrease factors respectively, Δ_{ij} is the update-value of the weight, w_{ij}, and $\Delta_{\max}$ and $\Delta_{\min}$ are the upper and lower limits of the update-values respectively.

VI. Feature Extraction and Generation

A. The Modified Census Transform

The modified census transform [13], [14] is a non-parametric local transform which is defined as an ordered set of comparison of pixel intensities in a local neighborhood. Let $N(x)$ and $\bar{I}(x)$ are a local spatial neighborhood of the pixel at x including it and the intensity mean on this neighborhood respectively. The modified census transform at x is defined using (2).

$$\Gamma(x)=\underset{y\in N}{\otimes}\xi(\bar{I}(x),I(y)) \tag{2}$$

where $\otimes$ is the concatenation operation and $\xi(I(x),I(x'))$ is a comparison function as shown using (3).

$$\xi(I(x),I(x'))=\begin{cases}1 & I(x)<I(x')\\ 0 & otherwise\end{cases} \tag{3}$$

B. 2D Gabor Filter

The Gabor filter captures salient visual properties such as spatial localization, orientation selectivity, and spatial frequency characteristics. Gabor filters were introduced to image analysis due to their biological relevance and computational properties [15], [16].

The 2D Gabor filters take the form of a complex plane wave modulated by a Gaussian envelope function [17]–[19]:

$$\psi_{u,v}(Z)=\frac{\|k_{u,v}\|^2}{\sigma^2}e^{-(\|k_{u,v}\|^2\|z\|^2/2\sigma^2)}\left[e^{ik_{u,v}z}-e^{-\frac{\sigma^2}{2}}\right] \tag{4}$$

where $k_{u,v}=\frac{k_{\max}}{f^{v}}e^{i\pi u/N}$, $z=(x,y)$, u and v define the orientation and scale of the Gabor filters, $k_{\max}$ is the maximum frequency, and f is the spacing factor between kernels in the frequency domain. $\sigma=2\pi$, $k_{\max}=\pi/2$ and $f=\sqrt{2}$.

Gabor features of a facial image are extracted based on its Gabor representations, which are obtained by convolving the facial image with Gabor filter. Let I(x,y) be an image, the convolution of I(x,y) and a Gabor filter is denoted as follows:

$$O_{u,v}(z)=I(z)*\psi_{u,v}(Z) \tag{5}$$

where $O_{u,v}(z)$ is the convolution result corresponding to the Gabor filter at orientation u and scale v. In this paper, for extracting discriminating information of different orientations and scales as much as possible, Wang *et al.* scheme was adopted [19]. In this scheme, the orientation and scale set is $U=\{u:o\le u\le 7\}$ and $V=\{v:o\le v\le 7\}$ respectively. Note that only the magnitude of Gabor filter response was used as the textural features.

C. The PCA Transform

Principal Component Analysis (PCA) has been successfully used in handprint recognition [20], face recognition [21]–[25], and robotics [26], [27]. Given a t-dimensional vector representation of each face image, the PCA [28] can be used to find a subspace whose basis vectors correspond to the maximum-variance directions in the original space. Let W represent the linear transformation that maps the original t-dimensional space onto a f-dimensional feature subspace where normally $f<<t$. The new feature vectors $y_i\in R^f$ are defined by $y_i=W^Tx_i$, i=1, …, N. The columns of W are the eigenvalues e_i obtained by solving the eigenstructure decomposition $\lambda_i e_i=Qe_i$, where $Q=XX^T$ is the covariance matrix of all N sample images $X=\{x_1,x_2,...x_N\}$ and λ_i the eigenvalue associated with the eigenvector e_i. Before obtaining the eigenvectors of Q: 1) the vectors are normalized such that $\|x_i\|=1$ and 2) the average of all images is subtracted from all normalized vectors.

VII. The Proposed Ensemble Technique and Mathematical Proof for its Efficiency

As mentioned before, our ensemble technique is the averaging operation on the best selected age estimators which

are diverse in error. The averaging operation is to take a linearly weighted summation of the individual estimator outputs. The output of the combination is:

$$f_{ens} = \sum_{i=1}^{L} w_i f_i(x_p) \tag{6}$$

Where L is the number of estimators, f_i is the output of the i th estimator that is referred to as the individual estimator, w_i is the corresponding non-negative real-valued combination weight, w_i is a testing datapoint, and f_{ens} is a convex combination of the individual estimators that is referred to as the ensemble estimator. The weights can be non-uniform but have the constraint that they sum to one: $\sum_{i=1}^{L} w_i = 1$; and also the uniform combination in (7), where all weights is equal at $w_i = \frac{1}{L}$.

$$\bar{f} = \frac{1}{L}\sum_{i=1}^{L} f_i(x_p) \tag{7}$$

If the ensemble is a uniformly weighted convex combination, we have the Bias-Variance-Covariance decomposition[29], [30]. So we have the mean squared error of the ensemble estimator $\bar{f}$ as (8).

$$\begin{aligned}
&E\{(\bar{f}-d)^2\} = \overline{bias}^2 + \frac{1}{L}\overline{\text{var}} + (1-\frac{1}{L})\overline{\text{cov}\,ar} \\
&\overline{bias} = \frac{1}{L^2}\sum_{i=1}^{L}\sum_{p=1}^{N}(E\{f_i(x_p)\} - d(x_p)) \\
&\overline{\text{var}} = \frac{1}{L^2}\sum_{i=1}^{L}\sum_{p=1}^{N}E\{(f_i(x_p) - E\{f_i(x_p)\})^2\} \\
&\overline{\text{cov}\,ar} = \frac{1}{L^3(L-1)}\sum_{\substack{i,j=1 \\ i\neq j}}^{L}\sum_{p=1}^{N}E\{f_i(x_p) - E\{f_i(x_p)\})(f_i(x_p) \\
&- E\{f_i(x_p)\})\}
\end{aligned} \tag{8}$$

where d the target value of an arbitrary testing datapoint and N is is the total number of training samples.

The generalization error of an individual estimator can be broken down into two components: bias and variance. These two usually work in opposition to each other. Attempts to reduce the bias component will cause an increase in variance and vice versa [30]. Techniques in the machine learning literature are often evaluated on how well they can optimize the trade-off between these two components. The bias can be characterized as a measure of how close your estimator is to its target. The variance is a measure of how stable the solution is. For example, in a slightly different training data, an estimator with high variance will tend to produce wildly varying performance. The covariance term is a measure of error correlation between individual estimators. We can see that the error of an ensemble of estimators depends critically on the amount of covariance term. As (8) shows, we would ideally like to decrease the covariance, without causing any increases in the $\overline{bias}^2$ or $\overline{\text{var}}$ terms. To achieve this goal, an averaging technique which is described in the following is proposed.

To improve the age estimation results by averaging technique in comparing to each individual age estimator, it is necessary that first, different representations of face images are provided to capture the error diversity between individual age estimators. Second, the best individual age estimators participate in averaging. The first decreases error covariance of ensemble estimator and the second decreases the error bias and variance of each individual age estimator. So based on the (8), the total error of ensemble estimator that comprises of these three terms decreases. From this point of view, different textural features from face images are evaluated. To generate textural features of face images, the total 64 Gabor filters and the Modified Census transform are used. The Gabor filter bank is selected based on minimum linear correlation between filters to reduce the redundancy of generated textural features. Selecting the Gabor filters based on the minimum linear correlation between filters is a factor to diversify the age estimation error between individual age estimators. Using each textural feature, a model of face appearance is built. The parameters of each model with Artificial Neural Networks (ANNs) classifier individually build the age estimators that have different errors in the age estimation results. Among the Gabor filters, some of them that have minimum age estimation errors are selected. The uniform average of the age estimation results of the individual age estimator that built using best Gabor filters and Modified Census transform make the age estimation system that improves the age estimation performance compared to all the component individual age estimators.

VIII. Experimental Results

To study of the automatic age estimation system, we used the FG-NET Aging Database [31]. Typical face images from this database are shown in Fig. 3. The age estimation performance are evaluated by the Mean Absolute Error (MAE) [2]–[5].

Fig. 3. Typical face images in the FG-NET aging database.

Sixty-eight landmark points and 18000 pixels from internal region of shape-free faces are used. In total, 130 textural features were applied to each of the ANNs. The ANNs architecture is as follows: 130 input units, 5 hidden units, which have been experimentally determined, and one output unit, representing the corresponding age of each face.

All neurons have linear threshold functions. Among the individual age estimators based on the Gabor filters, the best 2D Gabor filters $(u, v) = \{(5,4),(5,5),(5,6)\}$ achieves the mean absolute error of 5.82, 5.72, and 5.85 years respectively [32].

The MAE of the individual age estimator based on the Modified Census Transform is 5.49 years [33]. The ensemble age estimator improves the age estimation MAE compared to the individual age estimators up to 4.85 years.

IX. Conclusion

An ensemble age estimation algorithm which can improve individual age estimation algorithms is proposed. In the first step, several individual age estimators which are diverse in error are created. Different representation of face images using 2D Gabor filters and the Modified Census Transform are used to create diversity in error between the component individual age estimators. Among them, the four individual age estimators that have the best performance are selected. The four individual age estimators use a same classification model on different feature representations of same facial images. Averaging of the age estimation results of these best age estimators gives us an ensemble age estimator that has the better performace compared to all the individual estimators. Mathematical proofs garauntees further improvement of the enemble age estimation algorithm when the best individual age estimators are diverse in error.

Acknowledgment

The author wish to thank Dr. A. Lanitis for providing the FG-NET Aging Database.

References

[1] Y.H. Kwon and N. da Vitoria Lobo, "Age Classification from Facial Images," *Computer Vision and Image Understanding*, vol. 74, no. 1, pp. 1-21, 1999.

[2] A. Lanitis, C.J. Taylor, and T. Cootes, "Toward Automatic Simulation of Aging Effects on Face Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442-455, Apr. 2002.

[3] A. Lanitis, C. Draganova, and C. Christodoulou, "Comparing Different Classifiers for Automatic Age Estimation," *IEEE Trans. Systems, Man, and Cybernetics B*, vol. 34, no. 1, pp. 621-628, Feb. 2004.

[4] X. Geng, Z.-H. Zhou, K.S. Miles, "Automatic Age Estimation Based on Facial Aging Patterns," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2234-2240, Dec. 2007.

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

[6] G. Guo, Y. Fu, C. Dyer, and T.S. Huang, "Image-Based Human Age Estimation by Manifold Learning and Locally Adjusted Robust Regression," *IEEE Trans. Image Processing*, vol. 17, no. 7, pp. 1178-1188, July 2008.

[7] Y. Fu, G. Guo, and T.S. Huang, "Age Synthesis and Estimation via Faces: A Survey," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2234-2240, Nov. 2010.

[8] C. Goodall, "Procrustes methods in the statistical analysis of shape," *Journal of the Royal* Statistical Society B, pp. 285-339, 1991.

[9] J. C. Gower, "Generalized Procrustes analysis," *Psychometrika 40*, pp. 33-51, 1975.

[10] A. Goshtasby, "Piecewise linear mapping functions for image registration," *Pattern Recognition*, Vol. 19, no. 6, pp. 459-466, 1986.

[11] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, "Learning representations by back-propagation error," *Nature*, vol. 323, no. 9, pp. 533–536, 1986.

[12] M. Riedmiller and H. Braun, "A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm," *Proc. IEEE Int'l Conf. Neural Networks (ICNN '93), H. Ruspini, ed.*, pp.586-591, 1993.

[13] B. Froba and A. Ernst, "Face detection with the modified census transform," *Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 91–96, May 2004.

[14] Christian Kublbeck, and Andreas Ernst, "Face detection and tracking in video sequences using the modified census transformation," *Image Vis. Comput., vol*. 24, pp. 564-572, August 2005.

[15] S. Marcelja, "Mathematical description of the responses of simple cortical cells," *J. Opt. Soc. Amer.*, vol. 70, pp. 1297–1300, 1980.

[16] J. Jones and L. Palmer, "An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex," *J. Neurophys.*, pp. 1233–1258, 1987.

[17] J. G. Daugman, "Two-dimensional spectral analysis of cortical receptive field profiles," *Vis. Res.*, vol. 20, pp. 847–856, 1980.

[18] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. Comput.*, vol. 42, pp. 300–311, 1993.

[19] L.Wang, Y.Li, and C.Wang, "2D Gaborface representation method for face recognition with ensemble and multichannel model" *Image and Vision Computing*, vol. 26, pp. 820–828, 2008.

[20] H. Murase, F. Kimura, M. Yoshimura, and Y. Miyake, "An Improvement of the Autocorrelation Matrix in Pattern Matching Method and Its Application to Handprinted `HIRAGANA'," *Trans. IECE Japan*, vol. 64D, no. 3, 1981.

[21] L. Sirovich and M. Kirby, "A Low-Dimensional Procedure for the Characterization of Human Faces," *J. Optical Soc. Am. A*, vol. 4, no. 3, pp. 519-524, 1987.

[22] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103-108, Jan. 1990.

[23] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.

[24] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696-710, July 1997.

[25] A. M. Martinez, "Recognition of Partially Occluded and/or Imprecisely Localized Faces Using a Probabilistic Approach," *Proc. Computer Vision and Pattern Recognition*, vol. 1, pp. 712-717, June 2000.

[26] S.K. Nayar, N.A. Nene, and H. Murase, "Subspace Methods for Robot Vision," *IEEE Trans. Robotics and Automation*, vol. 12, no. 5, pp. 750-758, 1996.

[27] J.J. Weng, "Crescepton and SHOSLIF: Towards Comprehensive Visual Learning," *Early Visual Learning, S.K. Nayar and T. Poggio, eds.*, pp. 183-214, Oxford Univ. Press, 1996.

[28] K. Fukunaga, Introduction to Statistical Pattern Recognition, second ed. Academic Press, 1990.

[29] N. Ueda and R. Nakano, "Statistical analysis of the generalization error of ensemble estimators, " *in Int. Conf. on Neural Networks, ICNN96*, pp. 90-95, 1996.

[30] G. Brown, "Diversity in Neural Network Ensembles," Ph.D. Thesis, University of Birmingham, 2003.

[31] The FG-NET Aging Database, http://www.fgnet.rsunit.com/, http://www-prima.inrialpes.fr/FGnet/, 2010.

[32] A.K. Choobeh, "Human Age Estimation Using 2D Gabor Filter," *in* 3rd *International Conference on Mechanical and Electrical Technology, ICMET*, vol. 1, pp. 599-602, July 2011.

[33] A.K. Choobeh, and M.J.P. Jalali, "Automatic Human Age Estimation Based on Neural Networks and the Modified Face Model" *in 3rd International Conference on Software Technology and Engineering, ICSTE*, pp. 111-115, July 2011.

Alireza Keshavarz Choobeh was born in Karaj, Iran, on August 3, 1982. He received the B.S. degree in electronics engineering from Mazandaran University, Babol, Iran, in 2005. He received the M.S. degree in biomedical engineering from Tarbiat Modares University, Tehran, Iran, in 2009.

He is currently a researcher with the Young Researchers Club, Azad University, Iran. His research interests include computer vision and signal processing.