ORIGINAL ARTICLE

Human age estimation framework using different facial parts

Mohamed Y. El Dib *, Hoda M. Onsi

Faculty of Computers and Information, Cairo University, Cairo, Egypt

Abstract   Human age estimation from facial images has a wide range of real-world applications in human computer interaction (HCI). In this paper, we use the bio-inspired features (BIF) to analyze different facial parts: (a) eye wrinkles, (b) whole internal face (without forehead area) and (c) whole face (with forehead area) using different feature shape points. The analysis shows that eye wrinkles which cover 30% of the facial area contain the most important aging features compared to internal face and whole face. Furthermore, more extensive experiments are made on FG-NET database by increasing the number of missing pictures in older age groups using MORPH database to enhance the results.

© 2011 Faculty of Computers and Information, Cairo University.
Production and hosting by Elsevier B.V. All rights reserved.

1. Introduction

Human face serves as a knowledge base for a lot of useful information such as gender, expression and age. People can extract

* Corresponding author. Tel.: +20 101926931.
E-mail addresses: mohamed.y.eldib@gmail.com (M.Y. El Dib), drhoda2002a@hotmail.com (H.M. Onsi).

Production and hosting by Elsevier

gender or expression easily. But in the case of estimating human ages, they may fail to guess intelligently the age due the wide degree of variation in both shape and texture. Facial appearances differ on individual bones and are affected by deformation changes in expression, speaking and age progression, as well as lighting variations. Due to its numerous challenges, there has been relatively little work, to date, concerning automatic age estimation, despite having potential useful applications like an electronic consumer relationship management and demographic data collection. In this paper, age estimation framework is performed based on the biologically-inspired features (BIF) that showed promising results for age estimation problem [1,2] as well as gender recognition [3] problem.

Our main contributions include (1) analyzing and estimating human ages from eye wrinkles, whole internal face (without forehead area) and whole face (with forehead area); and (2) conducting extensive experiments on FG-NET database by increasing the number of missing pictures in older age groups using MORPH database to enhance the results.

The paper is organized as follows. In Section 2; we present the related work. Section 3 gives an overview of bio-inspired model. Section 4 explains the proposed age estimation framework. Section 5 describes our results on the publicly available FG-NET [4]. Section 6 shows result of increasing older age groups in FG-NET aging database using MORPH [5] dataset. Finally, Section 7 draws some conclusions and discusses our future work.

2. Related work

There is plentiful existing work on the facial aging progress, originating from psychological and biological studies. However, most of it aims at simulating the aging effects on human faces [6] (i.e., simulate how the face would look like at a certain age), which is the inverse procedure of age estimation. Age estimation approaches fall into two categories: (a) classification-based [7–10] and (b) regression-based [1,6,10–13]. An example of classification-based work is by [8] where an anthropometric model has been for age classification based on cranio-facial development theory and skin wrinkle analysis, with human faces finally classified into three groups: babies, young adults and senior adults. Geng et al. [7] defines the AGing pattErn Subspace (AGES) method which studies a subspace representation of aging sequences and estimating age by projecting the test face into the subspace.

In the second category of works, age estimation is viewed as a regression problem where facial features are extracted by the active appearance models (AAMs) [14] that incorporate shape and appearance information together. An input face image is then represented by a set of fitted model parameters. The regression coefficients are estimated from training data with an assumption of the regression function such as a quadratic model (QM) [9]. Yanet al. [12,13] also dealt with age uncertainty by formulating a semi-definite programming problem [13] or utilizing an EM-based algorithm [12] where they adopted traditional discriminative methods, using image intensities directly or other features exhaustively extracted from images. Suo et al. [6] presented compositional and dynamic models which decompose a face into parts and represent the face aging process dynamics as a first-order Markov chain on sparse graphs. On the front of features utilized in age estimation, the recently proposed BIF have shown very promising results [1,2]. Guo et al. [1] investigated the biologically-inspired features (BIF) for human age estimation from faces. Their main contributions included (1) Gabor filters with smaller sizes, (2) a new operator "STD" to encode the aging subtlety on faces. The main drawback of their work is that the facial landmarks for the face image were performed manually. Guo et al. [2] conducted a comprehensive study using BIF with manifold learning techniques and designed three frameworks for automatic age estimation that exhibit high performance that does not require manual separation of males and females prior to age estimation.

3. Bio-inspired model

Recent work suggests that visual processing in the cortex can be modeled as a hierarchy of increasingly sophisticated representations [15]. A recent theory suggests that the feed-forward
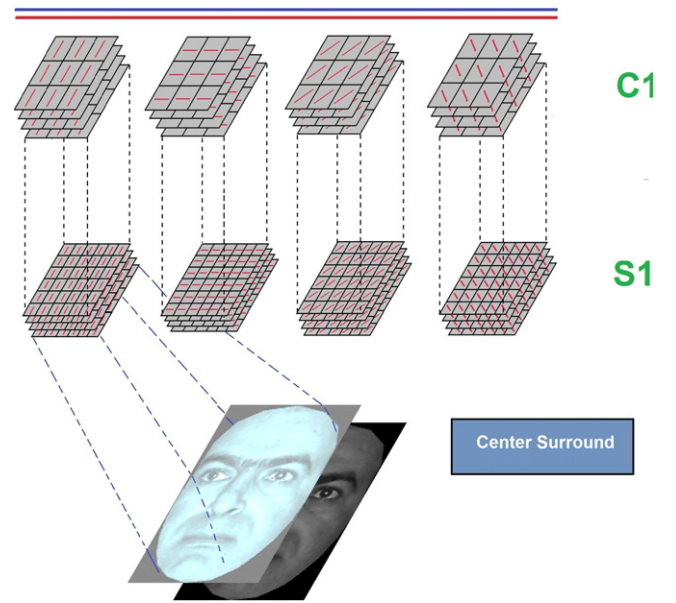

Figure 1 Bio-inspired model.

path of object recognition in the cortex accounts for the first few hundred milliseconds of visual processing in primate cortex which follows a mostly feed-forward hierarchy [15] [16]. Riesenhuber et al. [15] proposed a new hierarchical model derived from a feed-forward model of the primate visual object recognition pathway, called the "HMAX" model.

The standard framework as shown in Fig. 1 consists of different layers of computational units called simple (S) and complex (c) cell units creating increasing complexity as the layers progress from the primary visual cortex (V1) to inferior temporal cortex (IT) [1,15]. A notable property of the model is the nonlinear maximum operation "MAX" over the S units rather than the linear summation operation "SUM" in pooling inputs at the c layers. Specifically, the first layer of the model, called the S_1 layer, is created by convolving a pyramid of Gabor filters at 4 orientations and 16 scales, over the input gray-level image. Adjacent two scales of S_1 units are then grouped together to form 8 "bands" of units for each orientation. The second layer, called the c_1 layer, is then generated by taking the maximum values within a local spatial neighborhood and across the scales within a band. So the resulted c_1 representation contains 8 bands and 4 orientations. The advantage of taking the "MAX" operation within a small range of position and scale is to tolerate small shifts and scale changes [1].

4. Proposed age estimation framework

As shown in the diagram of Fig. 2, the proposed age estimation algorithm consists of two main stages, namely (1) preprocessing stage and (2) age estimation process.

4.1. Framework structure

In preprocessing stage, the facial landmarks for the whole face, internal face and eye wrinkles are detected automatically using Active Shape Model (ASM) block. The images are cropped to
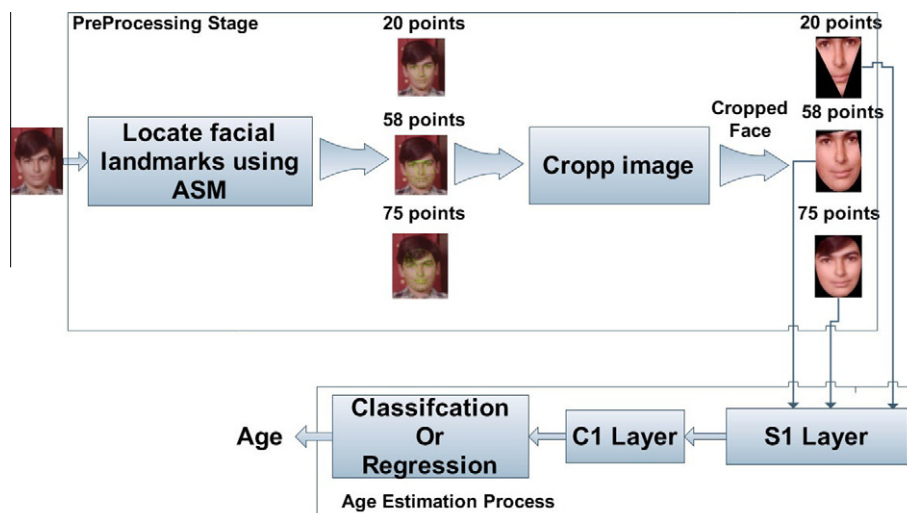
Figure 2   Age estimation framework.

the area covering a fixed number of points generated from the ASM stage (several numbers of points were tested experimentally). Then, in the age estimation process, the cropped images undergo filtering by a family of Gabor functions at different orientations and scales using S_1 block. The filtered outputs undergo a feature dimensionality reduction step by keeping the maximum (MAX) and standard deviations (STD) of the Gabor filtered outputs using c_1 block. Finally, both classification-based and regression-based models were used in the training phase (Support Vector Machine SVM and Support Vector Regression SVR in this case) to produce the final age model estimator.

4.2. Face feature localizer

In this step, we aim at accurately localizing the facial region to extract features only from the relevant parts of the input image. We use Active Shape Models [17] for automatic localization of facial landmark points in two main stages, namely training and fitting. In the training stage, we manually locate

landmark points for hundreds of images [18] in such a way that each landmark represents a distinguishable point presented on every example image. We aim at analyzing eye wrinkles using 20 points, internal face area using 58 points and whole face using 75 points which were provided by [18,19], respectively. The purpose of this analysis is to determine the locations of the most important aging features using eye wrinkles or internal face only rather than using the whole face area. Finally, we build three separate statistical shape models (1) 75 points shape model; (2) 20 points shape model; and (3) 58 points shape model by ASM based on the annotated images. Where, the used points that describe each facial part are uncorrelated to each other.

Having built the shape models, points on the incoming face image are fitted. First, we detect the whole face, eye wrinkles and internal face in the input image using 75, 20 and 58 points shape models, respectively. Second, we initialize the shape points and align images to obtain automatically the detected shape features. Finally the input image is cropped to the area covered by the ASM fitted landmark points. The difference

Figure 3   58, 20 and 75 points samples from FG-NET.

between using 75, 20 and 58 landmark points is illustrated in Fig. 3.

4.3. Face representation

Texture features have proven to be distinctive for the task of age estimation from facial images [1,20]. Particularly, the use of Gabor features has proven to be successful as in [1,2,20]. We use the bio-inspired features (BIF) to capture aging patterns which have two layers S_1 and c_1.

S_1 is created with a Gabor filtering on the cropped image output from the Active Shape Model (ASM) block [17–19] with 8 orientations and 16 scales. Gabor functions for a particular scale (sigma) and orientation (theta) are described by the following equation:

$$G(x, y) = \exp\left(-\frac{(X^2 + \gamma^2 Y^2)}{2\sigma^2}\right) \times \cos\left(\frac{2\pi}{\lambda} X\right) \quad (1)$$

where $X = x\cos\theta + y\sin\theta$ and $Y = -x\cos\theta + y\sin\theta$.

We adjusted θ to vary between 0 and π. The parameters λ (wavelength), σ (effective width), and γ (aspect ratio) = 0.3 are based on the work in [1]. The starting filter size is chosen to be (3×3) due to experiments selection. Fig. 4 shows the output of S_1 layer for eye wrinkles, internal face and whole face.

Gabor filtered outputs can serve as candidate features for the age estimation problem. However, they are of a very high dimension leading to difficulties in training. In addition, there are redundancies in the Gabor filter outputs. Hence, a usually adopted scheme is to summarize the outputs of the Gabor filters using some statistics measure. Here, we adopt the second layer C_1 that is used in [1] and proven to work quite well. It has two operations, namely the maximum "**MAX**" and standard deviations "**STD**" with a variation on the MAX definition by avoiding image subsampling to keep local variations which might be important for characterizing facial details (e.g., wrinkles, creases and muscles drop).
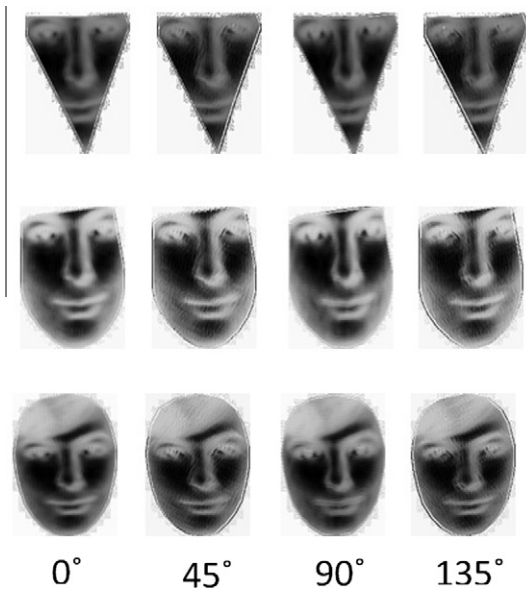

Figure 4 Gabor filtered results at band 1 with filter size 3×3 at four orientations for eye wrinkles, internal face and whole face.

Table 1 SVR and SVM model parameters.

SVR	SVM
$C = 1,5,30,60$ (ages < 30)	$C = 1000$
$C = 240,1000$ (ages > 30)	
$\gamma = 0.000308$, $\epsilon = 0.1$, $e = 0.002$	

$$F_i = \max(x_i^j, x_i^{j+1}) \quad (2)$$

where F_i corresponds to the maximum value of two adjacent filters in the same scale in the S_1 layer band at pixel i

$$\text{std} = \sqrt{\frac{1}{N_s \times N_s} \sum_{i=1}^{N_s \times N_s} (F_i - \overline{F})^2} \quad (3)$$

where x_i^j and x_i^{j+1} are the filtered values with scales j and $j + 1$ at pixel i. $\overline{F}$ is the mean value of the filtered values within pooling grid size of $N_s \times N_s$.

4.4. Classification or regression

Age estimation can be treated as a classification problem, when each age is considered as a class label. Alternatively, age estimation can be treated as a regression problem, where each age is considered a regression value. In our experiments, we use both SVR and SVM methods for age estimation on the FG-NET [4] and the MORPH [5] standard databases. The RBF SVR can address the three limitations of the traditional quadratic regression model [21]: (1) the simple quadratic function may not model the complex aging process, especially for a large span of years, e.g., 0–70; (2) the least square estimation is sensitive to outliers that come from incorrect labels when collecting a large image database; and (3) the least square estimate criterion only minimizes the empirical risk which may not generalize well for unseen examples [10,21].

A face feature localizer is used to find the facial landmarks of eye wrinkles, internal face and whole face in each image using Active Shape Model stage (ASM). Then, the images are cropped and resized to 59×80 gray-level images. For the face representation; we use the biologically-inspired features to encode the aging features on each facial part. We build six SVR models and one SVM model using the experimentally selected parameters provided in Table 1. Using SVR or SVM separately cannot adequately estimate age because of the diversity of the aging process across different ages. Hence, we combine SVR and SVM models by selecting which model to use over each age group, based on MSE results over the training. The age of the test image is predicted using a cascade of SVM and SVR models by taking the average over the estimated ages as shown in Fig 5. Then, based on the decision nodes, the final age is estimated.

5. Experiments

5.1. Evaluation measures

Two measures to evaluate age estimation performance: (1) Mean Absolute Error (MAE) and (2) Cumulative score (CS). The MAE is defined as the average of the absolute errors between the estimated ages and the ground truth ages.
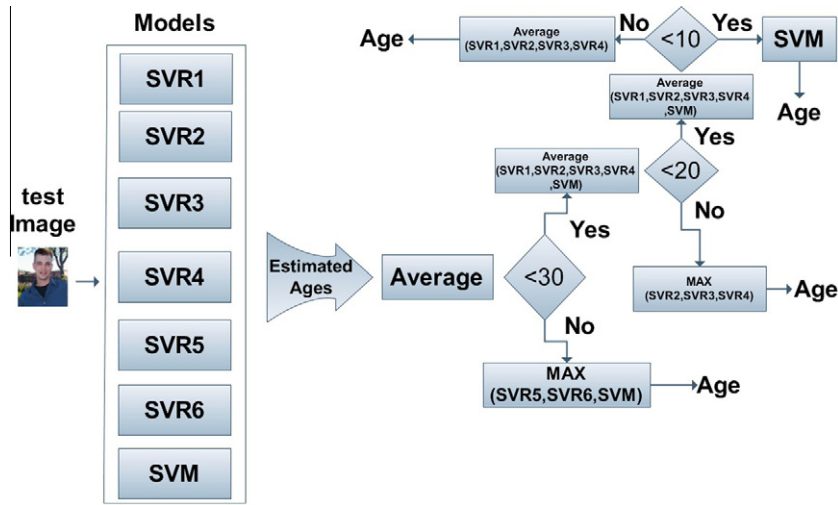
Figure 5 Age estimation process for test images using six SVR and one SVM model.

$$\text{MAE} = \sum_{k=1}^{N} \frac{|\hat{l}_k - l_k|}{N} \quad (4)$$

where l_k is the ground truth age for the test is image k and $\hat{l}_k$ is the estimated age and N is the total number of test images. The cumulative score CS(j) is defined as $\frac{N_{esi}}{N} \times 100\%$ where N_{esi} is the number of test images on which the age estimation makes an absolute error no higher than j years.

5.2. Datasets

The FG-NET [4] and MORPH [5] aging databases are used, which are publicly available to evaluate the performance of our approach. The FG-NET contains 1002 face images of 82 subjects with ages ranging from 0 to 69 with large variation of lighting, pose, and expression. While, The MORPH database is organized into albums, album 1 contains 1690 face images of 515 subjects ranging in age from 15 to 68 years for men and women of various ancestry groups. While, album 2 contains more than 20,000 face images obtained from more than 4000 individuals. We use the BioID Face Database [19] which has a large variety of illumination, background and face size. It contains 1521 gray-level images with a resolution of 384×286 pixels. Each one shows the frontal view of a face of one out of 23 different test persons with 20 landmark points covering the area under the eyes.

5.3. Results and analysis

A Leave-One-Person-Out (LOPO) test strategy is used on the FG-NET database to evaluate the performance of eye wrinkles, internal face and whole face.

Experimental results are shown in Table 2 on FG-NET database. The MAEs of eye wrinkles and internal face are 3.71 and 3.69, respectively, where they are close to the MAE of whole face 3.17. So, if 75 points cover the whole face with 100% coverage, internal face (58 points) covers 70% of the face and eye wrinkles (20 points) covers 30% of the face. It is clear that, eye wrinkles contain the most important aging features compared to internal face and whole as shown also from the values of MAEs.

Table 2 MAE (years) at different age groups for whole face, eye wrinkles and internal face on FG-NET.

Range level	Rng	#img.	Internal face	Eye	Whole face
1	0–9	372	1.67	2.11	1.73
2	10–19	338	1.09	1.20	1.42
3	20–29	144	3.31	3.74	3.58
4	30–39	79	7.87	9.17	9.19
5	40–49	46	13.02	15.47	16.53
6	50–59	15	18	21.78	14.00
7	60–69	8	26.25	36.43	33.85
Total		1002	**3.17**	**3.71**	**3.69**

Finally, In the FG-NET database, our facial analysis has MAE of 3.17, 3.71 and 3.69 years for whole face, eye wrinkles and internal face respectively. These average errors are substantially smaller than the RUN method [13] (5.78 years), BM [12] (5.33 years), LARR [10] (5.07 years), and even significantly lower than the very recent RPK [1] (4.95 years) and BIF [1] (4.77 years) approaches which were announced to be the best reported results so far. See Table 3 for more methods. CS curves are similarly shown in Fig. 6 for eye wrinkles, internal face and whole face.

Table 3 MAE (years) comparisons.

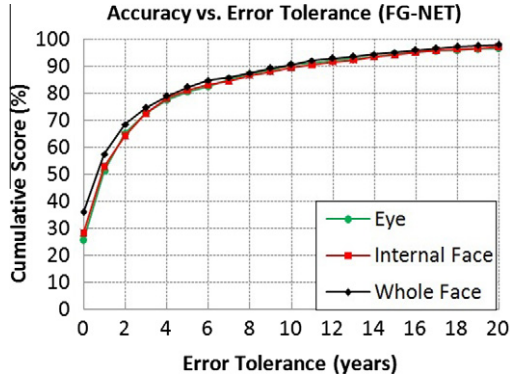| Method	FG-NET
WAS [5]	8.06
AGES [5]	6.77
QM [7]	6.55
AGES$_{ida}$ [5]	6.22
RUN [11]	5.78
BM [10]	5.33
LARR [8]	5.07
PFA [12]	4.97
RPK [1]	4.95
BIF [1]	4.77
Eye wrinkles (20 points)	**3.71**
Internal face (58 points)	**3.69**
Whole face (75 points)	**3.17**

Figure 6 Cumulative scores for eye wrinkles, internal face and whole face at absolute error levels from 0 to 20 years.

FG-NET has a small number of face images in the old range group (8 face images in the age range of 60–69 range level 7 and 15 face images in 50–59 range level 6), while it has a large number of face images in the younger range groups (range level 1–372 face images in the age range 0–9). This clarifies why classifiers like SVM or SVR cannot work well on the FG-NET database. This also tells why older age groups have much larger errors in those range levels as shown in Table 2.

We overcome this limitation by using the MORPH database to increase the number of missing pictures in each age group using the following assumptions:

1. Largest numbers of images exist in age group (0–9). So, we cover each age group with pictures equal to 372 images as in range level 1. By this way, the number of pictures are equally distributed in each age group.
2. We use MORPH dataset for range levels completion.
3. The images used to complete the missing age groups were chosen in a predefined manner. It is illustrated in the next section.
4. We complete all range levels except range level 1 and 2, because, subjects in MORPH are ranging in age from 15 to 68 years.
5. We cannot use ages from 15 to 19 in range level 2 due to the missing ages ranging from 10 to 14 in range level 2 which will affect on the integrity and fairness of the distribution of this range level.

6. MORPH pictures selection criteria

We use album 1 and 2 for selecting pictures for each range level. First, we group all pictures related to each age group separately. Then, the number of missing pictures in each age group in FG-NET is found by the following equation:

$$m_j = \mathrm{MAX}(AG_i) - AG_j \tag{5}$$

where m_j denotes the total number of missing images to be completed for Age Group j, $\mathrm{MAX}(AG_i)$ denotes maximum number of images for Age Group i and AG_j denotes the number of images for Age Group j.

Next, we need to find the number of missing pictures for each age inside each age group, which is described by the equation:

$$m_{nj} = \frac{m_j}{\mathrm{MAX}(A_j) - \mathrm{MIN}(A_j)} \tag{6}$$

where m_{nj} denotes the number of missing pictures to be completed for age n for Age Group j. $\mathrm{MAX}(A_j)$ denotes the maximum Age for Age Group j. While, $\mathrm{MIN}(A_j)$ denotes the minimum Age for Age Group j.

For every m_{nj}, we choose whites and blacks using the following equations:

$$W_{nj} = m_{nj}/2 \tag{7}$$

$$B_{nj} = m_{nj} - W_{nj} \tag{8}$$

where W_{nj} denotes the number of pictures for whites for Age n and Age Group j. B_{nj} denotes the number of pictures for blacks for Age n and Age Group j.

For every W_{nj} and B_{nj}, we choose males and females as described in the following equations:

$$\mathrm{MAW}_{nj}^l = W_{nj}/2 \tag{9}$$

$$\mathrm{FEW}_{nj}^l = \mathrm{MAW}_{nj}^l - W_{nj} \tag{10}$$

$$\mathrm{MAB}_{nj}^l = B_{nj}2 \tag{11}$$

$$\mathrm{FEB}_{nj}^l = \mathrm{MAB}_{nj}^l - B_{nj} \tag{12}$$

where MAW_{nj}^l and MAB_{nj}^l denote the number of males for whites and blacks l, Age n and Age Group j. FEW_{nj}^l and

Table 4 MAE (years) for face (75 points). Where F + M denote FG-NET combined with the injected pictures from MORPH. F indicates only FG-NET without MORPH. FACE includes all range levels except range level 1 and 2. While, FACE (All) include all range levels.

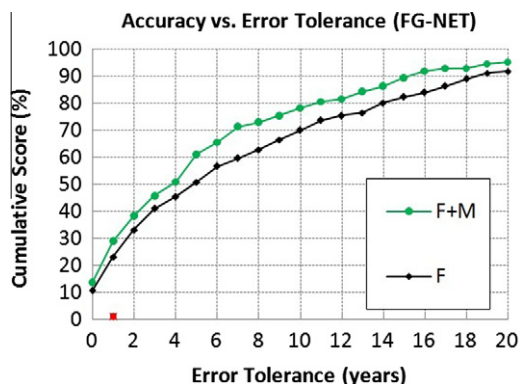| Range level	Range	F + M				F			
		FACE		FACE (All)		FACE		FACE (All)	
		#img.	MAE	#img.	MAE	#img.	MAE	#img.	MAE
1	0–9	–	–	372	2.17	372	–	372	1.69
2	10–19	–	–	338	3.38	338	–	338	1.09
3	20–29	372	2.86	372	2.86	144	3.31	144	3.31
4	30–39	372	6.32	372	6.32	79	7.87	79	7.87
5	40–49	372	11.32	372	11.32	46	13.02	46	13.02
6	50–59	360	16.86	360	16.86	15	18	15	18
7	60–69	262	21.12	262	21.12	8	26.25	8	26.25
Total		1738	**6.35**	2448	**3.31**	292	**7.80**	1002	**3.17**

Figure 7 Cumulative scores for FG-NET with MORPH datasets and FG-NET only at absolute error levels from 0 to 20 years.

FEB_{nj}^{l} denotes the number of females for white and blacks l, Age n and Age Group j.

6.1. Results of MORPH and FG-NET datasets

The total number of chosen images from MORPH dataset is 1446 images for range levels 3–7. Due to the random nature of the selection criteria and the limited number of pictures for each subject (2–3 images); MORPH is used as a training set only with other pictures from the LOPO strategy on the FG-NET dataset. For example, FG-NET has 82 subjects (1002 images) to use the LOPO strategy. We exclude one subject of 82 subjects who will be used as a testing set. This leaves the training set with 81 subjects + (1446 images from MORPH).

Table 4 gives the performance evaluation for using larger dataset by completing the missing images in age groups. F + M denote the combination of MORPH and FG-NET and F denote FG-NET only. FACE column contains all range levels except range level and 2 while FACE (ALL) contains all range levels.

In FACE (ALL), for F + M, the MAEs are decreased in range levels 3–7 and the performance is enhanced in these age groups significantly compared to the same age groups in F. The MAEs in range levels 1 and 2 for F are lower than the same range levels in F + M, because MORPH dataset lacks pictures in younger ages for range levels 1 and 2. In FACE, the MAE result on F + M is 6.35 which is significantly less than the 7.80 on F only. CS curves are shown in Fig. 7 for F + M (FG-NET + MORPH) and F (FG-NET only) for range levels 3–7 for whole face (75 points). In conclusion, completing the missing ranges with MORPH pictures enhanced the performance in FG-NET older age groups.

7. Conclusion and future work

We presented a detailed analysis for different facial parts such as eye wrinkles, internal face and whole face using different shape feature points. We used the bio-inspired features (BIF) to encode the aging features in each face part. The analysis shows that eye wrinkles contain almost as much information about the age compared to internal face or whole face areas despite the fact that such region constitutes 30% of the total facial area. This means that the study of the eye region can give

away the age almost as accurately as studying wrinkles in the whole face, which reflects positively on reducing the processing time. In addition, we enhanced the performance of our age estimator by increasing the number of missing images in FG-NET dataset age groups. From the extensive experiments, it is shown that the MAE of FG-NET + MORPH datasets was decreased by 20% compared to MAE of FG-NET only. As future work is concerned, we aim at exploring other more powerful modeling techniques that can statistically model the relation between numbers of variables such as Bayesians networks.

References

[1] Guo G, Mu G, Fu Y, Huang TS. Human age estimation using bio-inspired features. Proc IEEE Conf CVPR 2009:112–9.

[2] Guo G, Mu G, Fu Y, Dyer C, Huang TS. A study on automatic age estimation using a large database. Proc IEEE Conf ICCV 2009:1986–91.

[3] Guo, G. Mu, Y. Fu, C.R. Dyer, T. S. Huang, Is gender recognition influenced by age? Proc IEEE Int Workshop HCI, in conjunction with ICCV, 2009.

[4] FG-NET aging database. <http://www.fgnet.rsunit.com/>.

[5] Ricanek K, Tesafaye T. MORPH: a longitudinal image database of normal adult age-progression. Proc Automatic Face Gesture Recognit 2006:341–5.

[6] Suo J, Zhu S, Shan S, Chen X. A compositional and dynamic model for face aging. Proc IEEE Trans PAMI 2009;32(3):385–401.

[7] Geng X, Zhou Z-H, Smith-Miles K. Automatic age estimation based on facial aging patterns. Proc IEEE Trans PAMI 2007;29(12):2234–40.

[8] Kwon Y, Lobo N. Age classification from facial images. Proc IEEE Conf CVPR 1994:762–7.

[9] Lanitis A, Draganova C, Christodoulou C. Comparing different classifiers for automatic age estimation. Proc IEEE Trans SMC-B 2004;34(1):621–8.

[10] Guo G, Fu Y, Huang TS, Dyer C. Locally adjusted robust regression for human age estimation. Proc IEEE Conf WACV 2008:1–6.

[11] Lanitis A, Taylor C, Cootes T. Toward automatic simulation of aging effects on face images. Proc IEEE Trans PAMI 2002;24(4):442–55.

[12] Yan S, Wang H, Huang TS, Tang X. Ranking with uncertain labels. Proc IEEE Conf Multimedia Expo 2007:96–9.

[13] Yan S, Wang H, Tang X, Huang T. Learning auto-structured regressor from uncertain nonnegative labels. Proc IEEE Conf ICCV 2007:1–8.

[14] Fu Y, Xu Y, Huang TS. Estimating human ages by manifold analysis of face pictures and regression on aging features. Proc IEEE Conf Multimedia Expo 2007:1383–6.

[15] Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. Proc Nat Neurosci 1991;2(11):1019–25.

[16] Serre T, Wolf L, Bileschi S, Riesenhuber M, Poggio T. Robust object recognition with cortex-like mechanisms. Proc IEEE Conf Multimedia Expo 2007;29(3):411–26.

[17] Cootes TF, Taylor CJ, Cooper DH, Graham J. Active shape models – their training and application. Proc CVIU 1995;61(1):38–59.

[18] Wei Y. Research on facial expression recognition and synthesis. Master Thesis 2009.

[19] BIOID database. <http://www.bioid.com/>.

[20] Mutch J, Lowe D. Object class recognition and localization using sparse features with limited receptive fields. Proc IEEE Conf CVPR 2006:11–8.

[21] Vapnik VN. Statistical learning theory. New York: John Wiley; 1998.