Human Age Estimation Using Bio-inspired Features

Guodong Guo
NCCU
gdguo@nccu.edu

Guowang Mu
NCCU
gmu@nccu.edu

Yun Fu
BBN Technologies
yfu@bbn.com

Thomas S. Huang
UIUC
huang@ifp.uiuc.edu

Abstract

We investigate the biologically inspired features (BIF) for human age estimation from faces. As in previous bio-inspired models, a pyramid of Gabor filters are used at all positions of the input image for the S_1 units. But unlike previous models, we find that the pre-learned prototypes for the S_2 layer and then progressing to C_2 cannot work well for age estimation. We also propose to use Gabor filters with smaller sizes and suggest to determine the number of bands and orientations in a problem-specific manner, rather than using a predefined number. More importantly, we propose a new operator "STD" to encode the aging subtlety on faces. Evaluated on the large database YGA with 8,000 face images and the public available FG-NET database, our approach achieves significant improvements in age estimation accuracy over the state-of-the-art methods. By applying our system to some Internet face images, we show the robustness of our method and the potential of cross-race age estimation, which has not been explored by any studies before.

1. Introduction

Human face, as a window to the soul [37], conveys important perceptible information related to individual traits. Human age, as an important personal trait [10, 39], can be directly inferred by distinct patterns emerging from the facial appearance. People have the ability, developed early in life, to determine age between 20 and 60 years and conceive aging appearance from the face with high accuracy [1, 13]. Can machine do the same job? Recent studies have given a positive answer to this question by defining the *age estimation* technique which is to label a face image automatically by machine with the exact age (year) or the age group (year range) of the individual face.

Age estimation by machine is useful in scenarios where we don't need to specifically identify the individual, but want to know his or her age. Derived from rapid advances in computer vision and pattern recognition, computer-based age estimation via faces become a particularly interesting topic recently because of the explosively emergent real-world applications, such as electronic customer relationship management (ECRM) [2, 18], security control and surveillance monitoring [14, 20, 24], biometrics [25, 23], and entertainment.

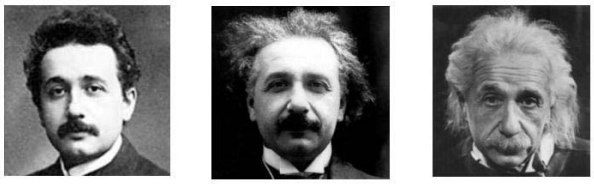

Figure 1. Age estimation on Einstein's faces (obtained from the Internet) using our method. The estimated ages below each face might be a little bit older than the true ages (unknown to us) but reasonable. Our training data are all Asian faces. This might be a good example to echo the phenomenon that Asian faces often aesthetically look younger than the Western.

Age estimation by machine has been reveled as a difficult and challenging problem by existing facts and attitudes from the perception field. Different people have different rates of aging process [15, 14], which is determined by not only the people's gene but also many factors, such as health condition, living style, working environment, and sociality [29, 5]. Aging shows different forms in different ages [4, 37]. From infancy to puberty, the greatest change is the craniofacial growth (shape change) [1]. Overall the face size is getting larger gradually during the craniofacial growth. From adulthood to old age, the most perceptible change becomes the skin aging (texture change). The shape change still continues, but less dramatically. Thus face aging is uncontrollable and personalized [24, 7, 11]. Furthermore, males and females may have different face aging patterns displayed in images due to the different extent in using makeups and accessories. Many female face images may potentially show younger appearances. It is still an open problem to extract general discriminative features for age estimation while reducing the negative influence of individual differences.

978-1-4244-3991-1/09/$25.00 ©2009 IEEE

年龄估计的方法:

用的4方向和16尺度？我们的VC代码中是用的8方向和16尺度？why？尺度和方向到底意味着什么呢？和后面的变换有什么关系呢？

The existing age estimation systems using face images typically consist of two concatenated modules: image representation and age estimation. For age image representation, the anthropometric model [19], active appearance model (AAM) [6], AGing pattErn Subspace (AGES) [12, 11], age manifold [8, 7], and patch-based appearance model [36, 33] are most popular ones. Given an aging feature representation, the next step is to estimate ages. The age estimation can be viewed as a multi-class classification problem [20, 30, 14], a regression problem [8, 35, 34, 7, 38], or a hybrid of the two [16] [14] [15].

Given the superior performance of human vision on general object recognition and age estimation, it is reasonable to look to biology for inspiration to improve the computer's performance for age estimation. This motivates us to explore features and methods from brain modelling and studies. Recently, the biologically inspired features [26] have shown excellent performance in some computer vision tasks such as object category recognition [28] [22] [27] and face recognition [21]. In this paper, we want to investigate the bio-inspired features (BIF) for the age estimation problem.

Our main contributions include that (1) the biologically inspired features are introduced to age estimation for the first time; (2) previous bio-inspired models are changed by proposing a novel "STD" operation in creating C_1 features; (3) some refinements to the bio-inspired models are proposed: The starting scale of the filter pyramid is changed smaller, the numbers of bands and orientations are adaptive to data; Each modification is shown to provide improvement in age estimation; and (4) our new method outperforms all of the state-of-the-art methods in age estimation.

2. Bio-inspired Models

Visual processing in cortex is modeled as a hierarchy of increasingly sophisticated representations. A recent theory of the feedforward path of object recognition in cortex accounts for the first 100-200 milliseconds of processing in the ventral stream of primate visual cortex [26] [27]. We briefly present some existing biologically inspired models for object recognition, and then describe our base model and improvement. Detailed implementation of our model is given in the next section.

2.1. Previous Models

Riesenhuber and Poggio [26] proposed a new set of features derived from a feed-forward model of the primate visual object recognition pathway, called the "HMAX" model. The framework of the model contains alternating layers 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). 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 an array of Gabor filters at four 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.

Serre et al. [28] [27] extended the "HMAX" model of Riesenhuber and Poggio [26] to include two higher level layers, called S_2 and C_2, for object recognition. In the S_2 layer, template matching is performed to match the patches of C_1 units with some pre-learned prototype patches that are extracted from natural images. This S_2 layer gets intermediate features that are more selective and thus useful for discriminating between classes of objects. These S_2 units are then convolved over an entire image and C_2 units are assigned the maximum response value on S_2. Mutch and Lowe [22] built on Serre et al.'s work [28] for object category recognition and proposed some improvements such as sparsify S_2 inputs (selecting dominating orientations from the four), suppress S_1 and C_1 outputs (reducing the number of output units), and select features that are highly weighted by the support vector machine (SVM) [31]. Meyers and Wolf [21] used the biologically inspired features for face recognition by concatenating the C_1 units to form a so-called S_2 facial features (S2FF) and used a relevant component analysis technique for feature dimension reduction.

2.2. Our Base Model

Our base model with the biological inspiration is particularly designed for human age estimation. In previous approaches to object category recognition [28] [22] [27], a number of prototypes (approximately 1,000 in [27]) are randomly selected from the learning images. These prototypes are stored for template matching in S_2 units. We found that the S_2 (and then C_2) features using pre-learned prototypes cannot work well for age estimation. Therefore we used only the S_1 and C_1 units. The outputs from the C_1 units are concatenated to form a long feature vector to represent each face image. After the biologically inspired C_1 features are extracted, we perform feature dimension reduction and statistical learning for age estimation.

2.3. Improvement

In using the S_1 and C_1 units, we found that the number of bands and orientations for S_1 units should not be predetermined by a fixed number, e.g., 4 orientations in [26] [28]

[27] [21], and 12 orientations in [22]. In addition, all of them used 8 bands. In our age estimation, the number of bands and orientations varies within a range. The best numbers are reported based on the age estimation results. The purpose is to show that these parameters should be adaptive to each specific problem, because there are different image sizes for different problems and various objects may contain different levels of details. In age estimation, we would like to know what are the proper numbers of bands and orientations for the S_1 units.

In applying the Gabor filters to S_1 units, we found that a smaller size of 5×5 can characterize the aging effects on faces better. So in our pyramid of Gabor filters, the starting filter size is from 5×5 rather than 7×7 as in [28] [27].

For C_1 units, we found that the "MAX" filtering can be improved further. We propose another nonlinear operator "STD" to each scale band (2 scales) of S_1 units after they are merged into one maximum map using the "MAX" operation.

We also do dimension reduction for the C_1 features to make them more efficient. All these improvements are necessary in order to obtain a high-performance age estimator.

3. Detailed Implementation

Our system for age estimation works in four steps. First, the face image goes through the S_1 layer and C_1 layer. The resulted C_1 features are about 7,000. Then a dimensionality reduction process reduces the dimensions greatly. Finally, classifiers or regressors are used for learning and the output is the estimated age. The system framework is illustrated in Figure 2. Before performing experimental evaluations, we describe the details for each step.

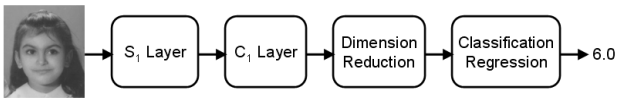

Figure 2. Our age estimation framework.

S_1 units: A grey level input image is first analyzed by an array of simple S_1 units that correspond to the classical simple cells of Hubel and Wiesel found in the primary visual cortex (V1) [17]. The Gabor functions [9] are usually taken for the S_1 units [26] [28] [27] [22] [21], which provide a good model of cortical simple cell receptive fields. The Gabor functions are of the following form

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

where $X = x\cos\theta + y\sin\theta$ and $Y = -x\sin\theta + y\cos\theta$ are the rotations of the Gabor filters with angle θ which varies between 0 and π. The aspect ratio is fixed as in [27], $\gamma = 0.3$, the effective width σ, the wavelength λ as well as the filter sizes s were adjusted as in [27], but in our case, the filter banks start from 5×5 rather than 7×7. Thus the pooling grid sizes and overlaps are changed accordingly. Details about the related parameters are shown in Table 1. The orientation θ varies from 0 to π uniformly with different intervals, resulting in different number of total orientations, such as 4, 6, 8, 10, and 12. The best number of orientations is determined from the data in our age estimation problem. As in the work of Serre, Wolf, and Poggio [28], the S_1 filters were arranged to form a pyramid of scales, spanning a range of sizes. But in our age estimation problem, we found that starting from a smaller size, such as 5×5, instead of 7×7, delivered better results (See experiments). The reasons might be that (1) the image size of 60×60 in our age estimation is much smaller than in object category recognition, e.g., in [27] all input images are down-sampled to 140 pixels high while maintaining the aspect ratio; and (2) small facial details are important to characterize subtle age variations. In addition, the number of bands was chosen and fixed as 8 (thus 16 scales in total) in Serre et al.'s work [27], but we let the data determine the best number of bands, selected from 2, 4, 6, and 8. The number of S_1 receptive field types is determined by the product of the number of scales and orientations. For example, it could be 192 types when 16 scales and 12 orientations $(0°, 15°, 30°, \ldots, 165°)$ are chosen by the algorithm.

Table 1. Parameters used in S_1 and C_1.

C_1 layer			S_1 layer		
Scale band S	Pool. grid	Overlap $\triangle_S$	filter size s	Gabor σ	Gabor λ
Band 1	6×6	3	5×5 7×7	2.0 2.8	2.5 3.5
Band 2	8×8	4	9×9 11×11	3.6 4.5	4.6 5.6
Band 3	10×10	5	13×13 15×15	5.4 6.3	6.8 7.9
Band 4	12×12	6	17×17 19×19	7.3 8.2	9.1 10.3
Band 5	14×14	7	21×21 23×23	9.2 10.2	11.5 12.7
Band 6	16×16	8	25×25 27×27	11.3 12.3	14.1 15.4
Band 7	18×18	9	29×29 31×31	13.4 14.6	16.8 18.2
Band 8	20×20	10	33×33 35×35	15.8 17.0	19.7 21.2

C_1 units: The C_1 units correspond to cortical complex cells that tend to have larger receptive fields [26, 27]. C_1 units pool over S_1 units from the previous layer with the same ori-

entation and from the same scale band. Each scale band has a pair of adjacent filter sizes. The scale band index of the S_1 units determines the size of the S_1 neighborhood $N_S \times N_S$ over which the C_1 units pool [27]. In previous models [26] [27] [22] [21], the maximum operation "MAX" was used as the pooling filter. Riesenhuber and Poggio [26] argued and demonstrated the advantages of using the nonlinear "MAX" operator over the linear summation operation "SUM." Here we propose another nonlinear operation – standard deviation "STD"

$$std = \sqrt{\frac{1}{N_S \times N_S} \sum_{i=1}^{N_S \times N_S} \left(F_i - \bar{F}\right)^2}, \qquad (2)$$

which reveals the variability in the data within a neighborhood $N_S \times N_S$ of S_1 units, where F_i is the maximum value of two consecutive S_1 units in the same scale band (but using different filers) at pixel index i,

$$F_i = \max\left(x_i^j,\, x_i^{j+1}\right), \qquad (3)$$

where x_i^j and x_i^{j+1} are the filtered values with scales j and $j+1$ at position i. $\bar{F}$ is the mean value of the filtered values within the neighborhood $N_S \times N_S$. The pooling of the "MAX" operation over two consecutive scales (i.e., in the same scale band) increases the tolerance to 2D transformations, such as scale changes with a small amount. The "MAX" operation merges two filtered images using filters of the same orientation but different scales into one, and then the "STD" operation is performed on the merged image within a neighborhood $N_S \times N_S$. In previous models [26] [28] [27] [22] [21], the operation on the merged image is another "MAX" filtering. While the second "MAX" can tolerate more shift and size changes, it cannot reveal the variability in the data. In age estimation, the description of local variability in data might be important for subtle age variation, such as wrinkles, on faces.

The process of "MAX" pooling and "STD" operation is done for each orientation and each scale band independently. For instance, consider the first band: $S = 1$. For each orientation, it contains two S_1 maps: the one obtained using a filter of size 5×5 and the one obtained using a filter of size 7×7. The two maps have the same size but different values because different filters are applied. First, the "MAX" pooling is performed on the two maps resulting in a maximum map. Then the "STD" operation is performed on the maximum map using a cell grid of size $N_S \times N_S = 6 \times 6$. From each grid cell, one single value is computed from the 36 elements. C_1 responses are only computed with an interval of half of the neighborhood size in one direction, i.e., $N_S/2$. As a result, there is only one half overlapping between two neighboring grid cells, and this will largely reduce the number of features in the C_1 layer and make the age estimation process more efficient.

Experimentally we found that the "STD" operation outperforms the pure "MAX" pooling for age estimation (See experiments). The reason is that the "STD" operation can reveal the local variations which might be important to characterize the subtlety of aging (e.g., wrinkles, creases, and eyelid bags) on faces. This can be observed in Figure 3 where the input face image is of size 60×60. The S_1 units at four orientations of band 4 (filter sizes of 17×17 and 19×19) are displayed. A pooling grid of 12×12 is drawn in each S_1 map. The local variation is significant (especially in the orientation of 45 degrees), while the pure "MAX" operation cannot reveal it.

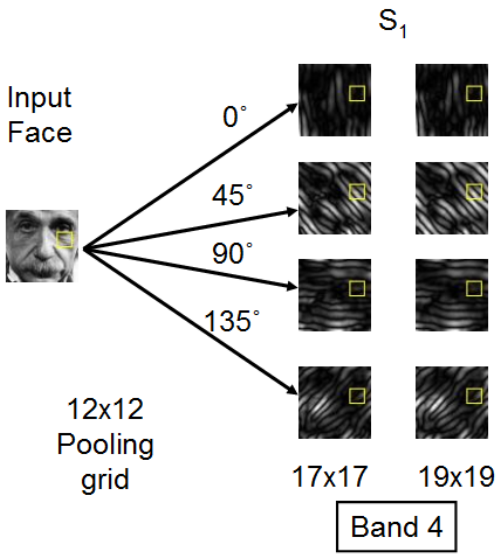

Figure 3. The S_1 layer of band 4 (two scales with filter sizes of 17×17 and 19×19) at four orientations. The pooling grid size for C_1 units is a 12×12 square shown in each S_1 map. Within each pooling grid, the "STD" operation characterizing the local variation on S_1 maps is very important in describing the subtlety of aging. For better viewing, please see enlarged color pdf file.

Feature Dimensionality Reduction: When the pre-learned prototypes are used for template matching in S_2 units [27] [22], the final feature dimension in C_2 is controlled by the number of prototypes used for comparison in S_2. A typical number of prototypes is set by $N = 1,000$ in [27]. Here we do not use the S_2 and C_2 units. The C_1 features are concatenated to form a feature vector for each face image. The resulted features have a high dimensionality. Reducing the feature dimension makes the algorithm more efficient. Previously, the C_2 features can be selected using those highly weighted by the SVM for object category recognition [22], or using a relevant component analysis for face recognition [21]. Here we show that the simple method of principal component analysis (PCA) [32] can work well for the bio-inspired C_1 features.

PCA is a linear transform technique. Let $\mathbf{X} =$

$[\mathbf{x}_1\ \mathbf{x}_2\ \cdots\ \mathbf{x}_n]$ be the feature vectors with dimension D, derived from the C_1 units on n training faces, $\bar{\mathbf{x}}$ be the mean vector of the training data. For dimensionality reduction, the purpose is to find an $n \times d$ matrix $\mathbf{P}$ satisfying $\mathbf{Y} = \mathbf{P}^T\mathbf{X}$, where $\mathbf{Y} = [\mathbf{y}_1\ \mathbf{y}_2\ \cdots\ \mathbf{y}_n]$ is the projected new features with dimension d, and $d \ll D$. The PCA method finds the embedding that maximizes the projected variance, $\mathbf{p} = \arg\max_{\|\mathbf{p}\|=1} \mathbf{p}^T\mathbf{S}\mathbf{p}$, where $\mathbf{S} = \sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$ is the scatter matrix.

The PCA technique can be used efficiently for dimensionality reduction on the bio-inspired C_1 features. Experimentally, we found that the age estimation errors do not change too much when the features with reduced dimensions are used. In some cases, the errors are even smaller than using all C_1 features.

Classification or Regression: Age estimation can be considered as a classification problem, when each age is taken as a class label. On the other hand, age estimation can also be considered as a regression problem, where each age is used as a regression value [14]. In our experiments, we use both the classification and regression approaches to age estimation but on different databases. We will analyze why sometimes the classification works better, while other times the regression is better. Following the suggestions in [14], we choose to use the linear SVM for age classification and support vector regressor (SVR) [31] for age regression.

4. Experiments

The performance of age estimation is measured by the mean absolute error (MAE) and the cumulative score (CS) [7, 12]. The MAE is defined as the average of the absolute errors between the estimated ages and the ground truth ages, $\text{MAE} = \sum_{k=1}^{N} |\hat{l_k} - l_k|/N$, where l_k is the ground truth age for the test image k, $\hat{l_k}$ is the estimated age, and N is the total number of test images. The cumulative score is defined as $\text{CS}(j) = N_{e\leq j}/N \times 100\%$, where $N_{e\leq j}$ is the number of test images on which the age estimation makes an absolute error no higher than j years.

4.1. Databases

Extensive age estimation experiments are performed on the large Yahama Gender and Age (YGA) database which contains 8,000 face images captured from 1,600 Asian subjects in an outdoor environment, 800 females and 800 males, in the age range from 0 to 93 years. The ground truth age for each face is provided. See Table 2 for the number of face images in each age group.

The second database is the FG-NET Aging Database [3] which is public available. The database contains 1,002 color or grey scale face images with large variation of lighting, pose, and expression. There are 82 subjects (multiple races) in total with the age ranges from 0 to 69 years, and each face image has 68 labelled points characterizing shape features. In our approach, the shape features are not used.

Table 3. MAE (years) at different age groups on FG-NET.

Range	#img.	Ours	RUN[35]	QM[20]	MLP[20]
0-9	371	2.99	2.51	6.26	11.63
10-19	339	3.39	3.76	5.85	3.33
20-29	144	4.30	6.38	7.10	8.81
30-39	70	8.24	12.51	11.56	18.46
40-49	46	14.98	20.09	14.80	27.98
50-59	15	20.49	28.07	24.27	49.13
60-69	8	31.62	42.50	37.38	49.13
Total	1002	**4.77**	5.78	7.57	10.39

4.2. Results

We perform a standard 4-fold cross validation test to evaluate the accuracy of our algorithms on the large YGA database. The test was executed on the female and male subsets separately, because age estimation is sensitive to gender on this large database. A leave-one-person-out (LOPO) test scheme is used on the FG-NET database. Each face image is cropped and resized to 60×60 in both databases, and only the grey level images are used. We use the classifiers of linear SVMs for the YGA database, and the regressor SVR for the FG-NET database, as suggested in [15, 14]. The parameters of the learners are adjust from a tuning data set (part of the training data). The general scheme is to use a small error tolerance value, e.g., 0.001, and the capacity value around 100 for the SVM. The width of the RBF kernel for regression [31] is between 0.1 and 0.01. Combining the classification and regression methods [15, 14, 16] might improve the results further, but we have not tried here. Our main focus is the new features based on biological inspiration.

The MAEs on the YGA database is shown in Table 2, and FG-NET in Table 3. The MAEs at each age group (about 10 years span) are also given. In the YGA database, our method has an MAE of 3.91 years for females, and 3.47 years for males. These average errors are substantially smaller than the RUN method [35] (9.79 and 10.36 years), and even significantly lower than the very recent RPK [36] approach (4.94 and 4.38 years) which announced to be the best reported result so far. See Table 6 for more methods.

In FG-NET, the MAE of our method is 4.77 years, which is significantly smaller than the 5.78 of the RUN method [35], and even lower than the best reported result of 4.95 in [36]. We notice that in the FG-NET database, there are only 8 face images in the age range of 60-69, and only 15 images of 50-59. The number of faces is too small to train classifiers for those ages. This interprets why classifiers like SVM cannot work well on the FG-NET database. This also

Table 2. MAE (years) at different age groups on the YGA database.

	Females in YGA					Males in YGA				
Range	#img.	Ours	RUN[35]	QM [20]	MLP[20]	#img.	Ours	RUN[35]	QM[20]	MLP[20]
0-9	500	1.84	11.21	11.97	14.33	500	1.24	9.86	13.42	14.08
10-19	500	2.46	6.23	9.58	8.85	500	2.67	7.52	10.33	9.46
20-29	500	3.36	7.95	9.29	9.70	500	3.62	8.85	10.21	9.35
30-39	500	5.22	8.17	9.85	9.66	500	3.37	7.76	9.35	8.60
40-49	500	5.71	8.64	10.45	8.78	500	4.06	8.67	11.71	9.10
50-59	500	4.35	9.43	10.15	9.53	500	4.61	11.10	13.38	10.08
60-69	500	3.82	11.12	13.49	10.88	500	3.92	12.49	15.99	13.44
70-93	500	4.08	15.56	19.66	16.52	500	3.87	16.60	20.44	19.69
Total	4000	**3.91**	9.79	11.80	11.03	4000	**3.47**	10.36	13.10	11.72

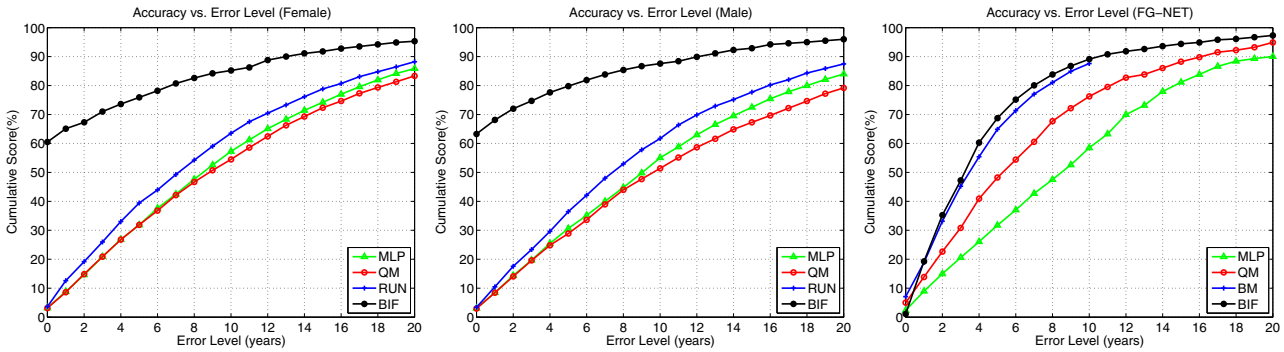

Figure 4. Cumulative scores on the YGA and FG-NET databases.

tells why all methods have much larger errors in those age ranges, shown in Table 3. Even though provided with a very small number of faces, our method is consistently better than other methods in comparison at those age ranges. This indirectly indicates that the bio-inspired features are quite robust to encode aging patterns on faces.

The SC curves are shown in Figure 4. The BM [34] method was used for comparison in FG-NET since it has a smaller MAE (5.33) than the RUN [35] method (5.78) on FG-NET.

4.3. Studies of Improvement

We would like to explore more details about our models, especially the improvement over previous models [28, 27, 22, 21] for the specific problem of age estimation.

C_1 vs. C_2 features We first verify if the C_2 features can work well for age estimation. Using almost the same setting as Serre et al.'s [27] for C_2 features with 2,000 prototypes randomly selected, we obtained an MAE of 10.62 and 10.73 years for females and males, separately. These errors are significantly higher than using C_1 features with any number of bands and orientations as shown in Table 4.

The number of bands and orientations In previous biologically inspired models [28, 27, 22, 21], the scale bands are fixed as 8, and the orientations are fixed as either 4 or 12. In age estimation, we want to show that the number of bands and orientations should be adaptive to the data, rather than using a pre-defined number. More bands or more orientations do not necessarily give better results. To check this in detail, we use the C_1 features with the starting filter size 7×7 and the pure "MAX" pooling for age estimation on the YGA database. The number of bands varies from 2 to 8, and the number of orientations varies from 4 to 12, with an interval of 2. The age estimation results are shown in Table 4. The smallest MAE for the female is 4.20 years obtained using 6 bands and 10 orientations, while the smallest MAE for the male is 3.61 years generated from 8 bands and 8 orientations. This suggests that for each specific problem, the number of bands and orientations should be determined by data, in order to obtain good results.

Filter banks starting from 5×5 vs. 7×7 In previous biologically inspired models [28, 27, 22, 21], the filter bank uses a fixed structure, starting from 7×7. We found that starting from a smaller size, such as 5×5, improves the age estimation results. To show this difference, we run the age estimation experiments using different C_1 features, one is starting from 5×5, and the other from 7×7. The MAEs are shown in Table 5. In all cases – females or males, original features or with dimensional-

Table 4. MAEs in terms of the number of bands and orientations in extracting the biologically inspired features. In each cell, the MAEs are for females and males, separately.

	# Orientations				
# Bands	4	6	8	10	12
2	5.26 5.48	4.73 4.56	4.85 4.66	4.81 4.60	4.90 4.57
4	4.96 4.74	4.44 3.72	4.56 3.85	4.64 3.87	4.52 3.82
6	4.85 4.76	4.44 3.72	4.37 3.68	**4.20** 3.70	4.21 3.68
8	4.77 4.48	4.38 3.79	4.31 **3.61**	4.29 3.78	4.31 3.77

Table 6. MAE (years) comparisons.

Method	YGA:F	YGA:M	FG-NET
WAS [12]	–	–	8.06
AGES [12]	–	–	6.77
QM [20]	9.96	10.51	6.55
MLPs [20]	10.99	12.00	6.98
RUN [35]	9.79	10.36	5.78
BM [34]	6.95	6.95	5.33
LARR [14]	5.25	5.30	5.07
PFA [16]	5.11	5.12	4.97
RPK [36]	4.94	4.38	4.95
BIF(Ours)	**3.91**	**3.47**	**4.77**

ity reduction, and "MAX" pooling or "STD" operation – the estimation errors are smaller when starting from 5×5.

"STD" vs. "MAX" In previous models [26, 28, 27, 22, 21], the "MAX" pooling is used to obtain the complex C layer features from the simple S layers. Riesenhuber and Poggio [26] argued the advantages of the nonlinear maximum operation over the linear summation "SUM." We propose another nonlinear operation "STD" which can characterize the aging subtlety on faces well. The results demonstrating the advantages of "STD" over "MAX" are given in Table 5 and Figure 5. In table 5, the age estimation errors based on "STD" are always smaller than the "MAX," no matter in which situation – females or males, original features or reduced, and different filter banks. In Figure 5, the "STD" also gives smaller errors in different feature dimensions.

Dimensionality reduction by PCA Although the PCA technique is a widely used method for feature dimensionality reduction, no previous work has shown that the PCA method is also useful for the bio-inspired high dimensional features. Here we want to show that the simple PCA method can work well for the biologically inspired features, reducing dimensions from the thousand level to hundred. From

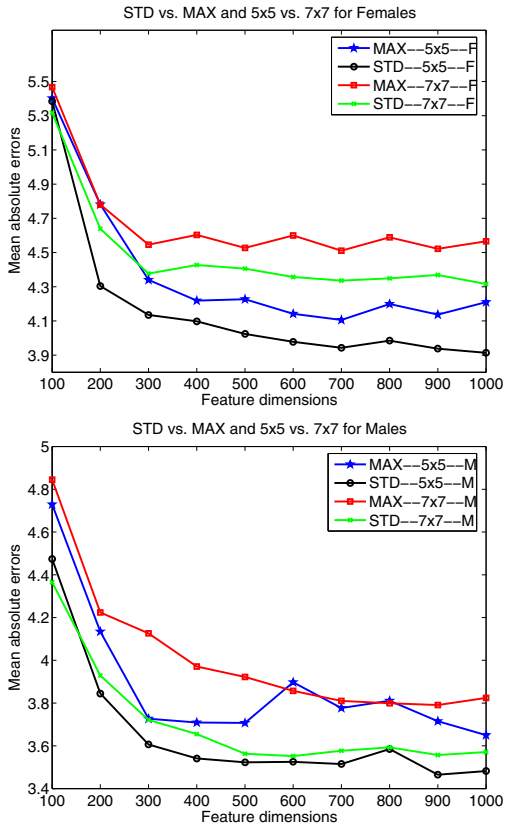

Figure 5. MAEs of "STD" vs. "MAX" and 5×5 vs. 7×7 as the starting filter size. The PCA method is used for dimensionality reduction.

Table 5, the errors do not change too much when the feature dimension is reduced using the PCA method. For example, the dimensions can be reduced from 6,976 to 900 without raising the error too much (from 3.47 to 3.44). In some cases, the errors are even smaller than using all C_1 features. In Figure 5, we show the error distributions over feature dimensions from 100 to 1,000. In sum, less than 1,000 features (selected from about 7,000) are sufficient for age estimation in our experiments.

4.4. Age Estimation on Internet Images

It may be interesting to see our age estimation performance on other face images. We used Internet image search to find some face images of Einstein, and extracted bio-inspired features on his face images. The training data are male Asian faces from the YGA database. Then we ran our method to estimate the ages from Einstein's faces. The age estimation results are shown in Figure 1. Although this is a very difficult case for cross-race and cross-database age estimation, the estimated ages, 32, 52, and 76, are very reasonable on the three faces. Since we do not have the ground truth on these faces, no MAEs are computed. This result

Table 5. MAE (#features) measures for the filters on S_1 units, starting from 5×5 vs. 7×7.

	Females in YGA				Males in YGA			
	Original features		Dim reduction		Original features		Dim reduction	
	5×5	7×7	5×5	7×7	5×5	7×7	5×5	7×7
MAX	4.15(6976)	4.53 (4376)	4.11(700)	4.51(700)	3.72(6976)	3.66(4376)	3.65 (1000)	3.79(900)
STD	3.94(6976)	4.32(4376)	**3.91**(1000)	4.32(1000)	**3.44**(6976)	3.57(4376)	3.47(900)	3.55(600)

shows the robustness of our approach to age estimation, and encourages us to do more evaluations in the future.

5. Conclusions

We have investigated the biologically inspired features for human age estimation. A couple of improvements over the original models have been proposed, which include introducing the new "STD" operation for pooling in C_1 units, changing the filter sizes in S_1 units, not using any pre-learned prototypes, and making the number of bands and orientations adaptive to data. All of these proposals to the bio-inspired models are important to obtain a high performance age estimator. Evaluated on two benchmark databases, our system performs better (smaller errors) than any published methods. Have not been explored by anyone before, we showed an interesting application of our method to Internet images of Einstein's faces, which demonstrates the robustness of our method and the potential of cross-race and cross-database age estimation. It also suggests a future work to further verify the method on more Internet images or surveillance camera inputs.

6. Acknowledgement

The work is partially supported by the NC Space New Investigation award to the first author. The authors thank S. Yan for providing his results for comparison.

References

[1] *Aging of the face*. http://www.face-and-emotion.com/dataface/facets/aging.jsp.

[2] *Electronic Customer Relationship Management (ECRM)*. http://en.wikipedia.org/wiki/ECRM.

[3] The fg-net aging database. In *http://www.fgnet.rsunit.com/*.

[4] A. M. Alberta, K. Ricanek, and E. Pattersonb. A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. *Forensic Science International*, 172(1):1–9, 2007.

[5] D. S. Berry and L. Z. McArthur. Perceiving character in faces: The impact of age-related craniofacial changes on social perception. *Psychological Bulletin*, 100(1):3–18, 1986.

[6] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. In *European Conference on Computer Vision*, pages 484–498, 1998.

[7] Y. Fu and T. S. Huang. Human age estimatin with regression on discriminative aging manifold. *IEEE Trans. on Multimedia*, 10(4):578–584, 2008.

[8] Y. Fu, Y. Xu, and T. S. Huang. Estimating human ages by manifold analysis of face pictures and regression on aging features. In *IEEE Conf. on Multimedia and Expo*, pages 1383–1386, 2007.

[9] D. Gabor. Theory of communication. *J. IEE*, 93:429–459, 1946.

[10] A. Gallagher and T. Chen. Estimating age, gender, and identity using first name priors. In *IEEE Conf. on CVPR*, 2008.

[11] X. Geng, Z.-H. Zhou, and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *IEEE Trans. on PAMI*, 29(12):2234–2240, 2007.

[12] X. Geng, Z.-H. Zhou, Y. Zhang, G. Li, , and H. Dai. Learning from facial aging patterns for automatic age estimation. In *ACM Conf. on Multimedia*, pages 307–316, 2006.

[13] P. George and G. Hole. Factors influencing the accuracy of age estimates of unfamiliar faces. *Perception*, 24(9):1059–1073, 1995.

[14] G. Guo, Y. Fu, C. Dyer, and T. Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression. *IEEE Trans. Image Proc.*, 17(7):1178–1188, 2008.

[15] G. Guo, Y. Fu, T. S. Huang, and C. Dyer. Locally adjusted robust regression for human age estimation. In *IEEE Workshop on Applications of Computer Vision*, 2008.

[16] G. Guo, Y. Fu, T. S. Huang, and C. Dyer. A probabilistic fusion approach to human age prediction. In *IEEE CVPR-SLAM workshop*, 2008.

[17] D. Hubel and T. Wiesel. Receptive fields, binocular interaction and function architecture in the cat's visual cortex. *Journal of Physiology*, 160:106–154, 1962.

[18] J. E. Kloeppel. *Step right up, let the computer look at your face and tell you your age*. http://news.illinois.edu/news/08/0923age.html, 2008.

[19] Y. Kwon and N. Lobo. Age classification from facial images. *Computer Vision and Image Understanding*, 74(1):1–21, 1999.

[20] A. Lanitis, C. Draganova, and C. Christodoulou. Comparing different classifiers for automatic age estimation. *IEEE Trans. on SMC-B*, 34(1):621–628, 2004.

[21] E. Meyers and L. Wolf. Using biologically inspired features for face processing. *Int. J. Comput. Vis.*, 76:93–104, 2008.

[22] J. Mutch and D. Lowe. Object class recognition and localization using sparse features with limited receptive fields. In *Conf. on Comput. Vision and Pattern Recognit.*, pages 11–18, 2006.

[23] E. Patterson, A. Sethuram, M. Albert, K. Ricanek, and M. King. Aspects of age variation in facial morphology affecting biometrics. In *IEEE Conf. on Biometrics: Theory, Applications, and Systems*, 2007.

[24] N. Ramanathan and R. Chellappa. Face verification across age progression. *IEEE Trans. on Image Processing*, 15(11):3349–3361, 2006.

[25] K. Ricanek and E. Boone. The effect of normal adult aging on standard pca face recognition accuracy rates. In *International Joint Conference on Neural Networks*, pages 2018–2023, 2005.

[26] M. Riesenhuber and T. Poggio. Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11):1019–1025, 1999.

[27] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio. Robust object recognition with cortex-like mechanisms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(3):411–426, 2007.

[28] T. Serre, L. Wolf, and T. Poggio. Object recognition with features inspired by visual cortex. In *Conf. on Comput. Vision and Pattern Recognit.*, 2005.

[29] A. Stone. *The aging process of the face techniques of rejuvenation*. http://www.aaronstonemd.com/Facial_Aging_Rejuvenation.shtm.

[30] K. Ueki, T. Hayashida, and T. Kobayashi. Subspace-based age-group classification using facial images under various lighting conditions. In *IEEE conf. on FGR*, 2006.

[31] V. N. Vapnik. *Statistical Learning Theory*. John Wiley, New York, 1998.

[32] A. R. Webb. *Statistical Pattern Recognition, 2nd Edition*. John Wiley, 2002.

[33] S. Yan, M. Liu, and T. Huang. Extracting age information from local spatially flexible patches. In *IEEE conf. on ICASSP*, pages 737–740, 2008.

[34] S. Yan, H. Wang, T. S. Huang, and X. Tang. Ranking with uncertain labels. In *IEEE conf. on Multimedia and Expo*, pages 96–99, 2007.

[35] S. Yan, H. Wang, X. Tang, and T. Huang. Learning auto-structured regressor from uncertain nonnegative labels. In *IEEE conf. on ICCV*, 2007.

[36] S. Yan, X. Zhou, M. Liu, M. Hasegawa-Johnson, and T. Huang. Regression from patch-kernel. In *IEEE conf. on CVPR*, 2008.

[37] L. A. Zebrowitz. *Reading Faces: Window to the Soul?* Westview Press, 1997.

[38] S. Zhou, B. Georgescu, X. Zhou, and D. Comaniciu. Image based regression using boosting method. In *IEEE conf. on ICCV*, pages 541–548, 2005.

[39] L. Zyga. *Intelligent Computers See Your Human Traits*. http://www.physorg.com/news131277135.html, PhysOrg.com, May 29, 2008.