Is Gender Recognition Affected by Age?

Guodong Guo
NCCU
gdguo@nccu.edu

Charles R. Dyer
UW-Madison
dyer@cs.wisc.edu

Yun Fu
BBN Technologies
yfu@bbn.com

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

Abstract

Gender recognition is important for many applications including human computer interaction (HCI). This paper shows that gender recognition accuracy is affected significantly by the age of the person. Our empirical studies on a large face database of 8,000 images with ages from 0 to 93 years show that gender classification accuracy on adult faces can be 10% higher than that on young or senior faces, evaluated using one of the state-of-the-art methods. We examine aging effects on human faces, which motivates us to investigate which features can incorporate shape and texture variations on faces together with gender encoding. Based on the aging effects, the local binary pattern (LBP) and histograms of oriented gradients (HOG) methods are evaluated for gender characterization with age variation. We also investigate a biologically-inspired method for gender recognition. Overall, no matter what methods are used, the accuracies on adult faces are consistently higher than on young or senior faces. This new finding suggests new efforts in both psychological studies and computational visual recognition for the purpose of HCI applications.

1. Introduction

Recognizing human gender plays an important role in many HCI areas. For example, search engines need an image filter to determine the gender of people in images from the Internet [4]; demographic research can use gender information extracted from images to count the number of men and women entering a shopping mall or movie theater; a "smart building" might use gender for surveillance and control of access to certain areas [18]. Besides these kinds of broad applications, gender recognition itself is an important research topic in both psychology [5, 28, 25] and computer vision [14, 18, 4, 30].

In psychology studies [5, 28, 25] for HCI, the main focus is about how humans discriminate between males and females and what kind of features are more discriminative. Wild et al. [25] showed that gender recognition is more difficult for subjects when children's faces are presented compared to adults' faces. In computational gender classification, Golomb et al. [12] trained a neural network, called SEXNET, to identify gender from 30×30 face images. An accuracy of 91.9% was obtained from the experiments on 90 photos (45 males and 45 females), which was also compared with an average accuracy of 88.4% from an average of five subjects. Cottrell [7] reported perfect gender classification using neural networks on a set of 160 face images (10 males and 10 females) of size 64×64. Brunelli and Poggio [6] developed HyperBF networks for gender classification and used 16 facial geometric features. An accuracy of 79% was reported on 168 images (21 males and 21 females). Wiskott et al. [26] used labeled graphs of two dimensional views to describe faces, and got an accuracy of 90.2% on a gallery of 112 face images. Gutta et al. [14] proposed a hybrid classifier based on neural networks and decision trees, and obtained an accuracy of 96% on 3,000 FERET faces of size 64×72. Recently, Moghaddam and Yang [18] used a nonlinear support vector machine (SVM) on raw pixel values with images of size 12×21, and reported an accuracy of 96.6% on 1,755 images (1,044 males and 711 females) from the FERET database [21]. Baluja and Rowley [4] used simple features from five types of pixel comparison operators and the AdaBoost classifier [10] for gender identification. They reported an accuracy of 94.4% on 2,409 face images (1,495 males and 914 females) from FERET [21]. The accuracy of AdaBoost is comparable with the nonlinear SVM. They also showed that 20×20 is better than 12×21 in terms of face size. Xu and Huang [27] used a variant of AdaBoost called SODA-Boosting for gender classification on raw pixels in 40×40 images. They reported an accuracy of 92.82% (slightly lower than the 93.34% of a nonlinear SVM) on 4,109 face images of 1,201 individuals (703 males and 498 females) from FERET. All three recent approaches [18, 4, 27] used five-fold cross validation in their experiments. Yang and Ai [30] used local binary patterns (LBP) for facial feature extraction and AdaBoost for age, gender, and ethnicity classification. They used a set of Chinese snapshot images for training and 3,540 images of 1,196 individuals from FERET for testing. An accuracy of 93.3% was reported for gender classification.

Although the large family of existing work has demonstrated the potential for biometric systems to achieve high accuracy for gender recognition, there is no existing work that answers how performance changes when multiple facial attributes are involved. Specifically, an interesting question is how human age affects gender recognition, which is a common case in real-world scenarios and databases. Our research in this paper focuses on the computational aspect of gender recognition with large age variation. We found, by extensive experimental analysis, that the performance of computational gender recognition via face image analysis is affected significantly by human age. Recognition accuracy on adult faces is shown to be 10% higher than on young and senior faces. These results can suggest new efforts in both psychological studies and computational visual recognition.

In this paper we first introduce a large face database for gender with a large span of ages (0 to 93 years) in Section 2. Then we evaluate gender recognition performance based on one of the state-of-the-art methods in Section 3. An examination of the aging effects on human faces is described in Section 4, which inspires us to investigate visual features other than raw pixels. Two recent operators, LBP and HOG, are chosen in Section 5. A new set of biologically-inspired features is introduced to gender recognition in Section 6. New evaluation results are presented in Section 7. Finally, discussion and conclusions are given.

2. The YGA Database

While there are a number of databases for face recognition, there has been no database designed specifically for evaluating gender recognition performance. Most previous research on gender recognition used part of the face recognition database, FERET [21], for experiments, and most faces from the FERET database are adult faces. Due to this limitation, there is no existing study investigating the effects of age on gender recognition performance.

In our research, the Yamaha Gender and Age (YGA) database, a recently-collected large face database, was used for our experiments. It contains 8,000 face images captured outdoors with half males and half females, in the age range from 0 to 93 years. The database has been used for human age estimation [11, 13, 29], but has never been used before for gender recognition. Here we used the YGA database for research on gender recognition influenced by age. Recognizing gender is also essential for age estimation where currently males and females are manually separated before estimating ages in [11, 13, 29]. A high performance gender recognizer should ideally do age estimation automatically to avoid manual separation. This is also one motivation for our research on gender recognition. The FGNET aging database [2] cannot be used here since it contains only a small span of ages.

Table 1. Gender recognition accuracies evaluated by the "Raw+SVM" approach [18] on the YGA database of 8,000 face images.

Method	All Ages (0-93)	Young (0-19)	Adult (20-60)	Senior (61-93)
Raw+SVM	89.28%	84.38%	94.56%	85.32%

3. An Evaluation and New Finding

Based on the overview of existing gender recognition methods in Section 1, the support vector machine (SVM) working on raw image pixels with small image sizes is still a state-of-the-art approach. We call it, used in [18], "Raw+SVM," and use it for evaluating gender recognition accuracy on the YGA database. The radial basis function (RBF) is used as the kernel for nonlinear SVM in all our experiments. Each face image is cropped and resized from 60×60 to 20×20 as the size of 20×20 is better than 12×21 as demonstrated in [4]. Since the database is large (8,000 images), we used two-fold cross validation to measure performance. A recognition accuracy of 89.28% was obtained in our experiments and listed in the first column of Table 1. This accuracy is significantly lower than many previously reported results, such as the 96.6% in [18] and 94.4% in [4].

Why is gender recognition accuracy so low (below 90%) on the YGA database? One reason may be that the database is much larger than most previous ones, especially those selected from the FERET database [21]. Another reason might be that we used two-fold cross validation while many previous approaches used five-fold cross validation where the number of training faces is larger than the testing faces. Even though these differences might contribute to the accuracy decrease to a small extent, we believe that there is something else that was ignored in previous research but is more important to explore.

Our conjecture is that gender recognition performance is affected by people's ages. In other words, the recognition accuracies might be different for people in different age groups. To verify this conjecture, we divided the 8,000 faces into three age groups: (1) "young" for ages from 0 to 19 years, containing 1,000 males and 1,000 females; (2) "adult" for ages 20 to 60, with 2,050 males and 2,050 females; and (3) "senior" for ages 61 to 93, with 950 males and 950 females. In each age group, there are half males and half females. The number of males and females in each age group is shown in Table 2.

We performed gender recognition on faces in each age group separately using the same "Raw+SVM" approach with two-fold cross validation. The results show gender recognition accuracy of 94.56% for "adults," but 84.38% for

Table 2. Numbers of male and female faces in the three age groups of the YGA face database: young, adult, and senior.

	Young (0-19)	Adult (20-60)	Senior (61-93)	All Ages (0-93)
Male	1,000	2,050	950	4,000
Female	1,000	2,050	950	4,000
Both	2,000	4,100	1,900	8,000

"young" people, and 85.32% for "seniors." These results are also given in columns 2 to 4 in Table 1. Clearly, the recognition accuracies are very different for different age groups. It is about 10% difference between the "adult" and "young" or between the "adult" and "senior." This experiment demonstrates that our conjecture is correct. Gender recognition is influenced significantly by people's ages.

Gender recognition usually consists of two parts: image representation and classification. The SVM has been demonstrated to be an effective method for learning classifiers both theoretically and practically [24], while the simple pixel intensity-based representation, used in most existing work, might not be good enough for characterizing gender attributes on young and senior faces. Hence, we investigate other features and at the same time explore the reasons why raw pixels cannot work well for young and senior faces. We believe that a good understanding of the aging effects on faces could help us identify new features that can extract and encode gender characteristics across age variations more effectively.

4. Aging on Human Faces

Human face aging is a slow and irreversible process. From infancy to adulthood, the greatest change is craniofacial growth – shape change. The eyes, nose, and mouth expand to fill a relatively greater area of the surface of the cranium; the relative area occupied by the forehead shrinks as the eyes move up into this area; and the chin tends to become larger and more protrusive [1]. From adulthood to old age, the most perceptible change is the skin texture change. The shape change continues, but less dramatically. The skin becomes darker, less flexible, rougher, and more leathery; lines, wrinkles, folds, pouches, and blemishes or discolorations gradually appear or become more pronounced; muscles and connective tissues change their elasticity; and fatty deposits and bone may be lost to produce pouches in the cheeks, bags under the eyes, and sagging under the chin [1]. Some young and senior faces are shown in Figure 1.

In short, shape changes are prominent on young faces, while facial texture changes are prevalent for senior people. Hence, features that are capable of characterizing facial shape or texture changes might be important for gender

Figure 1. Aging on human faces: Young and senior. Source images are from [1].

encoding over large age variations. Based on this consideration, we chose the local binary pattern (LBP) [3] and histograms of oriented gradients (HOG) [8] methods as new features for gender extraction from faces.

One the other hand, it might be interesting to explore how objects are recognized in the visual cortex. So we also investigate new features inspired by the biological visual systems for gender recognition in Section 6.

5. HOG and LBP

LBP features were originally developed for texture analysis [20] and later showed high performance for face identity recognition [3] and demographic classification [30]. HOG features were initially developed for pedestrian detection [8] in which they characterized local object shape by measuring the distribution of local intensity gradients. Here we use the LBP and HOG features for gender characterization. The LBP operator might capture the texture variation on faces of old ages, while HOG features might describe facial shape variations, especially for young faces. Since the HOG and LBP features are simple to compute and have become popular in computer vision, we do not describe them here. Readers may refer to [20] and [8] for more details.

6. Biologically-Inspired Features

It has been a long-term research goal to understand how objects are recognized in the visual cortex. Successful theories and algorithms from neuroscience and psychology will have great impact on the design of computational recognition algorithms. Here we introduce and investigate the bio-inspired features (BIF) for gender recognition. Since these features are relatively new, we will describe the BIF and some of our modifications in details.

6.1. Previous Models

Riesenhuber and Poggio [22] 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 "max" operation over the S units rather than "sum" in pooling inputs at the C layers. Specifically, the first layer of the model, called the S1 layer, is created by convolving an array of Gabor filters at four orientations and 16 scales, over the input image. Adjacent two scales of S1 units are then grouped together to form eight 'bands' of units for each orientation. The second layer, called the C1 layer, is then generated by taking the maximum values within a local spatial neighborhood and across the scales within a band. So the resulting C1 representation contains eight bands and four orientations. The advantage of taking the "max" operation within a small range of position and scale is to tolerate small shifts and changes in scale.

Serre et al. [23] extended the "HMAX" model of Riesenhuber and Poggio to include two higher level layers, called S2 and C2, for object recognition. In the S2 layer, template matching is performed to match the patches of C1 units with some pre-learned prototype patches that are extracted from natural images. This S2 layer gets intermediate features that are more selective and thus useful for discriminating between classes of objects. These S2 units are then convolved over an entire image and C2 units are assigned the maximum response value on S2. Mutch and Lowe [19] built on Serre et al.'s work for object category recognition and proposed some improvements such as sparsifying S2 inputs (selecting dominating orientations from the four), suppressing S1 and C1 outputs (reducing the number of output units), and selecting features that are highly weighted by the SVM. Meyers and Wolf [17] used biologically-inspired features for face recognition by concatenating the C1 units to form a so-called S2 facial features (S2FF) and used a relevant component analysis technique for feature dimension reduction.

6.2. Our Model

To the best of our knowledge, biologically-inspired features have not been investigated previously for gender recognition. To encode gender information from face images, our biologically-inspired approach has some difference from previous models.

First, we examine the issue of whether or not to use the S2 and C2 features (via pre-learned prototypes [23]) for the gender recognition problem. In object category recognition, e.g., cars, trees, and animals, the prototypes (approximately 1,000 in [23]) are learned and stored for S2 feature extraction via a template matching scheme. We found that using the prototypes for S2 and C2 features make gender recognition much worse, resulting in low accuracies, lower than all of the methods considered in this paper. So in our approach, only the C1 features are concatenated to create the final features for gender recognition.

Second, we examine how many bands and orientations are necessary for S1 feature extraction for the gender recognition problem. It might be better to determine by the data rather than using a predefined fixed number. In terms of the number of orientations, cells in visual cortex have much finer gradations of orientation than $\frac{\pi}{4}$ [15]. But how many orientations are necessary for gender encoding? Unlike using a fixed four orientations as in [22] [23] [17] or fixed 12 orientations in [19], we take an exhuastive search from a range of orientations. Our purpose is to find how many orientations are necessary for gender encoding from face images. From the experiments on the large database, we found that using ten orientations performed best (see results later). In terms of the number of bands, all previous work [22] [23] [19] [17] used eight bands (i.e., 16 scales). Again, we want to find how many bands are necessary for gender encoding. From our experiments, we found that using six bands gave the best performance (see results later).

6.3. Illustration of the Bio-inspired Features

The key idea of our biologically-inspired features using the S1 and C1 layers is shown in Figure 2. Only the first two bands and four orientations are shown for illustration. Given an input gray level face image, a pyramid of Gabor filters with various scales (up to 16) and orientations (up to 12) are first applied to the input image to obtain S1 units as in previous approaches [22] [23] [19] [17]. The S1 maps as shown in the "S1" column in Figure 2 have the same size as the input image. Then each scale band containing two adjacent scales uses a maximum pooling operation "max" at each orientation to derive the C1 units. The C1 maps as shown in the "C1" column in Figure 2 are much smaller than the input image and have different sizes for different bands [23]. Then the C1 maps are concatenated to form a feature vector as the representation of the face image. Note that we do not use the S2 and C2 units although the C2 features show superior performance in object category recognition [22] [23] [19]. The reason is that the C2 features give much worse performance in gender recognition. See later experiments for more details.

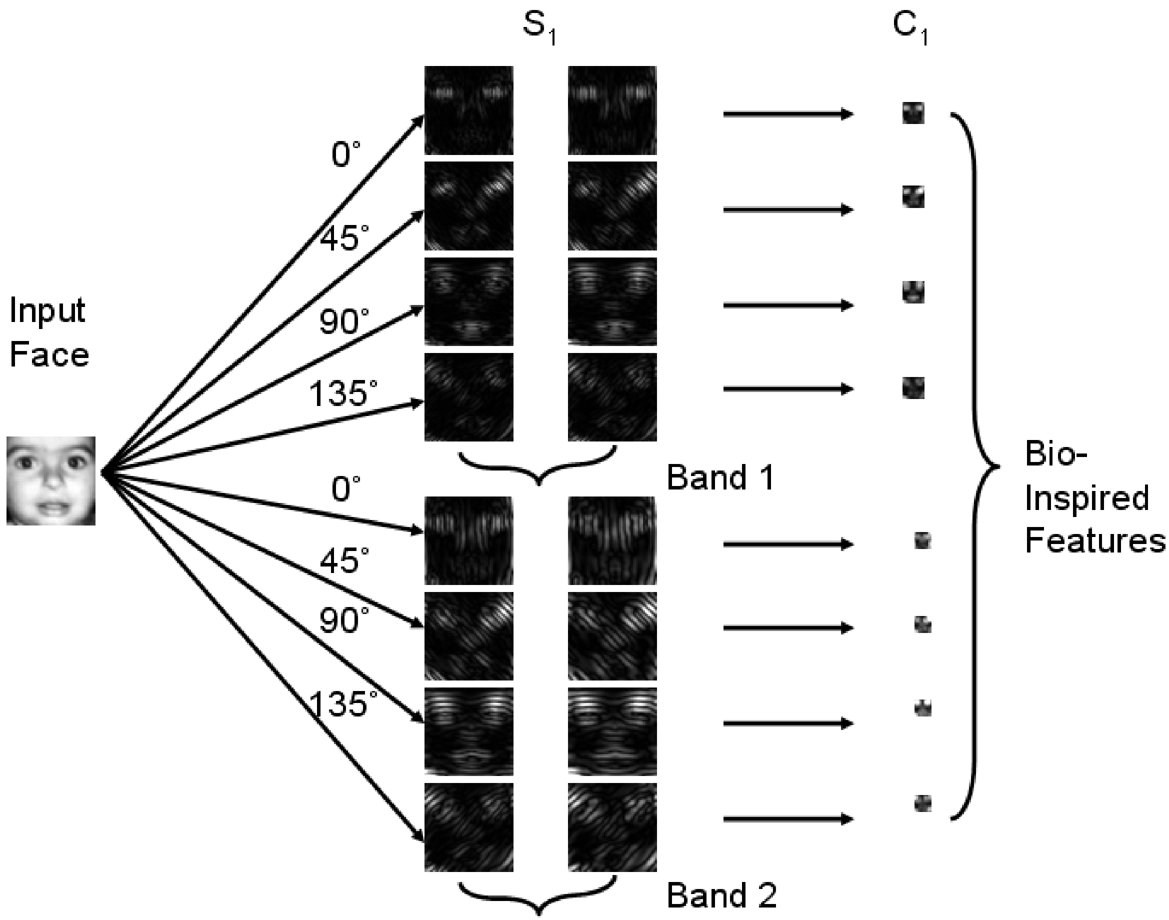

Figure 2. Visualization of the Bio-inspired features using two bands and four orientations.

7. New Evaluation Results

The LBP, HOG, and BIF features are used for gender recognition experimentally. Remember that our goal is to further evaluate the influence of age on gender recognition using the three representations.

Both LBP and HOG features were extracted from each face image at various patch positions for the three age groups. Face images are of size 60×60, and the patch size is 16×16 with an interval of eight pixels between neighboring patches. The HOG operator has eight directions as in [8], and the LBP operator uses the uniform pattern as in [3]. Since HOG features were initially used with a linear SVM for pedestrian detection [8], we also show gender classification results based on linear SVMs (labelled as L-SVM) in addition to nonlinear SVMs with the RBF kernel (denoted as N-SVM). The LBP operator can be applied to the whole face image (denoted as LBP(W)) or applied to small patches on the face (denoted as LBP(P)). The patch-based LBP is much better than the whole face-based LBP in gender recognition, as shown in Table 3. In each case, the parameters of the SVM are adjusted to optimal values on a tuning set (part of the training data).

7.1. HOG Results

When the HOG features are used with nonlinear SVMs, gender recognition accuracies were 86.44%, 94.03%, and 89.04%, for the young, adult, and senior groups, respectively, as shown in row 8 of Table 3. When compared with the "Raw+SVM" approach, the accuracies improved from 84.38% to 86.44% for the young faces, and improved from 85.32% to 89.04% for seniors, while the accuracy of 94.03% for adults is slightly lower than the 94.56% based on raw pixel representation. These results demonstrate that the HOG operator can characterize shape and improve recognition accuracies for young and senior faces. However, the two improved accuracies are still much lower than the 94.03% accuracy for adult faces. On the other hand, the results indicate that the "Raw+SVM" approach is still good for gender recognition on adult faces.

Table 3. Gender recognition with different representations: raw pixels, LBP, HOG, and BIF, using the linear SVM (L-SVM) or non-linear SVM (N-SVM) as classifiers.

Methods	Young (0-19)	Adult (20-60)	Senior (61-93)
Raw + L-SVM	78.59%	89.91%	81.17%
Raw + N-SVM	84.38%	94.56%	85.32%
LBP(W) + L-SVM	68.17%	72.33%	63.40%
LBP(W) + N-SVM	69.65%	77.08%	68.40%
LBP(P) + L-SVM	79.76%	92.65%	87.55%
LBP(P) + N-SVM	81.93%	94.96%	90.64%
HOG + L-SVM	75.83%	88.00%	77.13%
HOG + N-SVM	86.44%	94.03%	89.04%
BIF + L-SVM	83.01%	94.22%	91.81%
BIF + N-SVM	87.13%	96.03%	92.34%
Average(L-SVM)	80.52%	91.20%	84.42%
Average(N-SVM)	84.97%	94.90%	89.34%

We further explain the result as: (1) adult male and female faces have local shape differences that can be described by the HOG operator, and (2) shape changes in young faces and wrinkles in senior faces result in gradient variations that can be encoded by the HOG operator to some extent. However, the HOG performs much better for gender recognition on adult faces than on young and senior faces.

Also notice that linear SVMs performed much worse than kernel SVMs for each age group, as shown in rows 7 and 8 of Table 3.

7.2. LBP Results

When LBP features were used with kernel SVMs, gender recognition accuracies were 81.93%, 94.96%, and 90.64% for the young, adult, and senior groups, respectively, as shown in row 6 of Table 3. Here "P" represents patch-based LBP. When compared with the "Raw+SVM" approach, LBP features improved gender recognition accuracy for seniors (from 85.32% to 90.64%), but this is still lower than the accuracy of 94.96% for adult faces. More interestingly, the accuracy reduced to 81.93% for young faces, which is even lower than the 84.38% accuracy of the "Raw+SVM" approach, and much lower than the 94.96% accuracy for adult faces. Again, gender recognition performance is very different for the three age groups using LBP features: high performance for adult faces, lower performance for senior faces, and very low performance for young faces. Possible reasons for this phenomenon are: (1) adult male and female faces have local texture differences that can be described well by the LBP operator, and (2) complex textures (e.g., wrinkles) on senior faces can also be described well by the LBP operator. For young faces, facial textures are not very rich and the main changes are facial shapes where the LBP operator does not work well.

Table 4. Gender recognition using all ages.

Methods	All Ages (0-93)
Raw + L-SVM	82.33%
Raw + N-SVM	89.28%
LBP(P) + L-SVM	83.37%
LBP(P) + N-SVM	90.53%
HOG + L-SVM	81.93%
HOG + N-SVM	88.65%
BIF + L-SVM	87.88%
BIF + N-SVM	92.25%
Average(L-SVM)	83.88%
Average(N-SVM)	90.18%

It should be mentioned that linear SVMs with LBP features did not perform well for gender as shown in row 5 of Table 3. In addition, the LBP operator performed much worse when applied to whole faces, as shown in rows 3 and 4 of Table 3, no matter what classifier was used.

7.3. BIF Results

For the biologically-inspired features, we need to find the best structure and setting. To simplify the process, the gender recognition is performed over all ages first. A two-fold cross validation was used as the test scheme. The same divisions of training and test data are used for all algorithms in the paper, either over all ages or at separate age groups.

First, we evaluated the C2 features with a nonlinear SVM for gender classification over all ages. The feature extraction process is almost the same as that in [23]. The only difference is the number of prototypes to represent the gender. Since we have 8,000 images for the two-class classification problem, a small number of prototypes cannot work well (not shown here). We let the algorithm randomly select 2,000 prototypes from the female faces for S2 and C2 feature calculations. An accuracy of 81.05% was obtained, as shown in the first column of Table 5. This result is much worse than the 89.28% using the raw pixel representation, the 88.65% accuracy of the HOG method, and the 90.53% of the LBP, shown in Table 4. We also randomly selected 2,000 prototypes from the male faces, and the result was 81.00% – almost the same. Finally, we also let the algorithm randomly select 4,000 prototypes from both males and females, and got an accuracy of 83.00% – still very low. From this experiment, we believe that C2 features do not work well for gender recognition. We notice that Meyers and Wolf [17] did not use C2 features in their face recognition problem, but they did not show any results when C2

Table 5. Gender recognition accuracies over all ages when the pre-learned prototypes were used for S2 and C2 features. We took eight bands and four orientations and used the SVM for classification as in [23].

Method	# Prototypes		
	2,000 (F)	2,000 (M)	4,000 (Mix)
C2 + N-SVM	81.05%	81.00%	83.00%

Table 6. Gender recognition accuracies using the biologically inspired features, varying with the number of bands (Bds.) and orientations.

	# Orientations				
Bds.	4	6	8	10	12
2	88.90%	89.55%	89.88%	89.68%	89.90%
4	90.05%	91.17%	91.65%	91.60%	91.37%
6	90.70%	91.63%	92.17%	**92.25%**	92.15%
8	90.97%	91.67%	90.45%	90.83%	92.15%

features were used for face recognition. Based on our experience, C2 features are not a good choice for face-based gender classification, although these features demonstrated super performance on object category recognition [23] [19].

Next, we used only C1 features [22] for gender recognition. In extracting the S1 features, we varied the number of bands and orientations for the Gabor filters. The bands were 2, 4, 6, 8, and the orientations were 4, 6, 8, 10, 12. All combinations of the bands and orientations were searched to determine which was best. Nonlinear SVMs were used for classification. The results are shown in Table 6. The best is 92.25% corresponding to six bands and ten orientations. More bands or orientations resulted in a decrease of the accuracies, while less number of bands or orientations are not sufficient for encoding gender features. Remember our purpose is to find how many bands and orientations are necessary for gender encoding in using the BIF. By comparing the four different representations, the highest accuracy that we obtained for gender recognition over all ages was 92.25%, as shown in Table 4, using the bio-inspired features with our changes.

Finally, we also measured gender recognition accuracies in the three age groups separately, using the BIF with six bands and ten orientations. As shown in Table 3, nonlinear SVMs gave better results than linear SVMs for each age group. The best results were 87.13%, 96.03%, and 92.34% for the young, adult, and senior people, respectively. The accuracies obtained from the bio-inspired features improve over the raw pixel based approach by 3.3%, 1.6%, and 8.2%, for young, adult, and senior, respectively. The accuracies were also higher than the HOG or LBP features either over all ages or for each age group. Although the accuracies improved by using the bio-inspired features, they are still significantly affected by age.

7.4. Average Results and Using All Ages

We calculate the average gender recognition accuracies over the four representations (raw pixels, LBP, HOG, and BIF) in each age group using kernel SVMs as classifiers, excluding the whole face based LBP. The averages were 84.97%, 94.90%, and 89.34%, for young, adult, and senior, respectively. The average of gender recognition accuracy for adult is still about 10% higher than for young people. The LBP, HOG, and BIF encode the gender with local spatial processing, which are better than the raw pixel based approach, thus the gender recognition accuracies for senior are improved a lot. In average, however, there is still more than 5% difference between the adult and senior. The significant average differences explain the aging effect again.

Let us also look at the gender recognition accuracies over all ages when the four different representations are used. The training data in each age group are used altogether for the SVM learning, and all testing data are tested. It is still a two-fold cross validation test. The classification accuracies were 89.28%, 90.53%, 88.65%, and 92.25% for the raw pixels, LBP, HOG, and BIF features, respectively, as shown in Table 4. The BIF features are slightly better than the other three. The HOG operator gives lower accuracy than the other three, although it has good performance for young faces. Please note that the accuracies of the four different features over all ages are much lower than on adult faces alone. This indicates that it is difficult to get high accuracy for face-based gender recognition when all ages are processed altogether. Indirectly, this demonstrates that the aging process on faces truly affects gender recognition performance significantly.

8. Discussion

The biologically-inspired features use the Gabor filters [9] in S1 layer, but the S1 filters are not limited to Gabor filters. For example, some motion filters are used for the S1 layer in action recognition [16]. The essence of the BIF is the hierarchical structure that contains layers from simple to complex. The “max” operation is important to generate the complex layers. Readers should not confuse the BIF with the Gabor filters, although they are related in the current implementation.

Psychological studies show that human recognition of gender is different for children’s and adults’ faces [25]. In their experiments, the stimuli consisted of 20 children’s and 20 adults’ face photos (a very small database). Our computational results on a large database support the observation in that psychological study [25]. Moreover, we showed a difference between adult and senior faces as well. Our new finding from empirical studies may inspire more psychological studies on gender recognition by humans using a broader range of ages.

The best result that we obtained for gender recognition over all ages was 92.25%, by using the bio-inspired C1 features with six bands and ten orientations for S1. It is still unknown how helpful this gender recognizer will be for making age estimation [11, 13, 29] automatic rather than manually separating males and females. Considering the low accuracies in recognizing gender from young and senior faces, we doubt that gender classification and age estimation can be done sequentially, i.e., given a test face, the system first recognizes the gender and then estimates the age. A better way might be developing an intertwined process to estimate age and gender cooperatively.

9. Conclusions

We have presented empirical studies that show that gender recognition performance is affected significantly by human age. Based on the results on a large database of 8,000 faces with an age span from 0 to 93 years, we found that gender recognition accuracies can be 10% higher on adult faces than on young or senior faces, using one of the state-of-the-art methods, "Raw+SVM." We then examined the aging effects on human faces, and chose features based on the HOG and LBP operators to characterize shape changes and skin texture variations. We also investigated biologically-inspired features for gender recognition. Even if evaluated with the more advaced features, such as HOG, LBP, and BIF, gender recognition is still affected significantly by human age. Our results may inspire more psychological studies on gender recognition by humans using a broader range of ages. Computationally, more efforts are needed for gender classification with large age variations.

References

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

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

[3] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. In *the Eur. Conf. on Comput. Vision*, pages 469–481, 2004.

[4] S. Baluja and H. A. Rowley. Boosting sex identification performance. *Int. J. of Comput. Vision*, 71(1):111–119, 2007.

[5] V. Bruce, A. Burton, E. Hanna, P. Healey, and O. Mason. Sex discrimination: How do we tell the difference between male and female faces? *Perception*, 22:131–152, 1993.

[6] R. Brunelli and T. Poggio. Hyperbf networks for gender classification. In *Proc. DARPA Image Understanding Workshop*, pages 311–314, 1992.

[7] G. Cottrell. Empath: Face, emotion, and gender recognition using holons. In *Advances in Neural Information Processing Systems 3*, pages 564–571, 1991.

[8] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Conf. on Comput. Vision and Pattern Recognit.*, pages 886–893, 2005.

[9] J. Daugman. Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America-A*, 2(7):1160C1169, 1985.

[10] Y. Freund and R. Schapire. Experiments with a new boosting algorithm. In *Proc. the Thirteen International Conference on Machine Learning*, pages 148–156, 1996.

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

[12] B. Golomb, D. Lawrence, and T. Sejnowski. Sexnet: A neural network identifies sex from human faces. In *Advances in Neural Information Processing Systems 3*, pages 572–577, 1991.

[13] 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.

[14] S. Gutta, H. Wechsler, and P. Phillips. Gender and ethnic classification. In *Proc. the IEEE Int. workshop on Automatic Face and Gesture Recognit.*, pages 194–199, 1998.

[15] D. Hubel and T. Wiesel. Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology*, 148:574–591, 1959.

[16] H. Jhuang, T. Serre, L. Wolf, and T. Poggio. A biologically inspired system for action recognition. In *Int. Conf. on Comput. Vision*, 2007.

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

[18] B. Moghaddam and M.-H. Yang. Learning gender with support faces. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(5):707–711, 2002.

[19] 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.

[20] T. Ojala, M. Pietikainen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognit.*, 29:51–59, 1996.

[21] P. Phillips, H. Moon, S. Rizvi, and P. Rauss. The FERET evaluation methodology for face recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(10):1090–1104, 2000.

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

[23] 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.

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

[25] H. A. Wild, S. E. Barrett, M. J. Spence, A. J. O'Toole, Y. D. Cheng, and J. Brooke. Recognition and sex categorization of adults' and children's faces: examining performance in the absence of sex-stereotyped cues. *J. of Exp. Child Psychology*, 77:269–291, 2000.

[26] L. Wiskott, J.-M. Fellous, N. Kruger, and C. von der Malsburg. Face recognition and gender determination. In *Proc. Int'l Workshop Face and Gesture Recognition*, pages 92–97, 1995.

[27] X. Xu and T. S. Huang. SODA-Boosting and its application to gender recognition. In *LNCS 4778: 2007 IEEE International Workshop on Analysis and Modeling of Faces and Gestures*, pages 193–204, 2007.

[28] M. K. Yamaguchi, T. Hirukawa, and S. Kanazawa. Judgment of sex through facial parts. *Perception*, 24:563–575, 1995.

[29] S. Yan, X. Zhou, M. Liu, M. H-Johnson, and T. S. Huang. Regression from patch-kernel. In *Conf. on Comput. Vision and Pattern Recognit.*, 2008.

[30] Z. Yang and H. Ai. Demographic classification with local binary patterns. In *Int. Conf. on Biometrics*, pages 464–473, 2007.