Age Estimation Based on Complexity-Aware Features

Haoyu Ren, Ze-Nian Li

Vision and Media Lab
School of Computing Science
Simon Fraser University
8888 University Drive
Vancouver, BC, Canada

Abstract. The research related to age estimation using face images has become increasingly important. We propose an age estimator using two kinds of local features, the gradient features which well describe the local characteristic, and the Gabor wavelets which reflect the multi-scale directional information. The RealAdaBoost algorithm with a complexity penalty term in the feature selection module is applied to choose meaningful regions from human face for feature extraction, while balancing the discriminative capability and the computation cost at the same time. Furthermore, the hierarchical classifier, which is composed of an age group classification (e.g., 15-39 years old, 40-59 years old etc.) and a detailed age estimation (e.g. 19, 53 years old, etc.) are utilized to get the final age. Experimental results show that the proposed approach outperforms the methods using single feature on PAL and FG-NET database. It also achieves competitive accuracy with the state-of-the-art algorithms.

1 Introduction

Systems based pattern recognition have been proven to be very useful in many areas such as security and access control, human detection, human computer interaction, and brain computer interface. Recently, the research related to age estimation using face images is more important than ever. The potential applications include automatic guest enrollment, parent TV program control, video surveillance, etc.

In general, age estimation systems consist of two steps: feature extraction and classification/regression [25]. The features used in age estimation can be categorized into the local features and the global features. The local features are extracted on some regions which might contain specific facial characteristic, such as wrinkles and freckles. They have been used to classify people into age groups. Conversely, the global features are extracted based on whole face shape or all facial feature points. They are generally used to estimate the exact age. Some researchers also use hybrid features to improve the estimation accuracy, which is the combination of local features and global features.

After feature extraction, the classification/regression module is utilized to train the age estimator. The commonly-used algorithms include the age group classification, single-level estimation and the hierarchical age estimation. Age group classification is an approach that roughly predicts an age group, whereas single-level method focuses on detailed age prediction. The hierarchical method is a coarse-to-fine method which integrates the single-level and age group methods together.

Regarding the efficiency issue, local features based methods perform better compared to global features based methods utilizing ASM [1] or AAM [2]. Unfortunately, the use of the local features for age estimation has not been well investigated. The methods extracting a dense feature vector for each local region in the aligned face might lead to dimension redundant. In addition, using dense feature is relatively slow. Some other algorithms use AdaBoost to select key dimensions [3] from the dense features. But it is difficult to describe a specific pattern using single dimensional feature in complicate recognition tasks. It also leads to potential risk of weakening the discriminative power of the resulting classifier.

To solve this problem, we focus on selecting the meaningful regions in human face for feature extraction, while dealing with the accuracy and efficiency at the same time. This paper has two contributions. Firstly, we integrate two kinds of localized features together for age estimation, including SIFT, HOG, and Gabor. In addition, different from simply mixing or concatenating these features, we use the complexity-aware RealAdaBoost algorithm, which includes a complexity penalty term in the process of feature selection. As a result, both the discriminative power and the computation cost of the features are evaluated in the training procedure. We divide the training samples into 4 age groups, 0-15, 16-40, 41-59, and 60+. The above complexity-aware RealAdaBoost is applied to select the meaningful regions on each age group respectively. Then the support vector machine is utilized to train a hierarchical age estimator based on these selected features. Plenty of experiments on public datasets are used to evaluate our method. The experimental results show that our approach achieves significant improvement on the estimation accuracy compared to using single features. The result is also competitive with the state-of-the-arts approaches in PAL and FG-NET database.

The rest of this paper is organized as follows. Section 2 is the related work. Section 3 presents the features used in this paper. Section 4 introduces the RealAdaBoost algorithm with the complexity-aware criterion. Section 5 shows our experimental results. Conclusion is given in the last section.

2 RELATED WORK

There has been a great number of work about feature extraction for age estimation. Kwon and Lobo [4] classify facial images into three age groups using the distance ratio of facial components and the wrinkles. Hayashi et al. [5] use histogram equalization and Hough transform for skin extraction and wrinkle detection. A lookup table containing the wrinkle state against appearance at a given age and gender is utilized for age estimation. Fukai et al. [6] adopt fast Fourier transform to extract features from a face image by genetic algorithms. Gao et al. [7] integrate Gabor features and a fuzzy version of Linear Discriminant Analysis (LDA) to classify face into various age classes. Mu et al. [8] use biologically inspired features and introduced a new operator to model the aging process. Yan et al. [9] combine the local feature and global feature together and utilize a hierarchical classifier to improve the performance.

The problem of age estimation can be converted into a classification/regression problem. Classification can be in groups such as babies, teens, adults or 1-5, 5-10, 10-15, while the regression method predicts the exact age based on a set of coefficients learnt by using suitable loss functions. Lanitis et al. [10] approach the problem of age estimation in a regression way. They propose a quadratic function where age is dependent on feature vector extracted from the face. Ueki et al. [11] introduce a two phased approach based on LDA and 2D-LDA and have used only the first four dimensions of the extracted features to make Gaussian classifier to classify images in various age groups. Wang et al. [20] propose a novel data selection of the Furthest Nearest Neighbour (FNN) that generalizes the margin-based uncertainty to the multi-class case to handle large data efficiently in age classification. Guo et al. [23] and Liu et al. [15] solve the problem by Support Vector Machines (SVM) and Support Vector Regression(SVR). Ni. et al. [24] utilize a robust multi-instance regression learning algorithm to learn the kernel regression-based human age estimator in the presence of bag label noises. Kohli et al. [13] propose a technique which extracts features based on AAM and use a global classifier to obtain a rough estimate distinguishing between child/teen-hood and adulthood. An improved version of their work based on hierarchical classifier is published in [14]. Geng et al. [26] develop two algorithms, named IIS-LLD and CPNN, which make single face image not only contribute to the learning of its chronological age, but also to the learning of its adjacent ages.

3 FEATURES USED FOR FACE DESCRIPTION

In this section, we will introduce the three localized features used in our method, SIFT, HOG, and Gabor wavelets.

3.1 Gradient features

Scale Invariant Feature Transform (SIFT) is invariant to scaling, translation and rotation, and partially invariant to illumination changes and affine projection. Using these descriptors, objects can be reliably recognized even in the case of different views, low illumination or occlusion. In SIFT feature extraction, we first build a scale space by convolving it with multi-scale Gaussian kernels and then calculate the Difference of Gaussian (DoG) between each two adjacent scale spaces. The maximum and minimum of the DoG are selected as candidate interest points, from which elements with low contrast and edge responses are excluded.

After key points detection, we summarize information about local gradient around each key point, as shown in Fig. 1. The histogram of gradient orientation is computed as the resulting feature vector. 4×4 histograms with 8 orientation bins are extracted for each candidate region. The final dimension of SIFT feature is $4 \times 4 \times 8 = 128$.

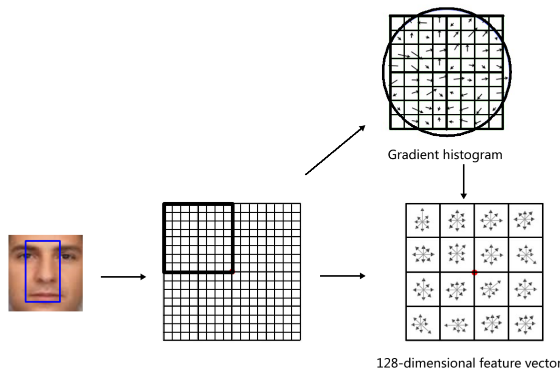


Fig. 1. SIFT feature extraction

Histogram of Oriented Gradient (HOG) divides the image region into a cell-block structure and generates histogram based on the gradient orientation and spatial location. The input region (block) is divided into small connected regions, called cells, and for each cell a histogram of edge orientations is computed. The histogram channels are evenly spread from 0 to 180 degrees. The histogram counts are normalized for illumination compensation. This can be done by accumulating a measure of local histogram energy over the somewhat larger connected regions and using the results to normalize all cells in the block. The concatenation of these histograms yields the final HOG descriptor. We extract 4 cells and 8 gradient orientation bins for each candidate block, as shown in Fig. 2. The dimension of HOG is $4 \times 8 = 32$.

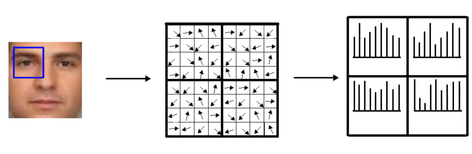


Fig. 2. HOG feature extraction

HOG is not invariant to rotation, but the computation cost is only 1/5 compared to SIFT. This will be considered in the complexity-aware process of the RealAdaBoost procedure.

3.2 Gabor filters

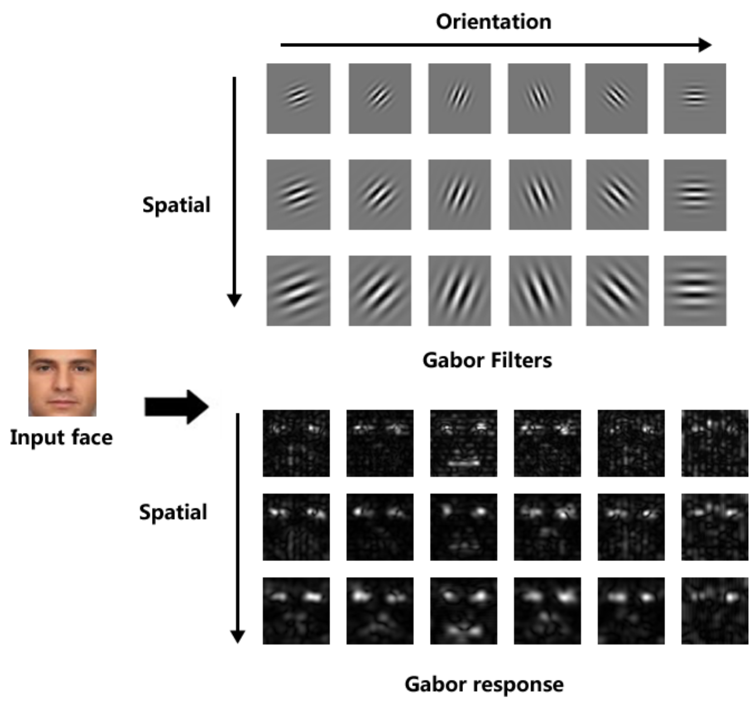


Fig. 3. Gabor filters using 3 scales and 6 orientations

The Gabor wavelets, whose kernels are similar to the 2D receptive field profiles of the mammalian cortical simple cells, exhibit desirable characteristics of spatial locality and orientation selectivity, and are optimally localized in the space and frequency domains. The Gabor wavelets are defined in equation (1)

$$\phi_{\boldsymbol{k}}(\boldsymbol{z}) = \frac{\boldsymbol{k}^2}{\sigma^2} e^{\frac{\boldsymbol{k}^2 \boldsymbol{z}^2}{2\sigma^2}} [e^{i\boldsymbol{k}\boldsymbol{z}} - e^{-\frac{\sigma^2}{2}}], \quad \ldots (1)$$

where σ decides the ratio of the window width and the wave length, z is the normalization vector, k controls the width of the Gaussian function, the wave length and direction of the shocking part, defined as follows:

$$\boldsymbol{k} = k_v e^{i\phi_u},$$

where $k_v = k_{max}/f_v$ and $\phi_u = \pi u/n$. k_{max} is the maximum frequency, f is the spacing factor between kernels in the frequency domain, n is the maximum orientation number.

The Gabor wavelets in (1) can be generated from the mother wavelet, by scaling and rotation via the wave vector $\boldsymbol{k}$. Each kernel is a product of a Gaussian envelope and a complex plane wave, while the first term in the square brackets in (1) determines the oscillatory part of the kernel and the second term compensates for the DC value. The effect of the DC term becomes negligible when the parameter σ, which determines the ratio of the Gaussian window width to wavelength, has sufficiently large values. In our case, we utilize three scales and six orientations to represent the components. And we set

$$\sigma = 2\pi \quad k_{max} = \frac{\pi}{2} \quad f = \sqrt{2}.$$

An example of the extracted Gabor features of an input face are illustrated in Fig. 3.

The feature dimension of dense Gabor feature depends on the size of the block, it will be quite high if we want to extract features in a large region. So we utilize a sub-sampling strategy, which applies a 2×2 to 6×6 sub-sampling based on the block size. The Gabor features are extracted only on the sub-sampled pixels. Using this strategy, the minimum feature dimension of Gabor is $3 \times 6 \times 16 = 256$ (8×8 block with 2×2 sub-sampling), and the maximum is $3 \times 6 \times 36 = 648$ (40×40 block with 6×6 sub-sampling).

4 LEARNING THE FEATURES USING REALADABOOST WITH COMPLEXITY PENALTY TERMS

We utilize RealAdaBoost to select the key features classifying each age group respectively. In RealAdaBoost, an image feature can be seen as a function from the image space to a real valued range $f : \mathbf{x} \rightarrow [f_{min}, f_{max}]$. The weak classifier based on f is a function from the feature vector $\mathbf{x}$ to a real valued classification confidence space. For the binary classification problem, suppose the training data as $(\mathbf{x_1}, y_1), \ldots, (\mathbf{x_n}, y_n)$ where $\mathbf{x}_i$ is the training sample and $y \in \{-1, 1\}$ is the class label, we first divide the sample space into N_b several equal sized sub-ranges B_j

$$X_j = \{\mathbf{x} | f(\mathbf{x}) \in B_j\}, j = 1, \ldots, N_b. \quad \ldots (2)$$

The weak classifier is defined as a piecewise function

$$h(\mathbf{x}) = \frac{1}{2} ln(\frac{W_+^j + \epsilon}{W_-^j + \epsilon}), \quad \ldots (3)$$

where ϵ is the smoothing factor, $W_\pm$ is the probability distribution of the feature value for positive/negative samples, implemented as a histogram

$$W_{\pm}^{j} = P(\mathbf{x} \in X_j, y \in \{-1, 1\}), j = 1, \ldots, N_b. \quad \ldots (4)$$

The best weak classifier is selected according to the classification error Z of the piecewise function in equation (5)

$$Z = 2\sum_{j} \sqrt{W_{+}^{j} W_{-}^{j}}. \quad \ldots (5)$$

We adopt RealAdaBoost to learn the key regions and the type of feature extraction methods. In consideration of the efficiency, we add a complexity-aware criteria into the decision term of RealAdaBoost, as shown in equation (6)

$$Z = 2\sum_{j} \sqrt{W_{+}^{j} W_{-}^{j}} + a \cdot fp \cdot C, \quad \ldots (6)$$

where C is the computation cost of the features, a is the complexity-aware factor to balance the discriminative capability and the computation complexity, fp is the false positive rate of current stage. In the training procedure, features with minimum Z are selected.

The equation (6) could be explained as follows, in the first stages of RealAdaBoost, the age group of faces are still easy to be classified, so efficient features are preferred. In the following stages, the patterns of the training samples are complicated. Then the features with high computation cost are considered.

We apply the above RealAdaBoost on the 4 age group classification tasks respectively. In each task, the positive samples are the samples in that age group, while the negative samples are the combination of all samples in other 3 groups. To evaluate the computation cost, we test the execution time of different feature extraction modules and set the C of SIFT to 10, HOG to 2, and Gabor to 3-6 based on its dimension. The complexity-aware factor a is set to 0.15 in our experiment. The diagram of the whole complexity-aware RealAdaBoost is illustrated in Fig. 4.

5 Experiments

5.1 Experiments setup

In the experiments, two databases are used to evaluate the performance of the proposed method: the PAL aging database and the FG-NET aging database. The PAL aging database contains 430 Caucasians with age range 18-93 years old [16]. The images in the database were captured using a digital camera with fixed light and position conditions. The resolution of the images is 640×480 pixels. This database includes various expressions such as smiling, sadness, anger, or neutral faces. In the experiments, we used only neutral faces in order to exclude the facial expression effect. Sample images used in our experiments are shown in Fig. 5.

Parameters
N number of training samples
M number of evaluated features each iteration
T maximum number of weak classifiers

Input: Training set
$\{(\mathbf{x}_i, y_i)\}, i = 1, \ldots, N, \mathbf{x}_i \in R^d, y_i \in \{-1, 1\}$

1. Initialize sample weight, classifier output, and false positive rate
$w_i = \frac{1}{N}, F(\mathbf{x}_i) = 0, i = 1, \ldots, N, fp_0 = 1$
2. Repeat for $t = 1, 2, \ldots, T$
   2.1 Update the sample weight w_i using the h^{th} weak classifier output
   $w_i = w_i e^{-y_i h_i(\mathbf{x}_i)}$
   2.2 For $m = 1$ to M
      2.2.1 Generate a random region with a specific feature extraction method (SIFT, HOG, or Gabor)
      2.2.2 Extract features and do least square to $y_i \in \{-1, 1\}$
      2.2.3 Build the predict distribution function W_+ and W_-
      2.2.4 Select the best feature which minimizes Z in equation (6)
   2.3 Update weak classifier using (3)
   2.4 Update strong classifier $F_{t+1}(\mathbf{x}_i) = F_t(\mathbf{x}_i) + h_t(\mathbf{x}_i)$
   2.5 Calculate current false positive rate fp_t
3. Output classifier $F(\mathbf{x}) = sign[\sum_{j=1}^{T} h_j(\mathbf{x})]$

Fig. 4. Learning the features using RealAdaBoost with complexity penalty term

Fig. 5. Sample images in PAL aging database

The FG-NET aging database [19] is one of the most frequently used database for estimating age in the previous works. The database has 1,002 images composed of 82 Europeans in the age range 0-69 years old. Individuals in the database have one or more images included at different ages. These Images were obtained by scanning. Therefore, there are extreme variations in lighting, expression, background, pose, resolution and noise from scanning. Sample images of the FG-NET aging database are shown in Fig. 6.

Fig. 6. Sample images in FG-NET aging database

With the PAL aging databases, five-folds cross validations are performed to evaluate the performance, which is similar to [17]. The age and gender are evenly distributed each fold. With the FG-NET aging database, Leave-One-Person-Out (LOPO) is performed because it contains a number of images of the same person. That means, 82-folds are used on the FG-NET aging database.

We divide the training samples into 4 age groups, 0-15, 16-40, 41-59, and 60+. All the faces are resized to 100×100. The complexity-aware RealAdaBoost is applied on each group classification to select the meaningful features. Moreover, a two-steps hierarchical classifier is further adopted to generate the final age estimator. Firstly, linear support vector machine based age group classification is trained based on the selected features. Then we use the support vector regression to estimate the exact age in each age group.

The evaluation is based on the Mean Absolute Error (MAE) and the Cumulative Score (CS). The MAE is defined as the mean of the absolute difference between the estimated age and the real age, as

$$MAE = \frac{\sum_{i=1}^{N} |e_i - g_i|}{N},$$

where N is the number of the test images, e_i is the estimated age of the test image i and g_i is the ground-truth age. The Cumulative Score(CS) is defined as the ratio of the number of data whose errors are lower than a threshold, as

$$CS = \frac{N_{error \leq threshold}}{N}.$$

5.2 Experimental results

We train 7 age estimators using the proposed framework, which includes the classifiers utilizing single feature (SIFT, HOG, and Gabor), the combination of

Table 1. MAE in PAL database. Units: years old

Approach	Mean Absolute Error (MAE)
SIFT	5.98
HOG	6.14
Gabor	5.88
SIFT + HOG	5.54
SIFT + Gabor	5.05
HOG + Gabor	5.57
All three features	**4.29**
[17]	5.36
[9]	4.33
[18]	4.52

two features, and all of the three features. There are no complexity-aware procedure if single feature is adopted. Table 1 and table 2 present the MAE of these age estimators in the PAL database and FG-NET database. It can be seen that using the complexity-aware feature combination, the estimation accuracy is significantly improved compared to using SIFT, HOG, or Gabor independently. Using all three features, the MAE is further reduced. The accuracy is also comparable with the state-of-the-art algorithms in both of the two datasets. This result show that the features evaluated by complexity-aware RealAdaBoost might be more effective than some artificial designed features.

Table 2. MAE in FG-NET database. Units: years old

Approach	Mean Absolute Error (MAE)
SIFT	5.97
HOG	5.86
Gabor	5.68
SIFT + HOG	5.27
SIFT + Gabor	5.09
HOG + Gabor	5.10
All three features	**4.49**
[21]	5.05
[9]	4.66
[22]	4.67

We also plot the curve of the cumulative scores for the above 7 age estimators in Fig. 7. It can been seen that the cumulative score moved up at a clear border on the PAL database and FG-NET database using the proposed complexity-aware method (black curve). This result also shows the effectiveness of our method.

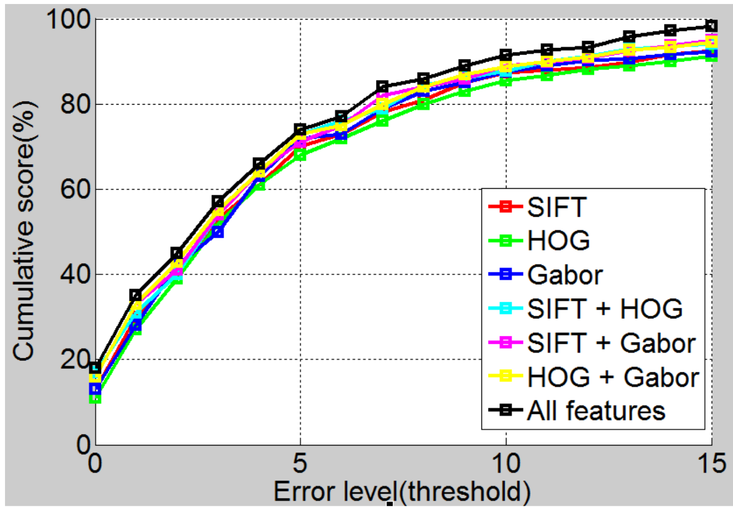
5.3 Analysis

We draw the first 7 features selected by the RealAdaBoost algorithm in LFW database. The LFW is used instead of any age databases because LFW contains a large number of faces in variable illuminations, ages, emotions, and ethnics. As shown in Fig. 8, there are 2 SIFT features, 3 HOG features, and 2 Gabor features. It could be seen that these features lay on the eyes, forehead and mouth region. This result is reasonable, because it is much easier to estimate the age from these regions rather than other face regions such as nose or eyebrow. For example, the wrinkles in the forehead and the shape of month describe the key characteristic for human ages.

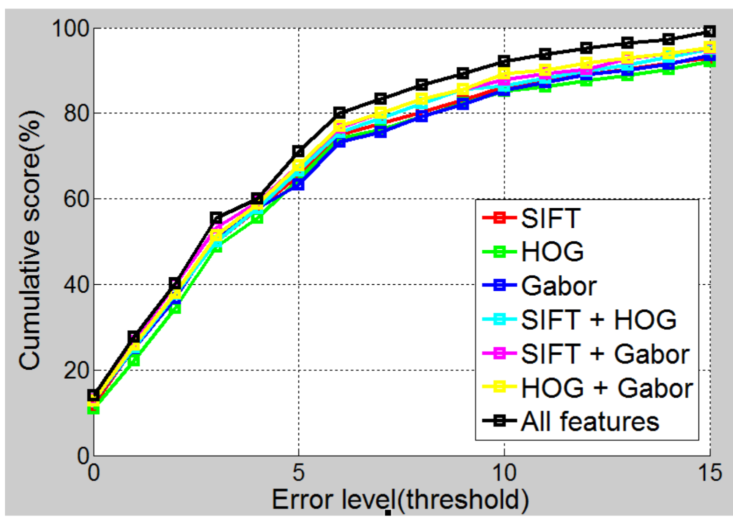
We test the resulting classifiers on a desktop PC with a 2.5 GHz I3 PC and 2 GB memory. The execution speed is shown in Table 3. We find that the estimator based on SIFT is relatively slow compared to the one using HOG or Gabor. If we combine these features together and use the complexity-aware strategy, the execution time will be reduced, shown as the rows with asterisks. Furthermore, if all three features are used, the speed is significantly improved from 26.88ms per face to 12.22ms per face using the complexity-aware RealAdaBoost. So we can get the conclusion that the proposed method contributes to both the accuracy and the efficiency of age estimation.

Table 3. Execution Speed of age estimators. Item with * denotes that the complexity-aware strategy is adopted.

Approach	Recognition time per face(ms)
SIFT	37.54
HOG	13.09
Gabor	20.13
SIFT + HOG (*)	22.11
SIFT + Gabor (*)	16.22
HOG + Gabor (*)	15.09
All three features	26.88
All three features (*)	**12.22**



(a) CS on PAL dataset



(b) CS on FG-NET dataset

Fig. 7. CS on PAL and FG-NET database

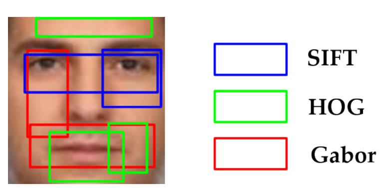


Fig. 8. The first 7 features selected by RealAdaBoost in human face

6 Conclusion

In this paper, we proposed a local feature based face representation for age estimation. We used the RealAdaBoost algorithm with a complexity penalty term to select the meaningful features, which successfully balances the accuracy and efficiency. High age estimation accuracy were reported in comparison to previously published results on two famous datasets. 4.29 MAE was achieved for PAL, and 4.49 was achieved for FG-Net.

The approach proposed in this paper could be further studied. We have already found that the proposed framework is also effective on other recognition tasks, such as gender recognition and emotion recognition.

Acknowledgement. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada under the Grant RGP36726.

References

1. Milborrow, Stephen and Nicolls, Fred: Locating facial features with an extended active shape model. European Conference on Computer Vision. (2008) 504–513.
2. Cootes, Timothy F and Edwards, Gareth J and Taylor, Christopher J and others: Active appearance models. IEEE Transactions on Pattern Analysis and Machine Intelligence. **23** (2001) 681–685
3. Shan, Caifeng: Learning local features for age estimation on real-life faces. Proceedings of the 1st ACM International Workshop on Multimodal Pervasive Video Analysis. (2010) 23–28
4. Kwon, Young Ho and da Vitoria Lobo, Niels: Age classification from facial images. Computer Vision and Pattern Recognition. (1994) 762–767
5. Hayashi, J and Yasumoto, M and Ito, H and Koshimizu, H: Method for estimating and modeling age and gender using facial image processing. Seventh International Conference on Virtual Systems and Multimedia. (2001) 439–448
6. Fukai, Hironobu and Takimoto, Hironori and Mitsukura, Yasue and Fukumi, Minoru: Apparent age estimation system based on age perception. Proc. of SICE. (2007) 2808–2812
7. Gao, Feng and Ai, Haizhou: Face age classification on consumer images with gabor feature and fuzzy lda method. Advances in biometrics. (2009) 132–141
8. Mu, Guowang and Guo, Guodong and Fu, Yun and Huang, Thomas S: Human age estimation using bio-inspired features. Computer Vision and Pattern Recognition. (2009) 112–119
9. Choi, Sung Eun and Lee, Youn Joo and Lee, Sung Joo and Park, Kang Ryoung and Kim, Jaihie: Age estimation using a hierarchical classifier based on global and local facial features. Pattern Recognition. **44** (2011) 1262–1281
10. Lanitis, Andreas and Taylor, Christopher J. and Cootes, Timothy F: Toward automatic simulation of aging effects on face images. IEEE Transactions on Pattern Analysis and Machine Intelligence. **24** (2002) 442–445
11. Subspace-based age-group classification using facial images under various lighting conditions: Ueki, Kazuya and Hayashida, Teruhide and Kobayashi, Tetsunori. 7th International Conference on Automatic Face and Gesture Recognition. (2006) 6–pp
12. Guo, Guodong and Fu, Yun and Huang, Thomas S and Dyer, Charles R: Locally adjusted robust regression for human age estimation. Urbana. **51** (2008) 61801

13. Luu, Khoa and Ricanek, Karl and Bui, Tien D and Suen, Ching Y: Age estimation using active appearance models and support vector machine regression. IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems. (2009) 1–5
14. Kohli, Sharad and Prakash, Surya and Gupta, Phalguni: Hierarchical age estimation with dissimilarity-based classification. Neurocomputing. **120** (2013) 164–176
15. Liu, Jianyi and Ma, Yao and Duan, Lixin and Wang, Fangfang and Liu, Yuehu: Hybrid constraint SVR for facial age estimation. Signal Processing. **94** (2014) 576–582
16. Minear, Meredith and Park, Denise C: A lifespan database of adult facial stimuli. Behavior Research Methods, Instruments, & Computers. **36** (2004) 630–633
17. Suo, Jinli and Wu, Tianfu and Zhu, Songchun and Shan, Shiguang and Chen, Xilin and Gao, Wen: Design sparse features for age estimation using hierarchical face model. 8th IEEE International Conference on Automatic Face & Gesture Recognition. (2008) 1–6
18. Guo, Guodong and Mu, Guowang and Fu, Yun and Dyer, Charles and Huang, Thomas: A study on automatic age estimation using a large database. IEEE 12th International Conference on Computer Vision. (2009) 1986–1991
19. FGNET. http:
www.fgnet.rsunit.com.
20. Wang, Jian-Gang and Sung, Eric and Yau, Wei-Yun: Active learning with the furthest nearest neighbor criterion for facial age estimation. Asian Conference on Computer Vision. (2010) 11–24
21. Kilinc, Merve and Akgul, Yusuf Sinan: Automatic Human Age Estimation Using Overlapped Age Groups. Computer Vision, Imaging and Computer Graphics. Theory and Application. (2013) 313–325
22. Chen, Ke and Gong, Shaogang and Xiang, Tao and Loy, Chen Change: Cumulative attribute space for age and crowd density estimation. Computer Vision and Pattern Recognition. (2013) 2467–2474
23. Guo, Guodong and Wang, Xiaolong: A study on human age estimation under facial expression changes. Computer Vision and Pattern Recognition. (2012) 2547–2553
24. Ni, Bingbing and Song, Zheng and Yan, Shuicheng: Web image and video mining towards universal and robust age estimator. IEEE Transactions on Multimedia. **13** (2011) 1217–1229
25. Fu, Yun and Guo, Guodong and Huang, Thomas S: Age synthesis and estimation via faces: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence. **32** (2010) 1955–1976
26. Geng, Xin and Yin, Chao and Zhou, Zhi-Hua: Facial age estimation by learning from label distributions. IEEE Transactions on Pattern Analysis and Machine Intelligence. **35** (2013) 2401–2472