2014 22nd International Conference on Pattern Recognition

Gender Recognition using Complexity-Aware Local Features

Haoyu Ren, Ze-Nian Li
Vision and Media Lab
School of Computing Science
Simon Fraser University
Vancouver, BC, Canada
Email: {hra15, li}@cs.sfu.ca

***Abstract*—We propose a gender classifier using two types of local features, the gradient features which have strong discrimination capability on local patterns, and the Gabor wavelets which reflect the multi-scale directional information. The RealAdaBoost algorithm with complexity penalty term 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. Linear SVM is further utilized to train a gender classifier based on the selected features for accuracy evaluation. Experimental results show that the proposed approach outperforms the methods using single feature. It also achieves comparable accuracy with the state-of-the-art algorithms on both controlled datasets and real-world datasets.**

I. Introduction

Gender recognition is a well-established issue for automatic face recognition. Successful gender recognition can boost a large number of advanced applications, such as customer information measurement, surveillance systems and interfaces, content-based indexing and searching, and demographic studies.

Gender recognition algorithms could be divided into two categories, geometric-based methods and appearance-based methods. The geometric-based methods consider using the geometric relationship of the facial features. While the geometric relationships are maintained, other information may be discarded [1]. The gender could be recognized by the accurately extracted facial feature points [2]. Brunelli and Poggio [30] trained a hyper basis function network classifier based on 18 point-to-point distances. Fellous [31] used 22 normalized fiducial distances of 40 manually extracted points.

The appearance-based methods utilize a classifier trained on the information extracted on the image pixels to get the gender information. Some researchers use efficient features such as the Local Binary Patterns (LBP) [5][6][9] and Weber Local Descriptor (WLD) [7]. Others adopt more complicate features including the gradient information or wavelet functions. Among these features, the Scale Invariant Feature Transform (SIFT) is one of the most commonly-used ones because it is invariant to image scaling, translation and rotation [10]. Using SIFT descriptor, the objects can be reliably recognized even from different views or under occlusion. Demirkus et al. [21] utilized a Markovian model to classify face gender from unconstrained video in natural scenes. Wang et al. [22] extracted SIFT descriptors at regular image grid points and combined it with global shape contexts of the face. Y. El-Din [19] proposed a decision-level fusion framework combining SIFT with LBP descriptor extracted from the whole face image.

Another commonly-used gradient feature is the Histogram of Oriented Gradient (HOG) features [8], which is able to capture local shape information from the gradient structure with easily controllable degree of invariance to translations [11]. HOG is often used in gender recognition with body information. Bourdev et al. [28] used a set of patches called poselets, represented by HOG features, color histogram, and skin features. The poselets were used to train attribute classifiers which were combined to infer gender using context information. Collins et al. [29] proposed Pixel HOG (PiHOG) descriptor computed from a custom edge map. In their method, color information was captured using a histogram computed based on the hue and saturation value of the pixels.

Gabor feature is another effective representation for face image analysis. Research in neurophysiology has shown that Gabor filters fit the spatial response profile of certain neurons in the visual cortex of the mammalian brain. Lian et al. [23] followed the method by Hosoi et al. [24], used Gabor wavelets from different facial points located using retina sampling. Leng and Wang [13] extracted Gabor wavelets of five different scales and eight orientations from each pixel of the image, which were then selected using AdaBoost. Scalzo et al. [26] extracted a large set of features using Gabor and Laplace filters which are used in a feature fusion framework of which the structure was determined by genetic algorithm.

Although the above methods show promising results on some datasets, the use of the local features for gender recognition has not been well investigated. The algorithms extracting a dense feature vector for each pixel 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 [6][13] from the dense features. In general, it is difficult to describe specific patterns using single dimensional feature, especially for some complicate object detection tasks. It will also lead 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

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.414

balance the efficiency and accuracy at the same time. The SIFT, HOG, and Gabor features are considered as candidates in feature selection. Different from simply mixing or concatenating these features, we use the complexity-aware RealAdaBoost algorithm, which includes a complexity penalty term in the procedure of feature selection. Both the discriminative power and the computation cost of the features are evaluated in the training procedure, and the features best balancing them will be selected. Linear SVM is further utilized to generate the final classifier using the selected features for accuracy evaluation. Plenty of experiments on public datasets are used to evaluate our method. The experimental results show that our approach achieves significant improvement on the recognition accuracy compared to using single features. The result is also comparable with the state-of-the-arts approaches in FERET, KinFace, and LFW datasets.

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

II. FEATURES USED FOR FACE DESCRIPTION

A. Gradient features

Scale Invariant Feature Transform (SIFT) is invariant to image scaling, translation and rotation, and partially invariant to illumination changes and affine projection. Using SIFT descriptor, objects can be reliably recognized even from different views, low illumination or under occlusion. Another advantage is that some preprocessing stages such as the accurate face alignment are not required using invariant features [20]. In the SIFT extraction, we first build a scale space of the input region by convolving it with a variable-scale Gaussian kernel. Then the Difference of Gaussian (DoG) between each two adjacent layers in the scale space are calculated. 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 keypoint detection, each keypoint is assigned a descriptor that summarizes information on local image gradient flows, as shown in Fig. 1(a). The final feature vector is the histogram of gradient orientation computed in an interest region around the keypoints. In our work, we extract 4×4 histograms with 8 orientation bins for each region. So the dimension of SIFT is $4 \times 4 \times 8 = 128$.

Histogram of Oriented Gradient (HOG) breaks the image region into a cell-block structure and generates histogram based on the gradient orientation and spatial location. The input region (block) is first divided into small connected regions, called cells, and for each cell a histogram of edge orientation is computed. The histogram channels are evenly spread from 0 to 180 degrees. Furthermore, 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 combination of these histograms represents the final HOG descriptor. In our implementation, we extract 4 cells and 8 gradient orientation

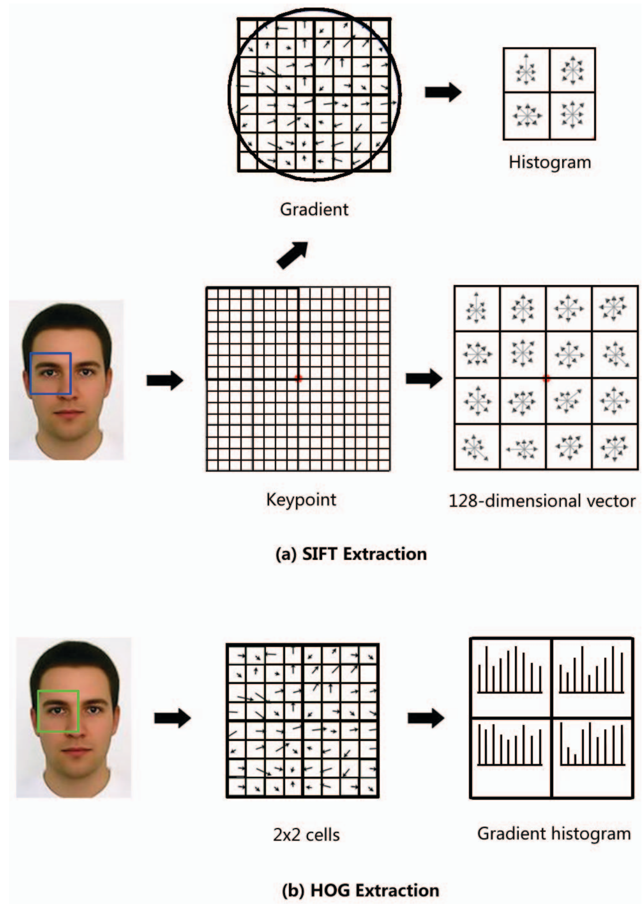

Fig. 1. Gradient feature extractions. (a) SIFT (b) HOG

bins for each block, as shown in Fig. 1(b). The dimension of HOG is $4 \times 8 = 32$.

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

B. Gabor filters

The Gabor filters, which could effectively extract the image local directional features at multiple scales, have been successfully and prevalently used in face recognition. The Gabor wavelets defined in equation (1), 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.

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

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:

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

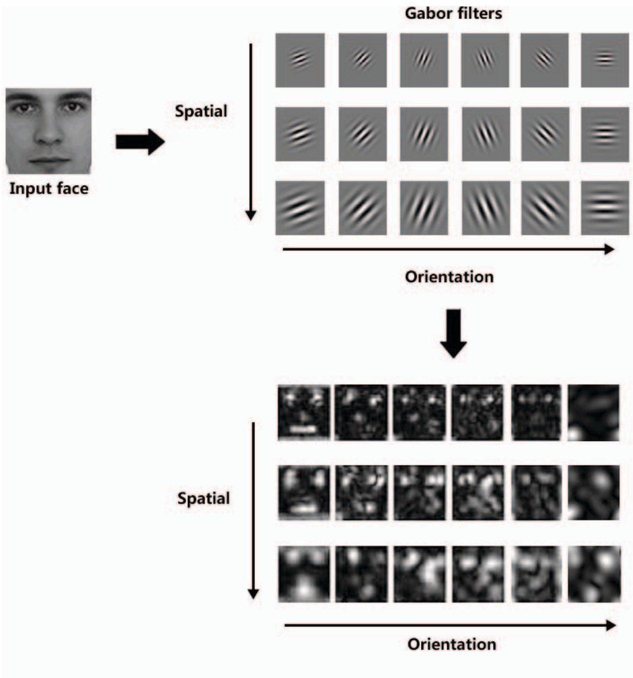

Fig. 2. Gabor filters using 3 scales and 6 orientations

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 kernels in (1) can be generated from the mother wavelet, by scaling and rotation via the wave vector $\vec{k}$. 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. 2.

The dimension of dense Gabor feature depends on the size of the block, so it will be quite high if we want to extract features in a large region. We utilize a sub-sampling strategy, which applies a 2×2 to 6×6 sub-sampling based on the block size. Using this strategy, the Gabor features will be extracted only on the sub-sampled pixels. As a result, the minimum dimension of Gabor is $3 \times 6 \times 9 = 162$ (6×6 block with 2×2 sampling), and the maximum is $3 \times 6 \times 30 = 540$ (32×40 block with 6×6 sampling).

III. LEARNING THE FEATURES USING REALADABOOST WITH COMPLEXITY PENALTY TERMS

In RealAdaBoost [26], 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 (5).

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

Features with smaller Z will be selected, which leads to better classification of positive samples and negative samples.

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 criterion into RealAdaBoost, which is similar to selecting the image strip features [16]. The discriminative criterion of RealAdaBoost is 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.

The equation (6) could be explained as follows, in the first stages of RealAdaBoost, the false positive rate is relatively high, and the gender of faces are still easy to be classified, so that efficient features are preferred. In the following stages, because of the lower false positive rate, the patterns of the training samples will be more complicated. Then the features with high computation cost are considered.

To decide the computation cost C, we test the execution time of different feature extraction methods. We set the C of SIFT to 10, HOG to 2, and Gabor to 2-5 according to its dimension. The complexity-aware factor a is set to 0.25. The diagram of the whole complexity-aware RealAdaBoost is illustrated in Fig. 3.

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 based on minimizing 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. 3. Learning the features using RealAdaBoost with complexity penalty term

IV. Experiments

A. Experiemtn on FERET dataset

The FERET database [4] contains gray scale images of 1,199 individuals with uniform illumination but different poses. Examples are shown in the first row of Fig. 4. Similar to Makinen and Raisamo's work [3], faces of one image per person the Fa subset were used and duplications were eliminated. Therefore, 199 female and 212 male images were used from the FERET database.

In our experiments, we adopted a 5-fold cross validation testing scheme, where the images are divided into five folds, keeping the same ratio between male and female faces. In the training procedure, all the faces are resized to 64×80 and aligned based on eye position.

We train 7 classifiers using RealAdaBoost, which includes the classifiers utilizing single feature (SIFT, HOG, and Gabor), the classifiers using the combination of two features, and the method proposed in this paper. After feature selection, linear SVM is utilized to train a classifier for accuracy evaluation. The items with (*) denotes that the complexity-aware RealAdaBoost is adopted. There is no complexity penalty term if single feature is used. Experimental results are shown in Table. 1. It can be seen that using the complexity-aware strategy, the average recognition rate is improved compared to using SIFT, HOG, or Gabor independently. Using all three features, the accuracy is also comparable with the state-of-the-art algorithms [18][19][27].

Fig. 4. Examples of databases used in our experiments. The first row- FERET. The second row-KinFace. The third row-LFW.

TABLE I. GENDER RECOGNITION ON FERET DATABASE

Approach	Recognition rate
SIFT	94.89%
HOG	94.64%
Gabor	94.89%
SIFT + HOG (*)	95.86%
SIFT + Gabor (*)	96.35%
HOG + Gabor (*)	95.62%
Y. El-Din [19]	97.11%
Tapia [18]	99.13%
L. Alexandre [27]	99.07%
All three features (*)	98.78%

B. Experiments on KinFace dataset

The UB KinFace dataset [14] offers a collection of faces captured from the web, representing a variation of expressions and lighting conditions. It contains 600 images of 3 groups. Each group is composed of child, adult, and senior images. Examples are shown in the second row of Fig. 4.

Similar to the experiments in FERET database, a 5-fold

TABLE II. GENDER RECOGNITION ON KINFACE DATABASE

Approach	Recognition rate
SIFT	92.50%
HOG	91.83%
Gabor	91.67%
SIFT + HOG (*)	95.17%
SIFT + Gabor (*)	95.50%
HOG + Gabor (*)	94.67%
Y. El-Din [19]	94.45%
All three features (*)	96.50%

TABLE III. GENDER RECOGNITION ON LFW DATABASE

Approach	Recognition rate
SIFT	94.62%
HOG	93.23%
Gabor	94.94%
SIFT + HOG (*)	96.10%
SIFT + Gabor (*)	96.97%
HOG + Gabor (*)	95.69%
Shan [6]	94.81%
Tapia [18]	98.01%
All three features (*)	98.01%

validation is applied on each group. Duplicate images of the same person are placed in the same fold. The experimental results of the average accuracy are listed in Table. 2. Using all features, the accuracy is 4% better compared to the methods using single feature. It also outperforms the state-of-the-art algorithm [19] in this database.

C. Experiments on LFW

We conduct experiments on the LFW database [17]. LFW is a database for studying the problem of unconstrained face recognition, which contains 13,233 color face photographs of 5,749 subjects collected from the web. Some examples are shown in the third row of Fig. 4. We manually labeled the groundtruth regarding gender for each face. The faces that are not (near) frontal, with rotation larger than 45 degree, small scale, and strong occlusion were not considered. In our experiments, we chose 4,500 males and 2,340 females, which is similar to [6]. All experimental results were obtained using 5-fold cross-validation. Similar to KinFace evaluation, duplicate images of the same person are placed in the same fold.

In Table 3, we compare our method with the methods using single feature, the combination of two features, and the state-of-the-art algorithms. It is observed that our method achieves comparable accuracy with the state-of-the-art algorithms [6][18]. The overall feature dimension of our classifier is around 5,000, which is much smaller than [6][18].

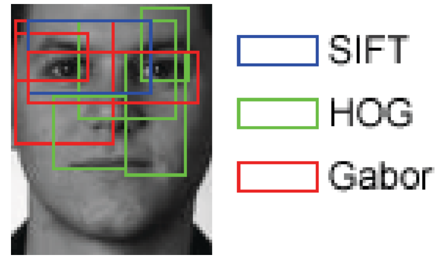

Fig. 5. The first 8 features selected by RealAdaBoost in human face

TABLE IV. EXECUTION SPEED OF THE GENDER CLASSIFIERS

Approach	Recognition time(ms)
SIFT	30.54
HOG	11.22
Gabor	18.83
SIFT + HOG (*)	19.77
SIFT + Gabor (*)	17.90
HOG + Gabor (*)	15.19
All three features	24.45
All three features (*)	11.89

D. Analysis

We draw the first 8 features selected by the RealAdaBoost algorithm in LFW database, as shown in Fig. 5. There are one SIFT feature, 4 HOG features, and 3 Gabor features. Only one SIFT feature is selected because of its heavy computation cost. In addition, it could be seen that most of the features lays on the upper part of the face. This circumstance is reasonable, because it is much easier to recognize the gender by eye, eyebrow and nose rather than using mouth, which is easily influenced by expression variation.

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 4. We find that SIFT is relatively slow compared to HOG and Gabor feature. 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 24.45ms per face to 11.89ms 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 gender recognition.

V. CONCLUSION

In this paper, we proposed a local feature-based representation for face gender recognition. We used a RealAdaBoost algorithm with the complexity penalty term to select the meaningful features, which successfully balances the accuracy and efficiency. High gender recognition rates were reported in comparison to previously published results on three famous datasets, where 98.8% was achieved for FERET, 96.5% achieved for KinFace, and 98.0% achieved for LFW.

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

VI. ACKNOWLEDGMENT

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

REFERENCES

[1] C. Benabdelkader and P. Griffin, A Local Region-based Approach to Gender Classification From Face Images. In CVPR Workshop, 2005.

[2] H. Kim, D. Kim, and Z. Ghahramani, Appearance-based Gender Classification with Gaussian Processes. In PRL Vol. 27, Page(s):618- 626, 2006.

[3] E. Makinen and R. Raisamo, Evaluation of Gender Classification Methods with Automatically Detected and Aligned Faces. In PAMI, Vol. 30, Page(s):541-547, 2008.

[4] P. Phillips, H. Moon, P. Rauss, and S. Rizvi, The FERET Evaluation Methodology for Face-Recognition Algorithms. In CVPR, 1997.

[5] T. Ojala and M. Pietikainen, Multiresolution Gray-scale and Rotation Invariant Texture Classification with Local Binary Patterns. In PAMI, Vol. 24, Page(s):971-987, 2002.

[6] C. Shan, Learning Local Binary Patterns for Gender Classification on Real-world Face Images. In PRL, Vol. 33, Page(s):431-437, 2013.

[7] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao, WLD: A Robust Local Image Descriptor. In PAMI, Vol. 32, Page(s):1705-1720, 2010.

[8] N. Dalal and B. Triggs, Histograms of Oriented gradients for human detection. In CVPR, 2005.

[9] A. Shobeirinejad and Y. Gao, Gender Classification Using Interlaced Derivative Patterns. In ICPR, 2010.

[10] D. Lowe, Object recognition from local scale-invariant features. In ICCV, 1999.

[11] L. Cao, M. Dikmen, Y. Fu, and T. Huang, Gender Recognition from Body. In ACM Multimedia, 2011.

[12] X. Leng, and Y. Wang, Improving Generalization for Gender Classification. In ICIP, 2008.

[13] J. Wang et al., Dense SIFT and Gabor Descriptors-based Face Representation with Applications to Gender Recognition. In ICARCV, 2010.

[14] S. Xia, M. Shao, J. Luo, and Y. Fu, Understanding Kin Relationships in a Photo. In IEEE Transactions on Multimedia, 2012.

[15] L. Cao, M. Dikmen, Y. Fu and T. Huang, Gender Recognition from Body. In Proceeding of the 16th ACM international conference on Multimedia, 2008.

[16] W. Zheng and L. Liang, Fast Car Detection using Image Strip Features. In CVPR, 2009.

[17] G. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. In Technical report, 07?9, University of Massachusetts, Amherst, 2007.

[18] J. Tapia and C. Perez, Gender Classification Based on Fusion of Different Spatial Scale Features Selected by Mutual Information From Histogram of LBP, Intensity, and Shape. In IFS, Vol. 8, Page(s):488-499, 2013.

[19] Y. El-Din, M. Moustafa, and H. Mahdi, Landmarks-SIFT Face Representation for Gender Classification. In ICIAP, 2013.

[20] R. Rojas-Bello, L. Lago-Fernandez, G. Martinez-Munoz, and M. Sdnchez-Montanes, A Comparison of Techniques for Robust Gender Recognition. In ICIP, 2011.

[21] M. Demirkus, M. Toews, J. Clark, and T. Arbel, Gender Classification from Unconstrained Video Sequences. In CVPR Workshop, 2010.

[22] J. Wang, J. Li, W. Yau, and E. Sung, Boosting Dense SIFT Descriptors and Shape Contexts of Face Images for Gender Recognition. In CVPR Workshop, 2010.

[23] H. Lian, B. Lu, and E. Takikawa, Gender Recognition Using a Min-Max Modular Support Vector Machine. In Advances in Natural Computation, 2005.

[24] S. Hosoi, E. Takikawa, and M. Kawade, Ethnicity estimation with facial image. In Automatic Face and Gesture Recognition, 2004.

[25] F. Scalzo, G. Bebis, M. Nicolescu, L. Loss, and A. Tavakkoli, Feature Fusion Hierarchies for Gender Classification. In ICPR, 2008.

[26] R. Schapire and Y. Singer, Improved Boosting Algorithms Using Confidence-rated Predictions. In Machine Learning, Vol. 37, Page(s):297-336, 1999.

[27] L. Alexandre., Gender recognition: Amultiscale Decision Fusion Approach. In PRL, Vol. 31, Page(s):1422-1427, 2010.

[28] L. Bourdev, S. Maji, and J. Malik, Describing People: A Poselet-Based Approach to Attribute Classification. In ICCV, 2011.

[29] M. Collins, J. Zhang, and P. Miller, Full Body Image Feature Representations for Gender Profiling. In ICCV Workshop, 2009.

[30] E. Brunelli, T. Poggio, Face Recognition: Features Versus Templates. In PAMI, Vol. 15, Page(s):1042-1052, 1993.

[31] J. Fellous, Gender Discrimination and Prediction on the Basis of Facial Metric Information. In Vision Research, Vol. 37, Page(s):1961-1973, 1997.