Adaptive Part-Level Model Knowledge Transfer for Gender Classification

Yongiang Gao, Zhifeng Li, *Senior Member, IEEE*, and Yu Qiao, *Senior Member, IEEE*

***Abstract*—In this letter, we propose an adaptive part-level model knowledge transfer approach for gender classification of facial images based on Fisher vector (FV). Specifically, we first decompose the whole face image into several parts and compute the dense FVs on each face part. An adaptive transfer learning model is then proposed to reduce the discrepancies between the training data and the testing data for enhancing classification performance. Compared to the existing gender classification methods, the proposed approach is more adaptive to the testing data, which is quite beneficial to the performance improvement. Extensive experiments on several public domain face data sets clearly demonstrate the effectiveness of the proposed approach.**

***Index Terms*—Domain adaption, Fisher vector (FV) faces, gender classification, least-square SVM, transfer learning.**

I. Introduction

GENDER classification [1], [2], also known as gender recognition or sex classification [3], is to tell a person's gender based on an input face image. It has many useful applications in real life, such as human–computer interaction, surveillance, content-based indexing and searching, biometrics, demographic studies, and targeted advertising [1], [4]. However, despite the advances in gender classification, it still remains a challenging problem. Especially, in many real-world application scenarios, it is very common that the training data are insufficient to learn a robust model. So, the performance of the learning-based methods on the limited training data would be limited. In this letter, we aim to improve the performance based on the adaptive transfer learning model for the insufficient training data.

A number of approaches have been proposed to address the representation and classification problems in gender classification of face images. Baluja and Rowley [3] presented a method based on AdaBoost to identify the gender of a person from a low resolution gray-scale face image. Rahman *et al.* [5] proposed an automatic facial feature extraction system for sex identification from color images. In [6], the authors used discrete cosine transform (DCT) to extract efficient features followed by a k-nearest neighbor classifier (KNN) for gender classification. Levi and Hassner [7] showed that by learning representations through convolutional neural networks (CNNs), a significant increased performance can be obtained. Makinen and Raisamo [8] carried out an experimental evaluation on gender classification, which contains four gender classification methods and four automatic alignment methods. In [9], they reported a method for gender classification based on mutual information and fusion of features extracted from intensity, shape, texture, and from three different spatial scales.

Recently, dense feature extraction is becoming increasingly popular in face image recognition and analysis, such as face recognition in the wild [10], [11], heterogeneous face recognition [12]–[14], and aging face recognition [15], [16]. The basic idea of dense feature extraction is to densely compute the local descriptors (such as [17]–[19], [24], and [41]), and then encode the dense descriptors into an appropriate feature vector. A popular encoding method in dense feature extraction is the bag-of-visual-words (BoVW) model [20] based on spatial–temporal local features. More recently, it has been shown that Fisher vector (FV) encoding method [21] is more effective in feature encoding and obtains superior performance over the other encoding ways [22]. Inspired by the good performance of the FV encoding method, we will use it as the baseline feature extraction and encoding method in this letter.

This letter deals with the situation that we have sufficient training examples in the source domain, but sparse examples in the testing target domain. Most of the existing learning-based methods rely heavily on a common assumption: the training data and testing data share a highly similar feature distribution. Otherwise, the performance of these methods would degrade notably. To alleviate this problem, we propose a new transfer learning method to improve the performance of gender classification. To this end, we first divide the whole face into several parts that are complementary to each other (the left of Fig. 2), and extract the FV features based on these parts for subsequent analysis. Then, a part-level transfer learning method is

Manuscript received November 13, 2015; revised April 05, 2016; accepted April 11, 2016. Date of publication April 20, 2016; date of current version May 13, 2016. This work was supported by the National Natural Science Foundation of China under Grant 61472410 and Grant 61103164, the Guangdong Innovative Research Program under Grant 2015B010129013 and Grant 2014B050505017, the Shenzhen Research Program under Grant KQCX2015033117354153, Grant JSGG20150925164740726, and Grant CXZZ20150930104115529, the Natural Science Foundation of Guangdong Province under Grant 2014A030313688, and the Key Laboratory of Human-Machine Intelligence-Synergy Systems through the Chinese Academy of Sciences. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Sanghoon Lee. *(Correpsonding author: Yu Qiao.)*

Y. Gao is with the Shenzhen Key Laboratory of Computer Vision and Pattern Recognition, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China, and also with the Shenzhen College of Advanced Technology, University of Chinese Academy of Sciences, Beijing 518055, China (e-mail: yq.gao@siat.ac.cn).

Z. Li and Y. Qiao are with the Shenzhen Key Laboratory of Computer Vision and Pattern Recognition, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 100864, China, and also with the Chinese University of Hong Kong, Hong Kong (e-mail: zhifeng.li@siat.ac.cn; yu.qiao@siat.ac.cn).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2016.2555480

developed to reduce the distribution gap between the training data and the testing data, combined with an adaptive SVM classifier for final decision. In this way, the proposed classification model can be more adapted to the testing data, which is quite beneficial to the performance improvement. Extensive experiments are conducted on several public domain data sets to show the effectiveness of the proposed model over the state of the art.

II. Proposed Approach

In this section, we first briefly review some related works including FV and LS-SVM, and then elaborate our proposed adaptive part-level model transfer learning approach.

The formal notation and necessary mathematical tools are introduced as follows. We use small and capital bold letters to denote the column vectors and matrices, respectively, e.g., $\mathbf{a} = [a_1, a_2, \ldots, a_N]^T \in R^N$ and $\mathbf{A} \in R^{M \times N}$ with $A_{i,j}$ corresponding to the (i, j) element. When only one subscripted index is present, it represents the column index, e.g., $\mathbf{A}_i$ is the ith column of the matrix $\mathbf{A}$.

A. Related Works

1) Fisher Vector: We basically follow the framework [23] to extract the dense features and then perform FV encoding, as described in the following.

First, we warp and crop the original face image into 116×80 facial image through five facial landmarks including left eye center, right eye center, nose tip, left mouth corner, and right mouth corner. We divide each facial image into six parts, corresponding to forehead area, left eye area, right eye area, mouth area, jaw area, and the whole facial area, see the left of Fig. 2. The SIFT [24], [25] features are extracted densely in scale for each part. Same as FV faces, 24×24 pixels patches are sampled with a stride of one pixel and this process is repeated at five scales with a scaling factors of $\sqrt{2}$ for each part.

Second, we encode the large pool of dense SIFT features into FV features. In general process, FV encoding starts by fitting a parametric generative model to the features, where the Gaussian mixture model (GMM) is the most commonly used, and then encoding the derivatives of the log-likelihood of the model with respect to its parameters. Also, the derivatives with respect to the Gaussian mean and variances are considered in our approach due to the gradient with respect to the weight parameters brings little additional information. We adopt L_2 normalization for each FV feature.

2) Least-Square Adaptation Method: Given a binary classification problem and a set of N samples $D = \{\boldsymbol{x}_i, y_i\}_{i=1}^N$, where $\boldsymbol{x}_i$ denotes the input feature vector of the ith sample with the corresponding label $y_i \in \{-1, 1\}$. The least-square SVM [26] is introduced by formulating the classification problem as

$$\begin{aligned} &\min_{\mathbf{w},b} \tfrac{1}{2}||\mathbf{w}||^2 + \tfrac{C}{2}\textstyle\sum_{i=1}^N \xi_i^2 \\ &\text{s.t.} \quad y_i = \mathbf{w}^T \cdot \boldsymbol{x}_i + \xi_i \quad \forall i \in 1, \ldots, N. \end{aligned} \tag{1}$$

In this equation, we set $\mathbf{w} = [\mathbf{w}, b]^T$ and $\boldsymbol{x}_i = [\boldsymbol{x}_i, 1]^T$ conveniently, b is the offset parameter, $\mathrm{Ker}(x, x') = x^T x'$, C is a parameter trading off between the empirical loss and regularization, and the slack variable ξ_i is used to measure the degree of the misclassification on the data $\boldsymbol{x}_i$.

Let us suppose that there are source data sets $D^S = \{\boldsymbol{x}_i^S, y_i^S\}_{i=1}^{N^S}$ obeying a distribution P^S, different with respect to the target data set P corresponding the data set $D = \{\boldsymbol{x}_i, y_i\}_{i=1}^N$. Note that $N \ll N^S$ is a small number. If the two distributions P, P^S are somehow related, the auxiliary knowledge can be transferred in guiding the learning process. After getting the optimal source knowledge/weights $\mathbf{w}$ by minimizing (1) which is shown as $\mathbf{w}^S$[1] in (2), we hope the target $\mathbf{w}$ chosen to be the optimal $\mathbf{w}^S$ with regularization term. The final model knowledge transfer of the least-square adaptation method [27] is

$$\begin{aligned} &\min_{\mathbf{w},b} \tfrac{1}{2}||\mathbf{w} - \mathbf{w}^S||^2 + \tfrac{C}{2}\textstyle\sum_{i=1}^N \zeta_i \xi_i^2 \\ &\text{s.t.} \quad y_i = \mathbf{w}^T \cdot \boldsymbol{x}_i + \xi_i \quad \forall i \in \{1, \ldots, N\}. \end{aligned} \tag{2}$$

B. Adaptive Part-Level Transfer Learning Model

In this section, we describe the proposed adaptive part model knowledge transfer approach. This model consists of K parts. In our task, the FV is adopted for dense feature extraction, and the linear kernel is adopted as $\mathrm{Ker}(x, x') = x^T x'$. Inspired by [27], Parameter $\mathbf{Z} = \mathrm{diag}\{\zeta_1^{-1}, \zeta_2^{-1}, \ldots, \zeta_N^{-1}\}$ is used to balance the different contributions of the different labeled samples, denoted by

$$\zeta_i = \begin{cases} \frac{N}{2N^+} & \text{if } y_i = +1 \\ \frac{N}{2N^-} & \text{if } y_i = -1 \end{cases} \tag{3}$$

which N^+ and N^- denote the number of positive and negative samples, respectively. Since there are multiple parts, we rewrite our adaptive transfer learning model (2) as

$$\min_{\mathbf{w},b} \frac{1}{2}\sum_{k=1}^K ||\mathbf{w}_k||^2 + \frac{1}{2}\sum_{k=1}^K ||a_k \mathbf{w}_k - a_k \mathbf{w}_k^S||^2 + \frac{C}{2}\sum_{i=1}^N \zeta_i \xi_i^2 \tag{4}$$

$$\text{s.t.} \quad a_k > 0, \quad \sum_{k=1}^K a_k = 1 \tag{5}$$

$$y_i = \sum_{k=1}^K \mathbf{w}_k^T \cdot \boldsymbol{x}_{i,k} + \xi_i \quad \forall i \in \{1, \ldots, N\}. \tag{6}$$

In (5), a_k denotes the weight for each part, and the sum of a_k ($k = \{1, 2, \ldots, K\}$) is set to 1 and the uppercase K denotes the numbers of parts.

Note that there are two unknown variables $\mathbf{a}(\mathbf{a} = [a_1, a_2, \ldots, a_K]^T)$ and $\mathbf{w}$, we first set $a_k = \frac{1}{K}$ for $k = \{1, 2, \ldots, K\}$ and the adaptive part-level least-square SVM model would become the regular model (2). The corresponding Lagrangian $\mathcal{L}_{\mathbf{w}}$ is

[1]The parameters with superscript S denote the source data sets and it is the same in the following letter.

$$\mathcal{L}_{\mathbf{w}} = \frac{1}{2}\sum_{k=1}^{K}||\mathbf{w}_k||^2 + \frac{1}{2}\sum_{k=1}^{K}||a_k\mathbf{w}_k - a_k\mathbf{w}_k^S||^2 + \frac{C}{2}\sum_{i=1}^{N}\zeta_i\xi_i^2 + \sum_{i=1}^{N}\lambda_i\left(y_i - \sum_{k=1}^{K}\mathbf{w}_k^T\boldsymbol{x}_{i,k} - \xi_i\right) \quad (7)$$

where $\boldsymbol{\lambda} \in R^N$ is the vector of Lagrange multipliers. The optimality (Karush–Kuhn–Tucker) condition with respect to w, ξ_i, and λ_i are

$$\frac{\partial\mathcal{L}_{\mathbf{w}}}{\partial\mathbf{w}_k} = 0 \Rightarrow \mathbf{w}_k = \frac{a_k^2}{1+a_k^2}\mathbf{w}_k^S + \frac{1}{1+a_k^2}\sum_{i=1}^{N}\lambda_i\boldsymbol{x}_{i,k} \quad (8)$$

$$\frac{\partial\mathcal{L}_{\mathbf{w}}}{\partial\xi_i} = 0 \Rightarrow C\zeta_i\xi_i = \lambda_i \quad (9)$$

$$\frac{\partial\mathcal{L}_{\mathbf{w}}}{\partial\lambda_i} = 0 \Rightarrow y_i = \sum_{k=1}^{K}\mathbf{w}_k^T\boldsymbol{x}_{i,k} + \xi_i. \quad (10)$$

By combining (8)–(10), we obtain

$$\mathbf{y}_i - \hat{\mathbf{y}}_i^S = \sum_{k=1}^{K}\frac{1}{1+a_k^2}\sum_{j=1}^{N}\lambda_j\boldsymbol{x}_{j,k}\boldsymbol{x}_{i,k} + \frac{\lambda_i}{C\zeta_i} \quad (11)$$

where $\hat{\mathbf{y}}_i^S = \sum_{k=1}^{K}\frac{a_k^2}{1+a_k^2}\mathbf{w}_k^{S^T}\boldsymbol{x}_{i,k}$ is the predicted label by the source model. By denoting the kernel matrix $\mathbf{Ker}$ with $\mathbf{K_{er}}_{ji,k} = \frac{1}{1+a_k^2}\sum_{j=1}^{N}\boldsymbol{x}_{j,k}^T\cdot\boldsymbol{x}_{i,k}$, (11) can be written in a matrix form, denoted by

$$\left(\mathbf{K_{er}} + \frac{\mathbf{Z}}{C}\right)\boldsymbol{\lambda} = \mathbf{y} - \hat{\mathbf{y}}^S. \quad (12)$$

Finally, the parameter $\boldsymbol{\lambda}$ and $\mathbf{w}$ are obtained as

$$\boldsymbol{\lambda} = \left(\mathbf{K_{er}} + \frac{\mathbf{Z}}{C}\right)^{-1}(\mathbf{y} - \hat{\mathbf{y}}^S) \quad (13)$$

$$\mathbf{w}_k = \frac{a_k^2}{1+a_k^2}\mathbf{w}_k^S + \frac{1}{1+a_k^2}\boldsymbol{\lambda}^T\boldsymbol{x}_k. \quad (14)$$

After getting $\mathbf{w}_k$, there is one unknown parameter $\mathbf{a}$. The corresponding Lagrange equation $\mathcal{L}_{\mathbf{a}}$ is

$$\mathcal{L}_{\mathbf{a}} = \frac{1}{2}\sum_{k=1}^{K}||a_k\mathbf{w}_k - a_k\mathbf{w}_k^S||^2 + \mu\left(\sum_{k=1}^{K}a_k - 1\right) + \text{const} \quad (15)$$

where μ is the Lagrange multipliers and const is the constant terms. The optimality condition with respect to a_k, μ are

$$\frac{\partial\mathcal{L}_{\mathbf{a}}}{\partial a_k} = 0 \Rightarrow a_k = -\mu Q_k \quad (16)$$

$$\frac{\partial\mathcal{L}_{\mathbf{a}}}{\partial\mu} = 0 \Rightarrow \sum_{k=1}^{K}a_k = 1 \quad (17)$$

where $Q_k = [(\mathbf{w}_k - \mathbf{w}_k^S)^T((\mathbf{w}_k - \mathbf{w}_k^S)]^{-1}$. By combining (16) and (17), we obtain

$$\mu = -\left(\sum_{k=1}^{K}Q_k\right)^{-1}. \quad (18)$$

Algorithm 1. Adaptive Part Level Model Knowledge Transfer

Input:
The set of source data, $D^S = \{X^S, Y^S\}$;
The set of small set of target data, $D = \{X, Y\}$;
The iterations, $iter$.
Output: The predicted label, Y.
Initialization : $a_k = \frac{1}{K}, k \in \{1, 2, \ldots, K\}$;
begin
  while $iter > 0$ **do**
    Compute $\mathbf{w}_k$ by Eq. 14;
    Compute a_k by Eq. 16;
    set $a_k = a_k$;
    set $iter = iter - 1$;
  Obtained $Y = \sum_{k=1}^{K}\mathbf{w}_k^T\cdot\boldsymbol{x}_k$.

Then, a_k can be obtained by (16)–(18).

In Algorithm 1, we summarize the adaptive part-level model knowledge transfer learning approach. Note that there are two unknown parameters $\mathbf{w}$ and $\mathbf{a}$, so we use an iterative learning strategy. In our experiment, we set $\#iter = 5$.

III. Experimental Results

In this section, we conduct extensive experiments to demonstrate the effectiveness of our adaptive part-level transfer learning model on two scenarios: 1) the transfer between unconstrained facial images and frontal facial images and 2) the transfer between optical facial images and infrared facial images.

The experimental details are as follows. In preprocessing, the facial images are warped and cropped by five landmarks and the resolution is 116×80. The cropped facial images are then divided into six parts and the FV features are obtained for each part. In the experiments, we fixed the hypeparameter C to 0.05 empirically.

A. Two Transfer Cases

Case 1: We evaluate the proposed model based on the transfer between unconstrained facial images and frontal facial images. The "LFW" data set [28] and the color "FERET" data set [29] are used in this case. The "LFW" contains more than 13 000 images of 5749 subjects collected from the web. These images satisfy "natural" distribution of faces in unconstrained environment, e.g., some of the facial images are under extreme lighting conditions, some of the range and diversity of pictures are very large, and so on. We choose the first picture for each subject as the image in our experiment. There are 5102 subjects with 3790 males and 1312 females. The color "FERET" data set consists of 2409 subjects with 1495 males and 914 females, and all the face images are very clean (noise free, fairly consistent lighting, no background clutter, etc.). For the two data sets, we randomly select the 80% images for training and the 20% images for testing.

Case 2: We then evaluate the proposed model based on the transfer between optical facial images and infrared facial

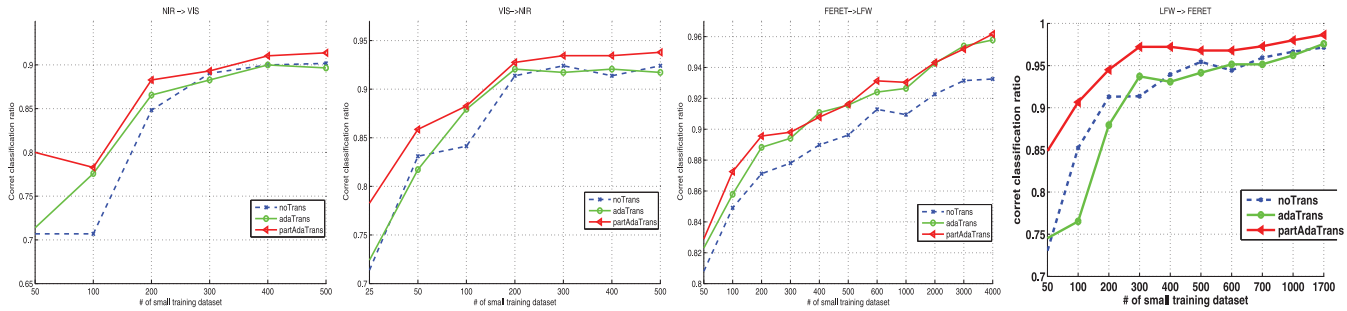

Fig. 1. Classification ration of comparison results along various small training data sets on two transfer types: the transfer between unconstrained facial images (LFW) and frontal facial images (FERET), and the transfer between optical facial images (VIS), and infrared facial images (NIR). We show the three comparison algorithms: classification without any transfer algorithm (NoTrans), classification with adaptive transfer learning (AdaTrans), and part-level adaptive transfer learning (PartAdaTrans).

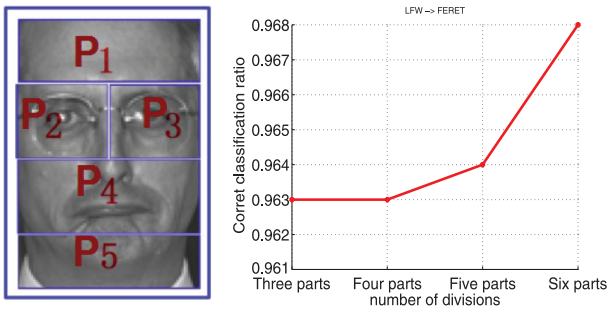

Fig. 2. Left: sample parts of one facial image. It is divided into five meaningful parts. Right: correct classification ration of comparison results along various divisions transferred from the "LFW" to the "FERET."

images. The CASIA NIR-VIS 2.0 face database [30] is used in this case. There are 728 subjects (407 males and 321 females), with each person having one optical (VIS) image and one corresponding infrared (NIR) image as a result of another two data sets in our experiment. This data set was collected in four recording sessions that are from 2007 to 2010, and the age distribution of the subjects is very broad (ranging from children to old people). For each data set, we randomly select the 500 subjects for training and the other 228 subjects for testing.

B. Evaluation of the Proposed Model

We set the source training data set and the target training data set as the whole training data, and the target testing data as the testing data. Fig. 1 shows the comparison results for these experiments, where "NoTrans" refers to the classification with only target data (1), "AdaTrans" refers to the classification with LS-SVM adaptation model (2), and "PartAdaTrans" is our proposed part-level LS-SVM adaptation model (4). From these results we have the following three observations.

1) The classification rates can be improved if the size of training data increases.
2) The performance of the adaptive models outperforms those with only target data used.
3) The "partAdaTrans" model can achieve the best result. This shows the advantage of the proposed model.

To investigate how the different numbers of divisions affects the performance of our approach, we design different types of division combinations. The "Three parts" type is ($P_1 + P_2 + P_3$, $P_4 + P_5$, the whole image), the "Four parts" type is ($P_1 + P_2 + P_3$, P_4, P_5, the whole image), the "Five parts" type is (P_1, $P_2 + P_3$, P_4, P_5, the whole image), and the "Six parts" type is (P_1, P_2, P_3, P_4, P_5, the whole image). Note that "P_k" is shown in the left of Fig. 2, and the comparison results are shown in the right of Fig. 2. We can see that "Six parts" type achieves the best performance. We choose the "Six parts" type in the following experiments.

TABLE I
COMPARISON RESULTS WITH THE STATE-OF-THE-ART METHODS

	LFW (%)		FERET (%)
Kumar *et al.* [31]	85.8	Baluja and Rawley [3]	94.4
Liu *et al.* [32]	94	Peret *et al.* [9]	99.1
Deep SPN [33]	92	Mirza *et al.* [37]	98.2
Liu *et al.* [36]	95.8	Li *et al.* [38]	95.8
Cao *et al.* [40]	96.2	Leng and Wang [39]	98.8
Danisman *et al.* [35]	91.9	Ren and Li [34]	98.8
FV (NoTrans)	94.8	FV (NoTrans)	98.4
AdaTrans	96.3	AdaTrans	98.6
PartAdaTrans	**96.8**	PartAdaTrans	**99.3**

Noting that the experimental constructions of "LFW" are presented different, we set 80% data for training and others for testing and five iterations are run and mean accuracy rates are reported.

C. Compared to the State of the Art

Finally, we compare the proposed model against the state-of-the-art methods on both the "LFW" and "FERET" data sets. Table I reports the comparative results. It is very encouraging to see that our proposed approach (partAdaTrans) consistently outperforms the existing ones on the two data sets by a clear margin. This confirms the effectiveness of the proposed new approach.

IV. CONCLUSION

In this letter, we have proposed a novel part-level model knowledge transfer approach for gender classification of facial images. Unlike the existing methods in gender classification, our model can reduce the distribution gap between the training data and the testing data by the model knowledge transfer learning. Extensive experiments have been conducted to demonstrate the effectiveness of our new approach.

REFERENCES

[1] C. Ng, Y. Tay, and B. Goi, "Vision-based human gender recognition: A survey," *CoRR*, vol. abs/1204.1611, 2012.

[2] P. Rai and P. Khanna, "Gender classification techniques: A review," *Adv. Comput. Sci. Eng Applicat.*, vol. 166, pp. 51–59, 2012.

[3] S. Baluja and H. Rowley, "Boosting sex identification performance," *Int. J. Comput. Vis.*, vol. 71, no. 1, pp. 111–119, 2007.

[4] C. Ng, Y. Tay, and B.-M. Goi, "Recognizing human gender in computer vision: A survey," Lecture Notes AI, vol. 7458, pp. 335–346, 2012.

[5] M. Rahman, T. Das, and M. Sarnaker, "Face detection and sex identification from color images using adaboost with SVM based component classifier," *Int. J. Comput. Appl.*, vol. 76, no. 3, pp. 1–6, 2013.

[6] M. Nazir, M. Ishtiaq, A. Batool, M. A. Jaffar, and A. M. Mirza, "Feature selection for efficient gender classification," in *Proc. Int. Conf. Neural Netw. Int. Conf. Evol. Comput. Fuzzy Syst.*, 2010, pp. 70–75.

[7] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Rec. Workshop*, 2015, pp. 34–42.

[8] E. Makinen and R. Raisamo, "Evaluation of gender classification methods with automatically detected and aligned faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 541–547, Mar. 2008.

[9] C. Perez, J. Tapia, P. Estévez, and C. Held, "Gender classification from face images using mutual information and feature fusion," *Int. J. Optomechat.*, vol. 6, no. 1, pp. 92–119, 2012.

[10] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recog.*, 2013, pp. 3025–3032.

[11] Z. Li, D. H. Gong, X. Li, and D. Tao, "Learning compact feature descriptor and adaptive matching framework for face recognition," *IEEE Trans. Image Process.*, vol. 24, no. 9, pp. 2736–2745, Sep. 2015.

[12] B. Klare, Z. Li, and A. K. Jain, "Matching forensic sketches to mugshot photos," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 639–646, Mar. 2011.

[13] Z. Li, D. Gong, Y. Qiao, and D. Tao, "Common feature discriminant analysis for matching infrared face images to optical face images," *IEEE Trans. Image Process.*, vol. 23, no. 6, pp. 2436–2445, Jun. 2014.

[14] Z. Li, D. Gong, Q. Li, D. Tao, and X. Li, "Mutual component analysis for heterogeneous face recognition," *ACM Trans. Intell. Syst. Tech.*, 2016.

[15] Z. Li, U. Park, and A. K. Jain, "A discriminative model for age invariant face recognition," *IEEE Trans. Informat. Forensics Secur.*, vol. 6, no. 3, pp. 1028–1037, Sep. 2011.

[16] Z. Li, D. Gong, X. Li, and D. Tao, "Aging face recognition: A hierarchical learning model based on local patterns selection," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2146–2154, May 2016.

[17] Y. Qiao, W. Wang, N. Minematsu, J. Liu, X. Tang, and M. Takeda, "A theory of phase singularities for image representation and its applications to object tracking and image matching," *IEEE Trans. Image Process.*, vol. 18, no. 10, pp. 2153–2166, Oct. 2009.

[18] X. Qi, R. Xiao, G. Li, Y. Qiao, J. Guo, and X. Tang, "Pairwise rotation invariant co-occurrence local binary pattern," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 11, pp. 2199–2213, Nov. 2014.

[19] Y. Gao, W. Huang, and Y. Qiao, "Local multi-grouped binary descriptor with ring-based pooling configuration and optimization," *IEEE Trans Image Process.*, vol. 24, no. 12, pp. 4820–4833, Dec. 2015.

[20] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2003, pp. 1470–1478.

[21] F. Perronnin, Y. Liu, J. Sánchez, and H. Poirier, "Large-scale image retrieval with compressed fisher vectors," in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn.*, 2010, pp. 3384–3391.

[22] Y. Huang, Z. Wu, L. Wang, and T. Tan, "Feature coding in image classification: A comprehensive study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 493–506, Mar. 2014.

[23] K. Simonyan, O. Parkhi, A. Vedaldi, and A. Zisserman, "Fisher vector faces in the wild," in *Proc. Br. Mach. Vis. Conf.*, 2013, pp. 1–12.

[24] D. Lowe, "Distinctive image features from scale-invariant key-points," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[25] A. Vedaldi and B. Fulkerson. (2008). *VLFeat: An Open and Portable Library of Computer Vision Algorithms* [Online]. Available: http://www.vlfeat.org/

[26] J. A. K. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural Process. Lett.*, vol. 9, no. 3, pp. 293–300, 1999.

[27] T. Tommasi, F. Orabona, and B. Caputo, "Learning categories from few examples with multi model knowledge transfer," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 5, pp. 928–941, May 2014.

[28] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts, Amherst, MA, USA, Tech. Rep. 07–49, 2007.

[29] P. Phillips, H. Moon, P. Rauss, and S. Rizvi, "The FERET evaluation methodology for face recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.

[30] S. Li, D. Yi, Z. Lei, and S. Liao, "The casia nir-vis 2.0 face database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2013, pp. 348–353.

[31] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Describable visual attributes for face verification and image search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 10, pp. 1962–1977, Oct. 2011.

[32] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," *IEEE Int. Conf. Comput. Vision (ICCV)*, 2015, pp. 3730–3738.

[33] P. Luo, X. Wang, and X. Tang, "A deep sum-product architecture for robust facial attributes analysis," in *Proc. IEEE Conf. Comput. Vis.*, 2013, pp. 2864–2874.

[34] H. Ren and Z. Li, "Gender recognition using complexity-aware local features," in *Proc. Int. Conf. Pattern Recog.*, 2014, pp. 2389–2394.

[35] T. Danisman, I. M. Bilasco, and C. Djeraba, "Cross-database evaluation of normalized raw pixels for gender recognition under unconstrained settings," in *Proc. Int. Conf. Pattern Recog.*, 2014, pp. 3144–3149.

[36] H. Liu, Y. Gao, and C. Wang, "Gender identification in unconstrained scenarios using self-similarity of gradients features," in *Proc. Int. Conf. Pattern Recog.*, 2014, pp. 5911–5915.

[37] A. Mirza, M. Hussian, H. Almuzaini, G. Muhammad, H. Aboalsamh, and G. Bebis, "Gender recognition using fusion of local and global facial features," *Adv. Visual Comput.*, 2013, pp. 493–502.

[38] B. Li, X. C. Lian, and B. L. Lu, "Gender classification by combining clothing, hair and facial component classifiers," *Neurocomputing*, vol. 76, no. 1, pp. 18–27, 2012.

[39] X. Leng and Y. Wang, "Improving generalization for gender classification," in *Proc. Int. Conf. Pattern Recog.*, 2008, pp. 1656–1659.

[40] D. Cao, R. He, M. Zhang, Z. Sun, and T. Tan, "Real-world gender recognition using multi-order LBP and localized multi-boost learning," in *Proc. IEEE Int. Conf. Identity Secur. Behav. Anal.*, 2015, pp. 1–6.

[41] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.