Learning discriminative subregions and pattern orders for facial gender classification

Learning Discriminative Subregions and Pattern Orders for Facial Gender Classification

Zhong Chen, Andrea Edwards, Yongsheng Gao, Kun Zhang

PII: S0262-8856(19)30095-2
DOI: https://doi.org/10.1016/j.imavis.2019.06.012
Reference: IMAVIS 3784

To appear in: *Image and Vision Computing*

Received date: 7 February 2018
Revised date: 30 November 2018
Accepted date: 27 June 2019

Please cite this article as: Z. Chen, A. Edwards, Y. Gao, et al., Learning Discriminative Subregions and Pattern Orders for Facial Gender Classification, Image and Vision Computing, https://doi.org/10.1016/j.imavis.2019.06.012

Learning Discriminative Subregions and Pattern Orders for Facial Gender Classification

Zhong Chen[a], Andrea Edwards[a], Yongsheng Gao[b], Kun Zhang[a,*]

[a] *Department of Computer Science, Xavier University of Louisiana, LA 70125, USA*
[b] *School of Engineering, Griffith University, QLD 4111, Australia*

Abstract

Facial image-based gender classification has been widely used in many real-world applications. Most of the existing work, however, focuses on designing sophisticated or specific feature descriptors for the entire face, which neglects the discriminative information carried by facial components and pattern order combinations. To address the issue, in this paper, we first propose a generalized texture operator, i.e. the multi-spatial multi-order interlaced pattern (MMIP) matrix, to represent the gender information possessed by the facial subregions with textural pattern orders. A chain-type support vector machine (CSVM) based feature vector selection scheme, is then developed to highlight the gender characteristics. As a result, the discriminative subregions and pattern orders are constructed as the feature representation for facial gender classification. We evaluate our proposed method on four benchmark datasets (i.e. FRGC 2.0, FERET, LFW and UND) for gender classification and show its interpretability, effectiveness and efficiency compared with state-of-the-art methods.

Keywords: Discriminative subregions, Pattern order, Chain-type SVM, Feature vector selection, Facial gender classification

*Corresponding author. Tel.: (+1) 504-520-6700. Fax: (+1) 504-520-7906.
Email address: kzhang@xula.edu (Kun Zhang)

1. Introduction

The human face is an information rich source of traits (e.g. gender, age, emotion, and ethnicity), which play an important role in many real-world identification and verification applications such as customer-oriented advertising, dynamic marketing surveys, video surveillance, human-computer interaction [1, 2]. Among them, gender clue is the most fundamental and important demographic attribute of human beings, which in the vast majority of cases remains unchanged through a lifetime. In the last two decades, face-based gender recognition has received much research interest in the scientific community due to its popularity for the security purpose such as face unlock, facial expression identification, and facial landmark detection, localization and alignment. In general, facial gender classification schemes can be classified into two main categories: geometry-based methods and appearance-based methods [3]. The former focuses on extracting the shape structure of the face to give prominence to gender differences, and the representative methods include Active Appearance Models (AAM) [4], 3-D-Geometric Features (3DGF) [5], and Discrete Cosine Transform (DCT) [6]. The latter uses texture or statistics-oriented descriptors to characterize the facial gender information, which have gained increasing attention due to the robustness to illumination variation and high discrimination capability.

Numerous texture and statistics-oriented descriptors have been proposed in the literature to effectively describe facial features for gender classification. The representative methods include Local Binary Patterns (LBP) [7], Local Derivative Patterns (LDP) [8], Interlaced Derivative Patterns (IDP) [9], Discrete Wavelet Transform (DWT) [10], and Pyramid Local Binary Patterns (PLBP) [11]. Among them, LBP, LDP, and IDP are multiple pattern orders based descriptors, which have been validated more efficient in extracting discriminative features for gender classification. In fact, LBP is regarded as the first-order circular derivation pattern, while LDP and IDP are high-order local patterns that encode distinctive spatial relationships for facial images. PLBP actually extends the conventional LBP to pyramid transform domain by cascading the LBP in-

formation of hierarchical spatial pyramids [11], however, the suitable scale of PLBP for each specific facial component is unclear. Although both texture and statistical-oriented descriptors are widely used in various face recognition applications, they often use a certain pattern order from the entire or partial face to tackle gender recognition problem, where the best pattern order is often determined by the best performance from different empirical studies [9]. Then, how to automatically choose a suitable pattern order for specific facial components is still an unsolved problem. In the literature, these methods are applied to facial gender classification under a global or local fashion. The global region based feature descriptors usually use the entire face to extract discriminative features without any distinctions among different facial components, leading to inferior performance even for slight variations in illumination and pose. On the contrary, local region based feature descriptors have obtained impressive results by investigating discriminative subregions (i.e. hair, forehead, cheeks, ears, chin, etc.) for gender classification. For instance, Ueki et al. [12] integrate facial, hairstyle, and clothing images for gender classification. Their empirical studies show that the integration strategy significantly reduces the classification error rate compared to the conventional approaches. Similar results are reported in [13], where they compare facial features from internal zones (i.e. eyes, nose, and mouth) and external zones (i.e. hair, chin, and ears). Experiments on the Face Recognition Grand Challenge (FRGC) database show that the external face zones contribute more useful information for gender classification. Later on, Lian & Lu [14] integrate facial features from eyes, nose, mouth, and hair information for accurate gender classification. Li et al. [15] utilize both local features on five facial components (i.e. forehead, eyes, nose, mouth and chin) and external features (i.e. hair and clothing) to improve the classification performance. These studies show that the external zones (i.e. clothing, head, shoulder, hair, and ears) give sufficient information for gender recognition, while only using facial information is more challenging for gender classification problem.

To address the facial gender classification problem, a large number of innovative, yet controversial methods have been introduced in past years. For

example, Lu et al. [16] investigate significance of different facial regions for gender classification. Experimental results show that the upper region of the face proves to be the most significant part for gender classification. On the contrary, Hasnat et al. [17] verify that the lower region of the face is more important for gender classification as mouth and chin carry more important facial gender information. Specifically, the male face is hairy and rough while the female face is non-hairy and smooth. Additionally, some more discriminative analysis on facial components are presented with controversy. For example, Merkow et al. [18] apply the information obtained from the periocular region to identify gender and achieve 85% accuracy on 936 low resolution images collected from the web. When combining other facial components (i.e. nose, lip, and chin), the dominant regions are clearly biased toward the ocular region. Therefore, they claim that brow and eyes contain more valuable information for reliably gender classification than the other facial components. Meanwhile, Özbudak et al. [19] conduct an experimental study on examining the effects of facial components for gender classification. To investigate the most influential component, parts of these facial components (i.e. forehead, eyebrows, eyes, nose, lip, and chin) are masked. Experimental results on the masked samples indicate that nose is the most influential part for gender classification.

In summary, existing studies on facial gender classification address that both texture- or statistics-oriented descriptors and facial components are essential to decrease redundant information and improve the classification accuracy. However, since the datasets used in the above-mentioned techniques have different resolutions, quality and resources, the discriminative facial components for gender classification span in a wide range of spectrum. Therefore, how to choose discriminative facial components and suitable texture operator is still a challenging issue for facial gender classification. Motivated by the issue, in this paper, we attempt to solve the following two challenges: 1) how to develop a unified and effective texture operator to describe facial components and multiple pattern orders for facial image; and 2) how to automatically select discriminative facial components and pattern orders for gender classification. To the end, we first

Figure 1: The flowchart of the proposed gender classification method.

introduce a generalized texture operator, i.e. the multi-spatial multi-order interlaced pattern (MMIP) matrix, to enhance facial gender information, and then propose a chain-type SVM (CSVM) based feature vector selection algorithm to investigate discriminative facial components and pattern orders combination for gender classification. As a result, the proposed approach addresses these challenges in the following aspects: 1) the proposed MMIP matrix that considers both facial components and pattern orders brings benefit to describing spatial and high-order facial gender information; and 2) the proposed feature vector selection algorithm can figure out indispensable information for gender classification.

The flowchart of the proposed gender classification system is shown in Figure 1, which consists of four seamless steps in the training and test stages: 1) partition facial images into non-overlapping facial subregions; 2) extract multi-spatial and multi-order texture information of facial images; 3) use the CSVM-based feature vector selection algorithm to form the feature representation; and 4) incorporate the feature representation into the standard SVM classifier for gender classification.

The main contributions of this paper are summarized as follows.

- A unified and efficient framework for describing facial subregions and pattern orders of facial images is proposed.
- A CSVM-based feature vector selection algorithm for acquiring the most informative feature vectors in a lower feature space is developed.
- The proposed method investigates the discriminative facial subregions and pattern orders for facial gender classification.
- Experimental results on four different datasets (i.e. FRGC 2.0 [20], Facial Recognition Technology (FERET) [21], Labeled Faces in the Wild (LFW)

[22], and University of Notre Dame (UND) [23]) clearly demonstrate the efficacy of the proposed method compared to the state-of-the-art baselines.

The remainder of this paper is organized as follows. A brief review of existing studies for facial gender classification is presented in Section 2. A detailed description of the proposed gender classification approach is elaborated in Section 3. Experimental results and analysis are reported in Section 4. Finally, a short summary and an outlook on future work are summarized in Section 5.

2. Related Work

Generally, gender classification encompasses two main steps: feature extraction and classification. The classification step demands efficient feature representation as it heavily affects the performance of a classifier. The key issue in feature representation is a well-designed feature description, which should have low computational cost, high robustness, and perform well on unseen test samples. State-of-the-art methods are summarized in Table 1, which compares the features, classifiers and datasets of recent studies for face-based gender classification.

These methods try to grasp the best features to distinguish female and male groups. Table 1 summarizes some geometric-based, appearance-based, and some hybrid methods for gender classification. For example, Moghaddam & Yang [24] utilize raw image pixels of facial images as appearance-based features and achieve an accuracy of 96.6% on the FERET database. Makinen & Raisamo [1] systematically evaluate diverse face alignment and gender classification approaches on the FERET database. Baluja & Rowley [25] propose an efficient gender classification method by boosting pixel comparisons of human face. Yang et al. [26] use the texture raw pixels for feature extraction and adopt the AdaBoost classifier for gender classification on the FERET database. All these methods are not suitable to learn discriminative information between female and male groups as they only use pixel intensity feature of facial image, leading to being sensitive to pose, illumination and expression variations [3].

Table 1: Overview of recent studies on gender classification.

Method	Feature	Classifier	Dataset
Moghaddam & Yang [24]	Raw pixels	SVM	FERET
Makinen & Raisamo [1]	Raw pixels	SVM	FERET
Baluja & Rowly [25]	Pixel comparison	AdaBoost	FERET
Yang et al. [26]	Texture	AdaBoost	FERET
Tapia & Pérez [27]	Fusion	SVM	FERET, LFW
Shan [7]	Boosted LBP	SVM	LFW
Gallagher & Chen [28]	Contextual Features	GML	Groups
Rai & Khanna [29]	Gabor+$(2D)^2$PCA	SVM	FERET
Mery & Bowyer [30]	Random Patch	Adaptive SRC	FERET, LFW, Groups
Hadid et al. [31]	LBP, LPQ, BSIF	SVM	Groups
Moeini et al. [32]	LGBP	SVM	FERET, LFW
Han et al. [33]	BIF	SVM	LFW
Jain & Huang [34]	ICA	LDA	FERET
Lu et al. [35]	PPBTF	SVM	FERET
Zheng & Lu [36]	LGBP, MLBP, LBP	SVMAC	CAS-PEAL, FERET, BCMI
Berbar [37]	DCT, GLCM, DWT	SVM	AT@T, Face94, UMIST, FERET
Andreu et al. [38]	LBP+PCA	SVM	FERET, AR
Our Method	MMIP+CSVM	SVM	FRGC 2.0, FERET, LFW, UND

To address the real-world facial gender classification problem, Tapia & Pérez [27] present mutual information for feature selection and fusion to improve gender classification. They achieve promising results on LFW and FERET databases by adopting the SVM classifier. Shan [7] proposes gender recognition on real-world facial images in which AdaBoost is utilized to select the discriminative LBP features. Finally, gender classification results are acquired by applying the SVM classifier with the boosted LBP features. Gallagher & Chen [28] propose contextual features and adopt Gaussian Maximum Likelihood (GML) for gender classification. They perform real-world gender classification on the Groups database and obtain 74.1% accuracy. Rai & Khanna [29] propose Gabor based $(2D)^2$PCA and adopt the SVM classifier for gender classification. They perform the experiments on LFW and FERET databases with promising results achieved. Mery & Bowyer [30] present random patch for feature extrac-

tion and use the sparse adaptive classification to classify the gender on FERET and Groups databases.

Furthermore, some attempts perform gender classification using different feature extraction methods. Hadid et al. [31] propose 13 variants of LBP for gender classification. They use the fusion of three features including LBP, Local Phase Quantization (LPQ) and Binarized Statistical Image Features (BSIF). Finally, they adopt the SVM classifier and obtain an accuracy of 89.85% on the Groups database. Moeini et al. [32] propose a feature fusion method to extract features from both texture and depth images for gender classification. Their method is evaluated on FERET and LFW databases by adopting the SVM classifier. Recently, Han et al. [33] present a novel method for demographic estimation (i.e. age, gender and ethnicity). Accordingly, they use Biologically Inspired Features (BIF) and adopt the SVM classifier for gender classification on the LFW database.

Table 1 also shows that a number of research studies that rely on LBP operator are proposed in the literature. For example, Jain & Huang [34] use Independent Component Analysis (ICA) to represent the image and the SVM classifier for gender classification, and achieve 96% classification accuracy on the FERET database. Shan [7] proposes discriminative LBP-Histogram (LBPH) bins for gender classification. The selected LBPH bins provide a compact facial representation and achieve 94.81% accuracy. Lu et al. [35] propose Pixel-pattern-based Texture Feature (PPBTF), and use AdaBoost and SVM classifiers for gender classification with 92.71% accuracy on the FERET database achieved. Zheng & Lu [36] use gray, Gabor, LBP, Multi-resolution Local Binary Pattern (MLBP), and Local Gabor Binary Pattern (LGBP) to extract gender features, and achieve more than 95% accuracy on multi-view facial images from the CAS-PEAL dataset and frontal images from FERET and BCMI datasets. Berbar [37] uses DCT, Gray-level Cooccurrence Matrix (GLCM) and DWT for feature extraction and the SVM classifier for gender classification. Their results show that merging of features extracted from DCT and GLCM degrades the classification accuracy on AT@T, FERET, UMIST, and Faces94 datasets, while using

2D-DWT, an improvement in classification accuracy (ranging from 96.18% to 99.6%) is observed for all datasets except FERET (92%). Andreu et al. [38] conduct a comprehensive experimental study of gender classification methods from neutral to distorted facial images. They compare local and global approaches for classifying the gender by using gray-level information, PCA, and LBP features as well as 1-NN, SVM, and PCA+SVM classifiers. Finally, experiments on the FERET database achieve 94.06% accuracy.

It is worth noting that our proposed method differs from the above-mentioned methods for facial gender classification by the following aspects: 1) conducting the multi-spatial multi-order based texture operator to represent the gender information; and 2) developing a feature vector selection scheme for investigating the discriminative facial subregions and pattern orders. As a result, the discriminative characteristics are obtained by the proposed MMIP matrix and the CSVM based feature vector selection scheme, leading to less redundant information for improving the classification accuracy.

3. Methodology

In this section, we present the MMIP distribution matrix to extract discriminative gender characteristics and the CSVM-based feature vectors elimination and ranking approach to measure the importance of facial subregions and textural pattern orders for gender classification. Specifically, we elaborate the design considerations and the implementation algorithms for these two parts in Sections 3.1 and 3.2, respectively. Table 2 summarizes the major notations used in this paper.

3.1. Feature Vector Extraction

We first introduce the MMP descriptor and then propose the MMIP matrix to emphasize the gender discriminative information.

Table 2: Major notations.

Notation	Description
$\mathcal{D}$	A textural image ($\mathcal{D} \subseteq \mathbb{R}^2$)
$\mathbf{MMP}(\mathcal{D})$	MMP matrix of $\mathcal{D}$
$R_{p,q}$	The p-th row and q-th column subregion of $\mathcal{D}$
$\mathbf{M}_{p,q}$	An element of $\mathbf{MMP}(\mathcal{D})$ ($p = 1, \ldots, P$; $q = 1, \ldots, Q$)
$MDP^i(R_{p,q})$	The i-th order MDP in $R_{p,q}$ ($i = 1, \ldots, N$)
$f^i(R_{p,q})$	The corresponding PDF of $MDP^i(R_{p,q})$
$Z_0^{p,q}$	An arbitrary pixel of $R_{p,q}$
$MDP^i(Z_0^{p,q})$	The i-th order MDP of $Z_0^{p,q}$
$MDP_\theta^i(Z_0^{p,q})$	The i-th order MDP in θ direction of $Z_0^{p,q}$
$\theta = \frac{2(k-1)\pi}{M}$	A direction between a certain neighbor and $Z_0^{p,q}$ ($k = 1, \ldots, \frac{M}{2}$)
$Z_j^{p,q}$	The j-th neighbor of $Z_0^{p,q}$ ($j = 1, \ldots, M$)
$I_\theta^i(Z_j^{p,q})$	The i-th order derivative in θ direction of $Z_j^{p,q}$
$Z_{l,j}^{p,q}$	The l-th neighbor of $Z_j^{p,q}$ ($l = 1, \ldots, M$)
$F(\cdot,\cdot)$	The encoding function for MDP
$\mathbf{MMIP}(\mathcal{D})$	MMIP matrix of $\mathcal{D}$
$\mathbf{N}_{p,q}$	An element of $\mathbf{MMIP}(\mathcal{D})$
$MIP^i(R_{p,q})$	The i-th order MIP in $R_{p,q}$
$g^i(R_{p,q})$	The corresponding PDF of $MIP^i(R_{p,q})$
$G(\cdot,\cdot)$	The encoding function for MIP
$\mathbf{X}_t,\ y_t$	The t-th training instance and class label ($t = 1, \ldots, T$)
$sign(\cdot),\ sum(\cdot)$	The sign and summation function
$\boldsymbol{w},\ b$	The weight vector and bias term
$E(\cdot)$	The transformation function that transforms a vector to a matrix
$\Phi,\ K(\cdot,\cdot)$	The mapping and kernel function
$\|\cdot\|,\ \|\cdot\|_F$	The L_2-norm and Frobenius norm
$\varepsilon_t,\ \alpha_t$	The slack variable and coefficient of $\mathbf{X}_t$
$C,\ \sigma$	The regularization and width parameter
$W^2(\boldsymbol{\alpha})$	The inverse-square of the classification margin

3.1.1. Multi-spatial Multi-order Pattern

Given a textural image $\mathcal{D}$, we propose a new texture operator, MMP distribution matrix $\mathbf{MMP}(\mathcal{D}) = [\mathbf{M}_{p,q}]_{P\times Q}$, to represent the intensity change with high-order derivative at different subregions of $\mathcal{D}$, where P and Q represent the number of horizontal and vertical subregions of $\mathcal{D}$, respectively. That is, $\mathcal{D} = [R_{p,q}]_{P\times Q}$, where $R_{p,q}$ represents the p-th row and q-th column subregion of $\mathcal{D}$. Each $\mathbf{M}_{p,q}$ in $\mathbf{MMP}(\mathcal{D})$ is defined as

$$\mathbf{M}_{p,q} = [f^1(R_{p,q}), \ldots, f^N(R_{p,q})], \tag{1}$$

where N is the number of derivative pattern order, $f^i(R_{p,q})$ $(i = 1, \ldots, N)$ is a feature vector that represents the probability density function (PDF) extracted from the i-th order multi-order derivative pattern (MDP), $MDP^i(R_{p,q})$, in subregion $R_{p,q}$, which is defined as

$$MDP^i(R_{p,q}) = \{MDP^i(Z_0^{p,q}) | Z_0^{p,q} \in R_{p,q}\}, \tag{2}$$

where $Z_0^{p,q} \in R_{p,q}$ is an arbitrary point of $R_{p,q}$, and $MDP^i(Z_0^{p,q})$ is the i-th order MDP of $Z_0^{p,q}$, which can be computed by

$$MDP^i(Z_0^{p,q}) = \{MDP^i_\theta(Z_0^{p,q}) | \theta = \tfrac{2(k-1)\pi}{M}, k = 1, \ldots, \tfrac{M}{2}\}, \tag{3}$$

where $MDP^i_\theta(Z_0^{p,q})$ is the i-th order MDP in θ direction of $Z_0^{p,q}$, $\theta = 0, \ldots, \frac{(M-2)\pi}{M}$ are the directions of all neighbors of $Z_0^{p,q}$, M (e.g. $M = 2, 4, 8, 16, \ldots$) is the number of the neighbors. $MDP^i_\theta(Z_0^{p,q})$ can be recursively computed by

$$MDP^i_\theta(Z_0^{p,q}) = \{F(I^{i-1}_\theta(Z_0^{p,q}), I^{i-1}_\theta(Z_1^{p,q})), \ldots, F(I^{i-1}_\theta(Z_0^{p,q}), I^{i-1}_\theta(Z_M^{p,q}))\}, \tag{4}$$

where $F(\cdot, \cdot)$ that encodes a certain order gradient transitions into binary patterns is defined as [8]

$$F(u, v) = \begin{cases} 0, & \text{if } uv > 0 \\ 1, & \text{if } uv \leq 0 \end{cases} \tag{5}$$

In addition, $I^{i-1}_\theta(Z_0^{p,q})$ and $I^{i-1}_\theta(Z_j^{p,q})$ $(j = 1, \ldots, M)$ that represent the $(i-1)$-th order derivative in θ direction of $Z_0^{p,q}$ and $Z_j^{p,q}$, respectively, are recursively

defined as

$$I_\theta^{i-1}(Z_0^{p,q}) = I_\theta^{i-2}(Z_0^{p,q}) - I_\theta^{i-2}(Z_j^{p,q}), \quad (6)$$

$$I_\theta^{i-1}(Z_j^{p,q}) = I_\theta^{i-2}(Z_j^{p,q}) - I_\theta^{i-2}(Z_{l,j}^{p,q}). \quad (7)$$

In particular, the first-order derivatives $I_\theta^1(Z_0^{p,q})$ and $I_\theta^1(Z_j^{p,q})$ are defined as

$$I_\theta^1(Z_0^{p,q}) = I(Z_0^{p,q}) - I(Z_j^{p,q}), \quad (8)$$

$$I_\theta^1(Z_j^{p,q}) = I(Z_j^{p,q}) - I(Z_{l,j}^{p,q}). \quad (9)$$

where $I(Z_0^{p,q})$, $I(Z_j^{p,q})$, and $I(Z_{l,j}^{p,q})$ represent the gray value of $Z_0^{p,q}$, $Z_j^{p,q}$, and $Z_{l,j}^{p,q}$, respectively.

In summary, $MDP_\theta^i(Z_0^{p,q})$ is a M-bit binary sequence that describes the gradient trend changes of $(i-1)$-th order directional derivatives of $Z_0^{p,q}$, while $MDP^i(Z_0^{p,q})$ is a $\frac{M^2}{2}$-bit binary sequence that describes $(i-1)$-th order derivatives of $Z_0^{p,q}$. Therefore, $\mathbf{MMP}(\mathcal{D})$ actually encodes all the points in $\mathcal{D}$ with all corresponding neighbors and directions, which is computationally time-consuming for non-trivial images. To address this problem and further efficiently capture the gender discriminative information, based on MMP matrix, we propose MMIP matrix as described below.

3.1.2. Multi-spatial Multi-order Interlaced Pattern

To develop a fast version of MMP and suppress the face information with gender information emerging, we only maintain the encodes on the most relevant neighbor for a given direction, i.e. discarding the MDP encodes on the other $(M-1)$-neighbors with less relevant information (Figure 2). In this way, we propose the MMIP distribution matrix, $\mathbf{MMIP}(\mathcal{D}) = [\mathbf{N}_{p,q}]_{P\times Q}$, as the gender textural descriptor. Each $\mathbf{N}_{p,q}$ in $\mathbf{MMIP}(\mathcal{D})$ is defined as

$$\mathbf{N}_{p,q} = [g^1(R_{p,q}), \ldots, g^N(R_{p,q})], \quad (10)$$

where $g^i(R_{p,q})$ $(i = 1, \ldots, N)$ is a feature vector that represents the PDF extracted from the i-th order multi-order interlaced pattern (MIP), $MIP^i(R_{p,q})$,

Figure 2: Illustration of the coding process of $MIP^i(Z_0^{p,q})$ (highlight in red).

in subregion $R_{p,q}$, which is defined as

$$MIP^i(R_{p,q}) = \{MIP^i(Z_0^{p,q}) | Z_0^{p,q} \in R_{p,q}\}, \tag{11}$$

where $MIP^i(Z_0^{p,q})$ is the i-th order MIP of $Z_0^{p,q}$ (Figure 2 illustrates each component of $MIP^i(Z_0^{p,q})$ in Eq. (12)), which is defined as

$$\begin{aligned} MIP^i(Z_0^{p,q}) = \{&G(I^{i-1}_{\frac{(M-2)\pi}{M}}(Z_0^{p,q}), I^{i-1}_{\frac{(M-2)\pi}{M}}(Z_1^{p,q})), \ldots, G(I^{i-1}_{\frac{(M-2k)\pi}{M}}(Z_0^{p,q}), I^{i-1}_{\frac{(M-2k)\pi}{M}}(Z_k^{p,q})), \\ &\ldots, G(I_0^{i-1}(Z_0^{p,q}), I_0^{i-1}(Z_{\frac{M}{2}}^{p,q})), G(I^{i-1}_{\frac{2\pi}{M}}(Z_0^{p,q}), I^{i-1}_{\frac{2\pi}{M}}(Z^{p,q}_{\frac{M}{2}+1})), \\ &\ldots, G(I^{i-1}_{\frac{2k\pi}{M}}(Z_0^{p,q}), I^{i-1}_{\frac{2k\pi}{M}}(Z^{p,q}_{\frac{M}{2}+k})), \ldots, G(I_0^{i-1}(Z_0^{p,q}), I_0^{i-1}(Z_M^{p,q}))\}, \end{aligned} \tag{12}$$

where $k = 1, 2, \ldots, \frac{M}{2}$, it is worth noting that if $k = \frac{M}{2}$, then $\theta = 0$ (not $\theta = \pi$ since $0 \leq \theta < \pi$) in the third row of Eq. (12); $G(\cdot, \cdot)$ is the encoding function for MIP, which is defined as [9]

$$G(u, v) = \begin{cases} 1, & \text{if } u \geq v \\ 0, & \text{if } u < v \end{cases} \tag{13}$$

In summary, in each direction, only the derivatives for the center point and its neighbor point in that particular direction will be calculated. This will dramatically decrease the length of the pixel representing code produced by MIP compared to the MDP operator. In this way, MIP keeps only the most important information hence makes the process much faster. It produces an M-bit representation of each point in $\mathcal{D}$, which makes the operator $\frac{M}{2}$ times faster than MDP. Moreover, compared to LBP [7], MIP contains a more detailed description by calculating the high-order derivative variations, while LBP only provides the first-order derivative information and is incapable of describing more detailed information. Compared with LDP [8] that encodes the entire face image using a specific pattern order, MDP encodes all the facial subregions using N different pattern orders textual information. Similarly, compared with IDP [9], MIP encodes all the facial subregions using N different pattern orders

to enhance the gender information. As such, the first-order or specific high-order feature descriptors (i.e. LBP [7], LDP [8], IDP [9]) are the special cases of MDP and MIP, instead, our proposed feature descriptor is more generic by considering both spatial and pattern order information of facial images and is able to depict more detailed information. It is worth mentioning that the differences between PLBP [11] and MIP are in the following: 1) compared to our proposed method deploying on different facial components, PLBP is designed for the entire face with different resolutions; 2) PLBP improves the discriminative power of LBP by cascading the LBP information of hierarchical spatial pyramids at the expense of computational resources and cost, and the suitable scale of PLBP for each specific facial component is unclear. Our proposed feature descriptor extends the LBP information to high-order derivative representation. Finally, based on MIP, we want to select the most discriminative feature vectors that encode the facial textural information inseparably for better gender classification performance, leading to a new feature vector selection problem. Therefore, we propose the CSVM based feature selection method to learn the discriminative facial subregions with textural pattern orders for gender classification as discussed below.

3.2. Feature Vector Selection

To investigate the discriminative facial subregions with textural pattern orders for gender classification, we extend the traditional feature selection techniques to the feature vector selection paradigm. The input training data compose MMIP distribution matrices, e.g. $\mathbf{X}_1 = \mathbf{MMIP}_1(\mathcal{D}), \ldots, \mathbf{X}_T = \mathbf{MMIP}_T(\mathcal{D})$ extracted from T facial images. In the following, we first propose the CSVM model and then discuss the CSVM-based feature vector selection technique.

3.2.1. Chain-type Support Vector Machine

For each input data $\mathbf{X}_t$ $(t = 1, \ldots, T)$, the corresponding feature dimension is $\mathbf{X}_t = \mathbf{MMIP}_t(\mathcal{D}) \in \mathbb{R}^{PD \times QN}$, where $D = 2^M$ (or $D = 2^{\tilde{M}}$ $(\tilde{M} < M)$) is the feature dimension of (or downsampling) MIP on each subregion, i.e. $g_t^i(R_{p,q}) \in$

$\mathbb{R}^D$. Since the feature representation of MIP, $g_t^i(R_{p,q})$, in each $\mathbf{X}_t$ is inseparable for gender classification, without an abuse of notation, we rearrange the elements of $\mathbf{X}_t$ and denote it by $\mathbf{X}_t = [g_t^1(R_{1,1}), \ldots, g_t^N(R_{1,1}), \ldots, g_t^1(R_{P,Q}), \ldots, g_t^N(R_{P,Q})]^T \in \mathbb{R}^{S\times D}$, where $S = PQN$. Let $y_t \in \{-1, +1\}$ be the class label of $\mathbf{X}_t$, whereas $y_t = +1$ represents $\mathbf{X}_t$ is male, otherwise is female in our study.

Since all the elements of an arbitrary row of $\mathbf{X}_t$ will integrally represent the gender information, we call them chain-type feature vectors. Hence, figuring out the most reliable classification boundary and investigating the most indispensable feature vectors for gender classification are necessary for the new learning paradigm. In the literature, since SVM-based feature selection and classification techniques are widely used with a strong interpretability, we extend the SVM technique to **chain-type SVM (CSVM)** for selecting the chain-type feature vectors. The propose CSVM aims to separate the training data with largest margin and capture the best generalization ability of the new hyper-plane (or decision function)

$$h(\mathbf{X}) = sign(sum(E(\boldsymbol{w}) \bigotimes \mathbf{X}) + b), \tag{14}$$

where $sign(\cdot)$ is the sign function that determines whether the predicted classification comes out positive $(+1)$ or negative (-1), $\bigotimes$ represents the Kronecker product of two matrices with the same size, $b \in \mathbb{R}$ is the bias term, an SVM classifier trained without a bias term will always pass through the origin, $sum(\cdot)$ is the summation function that sums all the elements of a matrix, $\boldsymbol{w} \in \mathbb{R}^S$ is the weight vector, and $E(\cdot)$ is a transformation function that transforms a vector to a matrix. In this paper, $E(\boldsymbol{w}) = \underbrace{[\boldsymbol{w}, \ldots, \boldsymbol{w}]}_{D} \in \mathbb{R}^{S\times D}$.

Suppose all the data are linearly separable, and the primary aim of the hyper-plane is to maximize the margin (or distance) of the closest positive and negative training data. To conduct the maximum margin or optimal hyper-plane, we must classify the matrices of the training set correctly with largest classification margin and simultaneously minimize the norm of the weight vector. For the non-linearly separable data, we use the mapping function Φ that maps

the training matrices into a high dimensional space $\mathcal{H}$, where the new hyper-plane with maximal margin is constructed. We propose the following quadratic optimization problem as the CSVM model

$$\begin{aligned} \boldsymbol{w}^* &= \text{argmin}_{\boldsymbol{w}}\{\tfrac{1}{2}\|\boldsymbol{w}\|^2 + C\textstyle\sum_{t=1}^{T}\varepsilon_t\} \\ s.t.\ y_t(sum(E(\boldsymbol{w}) \bigotimes \Phi(\mathbf{X}_t)) + b) &\geq 1 - \varepsilon_t, \varepsilon_t \geq 0, t = 1, \ldots, T, \end{aligned} \tag{15}$$

where $\Phi(\mathbf{X}_t) \triangleq [\Phi(\boldsymbol{x}_{t,1}), \ldots, \Phi(\boldsymbol{x}_{t,D})]$ ($\mathbf{X} = [\boldsymbol{x}_{t,1}, \ldots, \boldsymbol{x}_{t,D}], \boldsymbol{x}_{t,d} \in \mathbb{R}^S, d = 1, \ldots, D$), $\|\cdot\|$ is the L_2-norm, ε_t is the slack variable of $\mathbf{X}_t$, and C is a regularization parameter that controls trade-off between the margin maximization and classification error. As a result, the optimal hyper-plane becomes $h(\mathbf{X}) = sign(sum(E(\boldsymbol{w}^*) \bigotimes \Phi(\mathbf{X})) + b)$. To solve Eq. (15), we use the dual optimization technique [39] to transform it into its dual form

$$\begin{aligned} \boldsymbol{\alpha}^* &= \text{argmax}_{\boldsymbol{\alpha}}\{\textstyle\sum_{t=1}^{T}\alpha_t - \frac{1}{2}\sum_{t=1}^{T}\sum_{r=1}^{T}\alpha_t\alpha_r y_t y_r K(\mathbf{X}_t, \mathbf{X}_r)\} \\ s.t.\ &\textstyle\sum_{t=1}^{T}\alpha_t y_t = 0, 0 \leq \alpha_t \leq C, t = 1, \ldots, T, \end{aligned} \tag{16}$$

where α_t is the corresponding coefficient of $\mathbf{X}_t$ (if $\alpha_t \neq 0$, the corresponding matrix is a support matrix), $K(\mathbf{X}_t, \mathbf{X}_r) = \exp\{-\frac{\|\mathbf{X}_t - \mathbf{X}_r\|_F^2}{2\sigma^2}\}$ is the kernel function, σ is the width parameter and $\|\cdot\|_F$ is the Frobenius norm. We can further obtain b^* based on the solution $\boldsymbol{\alpha}^*$ of Eq. (16), leading to the new hyper-plane

$$h(\mathbf{X}) = sign(\textstyle\sum_{t=1}^{T}\alpha_t^* y_t K(\mathbf{X}_t, \mathbf{X}) + b^*). \tag{17}$$

3.2.2. CSVM-based Feature Vector Selection

Inspired by the theory of support matrix machine (SMM) [40], in this paper, we propose CSVM-based feature vector selection algorithm for discriminative gender analysis. Since the type of input data for the classifier is a feature matrix, it is equivalent to the traditional SVM (i.e. the input of the training data is a feature vector) when all rows of the matrix are connected as a string according to the column order. For the connected feature string, every element is actually a feature vector that represents the MIP operator of the facial subregions with a certain textural pattern order. Therefore, selecting different rows of the input

matrix is equivalent to selecting the chain-type string one feature vector by one feature vector.

Similar to the Mercer's Theorem [39], the kernel function can be derived by $K(\mathbf{X}_t, \mathbf{X}) = sum(\Phi(\mathbf{X}_t) \bigotimes \Phi(\mathbf{X}))$. Thus, the discrimination function of the CSVM used in Eq. (17) is

$$\begin{aligned} h(\mathbf{X}) &= sign(\textstyle\sum_{t=1}^{T} \alpha_t^* y_t sum(\Phi(\mathbf{X}_t) \bigotimes \Phi(\mathbf{X})) + b^*) \\ &= sign(sum(\mathbf{W}(\boldsymbol{\alpha}^*) \bigotimes \Phi(\mathbf{X})) + b^*), \end{aligned} \tag{18}$$

where $\mathbf{W}(\boldsymbol{\alpha}^*) = \sum_{t=1}^{T} \alpha_t^* y_t \Phi(\mathbf{X}_t)$. Therefore, the inverse-square of the classification margin is

$$\begin{aligned} W^2(\boldsymbol{\alpha}^*) = \|\mathbf{W}(\boldsymbol{\alpha}^*)\|_F^2 &= \textstyle\sum_{t=1}^{T} \sum_{r=1}^{T} \alpha_t^* \alpha_r^* y_t y_r sum(\Phi(\mathbf{X}_t) \bigotimes \Phi(\mathbf{X}_r)) \\ &= \textstyle\sum_{t=1}^{T} \sum_{r=1}^{T} \alpha_t^* \alpha_r^* y_t y_r K(\mathbf{X}_t, \mathbf{X}_r), \end{aligned} \tag{19}$$

therefore, to reduce the computational cost and simultaneously extract discriminative feature vectors for discriminative gender analysis, we need to select the most important feature vectors of the training matrices. Motivated by the feature selection techniques, e.g. support vector machine recursive feature elimination (SVM-RFE) [41] and additional regularization and embedded nonlinear feature selection [42], and to obtain the best generalization ability of CSVM with largest classification margin, we propose the recursive feature vector elimination scheme. We use the overall compound information among initial feature vectors and remove the group of feature vectors that has negative impact on the classification margin of CSVM (Eq. (19)), i.e. the feature vectors with deteriorating gender classification performance are discarded from the training matrices.

We further determinate the ranking order of the selected feature vectors according to their different contributions to the classification margin. This method is actually a greedy strategy to obtain the optimal feature vectors as the criterion tries to acquire optimal classification margin in each iteration. Algorithm 1 presents the detailed implementation of the method. Note that $\mathbf{X}_t^{-s}$ represents the matrix $\mathbf{X}_t$ with the s-th row removed, α_{-s}^* is the mapping solution when all

the training data $\{\mathbf{X}_t\}_{t=1}^{T}$ are replaced by $\{\mathbf{X}_t^{-s}\}_{t=1}^{T}$, $\{s_1, \ldots, s_{|\mathcal{F}|}\}$ is a certain permutation of $\mathcal{F}$ with $W^2_{-s_1}(\boldsymbol{\alpha}^*_{-s_1}) \leq \ldots \leq W^2_{-s_{|\mathcal{F}|}}(\boldsymbol{\alpha}^*_{-s_{|\mathcal{F}|}})$. The time complexities of Algorithm 1 are $O(S \times D)$ and $O(S \times \log S \times D)$ in terms of modeling CSVM and ranking process, respectively, where $S = PQN$ is the length of a chain, $D = 2^M$ (or $D = 2^{\tilde{M}}$) is the feature dimension of (or downsampling) MIP coding. Therefore, if we only aim to obtain the best feature vectors, Algorithm 1 would be more computationally efficient.

4. Experiments

We conduct experiments of gender recognition on four famous datasets to validate the efficiency of the proposed method. In the experiments we assume that the face images have been detected and normalized as discussed in [43]. Specifically, the experimental setup is presented in Section 4.1. The discriminative facial subregions with textural pattern orders are investigated on the FRGC 2.0 dataset in Section 4.2. We report the experimental results comparing six different feature descriptors and four different classifiers on FRGC 2.0 and FERET datasets in Section 4.3. Finally, the comparison results including average classification accuracy and running time with eight state-of-the-art baselines are discussed in Section 4.4 and Section 4.5, respectively.

4.1. Experimental Setup

We use four datasets, i.e. FRGC 2.0 [20], FERET [21], LFW [22], and UND [23], to evaluate our algorithm. In the following, we summarize the characteristics of these four datasets.

- The FRGC 2.0 is one of the most comprehensive datasets publicly available for face analysis. In this paper, we select 458 unique front individuals from the dataset, including 262 male and 196 female objects. All of them are normalized with the entire face and cropped to 128×144 images so that each image contains little or no hair information.

Algorithm 1 CSVM-based Feature Vector Elimination & Ranking Algorithm.

Input: T training matrices, $\mathbf{X}_t \in \mathbb{R}^{S\times D}$ $(t = 1,\ldots,T)$; the class label $y_t \in \{-1,+1\}$; the kernel function, $K(\cdot,\cdot)$; the width parameter, σ; the regularization parameter, C.

Output: the index set $\mathcal{F}$ of the optimal feature vectors; the index set $\mathcal{G}$ of the corresponding rankings of these feature vectors; new training matrices, $\mathbf{X}_t \in \mathbb{R}^{|\mathcal{F}|\times D}$ $(t = 1,\ldots,T)$.

Initialization: $\mathcal{F} = \{1,\ldots,S\}$, $\mathcal{G} = \emptyset$;
repeat
  Give a solution $\boldsymbol{\alpha}^*$ of Eq. (16), $\mathbf{W}(\boldsymbol{\alpha}^*) \leftarrow$ CSVM Training;
  Compute the inverse-square of the classification margin, $W^2(\boldsymbol{\alpha}^*)$, using Eq. (19);
  for each $s \in \mathcal{F}$ **do**
    Compute the inverse-square of the classification margin, $W^2_{-s}(\boldsymbol{\alpha}^*_{-s})$;
  end for
  Order these inverse-square of the classification margin, $W^2(\boldsymbol{\alpha}^*)$, $\{W^2_{-s}(\boldsymbol{\alpha}^*_{-s})\}_{s\in\mathcal{F}}$;
  if $W^2(\boldsymbol{\alpha}^*)$ is the smallest value of these $(|\mathcal{F}|+1)$ items **then**
    $\mathcal{F} = \mathcal{F}$;
    $\{\mathbf{X}_t\}_{t=1}^T = \{\mathbf{X}_t\}_{t=1}^T$;
  else if $W^2(\boldsymbol{\alpha}^*)$ is not the smallest value of these $(|\mathcal{F}|+1)$ items, while $W^2_{-s}(\boldsymbol{\alpha}^*_{-s})$ is the smallest value of them **then**
    $\mathcal{F} = \mathcal{F} - \{s\}$;
    $\{\mathbf{X}_t\}_{t=1}^T = \{\mathbf{X}_t^{-s}\}_{t=1}^T$;
  end if
until $\mathcal{F}$ no longer changes.
for each $s \in \mathcal{F}$ **do**
  Compute the inverse-square of the classification margin, $W^2_{-s}(\boldsymbol{\alpha}^*_{-s})$;
end for
Arrange these values, $\{W^2_{-s}(\boldsymbol{\alpha}^*_{-s})\}_{s\in\mathcal{F}}$, in an ascending order: $W^2_{-s_1}(\boldsymbol{\alpha}^*_{-s_1}),\ldots,W^2_{-s_{|\mathcal{F}|}}(\boldsymbol{\alpha}^*_{-s_{|\mathcal{F}|}})$;
Update $\mathcal{G} = \{1,\ldots,|\mathcal{F}|\}$ for $\mathcal{F} = \{s_1,\ldots,s_{|\mathcal{F}|}\}$.

- The FERET dataset is one of the most widely used database in facial gender classification studies. It consists of 14,126 gray-scale images representing 994 individuals (i.e. 591 male and 403 female objects). These images contain variations in lighting, facial expressions, pose angle, and aging effects. In this work, we collect 1,000 mug-shot face images with 500 male faces and rest 500 female faces. Meanwhile, 512×768 pixels face images are cropped and normalized to 80×80 pixels based on the true positions of two eyes and a mouth.

- The LFW dataset contains more than 13,000 facial images collected from the web, including 5,749 people. The faces have a large range of variation include lighting, expression, pose, race, gender, and background. Here, we only use the regular frontal facial images (i.e. the fa partition) in this database, containing 550 male and 450 female objects.

- There are 10,700 individuals, i.e. 5,646 male and 5,054 female objects, in the UND dataset. In the experiments, we use 5,000 male and 5,000 female images for gender classification.

For fair comparison, no other form of pre-processing is applied. The proposed MMIP matrix is compared with four different feature descriptors, i.e. LBP [7], LDP [8], IDP [9], DWT [10], and PLBP [11]. Meanwhile, we compare our method with four widely-used classifiers, i.e. k-NN [44], Naïve Bayes [45], AdaBoost [46], and SVM [47]. Through a series of preliminary studies, we set $M = 8$, $N = 4$ in the experiments. In addition, to compute spatially enhanced histograms, the images are divided into non-overlapping blocks of size 4×4 resulting in 16 blocks.

To evaluate gender classification performance, stratified five-fold cross validation is used for the datasets. The five-fold cross validation scheme randomly divides the data into five non-overlapping subsets and at a time, only one subset (20% of the data) is used as a test while other subsets (80% of the data) are used to train a classifier. This procedure is repeated five times so that each subset is used once as a test set. In addition, in order to highlight the

performance of different approaches, we assume the training and test datasets have the independent constraint condition (except [36]), which means that the same individual does not appear in both training and test set. Finally, the averages and standard variations of five different classification rates are reported. The parameters of SVMs are determined by five-fold cross validation for the Gaussian kernel function. The modified code for CSVM is based on LibSVM provided by [47]. The experiments are conducted on a computer with 4.00 GB RAM and Inter(R) Core (TM) i7-3770 with 3.40 GHz CPU.

4.2. Discriminative Feature Analysis

As a case study, in this section, we intend to find the most indispensable feature vectors for gender classification on the FRGC 2.0 dataset. In order to explore the significant differences between the female and male groups by highlighting meaningful facial components (i.e. forehead, eyes, nose, cheek, mouth, chin, or cross areas), we have designed two cases of experiments, i.e. symmetrically or asymmetrically segmenting the facial regions.

4.2.1. Discriminative Feature Analysis through Symmetrical Division

Every face sample is cropped to 128×144 and divided into 16 blocks of 32×36 pixels symmetrically. Four orders of MIP feature (i.e. first-order, second-order, third-order, and fourth-order MIP) are extracted. Totally we have 64×256 = 16,384 dimensions for each sample, using histogram down sampling, the dimension of each sample is reduced to 64×32 = 2,048. The grid search technique is used to figure out the best parameter value (C, σ) of SVM (Gaussian kernel function is utilized in the paper). We set the kernel parameter $\sigma = 0.1, 0.2, \ldots, 1$ and penalty parameter $C = 1, 10, \ldots, 10^4$. In grid search process, pairs of (C, σ) are tested, and the best parameter combination (i.e. $C = 10^2, \sigma = 0.6$) with average classification accuracy 95.58% is achieved. Meanwhile, the number of eliminated feature vectors is 26 and hence 38 feature vectors are optimally preserved.

Figure 3: The ranking list of the eliminated and selected feature vectors. The left-hand part is the input face image and subregions divided in a symmetrical way; the right-hand parts are: (a) sixteen subregions with four pattern orders of MIPs; (b) the selected and eliminated pattern orders in different subregions (the yellow parts represent the remaining feature vectors, e.g. 2-3(4) represents the second region with third-order MIP fourthly selected; the red parts represent eliminated feature vectors, e.g. 2-2(8) represents the second region with second-order MIP eighthly eliminated); (c) the optimal feature vectors combination; (d) the representation of selected subregions and pattern orders.

Figure 3 shows the ranking list of these eliminated and selected feature vectors, which indicates that the selected feature vectors are also approximately symmetrical. The CSVM based feature selection method not only increases the accuracy but also decreases the dimension of the feature space from $64\times32 = 2{,}048$ to $38\times32 = 1{,}216$ on the FRGC 2.0 dataset. To separate the male and female groups with optimal classification accuracy, Figure 3 shows that different subregions should be optimal combined with different textural pattern orders. For instance, in the region of forehead, we need two high-order patterns (i.e. third-order and fourth-order MIP), while in region of cheek, we need another two high-order patterns (i.e. second-order and fourth-order MIP) to figure out gender differences. Finally, the combination of all these feature vectors presents the largest classification margin, which validates that this combination is indispensable for gender classification. Furthermore, all the different importance for the SVM classifier can be quantitatively measured and ranked. According to the results of ranked feature vectors, the upper subregions with pattern orders combination have the highest priority to be selected for discriminative analysis and gender classification. Therefore, the upper regions of the face have been proved to be the most significant parts for the task of gender classification.

We also compare our results with first-order, second-order, third-order and fourth-order MIP, and all of them for gender classification. Table 3 shows that compared to the first-order or second-order MIP, the gender classification accuracy decreases significantly when using all pattern orders combination. However, all the gender classification rates are significantly less than our proposed method

Table 3: Gender classification rate comparison without feature vector selection mechanism.

Pattern order	First-order MIP	Second-order MIP	Third-order MIP	Fourth-order MIP	All orders MIP
Classification rate	73.80%	82.10%	59.83%	65.07%	71.62%

Figure 4: The ranking list of the eliminated and selected feature vectors. The left-hand part is the input face image and subregions divided in an asymmetrical way; the right-hand parts are: (a) sixteen subregions with four pattern orders of MIPs; (b) the selected and eliminated pattern orders in different subregions; (c) the optimal feature vectors combination; (d) the representation of selected subregions and pattern orders.

as we use facial subregions with textural pattern orders to form the optimal feature representation for gender classification, which also supports that gender differences exist in different facial subregions with different textural pattern orders, and also with different priorities to be selected.

4.2.2. Discriminative Feature Analysis through Asymmetrical Division

When dividing the facial images in an asymmetrical way, we explore the significant differences between female and male groups in the meaningful facial components, i.e. forehead, eyes, nose, cheek, mouth, and chin. In the experiments, the parameter combination (i.e. $C = 10^3, \sigma = 0.5$) with average classification accuracy 96.10% is selected. Meanwhile, the number of eliminated feature vectors is 30 and 34 feature vectors are optimally preserved.

Figure 4 shows the ranking list of these eliminated and selected feature vectors and indicates that the selected feature vectors are also approximately asymmetrical. The CSVM based feature selection method not only increases the accuracy but also decreases the dimension of the feature space from 64×32 = 2,048 to 34×32 = 1,088 on the FRGC 2.0 dataset. To separate the male and female groups with optimal classification accuracy, Figure 4 shows different subregions should be optimal combined with different textural pattern orders. For instance, in the forehead region, we need two high-order patterns (i.e., third-order and fourth-order MIP), while in eyes, nose, cheek and chin region, we need three high-order patterns (i.e. second-order, third-order and fourth-order MIP),

Table 4: Gender classification rate comparison without feature vector selection mechanism.

Pattern order	First-order MIP	Second-order MIP	Third-order MIP	Fourth-order MIP	All orders MIP
Classification rate	78.26%	84.61%	69.25%	72.43%	76.86%

in the mouth region, we need another two high-order patterns (i.e. second-order and third-order MIP) to figure out gender differences.

Finally, the combination of all these feature vectors presents the largest classification margin, which validates that this combination is indispensable for gender classification. Furthermore, the importance of the SVM classifier can be quantitatively measured and ranked. According to the results of ranked feature vectors, compared to the under subregions (i.e. mouth and chin), the upper subregions (i.e. forehead, eyes, nose and cheek) with selected pattern orders combination have the highest priority to be selected, which further strongly supports that the upper regions of the face are the most significant parts for the task of gender classification.

We also compare our results with first-order, second-order, third-order and fourth-order MIP, and all of them for gender classification, Table 4 shows that compared to the first-order or second-order MIP, the gender classification accuracy decreases significantly when using all pattern orders combination. However, all the gender classification rates are significantly less than our proposed method as we use facial subregions with textural pattern orders to form the optimal feature representation for gender classification.

In summary, our proposed approach has a significant improvement for gender classification by investigating gender differences in facial subregions and pattern orders. On one hand, when we symmetrically divide the facial images, the selected feature vectors are approximately symmetrical according to the selected feature vectors. Therefore, we can further apply the proposed method to gender classification even if either left or right half face is corrupted. These experimental results not only help determine the significant differences between female and male groups, but also obtain the symmetry of the selected feature vectors while reducing the classification error rate to 4.4%. One the other hand, when

Figure 5: The average classification accuracy and standard deviation by 5-fold cross validation are compared with six feature descriptions and four classifiers.

we asymmetrically divide the facial images, the proposed method not only selects significant different subregions and pattern orders for gender classification, but also explores the significant differences between female and male groups in meaningful facial components, i.e. forehead, eyes, nose, cheek, mouth, and chin. As a result, the gender classification accuracy has been improved from 95.58% to 96.10%.

4.3. Comparison with Different Feature Descriptors and Classifiers

The results of gender characteristics analysis show that selected feature vectors are indispensable for improving gender classification performance. For comprehensive analysis, in this section, we further compare our proposed method with five widely used feature descriptors and four classifiers for gender classification on FRGC 2.0 and FERET databases.

4.3.1. Comparative Studies on the FRGC 2.0 Dataset

The classification rates are listed in Figure 5. First, in all cases except the first one, the SVM classifier gives slightly better accuracy than the AdaBoost classifier, however, the AdaBoost classifier gives slightly better standard variations than that of the SVM classifier, one possible reason is that the SVM classifier has a greater capacity to fit attributes of individual face. Overall, the best accuracy for both SVM and AdaBoost classifiers seems to be significantly better than other classifiers when using the same feature descriptor. Second, the SVM classifier is superior in classification performance on the FRGC 2.0 dataset. The classification rate of MIP+SVM shows that the feature vectors extracted by MIP operator and selected by CSVM algorithm are discriminative for gender classification. Additionally, compared to LBP, LDP, IDP, DWT and PLBP operators, the proposed MMIP matrix is more powerful to integrate spatial and high-order information for enhancing gender characteristics.

Figure 6: The average classification accuracy and standard deviation by 5-fold cross validation are compared with six feature descriptions and four classifiers.

4.3.2. Comparative Studies on the FERET Dataset

Figure 6 lists the compared results of gender classification with different feature descriptors and classifiers on the FERET dataset. This is similar to the above compared results. The proposed MMIP feature descriptor consistently outperforms the other five baseline feature descriptors when using the same classifiers (i.e. Nave Bayes, AdaBoost, and SVM). It is notable that there is a counterexample: LBP achieves the best classification performance when using the k-NN classifier. However, in this case, MMIP still outperforms the other three feature descriptors, i.e. LDP, IDP and PLBP. In addition, the SVM classifier achieves better classification performance than the other classifier in most cases, where MMIP simultaneously shows the superiority to other five feature descriptors. Therefore, the effectiveness of proposed MMIP feature descriptor compared to the other five baseline feature descriptors has been verified. Moreover, the overall average accuracy on the FRGC 2.0 dataset is slightly better than that on the FERET dataset, one possible reason is that our experiments use the low resolution images from the FERET dataset.

In addition, we observe that given a feature extraction method, when using different classifiers, the SVM classifier achieves the best classification accuracy, the other classifiers perform from best to worst are AdaBoost, Naïve Bayes and k-NN. Given a classifier, in most cases, the MIP descriptor achieves the best classification accuracy, the other descriptors performance from best to worst PLBP, LBP, IDP, LDP and then DWT. Although the spatial pyramid information can improve the discriminative power of LBP, PLBP still performs less effective when compared with our proposed method. One possible reason is that, in our model, the process of training the SVM classifier is achieved by efficient eliminating a large number of feature vectors with low relevance and high redundancy. These results have been discussed in Section 4.2.

Table 5: Classification performance of state-of-the-art gender classification approaches on all datasets using 5-fold cross validation (the best result on each dataset is highlighted in bold).

Dataset/Method	Jain [34]	Makinen [1]	Lu [35]	Zheng [36]	Shan [7]	Berbar [37]	Andreu [38]	Mansanet [48]	Our Method
FRGC 2.0	94.86%	91.93%	90.32%	92.58%	95.26%	94.08%	95.69%	92.35%	**96.10%**
FERET	95.67%	92.86%	94.85%	**99.10%**	94.81%	92.00%	94.06%	91.58%	95.03%
LFW	93.15%	91.33%	93.68%	94.81%	92.56%	90.49%	92.25%	**95.16%**	94.10%
UND	86.33%	84.51%	86.19%	87.28%	84.68%	83.05%	85.72%	88.60%	**89.42%**

4.4. Comparison with State-of-the-art Approaches

For fair comparison, we conduct all the experiments using the four datasets described in Section 4.1. We report the gender classification performance (i.e. average classification accuracy, training and testing time) using 5-fold cross validation of all the methods except [36]. It is worth noting that the classification results in our experiments may be different from the reported results in the original papers, since we only use limited front images of the datasets.

Table 5 shows the classification performance of our proposed method on all datasets compared to other eight representative gender classification approaches. For example, on FRGC 2.0 and UND datasets, the classification accuracy of our method consistently leads the corresponding classification accuracy of other competitors. Specifically, the state-of-the-art performance is 95.69% by LBP+SVM [38] on the FRGC 2.0 dataset, and the best classification accuracy among the competitors is achieved by [48] (88.60% using local DNNs) on the UND dataset. Our proposed method is relatively more apt to achieve higher classification accuracy (i.e. 96.10% and 89.42%) than the best competitors on these two datasets. It is worth noting that, the differences in gender classification are with independent constraint conditions. However, in [36], facial images of individual people may appear in both the training and test sets (i.e. people have multiple images on the FERET dataset), which makes the task both easier and much less applicable to the problem we are interested in, i.e. recognizing the gender of people for whom we have not trained in the classifier. The results shown in Table 5 indicate that the performances of gender classification without independent constraint conditions is apparently higher than that

with independent constraint conditions. When using the method proposed by [36], the classification rate with independent constraint condition is reduced by about 7% compared to that without the independent constraint condition.

On the FERET dataset, the best performance is obtained by [36], however, they use a smaller dataset without independent constraint, which will actually make the problem easier to handle compared with other methods. As mentioned, the classification rate with independent constraint conditions is significantly reduced when using the method proposed by [36] on both FRGC 2.0 and UND datasets, this is also true on the LFW dataset (Table 5). Our proposed method is competitive with the methods proposed by [34, 48], and for comparison with other methods, our proposed method achieves best classification accuracy with independent constraint for front images. Although the better results are achieved by Jain [34] on the FERET dataset and Mansanet [48] on the LFW dataset, respectively, our proposed method is still comparative to these two methods with around 1% classification accuracy reduced. Furthermore, for the other three datasets, our proposed method consistently outperforms than the deep learning based method [48]. One possible reason is that the local DNNs method [48] is data hungry for gender classification while the volume of the datasets is relatively small in our study. Finally, the deep learning based method [48] often need more computational cost and lack interpretability for gender classification since it is designed with a black-box mechanism. Instead, the gender classification results achieved by our proposed method are significantly improved while the number of input feature vectors is drastically reduced, making it possible for practical applications when considering both effectiveness and interpretability.

In summary, the effectiveness of MMIP+CSVM for gender classification are three-fold: 1) the discriminative gender information has been captured by the proposed scheme, which has been proved to be outperforming other pattern orders on the entire face (Section 4.2); 2) the effectiveness of the proposed MMIP feature descriptor has been verified via different feature descriptors and classifiers comparison (Section 4.3); and 3) the proposed method achieves competitive

Table 6: The average running time for different methods over the four datasets. To focus on comprehensive comparisons among the practical gender classification systems, the running time for feature extraction is also included.

Time/Method	Jain [34]	Makinen [1]	Lu [35]	Zheng [36]	Shan [7]	Berbar [37]	Andreu [38]	Mansanet [48]	Our Method
Training time (minute)	49.17	54.52	11.60	30.41	7.52	34.81	47.31	351.20	46.34
Testing time (minute)	18.63	19.81	3.34	8.71	2.16	10.24	15.17	102.57	15.81

classification performance compared to eight state-of-the-art baselines for gender classification on a wide range of datasets (i.e. FRGC 2.0, FERET, LFW, and UND) (Section 4.4).

4.5. Running Time Comparison

The average training and testing time of 5-fold cross-validation over the four datasets are presented in Table 6. Averagely, on these four datasets, Shan [7] achieves the fastest running time among the other methods since the boosted LBP is utilized. The running time consumed by Mansanet [48] is merely 6 times more than the second slowest method [1] using the row pixels. The reason is that deep learning needs extremely more time to train the classification model. Our method is slower than the existing four methods (i.e. Shan [7], Lu [35], Zheng [36], and Berbar [37]), since they do not deploy any feature selection mechanism for gender classification. Compared to two feature selection methods, i.e. Andreu [38] and Jain [34], our proposed method consistently achieves less training and testing time. In addition, we can further speed up our proposed method for practical applications through the downsampling strategy for MMIP. Therefore, from the perspective of interpretability, effectiveness and efficiency for gender classification, our proposed method is the only one that can pinpoint the discriminative facial subregions and pattern orders and simultaneously achieve competitive classification accuracy and running time.

5. Conclusion

In this paper, we propose a novel technique for facial gender recognition by investigating the discriminative subregions and pattern orders. The proposed

method consists of two important components: 1) a generalized texture operator, i.e. the MMIP distribution matrix, for enhancing facial gender information; and 2) a CSVM based feature vector selection algorithm for measuring the importance of facial subregions and pattern orders. As a result, the redundant and irrelevant feature vectors represented by the MMIP matrix can be eliminated, and the discriminative gender information can be pinpointed for subsequent facial gender classification. The proposed method is technically effective and obtains higher classification performance with discriminative gender information obtained. Compared to the classical feature descriptors, the proposed MMIP descriptor achieves better classification accuracy. Finally, the proposed method achieves competitive classification performance compared to eight state-of-the-art baselines for gender classification. As part of future work, we will 1) apply the proposed method to classify gender from noisy facial images; and 2) investigate discriminative gender characteristics from corrupted facial images for robust gender classification.

Acknowledgements

The authors would like to thank the editor and the anonymous reviewers for their critical and constructive comments and suggestions. This work was made possible by funding from the DOD ARO grant #W911NF-15-1-0510, and the NIH grants 2G12MD007595 and P01 CA214091.

References

[1] E. Makinen, R. Raisamo, Evaluation of gender classification methods with automatically detected and aligned faces, IEEE Transactions on Pattern Analysis and Machine Intelligence 30 (3) (2008) 541–547.

[2] J. Bekios-Calfa, J. M. Buenaposada, L. Baumela, Revisiting linear discriminant techniques in gender recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (4) (2011) 858–864.

[3] H.-C. Shih, Robust gender classification using a precise patch histogram, Pattern Recognition 46 (2) (2013) 519–528.

[4] T. F. Cootes, G. J. Edwards, C. J. Taylor, Active appearance models, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (6) (2001) 681–685.

[5] L. Ballihi, B. B. Amor, M. Daoudi, A. Srivastava, D. Aboutajdine, Boosting 3-d-geometric features for efficient face recognition and gender classification, IEEE Transactions on Information Forensics and Security 7 (6) (2012) 1766–1779.

[6] Z. M. Hafed, M. D. Levine, Face recognition using the discrete cosine transform, International Journal of Computer Vision 43 (3) (2001) 167–188.

[7] C. Shan, Learning local binary patterns for gender classification on real-world face images, Pattern Recognition Letters 33 (4) (2012) 431–437.

[8] B. Zhang, Y. Gao, S. Zhao, J. Liu, Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor, IEEE Transactions on Image Processing 19 (2) (2010) 533–544.

[9] A. Shobeirinejad, Y. Gao, Gender classification using interlaced derivative patterns, in: International Conference on Pattern Recognition, IEEE, 2010, pp. 1509–1512.

[10] H. Hu, Variable lighting face recognition using discrete wavelet transform, Pattern Recognition Letters 32 (13) (2011) 1526–1534.

[11] X. Qian, X. Hua, P. Chen, L. Ke, Plbp: An effective local binary patterns texture descriptor with pyramid representation, Pattern Recognition 44 (3) (2011) 2502–2515.

[12] K. Ueki, H. Komatsu, S. Imaizumi, K. Kaneko, N. Sekine, J. Katto, T. Kobayashi, A method of gender classification by integrating facial,

hairstyle, and clothing images, in: International Conference on Pattern Recognition, IEEE, 2004, pp. 446–449.

[13] A. Lapedriza, M. J. Maryn-Jimenez, J. Vitria, Gender recognition in non controlled environments, in: International Conference on Pattern Recognition, IEEE, 2006, pp. 834–837.

[14] X.-C. Lian, B.-L. Lu, Gender classification by combining facial and hair information, in: International Conference on Neural Information Processing, Springer, 2008, pp. 647–654.

[15] B. Li, X.-C. Lian, B.-L. Lu, Gender classification by combining clothing, hair and facial component classifiers, Neurocomputing 76 (1) (2012) 18–27.

[16] L. Lu, Z. Xu, P. Shi, Gender classification of facial images based on multiple facial regions, in: WRI World Congress on Computer Science and Information Engineering, IEEE, 2009, pp. 48–52.

[17] A. Hasnat, S. Haider, D. Bhattacharjee, M. Nasipuri, A proposed system for gender classification using lower part of face image, in: International Conference on Information Processing, IEEE, 2015, pp. 581–585.

[18] J. Merkow, B. Jou, M. Savvides, An exploration of gender identification using only the periocular region, in: International Conference on Biometrics: Theory Applications and Systems, IEEE, 2010, pp. 1–5.

[19] Ö. Özbudak, M. Kirci, Y. Çakir, E. O. Güneş, Effects of the facial and racial features on gender classification, in: Mediterranean Electrotechnical Conference, IEEE, 2010, pp. 26–29.

[20] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, W. Worek, Overview of the face recognition grand challenge, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2005, pp. 947–954.

[21] P. J. Phillips, H. Moon, S. A. Rizvi, P. J. Rauss, The feret evaluation methodology for face-recognition algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 (10) (2000) 1090–1104.

[22] G. B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled faces in the wild: A database for studying face recognition in unconstrained environments, Tech. rep., University of Massachusetts, Amherst (2007).

[23] L. A. Alexandre, Gender recognition: A multiscale decision fusion approach, Pattern Recognition Letters 31 (11) (2010) 1422–1427.

[24] B. Moghaddam, M.-H. Yang, Learning gender with support faces, IEEE Transactions on Pattern Analysis and Machine Intelligence 24 (5) (2002) 707–711.

[25] S. Baluja, H. A. Rowley, et al., Boosting sex identification performance, International Journal of computer vision 71 (1) (2007) 111–119.

[26] Z. Yang, M. Li, H. Ai, An experimental study on automatic face gender classification, in: International Conference on Pattern Recognition, IEEE, 2006, pp. 1099–1102.

[27] J. E. Tapia, C. A. Perez, Gender classification based on fusion of different spatial scale features selected by mutual information from histogram of lbp, intensity, and shape, IEEE Transactions on Information Forensics and Security 8 (3) (2013) 488–499.

[28] A. C. Gallagher, T. Chen, Understanding images of groups of people, in: IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2009, pp. 256–263.

[29] P. Rai, P. Khanna, A gender classification system robust to occlusion using gabor features based (2d)2pca, Journal of Visual Communication and Image Representation 25 (5) (2014) 1118–1129.

[30] D. Mery, K. Bowyer, Automatic facial attribute analysis via adaptive sparse representation of random patches, Pattern Recognition Letters 68 (2015) 260–269.

[31] A. Hadid, J. Ylioinas, M. Bengherabi, M. Ghahramani, A. Taleb-Ahmed, Gender and texture classification: A comparative analysis using 13 variants of local binary patterns, Pattern Recognition Letters 68 (2015) 231–238.

[32] A. Moeini, K. Faez, H. Moeini, Real-world gender classification via local gabor binary pattern and three-dimensional face reconstruction by generic elastic model, IET Image Processing 9 (8) (2015) 690–698.

[33] H. Han, C. Otto, X. Liu, A. K. Jain, Demographic estimation from face images: Human vs. machine performance, IEEE Transactions on Pattern Analysis and Machine Intelligence 37 (6) (2015) 1148–1161.

[34] A. Jain, J. Huang, Integrating independent components and linear discriminant analysis for gender classification, in: International Conference on Automatic Face and Gesture Recognition, IEEE, 2004, pp. 159–163.

[35] H. Lu, Y. Huang, Y. Chen, D. Yang, Automatic gender recognition based on pixel-pattern-based texture feature, Journal of Real-Time Image Processing 3 (1-2) (2008) 109–116.

[36] J. Zheng, B.-L. Lu, A support vector machine classifier with automatic confidence and its application to gender classification, Neurocomputing 74 (11) (2011) 1926–1935.

[37] M. A. Berbar, Three robust features extraction approaches for facial gender classification, The Visual Computer 30 (1) (2014) 19–31.

[38] Y. Andreu, P. García-Sevilla, R. A. Mollineda, Face gender classification: A statistical study when neutral and distorted faces are combined for training and testing purposes, Image and Vision Computing 32 (1) (2014) 27–36.

[39] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, B. Scholkopf, Support vector machines, IEEE Intelligent Systems and their applications 13 (4) (1998) 18–28.

[40] L. Luo, Y. Xie, Z. Zhang, W.-J. Li, Support matrix machines, in: International Conference on Machine Learning, IMLS, 2015, pp. 938–947.

[41] I. Guyon, J. Weston, S. Barnhill, V. Vapnik, Gene selection for cancer classification using support vector machines, Machine Learning 46 (1) (2002) 389–422.

[42] J. Neumann, C. Schnörr, G. Steidl, Combined svm-based feature selection and classification, Machine Learning 61 (1) (2005) 129–150.

[43] S. Gutta, J. R. Huang, P. Jonathon, H. Wechsler, Mixture of experts for classification of gender, ethnic origin, and pose of human faces, IEEE Transactions on Neural Networks 11 (4) (2000) 948–960.

[44] F. Pernkopf, Bayesian network classifiers versus selective k-nn classifier, Pattern Recognition 38 (1) (2005) 1–10.

[45] S.-B. Kim, K.-S. Han, H.-C. Rim, S. H. Myaeng, Some effective techniques for naive bayes text classification, IEEE Transactions on Knowledge and Data Engineering 18 (11) (2006) 1457–1466.

[46] S. Mathanker, P. Weckler, T. Bowser, N. Wang, N. Maness, Adaboost classifiers for pecan defect classification, Computers and Electronics in Agriculture 77 (1) (2011) 60–68.

[47] C.-C. Chang, C.-J. Lin, Libsvm: a library for support vector machines, ACM Transactions on Intelligent Systems and Technology 2 (3) (2011) 27.

[48] J. Mansanet, A. Albiol, R. Paredes, Local deep neural networks for gender recognition, Pattern Recognition Letters 70 (15) (2016) 80–86.

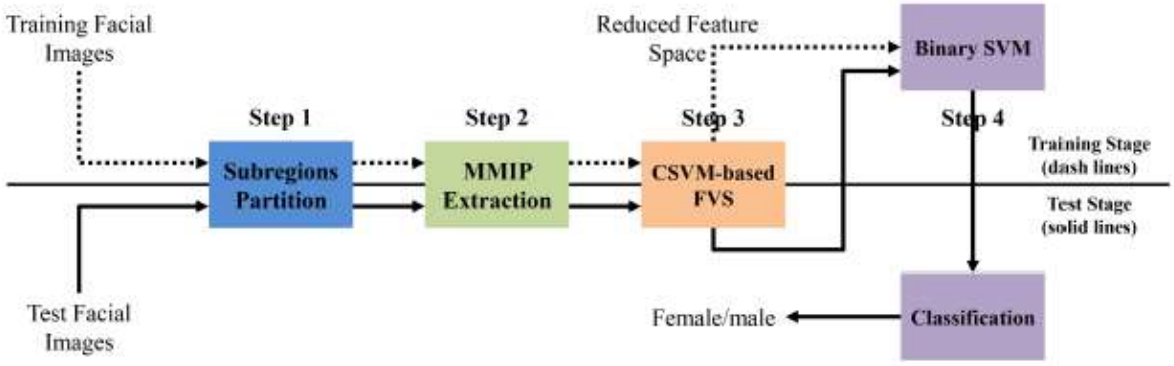

Fig. 1

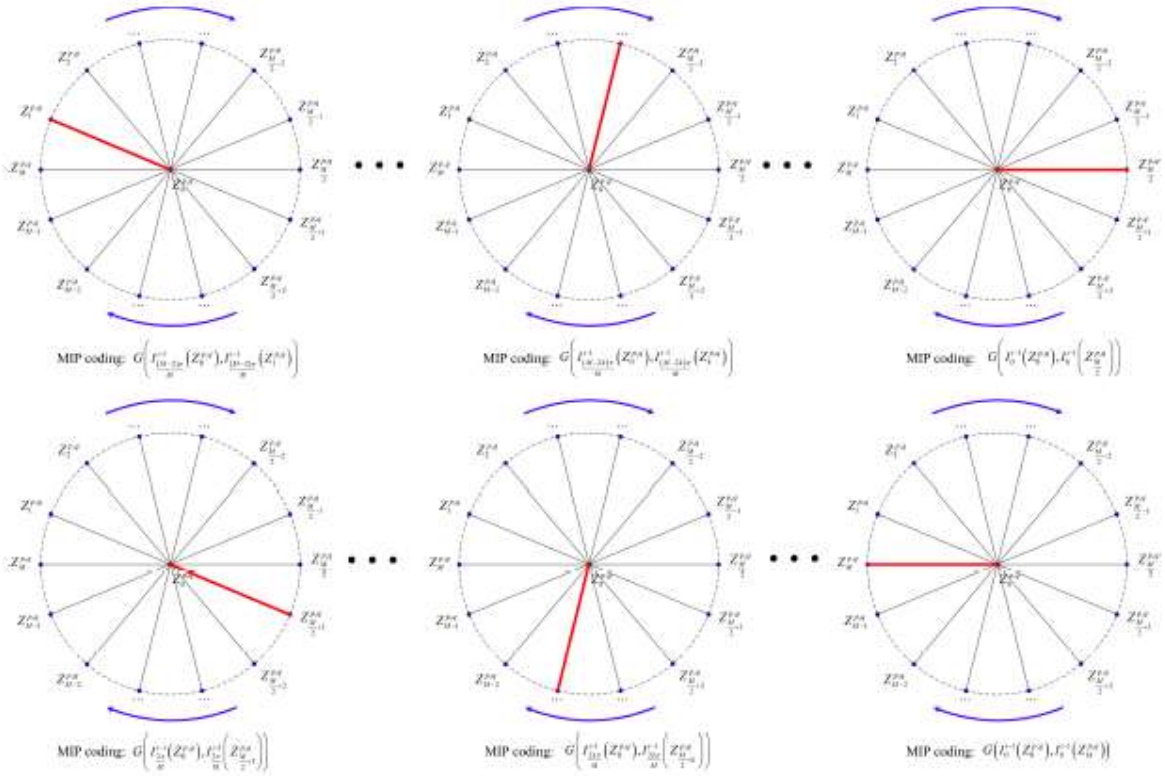

Fig. 2

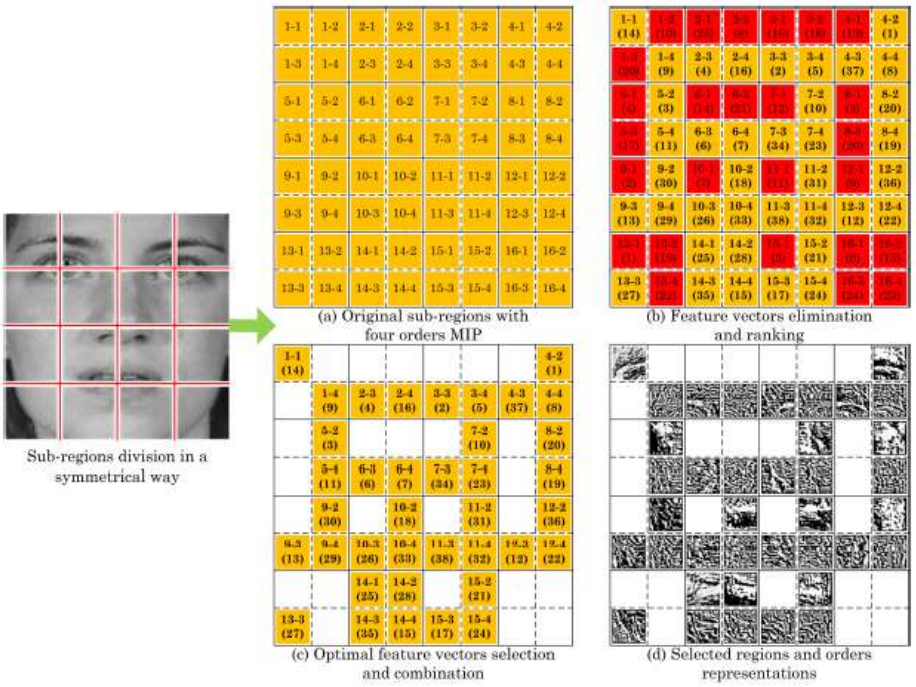

Fig. 3

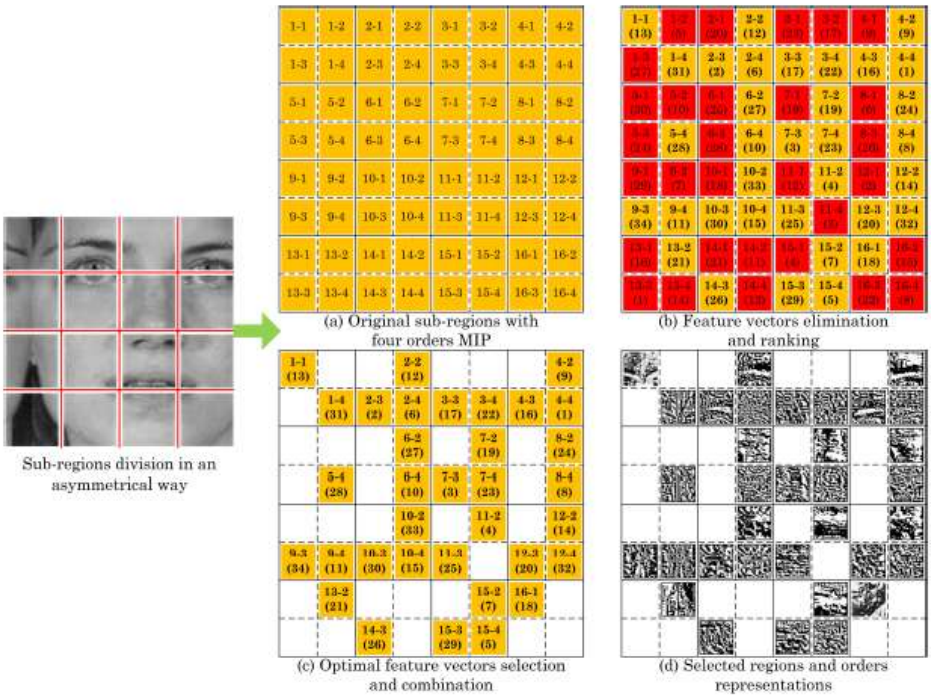

Fig. 4

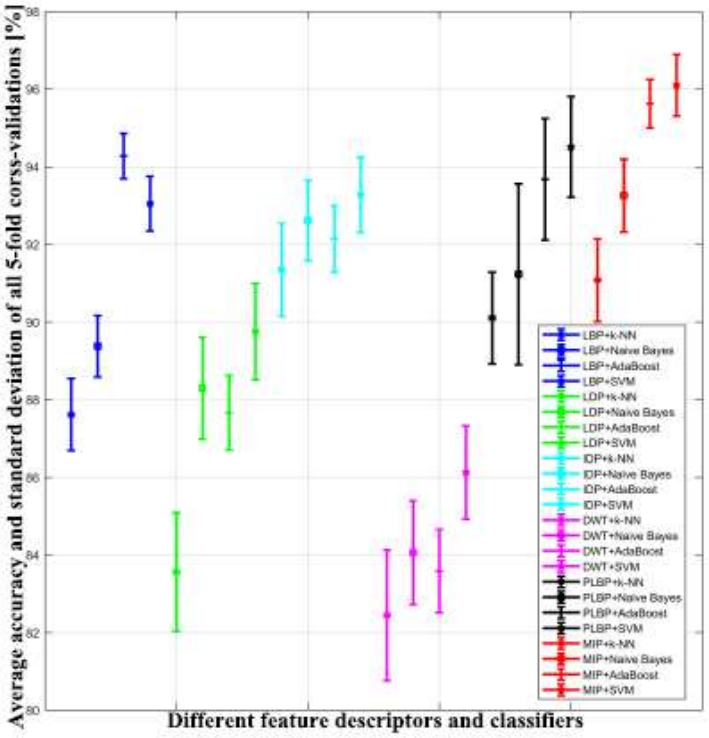

Fig. 5

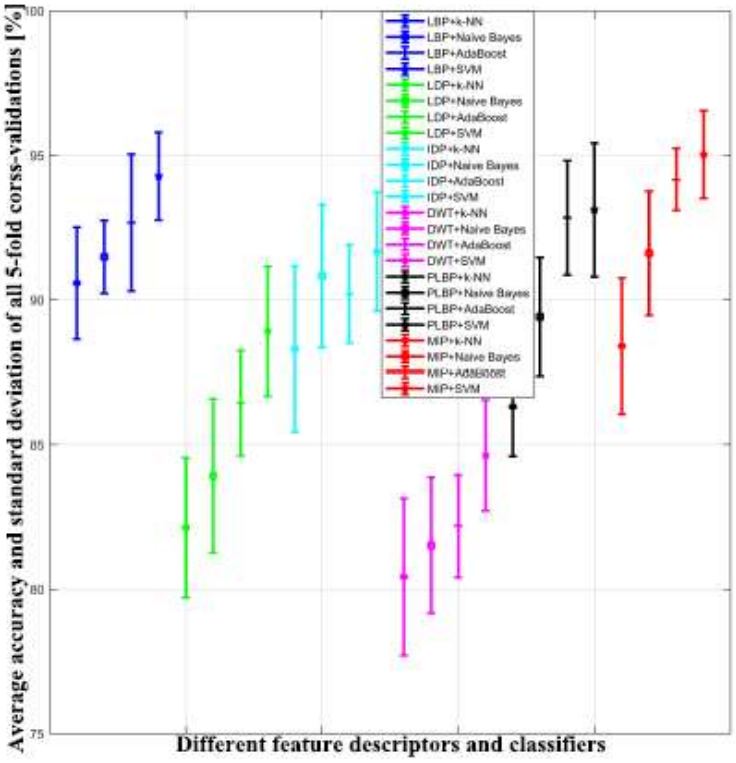

Fig. 6

Highlights

- We propose a generic multi-spatial multi-order feature descriptor;
- We develop a chain-type SVM based feature vector selection method;
- We study the roles of discriminative facial subregions and pattern orders for facial gender classification;
- Empirical studies on four widely used datasets demonstrate the efficacy of the proposed method.