Article

Face Recognition via Compact Second-Order Image Gradient Orientations

He-Feng Yin [1,2,*] , Xiao-Jun Wu [1,2,*] , Cong Hu [1,2] and Xiaoning Song [1,2]

1 School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China; conghu@jiangnan.edu.cn (C.H.); x.song@jiangnan.edu.cn (X.S.)
2 Jiangsu Provincial Engineering Laboratory of Pattern Recognition and Computational Intelligence, Jiangnan University, Wuxi 214122, China
* Correspondence: yin_hefeng@jiangnan.edu.cn (H.-F.Y.); wu_xiaojun@jiangnan.edu.cn (X.-J.W.)

Abstract: Conventional subspace learning approaches based on image gradient orientations only employ first-order gradient information, which may ignore second-order or higher-order gradient information. Moreover, recent researches on the human vision system (HVS) have uncovered that the neural image is a landscape or a surface whose geometric properties can be captured through second-order gradient information. The second-order image gradient orientations (SOIGO) can mitigate the adverse effect of noise in face images. To reduce the redundancy of SOIGO, we propose compact SOIGO (CSOIGO) by applying linear complex principal component analysis (PCA) in SOIGO. To be more specific, the SOIGO of training data are firstly obtained. Then, linear complex PCA is applied to obtain features of reduced dimensionality. Combined with collaborative-representation-based classification (CRC) algorithm, the classification performance of CSOIGO is further enhanced. CSOIGO is evaluated under real-world disguise, synthesized occlusion, and mixed variations. Under the real disguise scenario, CSOIGO makes 2.67% and 1.09% improvement regarding accuracy when one and two neutral face images per subject are used as training samples, respectively. For the mixed variations, CSOIGO achieves a 0.86% improvement in terms of accuracy. These results indicate that the proposed method is superior to its competing approaches with few training samples, and even outperforms some prevailing deep-neural-network-based approaches.

Keywords: face recognition; second-order gradient; image gradient orientations; collaborative-representation-based classification

MSC: 68T10

1. Introduction

As one of the most active research topics, face recognition (FR) has aroused great attention in the domain of pattern recognition and computer vision. Considerable progress has been made during the past decades and many successful methods have been proposed. Nevertheless, complicated variations in face images (e.g., occlusion, illumination, and expression) bring a great challenge for FR systems. To increase the robustness to occlusion, researchers have developed a variety of approaches. Sparse representation-based classification (SRC) [1] was developed for FR and shows robustness to occlusion and corruption in the test images when combined with the block partition technique. Naseem et al. [2] proposed a modular linear regression classification (Modular LRC) approach with a distance-based evidence fusion (DEF) algorithm to tackle the problem of contiguous occlusion. Dividing an image into different blocks is an effective way for feature extraction. Adjabi et al. [3] developed the multiblock color-binarized statistical image features (MB-C-BSIF) method for single-sample face recognition. Abdulhussain et al. [4] presented a method for fast calculation of features of overlapping image blocks. To further enhance

the performance of SRC, Li et al. [5] proposed a sparsity augmented weighted CRC approach for image recognition. Dong et al. [6] designed a low-rank Laplacian-uniform mixed (LR-LUM) model, which characterizes complex errors as a combination of continuous structured noises and random noises. Yang et al. [7] presented nuclear norm-based matrix regression (NMR), which employs two dimensional image-matrix-based error model rather than the one-dimensional pixel-based error model. The representation vector in NMR is imposed by the ℓ_2 norm, to make use of the discriminative property of sparsity, Chen et al. [8] proposed a sparse regularized NMR (SR-NMR) by replacing the ℓ_2 norm constraint on the representation vector with the ℓ_1 norm. However, the above approaches need uncorrupted training images. When providing corrupted training data, their performance will be deteriorated. To tackle the situation that both the training and test data are corrupted, low-rank matrix recovery (LRMR) can be applied. Chen et al. [9] proposed a discriminative low-rank representation (DLRR) method, which introduces the structural incoherence into the framework of low-rank representation (LRR) [10]. Gao et al. [11] proposed to learn robust and discriminative low-rank representation (RDLRR) by introducing low-rank constraint to simultaneously model the representation and each error term. Hu et al. [12] presented a robust FR method, which employs dual nuclear norm low-rank representation and a self-representation induced classifier. Yang et al. [13] developed a sparse low-rank component-based representation (SLCR) method for FR with low-quality images. Recently, Yang et al. [14] extended SLCR and proposed a FR technique named sparse individual low-rank component representation (SILR) for IoT-based systems. Inspired by LRR and deep learning techniques, Xia et al. [15] developed an embedded conformal deep low-rank autoencoder (ECLAE) neural network architecture for matrix recovery.

Recently, image gradient orientation (IGO) has attracted much attention due to its impressive results in occluded FR. Wu et al. [16] presented a gradient direction-based hierarchical adaptive sparse and low-rank (GD-HASLR) model, which performs in the image gradient direction domain rather than the image intensity domain. Li et al. [17] incorporated IGO into robust error coding and proposed an IGO-embedded structural error coding (IGO-SEC) model for FR with occlusion. Apart from the above two works, Zhang et al. [18] designed Gradientfaces for FR under varying illumination conditions. In essence, Gradientfaces is the IGO. Tzimiropoulos et al. [19] introduced the notion of subspace learning from IGO and developed approaches such as IGO-PCA and IGO-LDA. Vu [20] proposed a face representation approach called patterns of orientation difference (POD), which explores the relations of both gradient orientations and magnitudes. Zheng et al. [21] presented an online image alignment method via subspace learning from IGO. Qian et al. [22] presented a method called ID-NMR, in which the local gradient distribution is exploited to decompose the image into several gradient images. Wu et al. [23] proposed a new feature descriptor called the histogram of maximum gradient and edge orientation (HGEO) for the purpose of multispectral image matching.

The above IGO-based approaches only take the first-order gradient information into account, thus neglecting the second-order or higher-order gradient information. Latest researches on human vision have discovered that the neural image is a landscape or a surface whose geometric properties can be described by local curvatures of differential geometry through second-order gradient information [24,25]. Based on the second-order gradient, Huang et al. [24] presented a new local image descriptor called histograms of second-order gradient (HSOG). Li et al. [26] proposed a patterned fabric defect detection method based on the second-order, orientation-aware descriptor. Zhang et al. [27] designed a blind image quality assessment (IQA) method based on multiorder gradient statistics. Bastian et al. [28] developed a pedestrian detector utilizing both the first-order and the second-order gradient information in the image. Nevertheless, the above second-order-gradient-based approaches do not involve a dimensionality reduction technique, which results in redundant information. To alleviate this problem, we introduce PCA into the framework of SOIGO to extract more compact features. Moreover, we employ CRC as the final classifier due to its effectiveness and efficiency. Experimental results show that our

proposed method (CSOIGO) is robust to real disguise, synthesized occlusion, and mixed variations and is superior to some popular deep-neural-network-based approaches.

Our main contributions are outlined as follows:

1. We find that SOIGO is more robust to variations in face images compared with the first-order IGO. After extracting the SOIGO features of training samples, linear complex PCA is applied to reduce the redundancy of SOIGO.
2. The classic CRC algorithm is utilized to predict the identity of test samples, and it can further enhance the classification performance of CSOIGO.
3. Experiments on different scenarios demonstrate the efficacy and robustness of CSOIGO compared with other approaches.

The remainder of this paper is arranged as follows. Section 2 reviews some related work. In Section 3, we present our proposed approach. Section 4 conducts several experiments to demonstrate the efficacy of our proposed method. Finally, conclusions are drawn in Section 5.

2. Related Work

2.1. IGO-PCA

Given a set of images $\{\mathbf{Z}_i\}$ $(i = 1, 2, \ldots, N)$, where N denotes the number of training images and $\mathbf{Z}_i \in \mathbb{R}^{m \times n}$. Suppose that $\mathbf{I}(x, y)$ is the image intensities at pixel coordinates (x, y) of sample $\mathbf{Z}_i$, the horizontal and vertical gradient can be obtained by the following formulations:

$$\begin{aligned} \mathbf{G}_{i,x} &= h_x * \mathbf{I}(x, y) \\ \mathbf{G}_{i,y} &= h_y * \mathbf{I}(x, y), \end{aligned} \tag{1}$$

where $*$ expresses convolution, and h_x and h_y are filters employed to approximate the ideal differentiation operator along the image horizontal and vertical directions, respectively [29]. Image gradient contains edge information and is used to characterize the structure of an image. In [30], gradient feature map is extracted from the input image and exploited as a structural prior to guide the process of image reconstruction. However, the image data mostly distribute discretely in real-world scenarios; so, we usually use differences to compute the gradients, i.e., achieving the gradients through the difference between adjacent pixels' gray values. Thus, horizontal and vertical gradients can be reformulated as

$$\begin{aligned} \mathbf{G}_{i,x} &= \mathbf{I}(x + 1, y) - \mathbf{I}(x, y) \\ \mathbf{G}_{i,y} &= \mathbf{I}(x, y + 1) - \mathbf{I}(x, y). \end{aligned} \tag{2}$$

Then, the gradient orientation of the pixel location (x, y) is

$$\Phi_i(x, y) = \arctan \frac{\mathbf{G}_{i,y}}{\mathbf{G}_{i,x}}, i = 1, 2, ..., N. \tag{3}$$

For each image $\mathbf{Z}_i$ whose size is $m \times n$, we can obtain a corresponding gradient orientation matrix $\Phi_i \in [0, 2\pi)^{m \times n}$. Then, we can obtain the corresponding sample vectors by converting 2D images Φ_i into 1D vectors ϕ_i. Referring to [19], we also define the mapping from $[0, 2\pi)^K (K = m \times n)$ onto a subset of complex sphere with radius $\sqrt{K}$,

$$\boldsymbol{t}_i(\phi_i) = e^{j\phi_i}, \tag{4}$$

where $e^{j\phi_i} = [e^{j\phi_1}, e^{j\phi_2}, ..., e^{j\phi_K}]^T$ and $e^{j\theta}$ is Euler form, i.e., $e^{j\theta} = \cos\theta + j\sin\theta$. Then, we can apply complex linear PCA to the transformed t_i—that is, we seek for a set of $d < K$ orthonormal bases $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_d] \in \mathbb{C}^{K \times d}$ by solving the following problem:

$$\epsilon(\mathbf{U}) = \left\| \mathbf{X} - \mathbf{U}\mathbf{U}^H\mathbf{X} \right\|_F^2, \tag{5}$$

where $\mathbf{X} = [\boldsymbol{t}_1, \boldsymbol{t}_2, ..., \boldsymbol{t}_N] \in \mathbb{C}^{K \times N}$, $\mathbf{U}^H$ is the conjugate transpose of $\mathbf{U}$, and $\|.\|_F$ denotes the Frobenius norm. Equation (5) can be reformulated as

$$\mathbf{U}_o = \arg \max_{\mathbf{U}} tr(\mathbf{U}^H \mathbf{X} \mathbf{X}^H \mathbf{U}), \text{ s.t. } \mathbf{U}^H \mathbf{U} = \mathbf{I}. \tag{6}$$

The solution is given by the d eigenvectors of $\mathbf{X}\mathbf{X}^H$ corresponding to the d largest eigenvalues. Then, the d-dimensional embedding $\mathbf{Y} \in \mathbb{C}^{d \times N}$ of $\mathbf{X}$ is produced by $\mathbf{Y} = \mathbf{U}^H \mathbf{X}$.

2.2. Collaborative-Representation-Based Classification

During the past few years, the representation-based classification method (RBCM) has attracted lots of attention in the community of pattern recognition. The pioneering work is SRC [1]. In SRC, the ℓ_1 norm constraint is employed to attain the sparse coefficient of test data. Zhang et al. [31] argued that it is the collaborative representation mechanism rather than the ℓ_1 norm constraint that makes SRC successful for FR. Therefore, they developed the CRC method, which replaces the ℓ_1 norm constraint with the ℓ_2 norm. Afterwards, many improved methods were proposed to further boost the classification performance of CRC. Gou et al. [32] developed a class-specific mean vector-based weighted competitive and collaborative representation (CMWCCR) method, which fully employs the discrimination information in different ways. Motivated by the idea of linear representation, Gou et al. [33] proposed a representation coefficient-based k-nearest centroid neighbor (RCKNCN) method. Recently, Gou et al. [34] presented a hierarchical graph augmented deep collaborative dictionary learning (HGDCDL) model, which applies collaborative representation to the deepest-level representation learning. For simplicity, in this paper, we employ the original CRC as the classifier, and the objective function of CRC is formulated as follows:

$$\min_{\boldsymbol{\alpha}} \left\{ \|\boldsymbol{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_2^2 \right\}, \tag{7}$$

where $\boldsymbol{y}$ is the test sample, $\mathbf{D}$ is the dictionary that contains all the training data from C classes, and λ is a balancing parameter. Equation (7) has the following closed-form solution,

$$\boldsymbol{\alpha} = (\mathbf{D}^T \mathbf{D} + \lambda \mathbf{I})^{-1} \mathbf{D}^T \boldsymbol{y}. \tag{8}$$

In the classification stage, apart from the class-specific reconstruction error $\|\boldsymbol{y} - \mathbf{D}_j \boldsymbol{\alpha}_j\|_2$, $j = 1, 2, \ldots, C$, where $\boldsymbol{\alpha}_j$ is the coefficient vector corresponding to the jth class, Zhang et al. [31] found that $\|\boldsymbol{\alpha}_j\|_2$ also contains some discriminative information for classification. Thus, they presented the following regularized residuals for classification,

$$\text{identity}(\boldsymbol{y}) = \arg \min_{j} \frac{\|\boldsymbol{y} - \mathbf{D}_j \boldsymbol{\alpha}_j\|_2}{\|\boldsymbol{\alpha}_j\|_2}. \tag{9}$$

3. Proposed Method

Previous studies revealed that gradient information at different orders characterize different structural features of natural scenes. The first-order gradient information is related to the slope and elasticity of a surface, while the second-order gradient delivers the curvature-related geometric properties. Figure 1 depicts two images and their corresponding landscapes plotted as surfaces; one can see that these landscapes contain a variety of local shapes, such as cliffs, ridges, summits, valleys, and basins. Inspired by the above results, we propose a new FR method that exploits the SOIGO. The second-order gradient is obtained based on the first-order gradient information defined in Equation (2),

$$\begin{aligned} \mathbf{G}_{i,x}^2 &= \mathbf{G}_{i,x}(x+1, y) - \mathbf{G}_{i,x}(x, y) \\ \mathbf{G}_{i,y}^2 &= \mathbf{G}_{i,y}(x, y+1) - \mathbf{G}_{i,y}(x, y), \end{aligned} \tag{10}$$

where $\mathbf{G}_{i,x}^2$ and $\mathbf{G}_{i,y}^2$ are the second-order gradient along the horizontal and vertical directions, respectively. Therefore, the SOIGO is computed as follows:

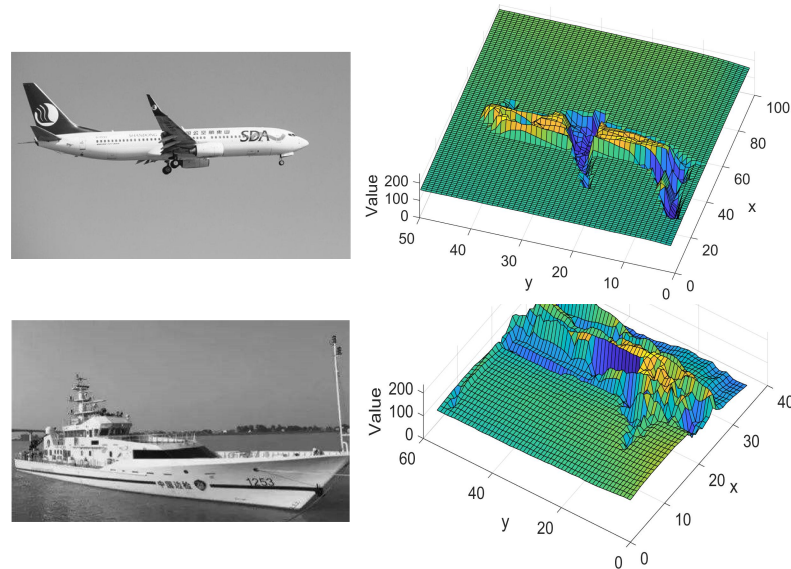$$\Phi_i^2(x,y) = \arctan\frac{\mathbf{G}_{i,y}^2}{\mathbf{G}_{i,x}^2}. \tag{11}$$

Figure 1. Original images (**left part**) and their surface plots (**right part**).

Figure 2 presents an original face image and its gradient orientations of the first and second orders; one can see that, compared with the first-order IGO, the SOIGO significantly depresses the noise in the orientation domain. Moreover, the SOIGO contains more fine information than the first-order IGO, e.g., areas around the eyes, nose, and mouth.
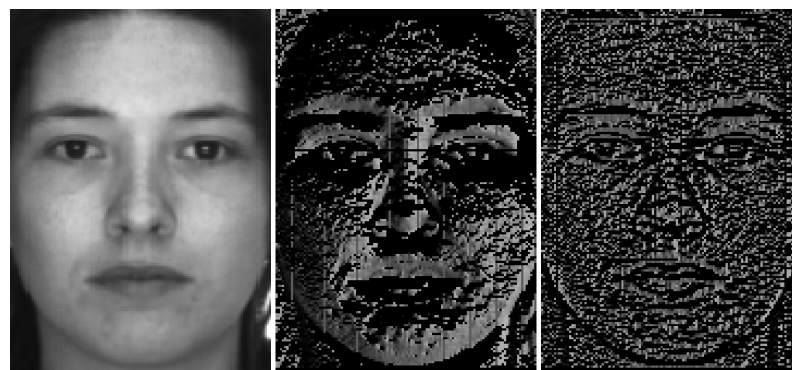

Figure 2. Original face image and its gradient orientations of the first and second orders, respectively.

To further illustrate the effectiveness of using the SOIGO, we visualize the original data, the first-order IGO, and the SOIGO on the AR database by employing the t-SNE algorithm [35] in Figure 3. These data are selected from the first ten subjects on the AR database; for each person, seven nonoccluded face images in Session 1 are used. Then, these images are occluded by a square baboon image with a percentage of 30%. For detailed experimental settings, please refer to Section 4.3. As can be seen from Figure 3, though the first-order IGO looks better compared with the original data, clusters of different classes are mixed together. In Figure 3c, the cluster of the same class is more compact than that of Figure 3b, which is beneficial for subsequent classification.
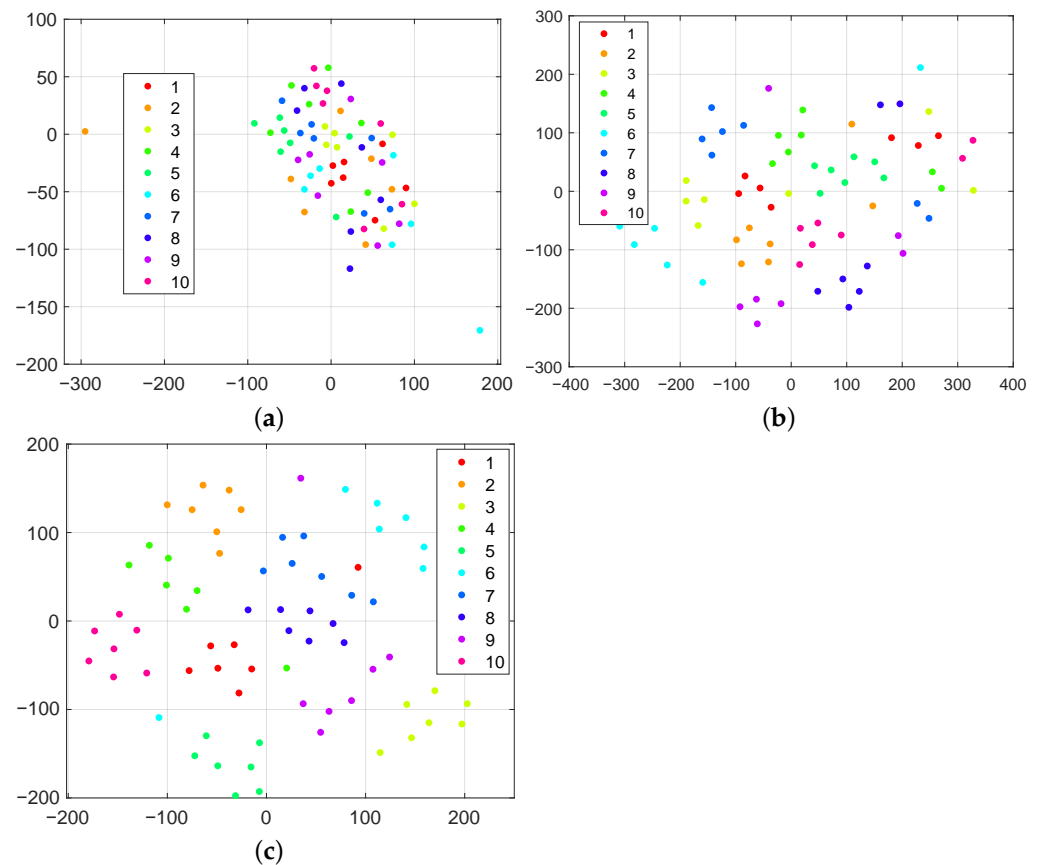
Figure 3. t-SNE visualization of (**a**) original data, (**b**) the first-order IGO with the mapping defined in Equation (4), and (**c**) the SOIGO with the mapping defined in Equation (4); each color represents a class. For better visualization, please refer to the electronic version of this paper.

The procedures of obtaining the projection matrix $\mathbf{U}$ is the same as in IGO-PCA. Then, for a test image $\mathbf{Z}_t$, we first compute its SOIGO and obtain t after the mapping defined by Equation (4). Embeddings of training and test images are derived as follows:

$$\mathbf{Y} = \mathbf{U}^H \mathbf{X}, \ z = \mathbf{U}^H t, \tag{12}$$

where $\mathbf{Y} \in \mathbb{C}^{d \times N}$ and $z \in \mathbb{C}^{d \times 1}$. To make the embeddings of training and test images suitable for CRC, we employ both the real and imaginary parts of $\mathbf{Y}$ and z as the input of CRC; let

$$\mathbf{D} = \begin{bmatrix} \mathrm{real}(\mathbf{Y}) \\ \mathrm{imag}(\mathbf{Y}) \end{bmatrix}, \ y = \begin{bmatrix} \mathrm{real}(z) \\ \mathrm{imag}(z), \end{bmatrix} \tag{13}$$

where $\mathrm{real}(\cdot)$ and $\mathrm{imag}(\cdot)$ are the real part and imaginary part of complex number, respectively. Then, we compute the representation coefficient vector of y over $\mathbf{D}$; this is followed by checking which class results in the least regularized residual. The pipeline of our proposed CSOIGO is illustrated in Figure 4, and the complete process of CSOIGO is outlined in Algorithm 1.

When assessing the performance of an algorithm, we should take its computational complexity into account. The major consumption of CSOIGO lies in the linear complex PCA and CRC, and they both involve the operation of matrix. It takes $\mathcal{O}(K^2 N)$ to compute the covariance matrix and $\mathcal{O}(K^3)$ for eigen-decomposition in the process of PCA, where $K = m \times n$ and N denote the dimensionality and total number of training images. From Equation (8), one can see that CRC contains matrix multiplication and matrix inversion, and it takes $\mathcal{O}(N^2 d)$ to compute $\mathbf{D}^T \mathbf{D}$ and $\mathcal{O}(N^3)$ for the inverse operation of matrix, where d is the reduced dimensionality. Suppose there are p test samples, CRC takes

$\mathcal{O}(N^2 d + N^3 + Ndp)$ to completely classify them. Therefore, the total computational complexity of CSOIGO is $\mathcal{O}(K^2 N + K^3 + N^2 d + N^3 + Ndp)$.
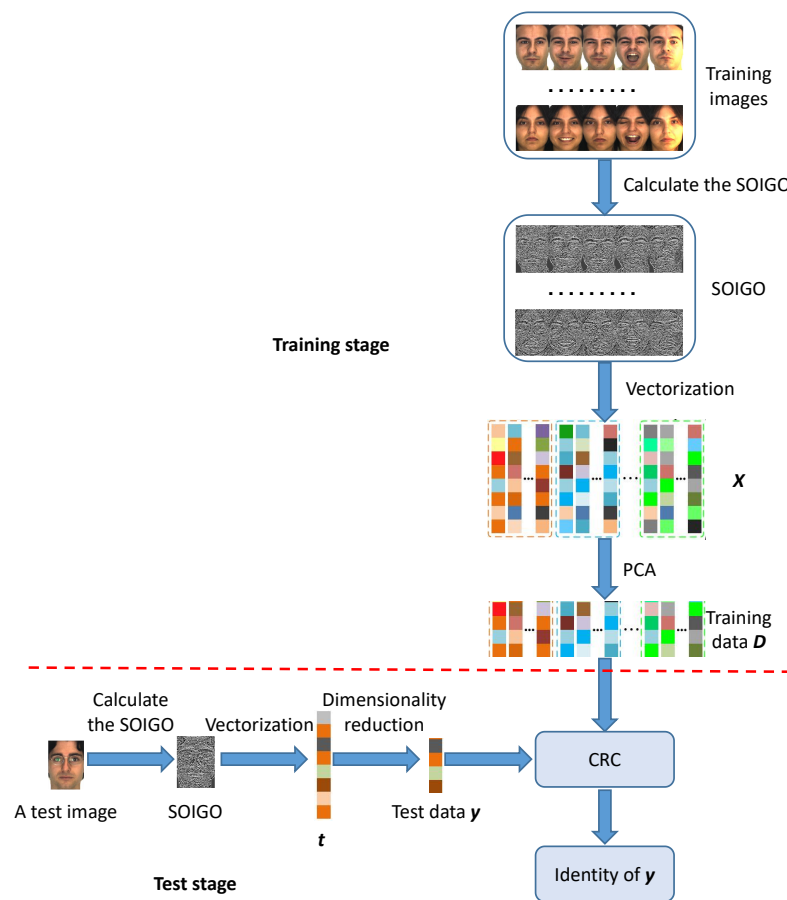

Figure 4. The pipeline of our proposed CSOIGO.

Algorithm 1 CSOIGO

Input: A set of N training images $\{\mathbf{Z}_i\}(i = 1, 2, \ldots, N)$ from C classes, test image $\mathbf{Z}_t$, the number of principal components d, and the regularization parameter λ for CRC.

    1. Obtain the SOIGO Φ_i^2 of training images and convert it to 1D vector ϕ_i^2.

    2. Compute $\mathbf{t}_i(\phi_i^2) = e^{j\phi_i^2}$; all the SOIGO of training images form the matrix $\mathbf{X} = [\mathbf{t}_1, \mathbf{t}_2, ..., \mathbf{t}_N]$.

    3. Obtain the projection matrix $\mathbf{U}$ via Equation (6).

    4. For the test image $\mathbf{Z}_t$, obtain its SOIGO Φ_t^2 and convert it to 1D vector ϕ_t^2; then, compute $\mathbf{t} = e^{j\phi_i^2}$.

    5. Obtain the embeddings of training and test images via Equation (12).

    6. Obtain $\mathbf{D}$ and $\mathbf{y}$ by Equation (13).

    7. Code $\mathbf{y}$ over $\mathbf{D}$ by Equation (8).

    8. Compute the regularized residuals $r_j = \frac{\|\mathbf{y} - \mathbf{D}_j \boldsymbol{\alpha}_j\|_2}{\|\boldsymbol{\alpha}_j\|_2}, j = 1, 2, \ldots, C.$

Output: $\text{identity}(\mathbf{Z}_t) = \arg \min_j r_j$.

---

4. Experimental Results and Analysis

In this section, experiments are conducted under different scenarios to validate the effectiveness of the proposed method. For reproduction, the source code of CSOIGO is available at https://github.com/yinhefeng/SOIGO.

4.1. Recognition with Real Disguise

The AR database contains over 4000 images of 126 subjects. For each individual, 26 images are taken in two separate sessions. There are 13 images for each session, in which three images with sunglasses, another three with scarves, and the remaining seven have different illumination and expression changes; the 13 images of one subject from Session 1 are shown in Figure 5. Each image is 165×120 pixels. For fair comparison, we use the same subset as in [16], which consists of 50 men and 50 women, and all images are resized to 42×30 pixels. The neutral face image of each subject is used as training data, and the sunglasses/scarf occluded images in each session for testing. The proposed method is compared with other state-of-the-art approaches, including HQPAMI [36], NR [37], ProCRC [38], F-LR-IRNNLS [39], EGSNR [40], LDMR [41], and GD-HASLR [16]. To better illustrate the superiority of CSOIGO, we also present the results of IGO-PCA-NNC [19], IGO-PCA-CRC, and SOIGO-PCA-NNC. Table 1 summarizes the experimental results; one can see that CSOIGO achieves the highest recognition accuracy under all cases except for the sunglasses scenario of session 1. Since the test images are partially occluded by sunglasses or scarf, HQPAMI, NR, ProCRC, and LDMR seem not very robust to contiguous occlusion. Due to the preprocessing step that separates outlier pixels and corruptions from the training samples, the overall classification accuracy of F-LR-IRNNLS is higher than that of EGSNR. IGO-PCA-CRC ranks second over all methods and achieves 5.66% higher accuracy than IGO-PCA-NNC, which validates the efficacy of CRC when coping with IGO features. GD-HASLR has competitive performance with SOIGO-PCA-NNC. However, the overall accuracy gain of CSOIGO over GD-HASLR and IGO-PCA-CRC is 4.5% and 2.67%, respectively. The above experimental results indicate that our proposed CSOIGO is robust to real disguise even when a single training sample per person is available.

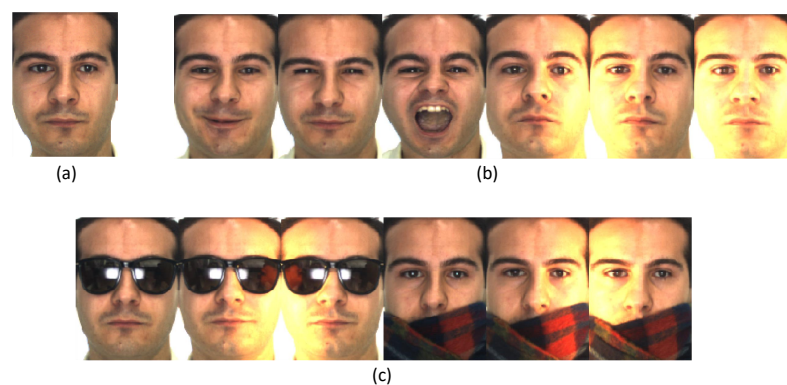(a)　　　　　　　　　　　　　　(b)

(c)

Figure 5. Some example face images from the AR database: (**a**) the neutral image of a subject from Session 1; (**b**) face images with illumination and expression variations; (**c**) images occluded by sunglasses/scarf.

Table 1. Recognition accuracy (%) of competing approaches on a subset of the AR database (test samples contain sunglasses occlusion or scarf occlusion) when only one neutral face image per subject from Session 1 is used as training sample. The dimension that leads to the best result for IGO- and SOIGO-based approaches is given in parentheses.

Methods	Sunglasses		Scarf		Overall
	Session 1	Session 2	Session 1	Session 2	
HQPAMI [36]	56.67	38.00	38.00	22.33	38.75
NR [37]	28.33	16.67	29.67	17.33	23.00
ProCRC [38]	53.07	31.00	18.67	7.33	27.52
F-LR-IRNNLS [39]	88.67	60.33	67.00	49.67	66.42
EGSNR [40]	84.00	54.00	70.33	48.33	64.16
LDMR [41]	68.33	45.67	59.67	34.00	51.92
GD-HASLR [16]	92.00	66.67	82.67	58.67	75.00
IGO-PCA-NNC [19]	89.00 (99)	69.00 (99)	73.33 (97)	53.33 (96)	71.17
IGO-PCA-CRC	**93.00 (85)**	74.33 (92)	81.67 (88)	58.33 (95)	76.83
SOIGO-PCA-NNC	88.67 (92)	73.33 (96)	80.33 (99)	61.00 (88)	75.83
CSOIGO	92.67 (89)	**76.67 (93)**	**83.33 (75)**	**65.33 (99)**	**79.50**

Bold values indicate the best recognition accuracy.

Next, we utilize two neutral face images per subject from Sessions 1 and 2 for training, and the test sets are identical with the first experiment. The results are reported in Table 2. As can be seen from Table 2, CSOIGO yields the best overall recognition accuracy and outperforms GD-HASLR by 2.92%. Again, IGO-PCA-CRC ranks second in all methods. SOIGO-PCA-NNC outperforms IGO-PCA-NNC, and CSOIGO achieves higher accuracy than IGO-PCA-CRC, which indicates that SOIGO is more robust to occlusion than IGO.

Table 2. Recognition accuracy (%) of competing approaches on a subset of the AR database (test samples contain sunglasses occlusion or scarf occlusion) when two neutral face images (from Sessions 1 and 2) per subject are used as training samples, the dimension that leads to the best result for IGO- and SOIGO-based approaches is given in parentheses.

Methods	Sunglasses		Scarf		Overall
	Session 1	Session 2	Session 1	Session 2	
HQPAMI [36]	61.33	59.33	44.67	48.00	53.33
NR [37]	34.00	33.33	33.00	35.67	34.00
ProCRC [38]	53.00	54.67	18.00	17.67	35.84
F-LR-IRNNLS [39]	90.33	87.67	78.67	76.00	83.17
EGSNR [40]	88.00	89.33	80.00	73.00	82.58
LDMR [41]	71.00	63.67	64.00	61.00	64.92
GD-HASLR [16]	93.00	93.33	82.67	84.00	88.25
IGO-PCA-NNC [19]	93.00 (182)	91.67 (191)	78.00 (199)	74.00 (193)	84.17
IGO-PCA-CRC	96.00 (128)	95.33 (116)	85.00 (190)	84.00 (160)	90.08
SOIGO-PCA-NNC	96.33 (187)	92.67 (197)	**86.33 (166)**	83.67 (189)	89.75
CSOIGO	**97.33 (144)**	**95.67 (124)**	86.00 (119)	**85.67 (198)**	**91.17**

Bold values indicate the best recognition accuracy.

4.2. Comparison with CNN-Based Approaches

In this subsection, we compare our proposed method with prevailing deep-learning-based approaches. The first one is VGGFace [42], which is based on the VGGNet [43] and has 16 convolutional layers, five max-pooling layers, three fully-connected layers, and a final linear layer with softmax layer. In our experiments, we employ FC6 and FC7 for feature extraction. The second one is Lightened CNN [44], which has a low computational complexity. Lightened CNN consists of two different models, i.e., Model A and Model B. Model A is based on the AlexNet [45], which contains four convolution layers using the max feature map (MFM) activation functions, four max-pooling layers, two fully-connected layers, and a linear layer with softmax activation in the output. Model B is based on

the Network in Network model [46] and consists of five convolution layers using the MFM activation functions, four convolutional layers for dimensionality reduction, five max-pooling layers, two fully-connected layers, and a linear layer with softmax activation in the output. For Lightened CNN, FC1 is used for feature extraction. All the features extracted by VGGFace and Lightened CNN are classified using the nearest neighbor classifier with cosine distance. When training VGGFace, the size of input image is 224×224, and the preprocessing operation involves subtracting the mean RGB value, computed on the training set, from each pixel. The batch size, number of epochs, and optimizer are 256, 74, and *sgdm*, respectively. The learning rate is initially set to 1×10^{-2} and then decreased by a factor of 10. For training Lightened CNN, the size of input image is 144×144, and the input image is cropped into 128×128 and mirrored. The batch size, number of epochs, and optimizer are 20, 150, and *rmsprop*, respectively. The learning rate is set to 1×10^{-3} initially and reduced to 5×10^{-5} gradually.

As in Section 4.1, the first experiment is one neutral face of each subject for training on the AR database, and the experimental results are summarized in Table 3. Table 4 lists the results when two neutral faces are used for training. From Tables 3 and 4, we can see that VGGFace performs better in the scarf scenario than in the sunglasses scenario. This indicates that VGGFace has difficulty tackling the upper face occlusion, and this phenomenon is also observed in [47]. Moreover, when using more training samples, the performance of VGGFace does not improve. Hence, to increase robustness to upper face occlusion, VGGFace may need much more training data. By comparison, our proposed CSOIGO can achieve better results even with few training samples. In practical applications, training data may be insufficient. In this situation, CSOIGO is more appropriate to realize robust face recognition than VGGFace.

Similar to the results of VGGFace, Lightened CNN performs worse in the sunglasses scenario than in the scarf scenario. Additionally, Model A outperforms Model B, and Model A also achieves higher accuracy than VGGFace. However, whether one or two neutral face images per subject are used for training, our proposed CSOIGO achieves the best overall recognition accuracy.

Table 3. Comparison with CNN-based approaches on a subset of the AR database (test samples contain sunglasses occlusion or scarf occlusion) when only one neutral face image per subject from Session 1 is used as training samples. The dimension that leads to the best result for IGO- and SOIGO-based approaches is given in parentheses.

Methods	Sunglasses		Scarf		Overall
	Session 1	Session 2	Session 1	Session 2	
VGGFace FC6 [42]	54.00	45.00	**91.67**	**88.00**	69.67
VGGFace FC7 [42]	45.67	40.00	88.67	84.00	64.59
Lightened CNN (A) [44]	67.33	56.00	87.00	82.33	73.17
Lightened CNN (B) [44]	36.33	31.33	80.67	73.67	55.50
GD-HASLR [16]	92.00	66.67	82.67	58.67	75.00
IGO-PCA-NNC [19]	89.00 (99)	69.00 (99)	73.33 (97)	53.33 (96)	71.17
IGO-PCA-CRC	**93.00 (85)**	74.33 (92)	81.67 (88)	58.33 (95)	76.83
SOIGO-PCA-NNC	88.67 (92)	73.33 (96)	80.33 (99)	61.00 (88)	75.83
CSOIGO	92.67 (89)	**76.67 (93)**	83.33 (75)	65.33 (99)	**79.50**

Bold values indicate the best recognition accuracy.

Table 4. Comparison with CNN-based approaches on a subset of the AR database (test samples contain sunglasses occlusion or scarf occlusion) when two neutral face images (from Sessions 1 and 2) per subject are used as training samples. The dimension that leads to the best result for IGO- and SOIGO-based approaches is given in parentheses.

Methods	Sunglasses		Scarf		Overall
	Session 1	Session 2	Session 1	Session 2	
VGGFace FC6 [42]	44.67	51.00	**91.67**	**93.33**	70.17
VGGFace FC7 [42]	41.67	44.67	88.67	89.33	66.08
Lightened CNN (A) [44]	64.67	58.33	86.67	85.33	73.75
Lightened CNN (B) [44]	38.67	38.00	81.67	79.33	59.42
GD-HASLR [16]	93.00	93.33	82.67	84.00	88.25
IGO-PCA-NNC [19]	93.00 (182)	91.67 (191)	78.00 (199)	74.00 (193)	84.17
IGO-PCA-CRC	96.00 (128)	95.33 (116)	85.00 (190)	84.00 (160)	90.08
SOIGO-PCA-NNC	96.33 (187)	92.67 (197)	86.33 (166)	83.67 (189)	89.75
CSOIGO	**97.33 (144)**	**95.67 (124)**	86.00 (119)	85.67 (198)	**91.17**

Bold values indicate the best recognition accuracy.

4.3. Random Block Occlusion

Here, we conduct other experiments using synthesized occluded face data as testing data. For each subject, seven nonoccluded face images in the AR dataset in Session 1 are used for training and the other seven nonoccluded images in Session 2 for testing, the image size is 42×30 pixels. Block occlusion is tested by placing the square baboon image on each test image. The location of the occlusion is randomly chosen and is unknown during training. We consider different sizes of the object such that the face is covered with the occluded object from 30% to 50% of its area; some occluded face images are shown in Figure 6. The above experimental results indicate that GD-HASLR is superior to other competing approaches; therefore, in this subsection and the following subsection, we report the result of GD-HASLR for comparison. Recognition results for different levels of occlusion are shown in Table 5. One can see that CSOIGO outperforms GD-HASLR by a large margin, and the performance gain is significant with the increasing percentage of occlusion. Moreover, SOIGO-PCA-NNC outperforms IGO-PCA-NNC and CSOIGO performs better than IGO-PCA-CRC, which demonstrates that SOIGO is more robust than IGO when dealing with artificial occlusion.
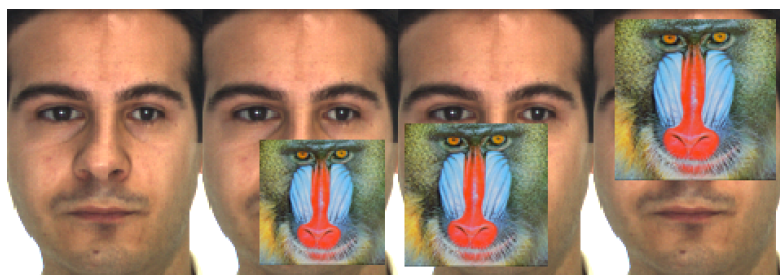

Figure 6. Original face image and its occluded images with different occlusion percentages; from the second to the last, the percentage is 30%, 40%, and 50%, respectively.

Table 5. Recognition accuracy (%) of competing methods under different percentages of occlusion on a subset of the AR database (original training and test samples have no sunglasses occlusion or scarf occlusion). The dimension that leads to the best result for IGO- and SOIGO-based approaches is given in parentheses.

Occlusion Percentage	30%	40%	50%
GD-HASLR [16]	81.29	71.14	56.14
IGO-PCA-NNC [19]	86.14 (588)	80.57 (606)	66.29 (321)
IGO-PCA-CRC	89.14 (205)	80.14 (185)	71.29 (569)
SOIGO-PCA-NNC	88.86 (458)	84.57 (575)	73.29 (693)
CSOIGO	**93.57 (423)**	**87.00 (533)**	**76.57 (698)**

Bold values indicate the best recognition accuracy.

To vividly show the performance of IGO- and SOIGO-based approaches under different numbers of features, in Figure 7 we plot the recognition accuracy against the number of features when the percentage of occlusion is 30%. We can clearly see that with the increasing number of features, CSOIGO consistently outperforms the other three competing approaches.
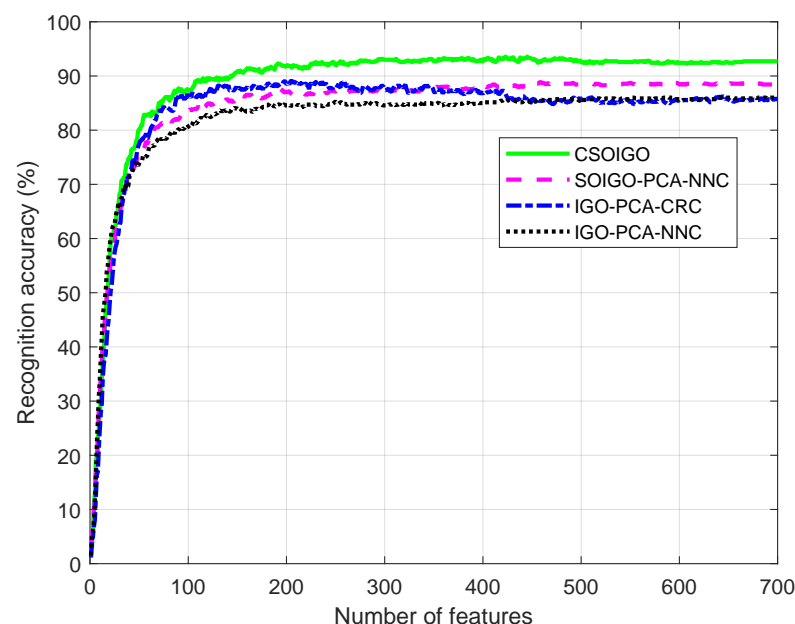

Figure 7. Recognition accuracy versus different numbers of features when the percentage of occlusion is 30%.

4.4. Recognition with Mixed Variations

In this subsection, we evaluate our proposed CSOIGO and other compared approaches under the mixed variations. As shown in Figure 5a,b, the first seven images per subject in Session 1 have variations of expression and illumination; thus, seven nonoccluded images from Session 1 of the AR database are selected for training and another seven undisguised images from Session 2 are used for testing. Recognition accuracy and testing time of compared methods are shown in Table 6. It should be noted that the testing time refers to the time that classifies all the test samples. All experiments are performed on a laptop with Windows 10, an Intel Core i9-8950HK CPU at 2.90 GHz, and 32.00 GB RAM. The implementation software is MATLAB R2022a. From Table 6, we can see that CSOIGO has the best classification performance. Specifically, it makes 1.86% and 0.86% improvement in terms of accuracy over GD-HASLR and IGO-PCA-CRC, respectively. Due to the complex optimization process, GD-HASLR consumes much more time than the other approaches. The testing time is almost the same for both IGO-PCA-NNC and SOIGO-PCA-NNC. NNC

is a simple and efficient classifier, while CRC involves the computations of coefficient vector and classwise residual. As a result, CSOIGO takes a little longer than SOIGO-PCA-NNC. However, CSOIGO is much faster than GD-HASLR.

Table 6. Recognition accuracy (%) and testing time (s) of compared approaches with mixed variations on a subset of the AR database (training and test samples have expression and illumination changes). The dimension that leads to the best result for IGO- and SOIGO-based approaches is given in parentheses.

Methods	Accuracy (%)	Testing Time (s)
GD-HASLR [16]	96.71	414.29
IGO-PCA-NNC [19]	93.14 (478)	0.50
IGO-PCA-CRC	97.71 (100)	1.92
SOIGO-PCA-NNC	94.71 (371)	0.45
CSOIGO	**98.57 (171)**	2.43

Bold values indicate the best recognition accuracy.

As in the previous subsection, we show the recognition accuracy against the number of features in Figure 8. It can be seen that as the number of features increases, the recognition accuracies of IGO-PCA-NNC, SOIGO-PCA-NNC, and CSOIGO also increase. The recognition accuracy of IGO-PCA-CRC firstly increases, then decreases to some extent, and then it increases again. When the number of features exceeds 108, CSOIGO always achieves higher accuracy than its competing methods. This again demonstrates that CSOIGO is robust to mixed variations in face images.
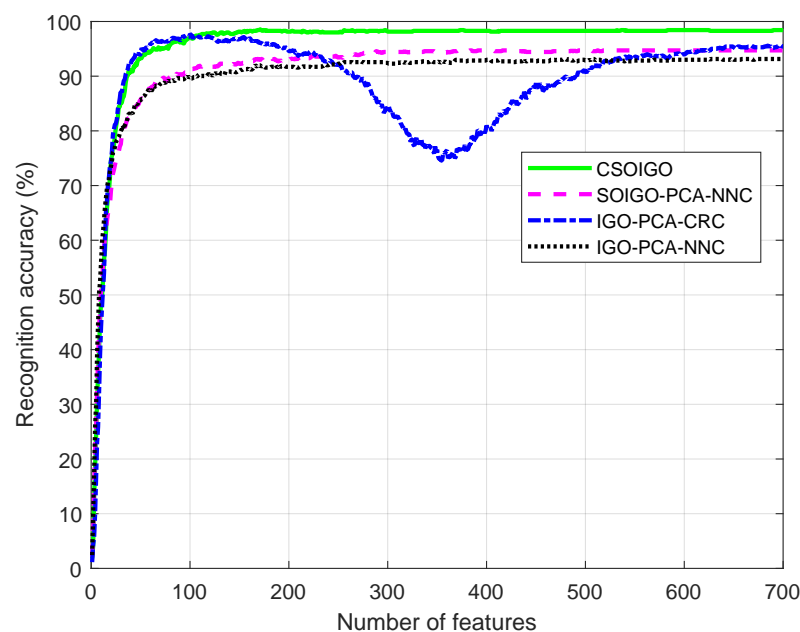

Figure 8. Recognition accuracy versus different number of features under mixed variations.

5. Conclusions

In this paper, we present a new method for occluded face recognition, namely, CSOIGO, by exploiting the second-order gradient information. SOIGO is robust to real disguise, synthesized occlusion, and mixed variations. By employing CRC as the final classifier, our proposed method achieves impressive results in various scenarios and even outperforms some deep-neural-network-based approaches. Taking the real disguise experiment as an example, when one and two neutral face images per subject are used as training samples, CSOIGO attains an overall accuracy of 79.50% and 91.17%, respectively. Therefore, our proposed CSOIGO is superior to its competing approaches.

The limitation of CSOIGO is that it needs registered images for training and testing, i.e., when classifying face images with pose changes, its recognition performance will be degraded. Consequently, CSOIGO can be applied to applications of access control, automatic teller machines, or other security facilities. In these circumstances, we can obtain controlled training images in advance and the test images will be collected under similar scenarios. However, if registered face images cannot be collected during either the training or test stage, one can employ image registration methods to remedy the above limitation to some extent.

In future work, we will introduce SOIGO into other popular subspace learning approaches, e.g., linear discriminant analysis (LDA), to extract more discriminative features. Moreover, other variants of CRC will also be investigated to further enhance the performance of recognition.

Author Contributions: Conceptualization, H.-F.Y.; methodology, H.-F.Y. and X.-J.W.; software, H.-F.Y.; validation, H.-F.Y., X.-J.W., C.H. and X.S.; formal analysis, H.-F.Y. and X.-J.W.; investigation, H.-F.Y. and X.-J.W.; resources, H.-F.Y.; data curation, H.-F.Y., X.-J.W. and X.S.; writing—original draft preparation, H.-F.Y.; writing—review and editing, X.-J.W., C.H. and X.S.; visualization, H.-F.Y.; supervision, X.-J.W.; project administration, X.-J.W. and X.S.; funding acquisition, X.-J.W., C.H. and X.S. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 210–227.
2. Naseem, I.; Togneri, R.; Bennamoun, M. Linear regression for face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 2106–2112.
3. Adjabi, I.; Ouahabi, A.; Benzaoui, A.; Jacques, S. Multi-block color-binarized statistical images for single-sample face recognition. *Sensors* **2021**, *21*, 728.
4. Abdulhussain, S.H.; Mahmmod, B.M.; Flusser, J.; AL-Utaibi, K.A.; Sait, S.M. Fast Overlapping Block Processing Algorithm for Feature Extraction. *Symmetry* **2022**, *14*, 715.
5. Li, Z.Q.; Sun, J.; Wu, X.J.; Yin, H. Sparsity augmented weighted collaborative representation for image classification. *J. Electron. Imaging* **2019**, *28*, 053032.
6. Dong, J.; Zheng, H.; Lian, L. Low-rank laplacian-uniform mixed model for robust face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 11897–11906.
7. Yang, J.; Luo, L.; Qian, J.; Tai, Y.; Zhang, F.; Xu, Y. Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 156–171.
8. Chen, Z.; Wu, X.J.; Kittler, J. A sparse regularized nuclear norm based matrix regression for face recognition with contiguous occlusion. *Pattern Recognit. Lett.* **2019**, *125*, 494–499.
9. Chen, J.; Yi, Z. Sparse representation for face recognition by discriminative low-rank matrix recovery. *J. Vis. Commun. Image Represent.* **2014**, *25*, 763–773.
10. Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; Ma, Y. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *35*, 171–184.
11. Gao, G.; Yang, J.; Jing, X.Y.; Shen, F.; Yang, W.; Yue, D. Learning robust and discriminative low-rank representations for face recognition with occlusion. *Pattern Recognit.* **2017**, *66*, 129–143.

12. Hu, Z.; Gao, G.; Gao, H.; Wu, S.; Zhu, D.; Yue, D. Robust Face Recognition Via Dual Nuclear Norm Low-rank Representation and Self-representation Induced Classifier. In Proceedings of the 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), Nanjing, China, 23–25 November 2018; pp. 920–924.

13. Yang, S.; Zhang, L.; He, L.; Wen, Y. Sparse low-rank component-based representation for face recognition with low-quality images. *IEEE Trans. Inf. Forensics Secur.* **2018**, *14*, 251–261.

14. Yang, S.; Wen, Y.; He, L.; Zhou, M.; Abusorrah, A. Sparse Individual Low-Rank Component Representation for Face Recognition in the IoT-Based System. *IEEE Internet Things J.* **2021**, *8*, 17320–17332.

15. Xia, H.; Feng, G.; Cai, J.x.; Tang, X.; Chi, H. Embedded conformal deep low-rank auto-encoder network for matrix recovery. *Pattern Recognit. Lett.* **2020**, *132*, 38–45.

16. Wu, C.Y.; Ding, J.J. Occluded face recognition using low-rank regression with generalized gradient direction. *Pattern Recognit.* **2018**, *80*, 256–268.

17. Li, X.X.; Hao, P.; He, L.; Feng, Y. Image gradient orientations embedded structural error coding for face recognition with occlusion. *J. Ambient. Intell. Humaniz. Comput.* **2020**, *11*, 2349–2367.

18. Zhang, T.; Tang, Y.Y.; Fang, B.; Shang, Z.; Liu, X. Face recognition under varying illumination using gradientfaces. *IEEE Trans. Image Process.* **2009**, *18*, 2599–2606.

19. Tzimiropoulos, G.; Zafeiriou, S.; Pantic, M. Subspace learning from image gradient orientations. *IEEE Trans. Pattern Anal. Mach. Intell.* **2012**, *34*, 2454–2466.

20. Vu, N.S. Exploring patterns of gradient orientations and magnitudes for face recognition. *IEEE Trans. Inf. Forensics Secur.* **2012**, *8*, 295–304.

21. Zheng, Q.; Wang, Y.; Heng, P.A. Online Subspace Learning from Gradient Orientations for Robust Image Alignment. *IEEE Trans. Image Process.* **2019**, *28*, 3383–3394.

22. Qian, J.; Yang, J.; Xu, Y.; Xie, J.; Lai, Z.; Zhang, B. Image decomposition based matrix regression with applications to robust face recognition. *Pattern Recognit.* **2020**, *102*, 107204.

23. Wu, Q.; Zhu, S. Multispectral Image Matching Method Based on Histogram of Maximum Gradient and Edge Orientation. *IEEE Geosci. Remote. Sens. Lett.* **2022**, *19*, 1–5.

24. Huang, D.; Zhu, C.; Wang, Y.; Chen, L. HSOG: a novel local image descriptor based on histograms of the second-order gradients. *IEEE Trans. Image Process.* **2014**, *23*, 4680–4695.

25. Morgan, M.J. Features and the primal sketch. *Vis. Res.* **2011**, *51*, 738–753.

26. Li, C.; Gao, G.; Liu, Z.; Huang, D.; Xi, J. Defect detection for patterned fabric images based on GHOG and low-rank decomposition. *IEEE Access* **2019**, *7*, 83962–83973.

27. Zhang, Y.; Bai, X.; Yan, J.; Xiao, Y.; Chatwin, C.R.; Young, R.; Birch, P. No-reference image quality assessment based on multi-order gradients statistics. *J. Imaging Sci. Technol.* **2020**, *64*, 10505-1.

28. Bastian, B.T.; Jiji, C. Pedestrian detection using first-and second-order aggregate channel features. *Int. J. Multimed. Inf. Retr.* **2019**, *8*, 127–133.

29. Abdulhussain, S.H.; Ramli, A.R.; Hussain, A.J.; Mahmmod, B.M.; Jassim, W.A. Orthogonal polynomial embedded image kernel. In Proceedings of the International Conference on Information and Communication Technology, Nanning, China, 11–13 January 2019; pp. 215–221.

30. Chen, J.; Huang, D.; Zhu, X.; Chen, F. Gradient-Guided and Multi-Scale Feature Network for Image Super-Resolution. *Appl. Sci.* **2022**, *12*, 2935.

31. Zhang, L.; Yang, M.; Feng, X. Sparse representation or collaborative representation: Which helps face recognition? In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–11 November 2011; pp. 471–478.

32. Gou, J.; He, X.; Lu, J.; Ma, H.; Ou, W.; Yuan, Y. A class-specific mean vector-based weighted competitive and collaborative representation method for classification. *Neural Netw.* **2022**, *150*, 12–27.

33. Gou, J.; Sun, L.; Du, L.; Ma, H.; Xiong, T.; Ou, W.; Zhan, Y. A representation coefficient-based k-nearest centroid neighbor classifier. *Expert Syst. Appl.* **2022**, *194*, 116529.

34. Gou, J.; Yuan, X.; Du, L.; Xia, S.; Yi, Z. Hierarchical Graph Augmented Deep Collaborative Dictionary Learning for Classification. *IEEE Trans. Intell. Transp. Syst.* **2022**. https://doi.org/10.1109/TITS.2022.3177647.

35. Maaten, L.v.d.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.

36. He, R.; Zheng, W.S.; Tan, T.; Sun, Z. Half-quadratic-based iterative minimization for robust sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *36*, 261–275.

37. Qian, J.; Luo, L.; Yang, J.; Zhang, F.; Lin, Z. Robust nuclear norm regularized regression for face recognition with occlusion. *Pattern Recognit.* **2015**, *48*, 3145–3159.

38. Cai, S.; Zhang, L.; Zuo, W.; Feng, X. A probabilistic collaborative representation based approach for pattern classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2950–2959.

39. Iliadis, M.; Wang, H.; Molina, R.; Katsaggelos, A.K. Robust and low-rank representation for fast face identification with occlusions. *IEEE Trans. Image Process.* **2017**, *26*, 2203–2218.

40. Zhang, C.; Li, H.; Chen, C.; Qian, Y.; Zhou, X. Enhanced group sparse regularized nonconvex regression for face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *44*, 2438–2452.

41. Zhang, C.; Li, H.; Qian, Y.; Chen, C.; Zhou, X. Locality-constrained discriminative matrix regression for robust face identification. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 1254–1268.
42. Parkhi, O.M.; Vedaldi, A.; Zisserman, A.; et al. Deep face recognition. In Proceedings of the British Machine Vision Conference, Swansea, UK, 7–11 September 2015.
43. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
44. Wu, X.; He, R.; Sun, Z. A lightened cnn for deep face representation. *arXiv* **2015**, arXiv:1511.02683.
45. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 3–6 December 2012; pp. 1097–1105.
46. Lin, M.; Chen, Q.; Yan, S. Network in network. *arXiv* **2013**, arXiv:1312.4400.
47. Mehdipour Ghazi, M.; Kemal Ekenel, H. A comprehensive analysis of deep learning based representation for face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 27–30 June 2016; pp. 34–41.