 doi:10.20944/preprints202012.0237.v1

Article

Multi-Block Color-Binarized Statistical Images for Single Sample Face Recognition

Insaf Adjabi [1], Amir Benzaoui [2], Abdeldjalil Ouahabi [1,3,*] and Sébastien Jacques [4]

[1] Department of Computer Science, LIMPAF, University of Bouira, 10000 Bouira, Algeria; i.adjabi@univ-bouira.dz (I.A.)
[2] Department of Electrical Engineering, University of Bouira, 10000 Bouira, Algeria; a.benzaoui@univ-bouira.dz (A.B.)
[3] Polytech Tours, Imaging and Brain, INSERM U930, University of Tours, 37200 Tours, France; abdeldjalil.ouahabi@univ-tours.fr (A.O.)
[4] GREMAN UMR 7347, University of Tours, CNRS, INSA Centre Val-de-Loire, 37200 Tours, France; sebastien.jacques@univ-tours.fr (S.J.)
* Correspondence: abdeldjalil.ouahabi@univ-tours.fr; Tel.: +33-2-4736-1323

Abstract: Single sample face recognition (SSFR) is a computer vision challenge. In this scenario, there is only one example from each individual on which to train the system, making it difficult to identify persons in unconstrained environments, particularly when dealing with changes in facial expression, posture, lighting, and occlusion. This paper suggests a different method based on a variant of the Binarized Statistical Image Features (BSIF) descriptor called Multi-Block Color-Binarized Statistical Image Features (MB-C-BSIF) to resolve the SSFR Problem. First, the MB-C-BSIF method decomposes a facial image into three channels (e.g., red, green, and blue), then it divides each channel into equal non-overlapping blocks to select the local facial characteristics that are consequently employed in the classification phase. Finally, the identity is determined by calculating the similarities among the characteristic vectors adopting a distance measurement of the k-nearest neighbors (K-NN) classifier. Extensive experiments on several subsets of the unconstrained Alex & Robert (AR) and Labeled Faces in the Wild (LFW) databases show that the MB-C-BSIF achieves superior results in unconstrained situations when compared to current state-of-the-art methods, especially when dealing with changes in facial expression, lighting, and occlusion. Furthermore, the suggested method employs algorithms with lower computational cost, making it ideal for real-time applications.

Keywords: Biometrics, Face Recognition, Single Sample Face Recognition, Binarized Statistical Image Features, K-Nearest Neighbors.

1. Introduction

Generally speaking, biometrics aims to identify or verify an individual's identity according to some physical or behavioral characteristics [1]. Biometric practices are replacing conventional knowledge-based solutions, such as passwords or PINs, and possession-based strategies, such as ID cards or badges [2]. Several biometric methods have been developed to varying degrees and are being implemented and used in numerous commercial applications [3]. Biometric recognition is a continually growing area of study and one that is continually searching for new high-performance methods.

Fingerprints are the biometric features most commonly used to identify criminals [4]. The first automated fingerprint authentication device was commercialized in the early 1960s. In actual fact, multiple studies have shown that the iris of the eye is the most accurate modality since its texture remains stable throughout a person's life [5]. However, those techniques have the significant drawback of being invasive, which significantly restricts their applications. In addition, use of iris recognition remains problematic for users who do not wish to put their eyes in front of a sensor. On

 doi:10.20944/preprints202012.0237.v1

the contrary, biometric recognition based on facial analysis does not pose any such user constraints: in contrast to other biometric modalities, face recognition is a modality that can be employed without any user-sensor cooperation and can be applied discreetly in surveillance applications. Face recognition has a number of advantages: the sensor device (i.e., the camera) is simple to mount; it is not costly; it does not require subject cooperation; there are no hygiene issues; and, being passive, people much prefer this modality [6].

2D face recognition with single sample face recognition (SSFR) (i.e., using a single sample per person (SSPP) in the training set) has already matured as a technology. Although the latest studies on the Face Recognition Grand Challenge (FRGC) [7] project have shown that computer-vision systems [8] offer better performance than human-visual systems in controlled conditions [9], research into face recognition, however, needs to be geared towards, more realistic uncontrolled conditions. In an uncontrolled scenario, human visual systems are more robust when dealing with the numerous possibilities that can impact the recognition process [10], such as variations in lighting, facial orientation, facial expression, and facial appearance due to the presence of sunglasses, a scarf, a beard, or makeup. Solving such challenges will make 2D face recognition techniques a much more important technology for identification or the verification of an identity.

Several methods and algorithms have been suggested in the face recognition literature. They can be subdivided into 4 fundamental approaches [11] depending on the method used for feature extraction and classification: holistic, local, hybrid, and deep learning-based approaches. The deep learning class [12], which applies consecutive layers of information processing arranged hierarchically for representation, learning, and classification, has dramatically increased state-of-the-art performance, especially with unconstrained large-scale databases, and encouraged real-world applications [13].

Most current methods in the literature use several facial images (samples) per person in the training set. Nevertheless, in real-world systems (e.g., in fugitive tracing, ID cards, immigration management, or passports), only SSFR systems are used, which employ a single sample per person in the training stage, i.e., just one example of the person to be recognized is recorded in the database and accessible for the recognition task [14]. Since there are insufficient data (i.e., we do not have several samples per person) to perform supervised learning, many well-known algorithms may not work particularly well. For instance, deep neural networks (DNNs) [13] can be used in powerful face recognition techniques. Nonetheless, they necessitate a considerable volume of training data to work well. Vapnik and Chervonenkis [15] showed that vast training data are required to ensure learning systems generalization in their statistical learning theorem. We can infer that SSFR remains an unsolved topic in academia and business, particularly in terms of the great efforts being made and the growth in face recognition.

In this paper, we tackle the SSFR issue in unconstrained conditions by proposing an efficient method based on a variant of the local texture operator Binarized Statistical Image Features (BSIF) [16] called Multi-Block Color-binarized Statistical Image Features (MB-C-BSIF). It employs local color texture information to obtain honest and precise representation. The BSIF descriptor has been widely used in texture analysis [17, 18] and has proven its utility in many computer vision tasks. In the first step, the proposed method uses preprocessing to enhance the quality of facial photographs and remove noise [19-21]. The color image is then decomposed into three channels (e.g., red, green, and blue for the RGB color-space). Next, to find the optimum configuration, several multi-block decompositions are checked and examined under various color-spaces (i.e., we tested RGB, HSV, in addition to the YCbCr color-spaces). Finally, classification is undertaken using the distance measurement of the k-nearest neighbors (K-NN) classifier.

The rest of the paper is structured as follows. We discuss relevant research about SSFR in Section 2. Section 3 describes our suggested method. In Section 4, the experimental study, key findings, and comparisons are performed and presented to show our method's superiority. Section 5 of the paper presents some conclusions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2020 doi:10.20944/preprints202012.0237.v1

2. Related Work

Current methods designed to resolve the issue of SSFR can be categorized into four fundamental classes [22], namely: virtual sample generating, generic learning, image partitioning, and deep learning methods.

2.1. Virtual Sample Generating Methods

This class's methods produce some additional virtual training-samples for each individual to augment the gallery (i.e., data augmentation), so that discriminative sub-space learning can be employed to extract features. For example, Vetter (1998) [23] proposed a robust SSFR algorithm by generating 3D facial models through the recovery of high-fidelity reflectance and geometry. Zhang et al. (2005) [24] and Gao et al. (2008) [25] developed two techniques to tackle the issue of SSFR based on the singular value decomposition (SVD). Hu et al. (2015) [26] suggested a different SSFR system based on the lower-upper (LU) algorithm. In their approach, each single subject was decomposed and transposed employing the LU procedure and each raw image was rearranged according to its energy. Dong et al. (2018) [27] proposed an effective method for the completion of SSFR tasks called K-Nearest Neighbors virtual image set-based Multimanifold Discriminant Learning (KNNMMDL). They also suggested an algorithm named K-Nearest Neighbor-based Virtual Sample Generating (KNNVSG) to augment the information of intra-class variation in the training samples. They also proposed the Image Set-based Multimanifold Discriminant Learning algorithm (ISMMDL) to exploit intra-class variation information. While these methods can somewhat alleviate the SSFR problem, their main disadvantage lies in the strong correlation between the virtual images, which cannot be regarded as independent examples for the selection of features.

2.2. Generic Learning Mehods

This class's methods first extract discriminant characteristics from a supplementary generic training set that includes several examples per individual and then use those characteristics for SSFR tasks. Deng et al. (2012) [28] developed the Extended Sparse Representation Classifier (ESRC) technique in which the intra-class variant dictionary is created from generic persons not incorporated in the gallery set to increase the efficiency of the identification process. In a method called Sparse Variation Dictionary Learning (SVDL), Yang et al. (2013) [29] trained a sparse variation dictionary by considering the relation between the training set and the outside generic set, disregarding the distinctive features of various organs of the human face. Zhu et al. (2014) [30] suggested a system for SSFR based on local generic representation (LGR), which leverages the benefits of both image partitioning and generic learning and takes into account the fact that the intra-class face variation can be spread among various subjects.

2.3. Image Partitioning Methods

This class's methods divide each person's images into local blocks, extract the discriminant characteristics, and, finally, perform classifications based on the selected discriminant characteristics. Zhu et al. (2012) [31] developed a patch-based CRC (PCRC) algorithm that applies the original method proposed by Zhang et al. (2011) [32], named Collaborative Representation-based Classification (CRC), to each block. Lu et al. (2013) [33] suggested a technique called Discriminant Multimanifold Analysis (DMMA) that divides any registered image into multiple non-overlapping blocks and then learns several feature spaces to optimize the various margins of different individuals. Zhang et al. (2018) [34] developed local histogram-based face image operators. They decomposed each image into different non-overlapping blocks. Next, they tried to derive a matrix to project the blocks into an optimal subspace to maximize the different margins of different individuals. Each column was then redesigned to an image filter to treat facial images and the filter responses were binarized using a fixed threshold. Gu et al. (2018) [35] proposed a method called Local Robust Sparse Representation (LRSR). The main idea of this technique is to merge a local sparse representation model with a block-based generic variation dictionary learning model to determine the possible facial

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2020 doi:10.20944/preprints202012.0237.v1

intra-class variations of the test images. Zhang et al. (2020) [36] introduced a novel nearest neighbor classifier (NNC) distance measurement to resolve SSFR problems. The suggested technique, entitled Dissimilarity-based Nearest Neighbor Classifier (DNNC), divides all images into equal non-overlapping blocks and produces an organized image block-set. The dissimilarities among the given query image block-set and the training image block-sets are calculated and considered by the NNC distance metric.

2.4. Deep Learning Methods

This class's methods employ consecutive hidden layers of information-processing arranged hierarchically for representation, learning, and classification. They can automatically determine complex non-linear data structures [37]. Zeng et al. (2017) [38] proposed a method that uses deep convolutional neural networks (DCNN). Firstly, they propose using an expanding sample technique to augment the training sample set and then a trained DCNN model is implemented and fine-tuned by those expanding samples to be used in the classification process. Ding et al. (2017) [39] developed a deep learning technique centered on a kernel principal component analysis network (KPCANet) and a novel weighted voting technique. First, the aligned facial image is segmented into multiple non-overlapping blocks to create the training set. Then, a KPCANet is employed to get filters and banks of features. Lastly, recognition of the unlabeled probe is achieved by applying the weighted voting form. Zhang and Peng (2017) [40] introduced a different method to generate intra-class variances using a deep auto-encoder. They then used these intra-class variations to expand the new examples. First, a generalized deep auto-encoder is used to train facial images in the gallery. Second, a class-specific deep auto-encoder (CDA) is fine-tuned with a single example. Finally, the corresponding CDA is employed to expand the new samples. Du and Da (2020) [41] proposed a method entitled block dictionary learning (BDL) that fuses sparse representation (SR) with CNNs. SR is implemented to augment CNN efficiency by improving the inter-class feature variations and creating a global-to-local dictionary learning process to increase the method's robustness.

It is clear that the deep learning-based approach for face recognition has gained particular attention in recent years, but it suffers considerably with SSFR systems as they still require a significant amount of information in the training set.

Motivated by the successes of the third approach, "image partitioning," and the reliability of the local texture descriptor BSIF, in this paper we propose an image partitioning method to address the problems of SSFR. The proposed method, called MB-C-BSIF, decomposes each image into several color channels, divides each color component into various equal non-overlapping blocks, and applies the BSIF descriptor to each block-component to extract the discriminative features. In the following section, the framework of the MB-C-BSIF is explained in detail.

3. Proposed Method

The suggested feature extraction and classification rules, which compose the essential steps in our proposed SSFR, are listed in this section. The facial image is enhanced by applying histogram normalization and then filtered with a non-linear filter. The median filter [42] was adopted to minimize noise while preserving facial appearance and enhancing the operational outcomes [43].

3.1. MB-C-BSIF-based Feature Extraction

Our advanced feature extraction technique is based on the multi-block color representation of the BSIF descriptor, entitled Multi-Block Color BSIF (MB-C-BSIF). The BSIF operator proposed by Kannala and Rahtu [16] is an efficient and robust descriptor for texture analysis [44]. BSIF focuses on creating local image descriptors that powerfully encode texture information and are appropriate for describing image regions in the form of histograms. The method calculates a binary code for all pixels by linearly projecting local image blocks onto a subspace whose basis vectors are learned from natural pictures through independent component analysis (ICA) [45] and by binarizing the coordinates through thresholding. The number of basis vectors defines the length of the binary code string. Image regions can be conveniently represented with histograms of the pixels' binary codes. Other

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2020 doi:10.20944/preprints202012.0237.v1

descriptors that generate binary codes, such as the local binary pattern (LBP) [46] and the local phase quantization (LPQ) [47], have inspired the BSIF process. However, the BSIF is based on natural image statistics, rather than heuristic or handcrafted code constructions, enhancing its modeling capabilities.

Technically speaking, the s_i filter response is calculated, for a given picture patch X of size $l \times l$ pixels and a linear filter W_i of the same size, by:

$$s_i = \sum_{u,v} W_i(u,v)X(u,v) \quad (1)$$

where the index i in W_i indicates the i^{th} filter.
The binarized b_i feature is calculated as follows:

$$b_i = \begin{cases} 1 & \text{if } s_i > 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

The BSIF descriptor has two key parameters: the filter size $l \times l$ and the bit string length n. Using ICA, W_i filters are trained by optimizing s_i's statistical independence. The training of W_i filters is based on different choices of parameter values. In particular, each filter set was trained using 50,000 image patches. Figure 1 displays some examples of the filters obtained with $l \times l = 7 \times 7$ and $n = 8$. Figure 2 provides some examples of facial images and their respective BSIF representations (with $l \times l = 7 \times 7$ and $n = 8$).

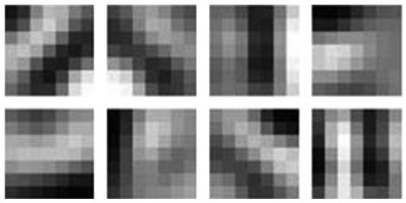


Figure 1. Examples of 7×7 BSIF filter banks learned from natural pictures.

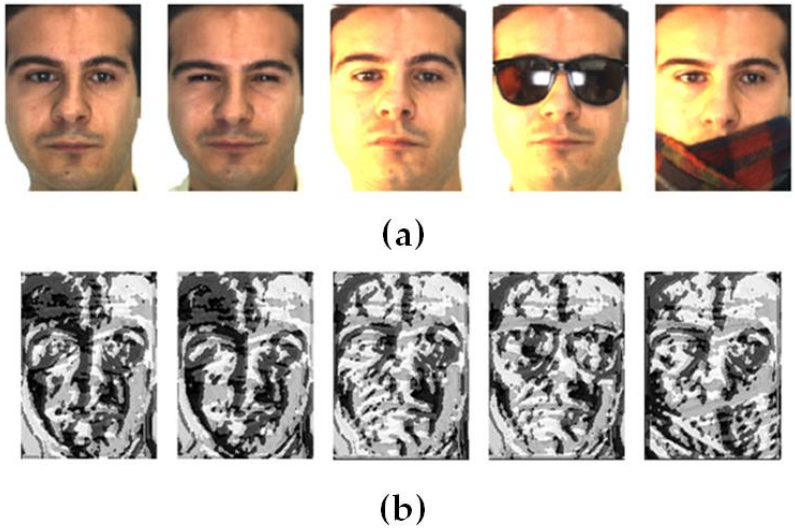


Figure 2. (a) Examples of facial images, and (b) their parallel BSIF representations.

Like LBP and LPQ methodologies, the BSIF codes' co-occurrences are collected in a histogram $H1$, which is employed as a feature vector.

However, the simple BSIF operator based on a single block does not possess information that dominates the texture characteristics, which is forceful for the image's occlusion and rotation. To address those limitations, an extension of the basic BSIF, the Multi-Block BSIF (MB-BSIF), is used. The concept is based on partitioning the original image into non-overlapping blocks. An undefined facial

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2020 doi:10.20944/preprints202012.0237.v1

image may be split equally along the horizontal and vertical directions. As an illustration, we can derive 1, 4, or 16 blocks by segmenting the image into grids of 1×1, 2×2, or 4×4, as shown in Figure 3. Each block possesses details about its composition, such as the nose, eyes, or eyebrows. Overall, these blocks provide information about position relationships, such as nose to mouth and eye to eye. The blocks and the data between them are thus very important for SSFR tasks.

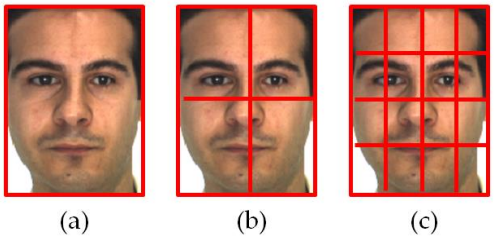


Figure 3. Examples of multi-block (MB) image decomposition: (a) 1×1, (b) 2×2, and (c) 4×4.

Our idea was to segment the image into equal non-overlapping blocks and calculate the BSIF operator's histograms related to the different blocks. The histogram $H2$ represents the fusion of the regular histograms calculated for the different blocks, as shown in Figure 4.

In the face recognition literature, a number of works have concentrated solely on analyzing the luminance details of facial images (i.e., grayscale). This paper suggests a different and exciting technique that exploits color texture information and shows that analysis of chrominance can be beneficial to SSFR systems. To prove this idea, we can separate the RGB facial image into three channels (i.e., red, green, and blue) and then compute the MB-BSIF separately for each channel. The final feature vector is the concatenation of their histograms in a global histogram $H3$. This approach is called Multi-Block Color BSIF (MB-C-BSIF). Figure 4 provides a schematic illustration of the proposed MB-C-BSIF framework.

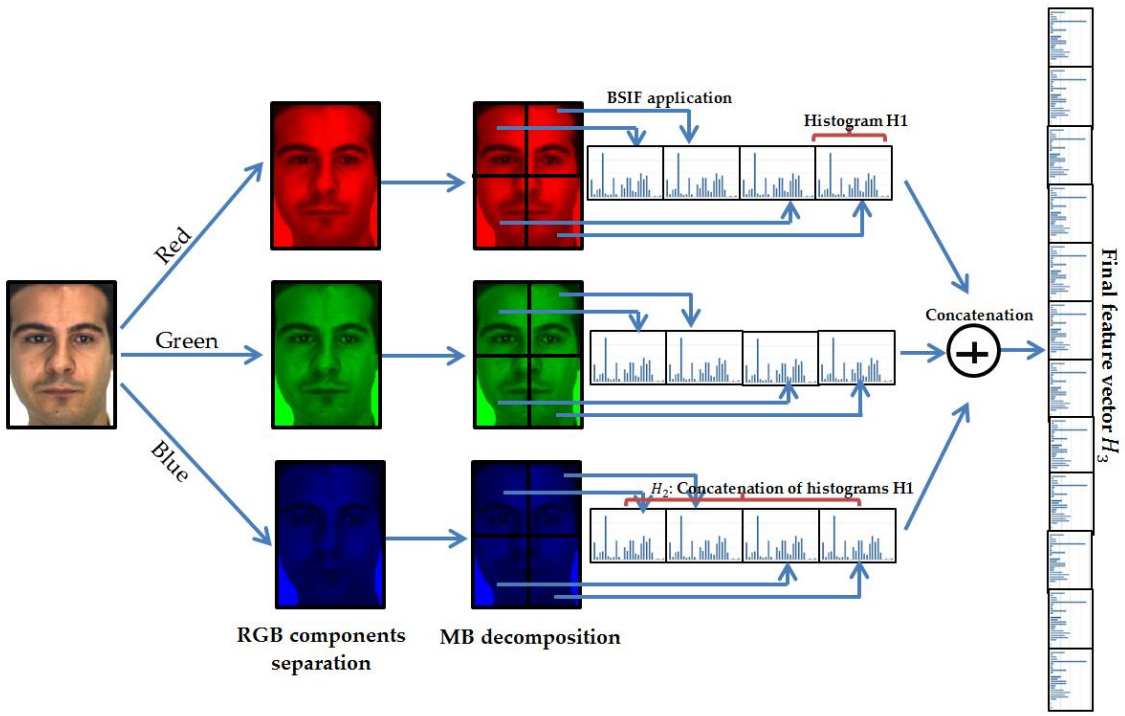


Figure 4. Structure of the proposed feature extraction approach: MB-C-BSIF.

We note that the RGB is the most commonly employed color-space for detecting, modeling and displaying color images. Nevertheless, its use in image interpretation is restricted due to the broad connection between the three color channels (i.e., red, green, and blue) and the inadequate separation

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2020 doi:10.20944/preprints202012.0237.v1

of details in terms of luminance and chrominance. To identify captured objects, the various color channels can be highly discriminative and offer excellent contrast for several visual indicators from natural skin tones. In addition to the RGB, we studied and tested two additional color-spaces—HSV and YCbCr—to exploit color texture details. These color-spaces are based on separating components of the chrominance and luminance. For the HSV color-space, the dimensions of hue and saturation determine the image's chrominance while the dimension of brightness (v) matches the luminance. The YCbCr color-space divides the components of the RGB into luminance (Y), chrominance blue (Cb), and chrominance red (Cr). We should note that the representation of chrominance components in the HSV and YCbCr domains is dissimilar and, consequently, they can offer additional color texture descriptions for SSFR systems.

3.2. K-Nearest Neighbors (K-NN) Classifier

During the classification process, each tested facial image is compared with those saved in the dataset. To assign the corresponding label (i.e., identity) to the tested image, we used the k-nearest neighbors (K-NN) classifier [48] associated with a distance metric. In scenarios of general usage, K-NNs show excellent flexibility and usability in substantial applications.

Technically speaking, for a presented training set $\{(x_i, y_i) \mid i = 1,2, \dots, s\}$, where $x_i \in R^D$ denotes the i^{th} person's feature vector, y_i denotes this person's label, D is the dimension of the characteristic vector, and s represents the number of persons. For a test person $x' \in R^D$ that is expected to be classified, the K-NN is used to determine a training person x^* resembling to x' based on the distance rate and then attribute the label of x^* to x'.

K-NN can be implemented with various distance measurements. We evaluated and compared three widely used distance metrics in this work: Hamming, Euclidean, and city block (also called Manhattan distance).

The Hamming distance between $x'$and x_i is calculated as follows:

$$d(x', x_i) = \sum_{j=1}^{D} \left| (x'_j - x_{ij})^2 \right| \quad (3)$$

The Euclidean distance between $x'$and x_i is formulated as follows:

$$d(x', x_i) = \sqrt{\sum_{j=1}^{D} (x'_j - x_{ij})^2} \quad (4)$$

The city block distance between $x'$and x_i is measured as follows:

$$d(x', x_i) = \sum_{j=1}^{D} \left| (x'_j - x_{ij}) \right| \quad (5)$$

where x' and x_i are two vectors of dimension D, while x_{ij} is the j^{th} feature of x_i, and x'_j is the j^{th} feature of x'.

The corresponding label of x' can be determined by:

$$y' = y_{i^*} \quad (6)$$

where

$$i^* = arg_{i=1,\dots,s}\big(min(d(x', x_i))\big) \quad (7)$$

The distance metric in SSFR corresponds to calculating the similarities between the test example and the training examples.

The following algorithm sums up our proposed method of SSFR recognition.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2020 doi:10.20944/preprints202012.0237.v1

Algorithm 1 SSFR based on MB-C-BSIF and K-NN

Input: Facial image X

1. Apply histogram normalization on X
2. Apply median filtering on X
3. Divide X into three components (red, green, blue): $C^n; n = 1,2,3$
4. **for** $n = 1$ to 3
5. Divide C^n into K equivalent blocks: $C_k^n; k = 1, \dots, K$
6. **for** $k = 1$ to K
7. Compute BSIF on the block-component C_k^n: $H1_{(k)}^{(n)}$
8. **end for**
9. Concatenate the computed MB-BSIF features of the component C_n:

$$H2^{(n)} = H1_{(1)}^{(n)} + H1_{(2)}^{(n)} + \cdots + H1_{(K)}^{(n)}$$

10. **end for**
11. Concatenate the computed MB-C-BSIF features: $H3 = H2^{(1)} + H2^{(2)} + H2^{(3)}$
12. Apply K-NN associated with a metric distance

Output: Identification decision

4. Experimental Analysis

The proposed SSFR was evaluated using the unconstrained Alex & Robert (AR) [48] and Labeled Faces in the Wild (LFW) [49] databases. In this section, we present the specifications of each utilized database and their experimental setups. Furthermore, we analyze the findings obtained from our proposed SSFR method and compare the accuracy of recognition with other current state-of-the-art approaches.

4.1. Experiments on the AR Database

4.1.1. Database Description

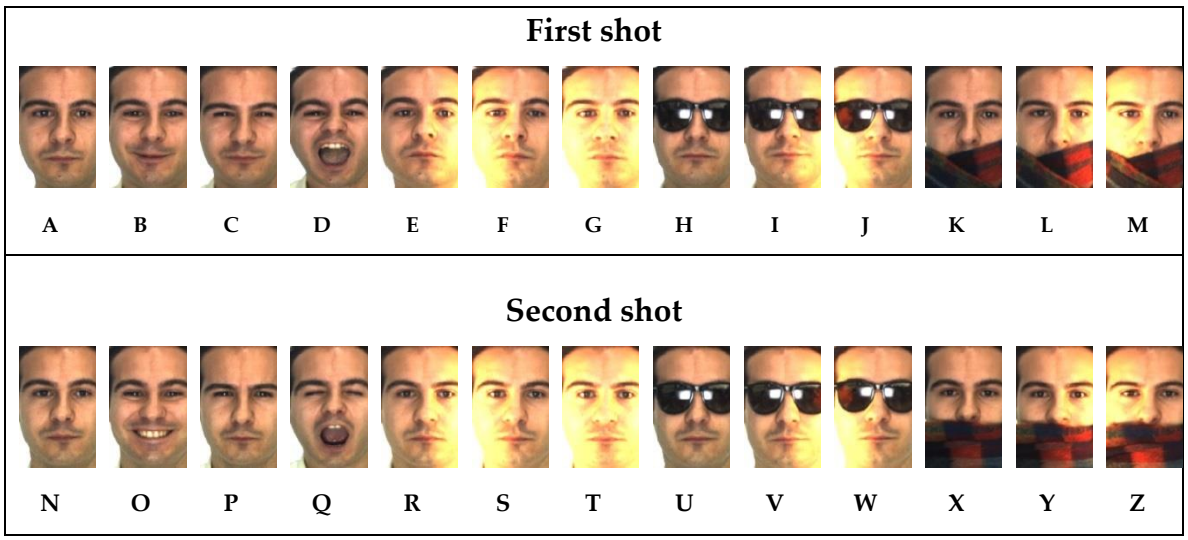
The Alex & Robert (AR) face database [48] includes more than 4,000 colored facial photographs of 126 individuals (56 females & 70 males); each individual has 26 different images with a frontal face taken with several facial expressions, lighting conditions, and occlusions. These photographs were acquired at an interval of two-weeks and their analysis was in two sessions (shots 1 & 2). Each session comprised 13 facial photographs per subject. A subset of facial photographs of 100 distinct individuals (50 males & 50 females) was selected in the subsequent experiments. Figure 5 displays the 26 facial images of the first individual from the AR database along with detailed descriptions of them.

4.1.2. Setups

To determine the efficiency of the proposed MB-C-BSIF in dealing with changes in facial expression, subset A (normal-1) was used as the training set and subsets B (smiling-1), C (angry-1), D (screaming-1), N (normal-2), O (smiling-2), P (angry-2), and Q (screaming-2) were employed for the test set. The facial images from the 8 subsets display different facial expressions and were used in two different sessions. For the training set, we employed 100 images of the normal-1 type (100 images for 100 persons, i.e., one image per person). Moreover, we employed 700 images in the test set (smiling-1, angry-1, screaming-1, normal-2, smiling-2, angry-2, and screaming-2). These 700 images were divided into 7 subsets for testing, with each subset containing 100 images.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2020 doi:10.20944/preprints202012.0237.v1

As shown in Figure 5, two forms of occlusion are found in 12 subsets. The first is occlusion by sunglasses, as seen in subsets H, I, J, U, V, and W; while the second is occlusion by scarf in subsets K, L, M, X, Y, and Z. In these 12 subsets, each individual's photographs have various illumination conditions and were acquired in two distinct stages. There are 100 different items in each subset and the total number of facial photographs used in the test set was 1,200. To examine the performance of the suggested MB-C-BSIF under conditions of object occlusion, we considered subset A as the training set and the 12 occlusion subjects as the test set, which was similar to the initial setup.



A: Normal
B: Smiling
C: Angry
D: Screaming
E: Normal + right light
F: Normal + left light
G: Normal + frontal light
H: With sunglasses + frontal light
I: With sunglasses + left light
J: With sunglasses + left light
K: With scarf + without light
L: With scarf + right light
M: With scarf + left light
N: Normal
O: Smiling
P: Angry
Q: Screaming
R: Normal + right light
S: Normal + left light
T: Normal + frontal light
U: With sunglasses + without light
V: With sunglasses + right light
W: With sunglasses + left light
X: With scarf + without light
Y: With scarf + right light
Z: With scarf + left light

Figure 5. The 26 facial images of the first individual from the AR database and their detailed descriptions.

4.1.3. Experiment #1 (Effects of BSIF Parameters)

As stated in subsection 3.1, the BSIF operator is based on two parameters: filter kernel size $l \times l$ and bit string length n. In this test, we assessed the proposed method by testing various BSIF parameters to obtain the best configuration, i.e., the one that yielded the best recognition accuracy. We transformed the image into a grayscale level, we did not segment the image into non-overlapping blocks (i.e., 1×1 block), and we used the city block distance associated with K-NN. Tables 1, 2, and 3 show comprehensive details and comparisons of results obtained using some (key) BSIF configurations for facial expression variation subsets, occlusion subsets for sunglasses, and occlusion subsets for scarfs, respectively. The best results are in bold.

We note that using the parameters $l \times l = 17 \times 17$and $n = 12$ for the BSIF operator achieves the best performance in identification compared to other configurations considered in this experiment. Furthermore, an increase in the identification rate appears when we augment the values of l or n.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2020 doi:10.20944/preprints202012.0237.v1

The implemented configuration can achieve better accuracy for changes in facial expression with all 7 subsets. However, for subset Q, which is characterized by considerable variation in facial expression, the accuracy of recognition was very low (71 %). Lastly, the performance of this implemented configuration under conditions of occlusion by an object is unsatisfactory, especially with occlusion by scarf, and needs further improvement.

Table 1. Comparison of the results obtained using six BSIF configurations with changes in facial expression.

$l \times l$ (pixels)	n (bits)	Accuracy (%)							Average Accuracy (%)
		B	C	D	N	O	P	Q	
3×3	**5**	70	72	38	36	20	24	14	39.14
5×5	**9**	94	97	59	75	60	66	30	68.71
9×9	**12**	100	100	91	95	90	92	53	88.71
11×11	**8**	97	99	74	85	70	75	43	77.57
15×15	**12**	100	100	96	97	96	96	73	94.00
17×17	**12**	100	100	98	97	96	97	71	94.14

Table 2. Comparison of the results obtained using six BSIF configurations with occlusion by sunglasses.

$l \times l$ (pixels)	n (bits)	Accuracy (%)						Average Accuracy (%)
		H	I	J	U	V	W	
3×3	**5**	29	8	4	12	4	3	10.00
5×5	**9**	70	24	14	28	14	8	26.50
9×9	**12**	98	80	61	80	38	30	61.50
11×11	**8**	78	34	23	48	26	15	37.33
15×15	**12**	100	84	85	87	50	46	75.33
17×17	**12**	100	91	87	89	58	46	78.50

Table 3. Comparison of the results obtained using six BSIF configurations with occlusion by scarf.

$l \times l$ (pixels)	n (bits)	Accuracy (%)						Average Accuracy (%)
		K	L	M	X	Y	Z	
3×3	**5**	7	4	2	3	2	2	3.33
5×5	**9**	22	9	6	12	6	2	9.50
9×9	**12**	88	54	34	52	31	15	45.67
11×11	**8**	52	12	90	22	9	7	32.00
15×15	**12**	97	69	64	79	48	37	65.67
17×17	**12**	98	80	63	90	48	31	68.33

4.1.4. Experiment #2 (Effects of Distance)

In this experiment, we evaluated the last configuration (i.e., grayscale level image, 1×1 block $l \times l = 17 \times 17$, and $n = 12$) by checking various distances associated with K-NN for classification. Tables 4, 5, and 6 compare the results achieved by adopting the city block distance and other well-known distances with facial expression variation subsets, occlusion subsets for sunglasses, and occlusion subsets for scarfs, respectively. The best results are in bold.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2020 doi:10.20944/preprints202012.0237.v1

We note that the city block distance produced the most reliable recognition performance compared to the other distances analyzed in this test, such as the Hamming and Euclidean distances. As such, we can say that the city block distance is the most suitable for our method.

Table 4. Comparison of the results obtained using different distances with changes in facial expression.

Distance	Accuracy (%)							Average Accuracy (%)
	B	C	D	N	O	P	Q	
Hamming	63	79	9	69	23	40	6	41.29
Euclidean	99	100	80	90	83	82	43	82.43
City block	100	100	98	97	96	97	71	94.14

Table 5. Comparison of the results obtained using different distances with occlusion by sunglasses.

Distance	Accuracy (%)						Average Accuracy (%)
	H	I	J	U	V	W	
Hamming	37	5	6	11	4	2	10.83
Euclidean	96	68	42	68	31	17	53.67
City block	100	91	87	89	58	46	78.50

Table 6. Comparison of the results obtained using different distances with occlusion by scarf.

Distance	Accuracy (%)						Average Accuracy (%)
	K	L	M	X	Y	Z	
Hamming	34	5	8	20	4	4	12.50
Euclidean	79	32	16	41	22	5	32.50
City block	98	80	63	90	48	31	68.33

4.1.5. Experiment #3 (Effects of Image Segmentation)

To improve recognition accuracy, especially under conditions of occlusion, we proposed decomposing the image into several non-overlapping blocks, as discussed in subsection 3.1. The objective of this test was to estimate identification performance when MB-BSIF features are used instead of their global computation over an entire image. In this paper, three methods for image segmentation are considered and compared. Each original image was divided into 1×1 (i.e., global information), 2×2, and 4×4 blocks (i.e., local information). In other terms, an image was divided into 1 block (i.e., the original image), 4 blocks, and 16 blocks. For the last two cases, the feature vectors (i.e., histograms H1) derived from each block were fused to create the entire image's extracted feature vector (Histogram H2). Tables 7, 8, and 9 present and compare the recognition accuracy of the tested MB-BSIF for various blocks with subsets of facial expression variation, occlusion subsets for sunglasses, and occlusion subsets for scarfs, respectively (with grayscale images, city block distance, $l \times l = 17 \times 17$, and $n = 12$). The best results are in bold.

From the resulting outputs, we can observe that:

- For subsets of facial expression variation, a small change arises because the results of the previous experiment were already reasonable (e.g., subsets A, B, D, N, and P). However, the accuracy rises from 71 % to 76 % for subset Q, which is characterized by significant changes in facial expression.
- For occluded subsets, there was a significant increase in recognition accuracy when the number of blocks was augmented. As an illustration, when we applied 1 to 16 patches, the accuracy grew from 31 % to 71 % for subset Z, from 46 % to 79 % for subset W, and from 48 % to 84 % for subset Y.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2020 doi:10.20944/preprints202012.0237.v1

- As such, in the case of partial occlusion, we may claim that local information is essential. It helps to go deeper in extracting relevant information from the face like details about the facial structure, such as the nose, eyes, or mouth, and information about position relationships, such as nose to mouth, eye to eye, and so on.
- Finally, we note that the optimum configuration with the best accuracy was provided by 4×4 blocks for subsets of facial expression, occlusion by sunglasses, and occlusion by scarf.

Table 7. Comparison of the results obtained using different divided blocks with changes in facial expression.

Segmentation	Accuracy (%)							Average Accuracy (%)
	B	C	D	N	O	P	Q	
(1×1)	100	100	98	97	96	97	71	94.14
(2×2)	100	100	95	98	92	91	60	90.86
(4×4)	100	100	99	98	92	97	76	94.57

Table 8. Comparison of the results obtained using different divided blocks with occlusion by sunglasses.

Segmentation	Accuracy (%)						Average Accuracy (%)
	H	I	J	U	V	W	
(1×1)	100	91	87	89	58	46	78.50
(2×2)	100	99	98	91	83	71	90.33
(4×4)	100	99	99	93	81	79	91.83

Table 9. Comparison of the results obtained using different divided blocks with occlusion by scarf.

Segmentation	Accuracy (%)						Average Accuracy (%)
	K	L	M	X	Y	Z	
(1×1)	98	80	63	90	48	31	68.33
(2×2)	98	95	92	92	79	72	88.00
(4×4)	99	98	95	93	84	77	91.00

4.1.6. Experiment #4 (Effects of Color Texture Information)

For this analysis, we evaluated the performance of the last configuration (i.e., segmentation of the image into 4×4 blocks, K-NN associated with city block distance, $l \times l = 17 \times 17$, and $n = 12$) by testing three color-spaces, namely RGB, HSV, and YCbCr, instead of transforming the image into grayscale. This feature extraction method is called MB-C-BSIF, as described in subsection 3.1. The AR database images are already in RGB and so do not need transformation of the first color-space. For the other color-spaces, however, the images must be converted from RGB to HSV and RGB to YCbCr. Tables 10, 11, and 12 display and compare the recognition accuracy of the MB-C-BSIF using several color-spaces with subsets of facial expression variations, occlusion by sunglasses, and occlusion by scarf, respectively. The best results are in bold.

From the resulting outputs, we can see that:

- The results are almost identical for subsets of facial expression variation with all checked color-spaces. In fact, with the HSV color-space, a slight improvement is reported, although slight degradations are observed with both RGB and YCbCr color-spaces.
- All color-spaces see enhanced recognition accuracy compared to the grayscale standard for sunglasses occlusion subsets. RGB is the color-space with the highest output, seeing an increase from 91.83 % to 93.50 % in terms of average accuracy.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2020 doi:10.20944/preprints202012.0237.v1

- HSV shows some regression for scarf occlusion subsets, but both the RGB and YCbCr color-spaces display some progress compared to the grayscale norm. Additionally, RGB remains the color-space with the highest output.
- The most significant observation is that the RGB color-space saw significantly improved performance in the V, W, Y, and Z subsets (from 81 % to 85 % with V; 79 % to 84 % with W; 84 % to 88 % with Y; and 77 % to 87 % with Z). Note that images of these occluded subsets are characterized by light degradation (either to the right or left, as shown in Figure 5).
- Finally, we note that the optimum color-space, providing a perfect balance between lighting restoration and improvement in identification, was the RGB.

Table 10. Comparison of the results obtained using different color-spaces with changes in facial expression.

Color-space	Accuracy (%)							Average Accuracy (%)
	B	C	D	N	O	P	Q	
Gray Scale	100	100	99	98	92	97	76	94.57
RGB	100	100	95	97	92	93	67	92.00
HSV	100	100	99	97	96	95	77	94.86
YCbCr	100	100	96	98	93	93	73	93.29

Table 11. Comparison of the results obtained using different color-spaces with occlusion by sunglasses.

Color-space	Accuracy (%)						Average Accuracy (%)
	H	I	J	U	V	W	
Gray Scale	100	99	99	93	81	79	91.83
RGB	100	99	100	93	85	84	93.50
HSV	100	97	99	96	82	80	92.33
YCbCr	100	99	98	93	81	80	91.83

Table 12. Comparison of the results obtained using different color-spaces with occlusion by scarf.

Color-space	Accuracy (%)						Average Accuracy (%)
	K	L	M	X	Y	Z	
Gray Scale	99	98	95	93	84	77	91.00
RGB	99	97	97	94	88	81	92.67
HSV	99	96	90	95	75	74	88.17
YCbCr	98	98	96	93	87	78	91.67

4.1.7. Comparison #1

To confirm that our suggested method produces superior recognition performance with variations in facial expression, we compared the collected results with several state-of-the-art methods recently employed to tackle the SSFR issue. Table 13 presents the highest accuracies obtained using the same subsets and the same assessment protocol with Subset A as the training set and subsets of facial expression variations B, C, D, N, O, and P constituting the test set. The results presented in Table 13 are taken from a number of references [30, 33, 45, and 46]. "- -" signifies that the considered method has no experimental results. The best results are in bold.

The outcomes obtained validate the robustness and reliability of our proposed SSFR system compared to state-of-the-art methods when assessed with identical subsets. We suggest that this is a

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2020 doi:10.20944/preprints202012.0237.v1

competitive technique that has achieved a desirable level of identification accuracy with the six subsets of up to: 100.00 % for B and C; 95.00 % for D; 97.00 % for N; 92.00 % for O; and 93.00 % for P.

For all subsets, our suggested technique surpasses the state-of-the-art methods analyzed in this paper, i.e., the proposed MB-C-BSIF can achieve excellent identification performance under the condition of variation in facial expression.

Table 13. Comparison of 16 methods of facial expression variation subsets.

Authors	Year	Method	Accuracy						Average accuracy (%)
			B	C	D	N	O	P	
Turk, Pentland [52]	1991	PCA	97.00	87.00	60.00	77.00	76.00	67.00	77.33
Wu and Zhou [53]	2002	$(PC)^2A$	97.00	87.00	62.00	77.00	74.00	67.00	77.33
Chen et al. [54]	2004	$E(PC)^2A$	97.00	87.00	63.00	77.00	75.00	68.00	77.83
Yang et al. [55]	2004	2DPCA	97.00	87.00	60.00	76.00	76.00	67.00	77.17
Gottumukkal and Asari [56]	2004	Block-PCA	97.00	87.00	60.00	77.00	76.00	67.00	77.33
Chen et al. [57]	2004	Block-LDA	85.00	79.00	29.00	73.00	59.00	59.00	64.00
Zhang , Zhou [58]	2005	$(2D)^2PCA$	98.00	89.00	60.00	71.00	76.00	66.00	76.70
Tan et al. [59]	2005	SOM	98.00	88.00	64.00	73.00	77.00	70.00	78.30
He et al. [60]	2005	LPP	94.00	87.00	36.00	86.00	74.00	78.00	75.83
Zhang et al. [24]	2005	SVD-LDA	73.00	75.00	29.00	75.00	56.00	58.00	61.00
Deng et al. [61]	2010	UP	98.00	88.00	59.00	77.00	74.00	66.00	77.00
Lu et al. [33]	2013	DMMA	99.00	93.00	69.00	88.00	85.00	85.50	79.00
Mehrasa et al. [50]	2017	SLPMM	99.00	94.00	65.00	- -	- -	- -	- -
Ji et al. [51]	2017	CPL	92.22	88.06	83.61	83.59	77.95	72.82	83.04
Zhang et al. [34]	2018	DMF	100.00	99.00	66.00	- -	- -	- -	- -
Chu et al. [62]	2019	MFSA+	100.00	100.00	74.00	93.00	85.00	86.00	89.66
Zhang et al. [36]	2020	DNNC	100.00	98.00	69.00	92.00	76.00	85.00	86.67
Our method	**2020**	**MB-C-BSIF**	**100.00**	**100.00**	**95.00**	**97.00**	**92.00**	**93.00**	**96.17**

4.1.8. Comparison #2

To further demonstrate the efficacy of our proposed SSFR system, we also compared the best configuration of the MB-C-BSIF (i.e., RGB color-space, segmentation of the image into 4×4 blocks, city block distance, $l \times l = 17 \times 17$, and $n = 12$) with recently published work under unconstrained conditions. We followed the same experimental protocol described in [30, 36]. Table 14 displays the accuracies of the works compared on the tested subsets H + K (i.e., occlusion by sunglasses and scarf) and subsets J + M (i.e., occlusion by sunglasses and scarf with variations in lighting). The best results are in bold.

In Table 14, we can observe that the work presented by Zhu et al. [30], called LGR, shows a comparable level, but the identification accuracy of our MB-C-BSIF procedure is much higher than all the methods considered for both test sessions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2020 doi:10.20944/preprints202012.0237.v1

Compared to related SSFRs, which can be categorized as either generic learning methods (e.g., ESRC [28], SVDL [29], and LGR [30), image partitioning methods (e.g., CRC [29], PCRC [30], and DNNC [36]), or deep learning methods (e.g., DCNN [38] and BDL [41]), the capabilities of our method can be explained in terms of its exploitation of different forms of information. This can be summarized as follows:

- The BSIF descriptor scans the image pixel by pixel, i.e., we consider the benefits of local information.
- The image is decomposed into several blocks, i.e., we exploit regional information.
- BSIF descriptor occurrences are accumulated in a global histogram, i.e., we manipulate global information.
- The MB-BSIF is applied to all RGB image components, i.e., color texture information is exploited.

Table 14. Comparison of 9 methods on occlusion and lighting-occlusion sessions.

Authors	Year	Method	Occlusion (%)	Lighting + Occlusion (%)	Average accuracy (%)
Zhang et al. [32]	2011	CRC	58.10	23.80	40.95
Deng et al. [28]	2012	ESRC	83.10	68.60	75.85
Zhu et al. [31]	2012	PCRC	95.60	81.30	88.45
Yang et al. [29]	2013	SVDL	86.30	79.40	82.85
Lu et al. [33]	2013	DMMA	46.90	30.90	38.90
Zhu et al. [30]	2014	LGR	98.80	96.30	97.55
Ref. [63]	2016	SeetaFace	63.13	55.63	59.39
Zeng et al. [37]	2017	DCNN	96.5	88.3	92.20
Chu et al. [62]	2019	MFSA+	91.3	79.00	85.20
Cuculo et al. [64]	2019	SSLD	90.18	82.02	86.10
Zhang et al. [36]	2020	DNNC	92.50	79.50	86.00
Du and Da [41]	2020	BDL	93.03	91.55	92.29
Our method	**2020**	**MB-C-BSIF**	**99.5**	**98.5**	**99.00**

4.2. Experiments on the LFW Database

4.2.1. Database Description

The Labeled Faces in the Wild (LFW) database [49] comprises more than 13,000 photos collected from the World Wide Web of 5,749 diverse subjects in challenging situations, of which 1,680 subjects possess two or more shots per individual. In our tests, we employed the LFW-a, a variant of the standard LFW where the facial images are aligned with a commercial normalization tool. It can be observed that the intra-class differences in this database are very high compared to the well-known constrained databases and face normalization has been carried out. The size of each image is 250×250 pixels and uses the jpeg extension. LFW is a very challenging database: it aims to investigate the unconstrained issues of face recognition, such as changes in lighting, age, clothing, focus, facial expression, color saturation, posture, race, hairstyle, background, camera quality, gender, ethnicity, and other factors, as presented in Figure 6.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2020 doi:10.20944/preprints202012.0237.v1

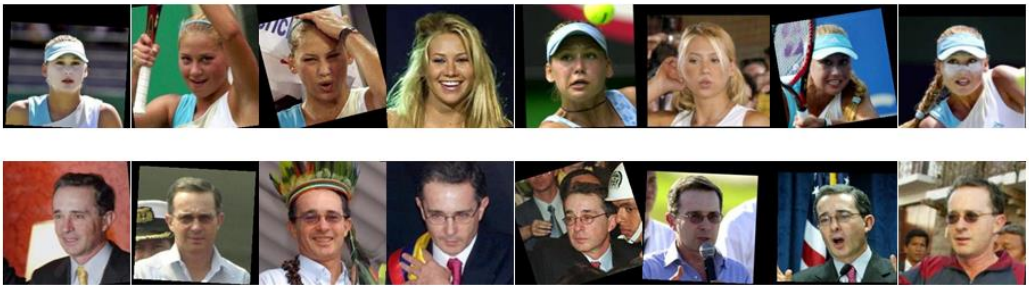


Figure 6. Examples of two different subjects from the LFW-a database [44].

4.2.2. Experimental Protocol

This study followed the experimental protocol presented in [27, 29-31]. From the LFW-a database, we selected only those subjects possessing more than 10 images to obtain a subset containing the facial images of 158 individuals. We cropped each image to a size of 120×120 pixels and then resized it to 80×80 pixels. We considered the first 50 subjects' facial photographs to create the training set and the test set. We randomly selected one shot from each subject for the training set, while the remaining images were employed in the test set. This process was repeated for five permutations and the average result for each was taken into consideration.

4.2.3. Limitations of SSFR Systems

In this section, the SSFR systems, and in particular the method we propose, will be voluntarily tested in a situation that is not adapted to their application: they are applicable in the case where only one sample is available and very often this sample is captured in very poor conditions.

We are particularly interested in cases where hundreds of samples are available, as in the LFW database, or when the training stage is based on millions of samples. In such situation, deep learning approaches must be obviously chosen.

The objective of this section is therefore to assess the limits of our approach.

Table 15 summarizes the performance of several rival approaches in terms of identification accuracy. Our best result was obtained by adopting the following configuration:

- BSIF descriptor with filter size $l \times l = 17 \times 17$and bit string length $n = 12$.
- K-NN classifier associated with city block distance.
- Segmentation of the image into blocs of 40×40 and 20×20 pixels.
- RGB color-space.

We can observe that the traditional approaches did not achieve particularly good identification accuracies. This is primarily because the photographs in the LFW database have been taken in unregulated conditions, which generates facial images with rich intra-class differences and increases face recognition complexity. As a consequence, the efficiency of the SSFR procedure is reduced. However, our recommended solution is better than the other competing traditional approaches. The superiority of our method can be explained by its exploitation of different forms of information, namely: local, regional, global, and color texture information. SVDL [29] and LGR [30] also achieved success in SSFR because the intra-class variance information obtained from other subjects in the standardized training set (i.e., augmenting the training-data) helped boost the performance of the system. Additionally, KNNMMDL [27] achieved good performance because it uses the Weber-face algorithm in the preprocessing step, which handles the illumination variation issue, and employs data augmentation to enrich the intra-class variation in the training set.

In another experiment, we implemented and tested the successful algorithm DeepFace [12], the weights of which have been trained on millions of images taken from the ImageNet database that are close to real-world situations. As presented in Table 15, the DeepFace algorithm shows significant superiority to the compared methods. This success is down to the profound and specific training of the weights in addition to the significant number of images employed in its operation.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 December 2020 doi:10.20944/preprints202012.0237.v1

In a recent work by Zeng et al. [67], the authors combined traditional (handcrafted) and deep learning (TDL) characteristics to overcome the limitation of each class. They reached an identification accuracy of near 74 %, which is something of a quantum leap in this challenging topic.

In the comparative study presented in [68], we can see that current face recognition systems employing several examples in the training set achieve very high accuracy with the LFW database, especially with deep-learning-based methods. However, SSFR systems suffer considerably when using the challenging LFW database and further research is required to improve their reliability.

In the situation where the learning stage is based on millions of images, the proposed SSFR technique cannot be used. In such a situation, the references [12] and [67], which use deep learning techniques, allow to obtain better accuracy.

Finally, the proposed SSFR method is reserved for the case where only one sample per person is available, which is the most common case in real world through remote surveillance or unmanned aerial vehicles' shots. In these applications, faces are most often captured under harsh conditions, such as changing lighting, posture, or if the person is wearing accessories such as glasses, masks or disguises. In these cases, the method proposed here is by far the most accurate.

Table 15. Identification accuracies using the LFW database.

Authors	Year	Method	Accuracy (%)
Chen et al. [57]	2004	Block LDA	16.40
Zhang et al. [24]	2005	SVD-FLDA	15.50
Wright et al. [65]	2009	SRC	20.40
Su et al. [66]	2010	AGL	19.20
Zhang et al. [32]	2011	CRC	19.80
Deng et al. [28]	2012	ESRC	27.30
Zhu et al. [31]	2012	PCRC	24.20
Yang et al. [29]	2013	SVDL	28.60
Lu et al. [33]	2013	DMMA	17.80
Zhu et al. [30]	2014	LGR	30.40
Ji et al. [51]	2017	CPL	25.20
Dong et al. [27]	2018	KNNMMDL	32.30
Chu et al. [62]	2019	MFSA+	26.23
Our method	**2020**	**MB-C-BSIF**	**38.01**
Parkhi et al. [12]	2015	Deep-Face	62.63
Zeng et al. [67]	2018	TDL	74.00

5. Conclusion and Perspectives

In this paper, we have presented a novel method for single sample face recognition (SSFR) based on the Multi-Block Color-binarized Statistical Image Features (MB-C-BSIF) descriptor to extract features and the k-nearest neighbors (K-NN) classifier associated with the city block (CTB) distance for classification. The proposed method, which is suitable for real-time applications thanks to its algorithms with reduced calculation costs, is characterized by its exploitation of different forms of information, including local, regional, global, and color texture information. In our experiments, the MB-C-BSIF has been evaluated on several subsets of images from the unconstrained AR and LFW databases. The experimental results illustrate the proposed technique's effectiveness and superiority when dealing with several variations of facial expression from the AR database. When tested on the

challenging LFW database, the results outperform traditional state-of-the-art methods and validate this work's effectiveness.

Even if the proposed method is used in applications where it is not initially intended, such as in cases where hundreds of samples are available, or where the training phase is based on millions of samples, it still has higher facial recognition accuracies than other SSFR techniques described in the literature.

In the future, we aim to explore the effectiveness of combining both deep learning and traditional methods in addressing the SSFR issue. We also aim to investigate and analyze the SSFR issue in unconstrained environments using large-scale databases that hold millions of facial images.

Author Contributions: *Investigation, Software, writing original draft, I.A.; methodology, validation, writing, review and editing, A.B.; project administration, supervision, validation, writing, review and editing, A.O.; validation, writing, review and editing, S.J. All authors have read and agreed to the published version of the manuscript.*

Funding: This research received no external funding.

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

References

1. Alay, N.; Al-Baity, H.H. Deep Learning Approach for Multimodal Biometric Recognition System Based on Fusion of Iris, Face, and Finger Vein Traits. *Sensors* **2020**, 20, 5523.
2. Pagnin, E.; Mitrokotsa, A. Privacy-Preserving Biometric Authentication: Challenges and Directions. *Security and Communication Networks* **2017**, 1-9.
3. Mahfouz, A.; Mahmoud, T.M.; Sharaf Eldin, A. A survey on behavioral biometric authentication on smartphones. *Journal of Information Security and Applications* **2017**, 37, 28-37.
4. Ferrara, M.; Cappelli, R.; Maltoni, D. On the Feasibility of Creating Double-Identity Fingerprints. *IEEE Transactions on Information Forensics and Security* **2017**, 12(4), 892-900.
5. Thompson, J.; Flynn, P.; Boehnen, C.; Santos-Villalobos, H. Assessing the Impact of Corneal Refraction and Iris Tissue Non-Planarity on Iris Recognition. *IEEE Transactions on Information Forensics and Security* **2019**, 14(8), 2102-2112.
6. Benzaoui, A.; Bourouba, H.; Boukrouche, A. System for automatic faces detection. In Proceedings of the 3rd International Conference on Image Processing, Theory, Tools, and Applications (IPTA), Istanbul, Turkey, 2012, pp. 354-358.
7. Phillips, P.J.; Flynn, P.J.; Scruggs, T.; Bowyer, K.W.; Chang, J.; Hoffman, K.; Marques, J.; Min; J.; Worek, W. Overview of the face recognition grand challenge. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), San Diego, CA, USA, 2005, pp. 947-954.
8. Femmam, S.; M'Sirdi, N. K.; Ouahabi, A. Perception and characterization of materials using signal processing techniques. *IEEE Transactions on Instrumentation and Measurement* **2001**, 50(5), 1203-1211.
9. Ring, T. Humans vs machines: the future of facial recognition. *Biometric Technology Today* **2016**, 4, 5-8.
10. Phillips, P.J.; Yates, A.N.; Hu, Y.; Hahn, A.C.; Noyes, E.; Jackson, K.; Cavazos, J.G.; Jeckeln, G.; Ranjan, R.; Sankaranarayanan, S.; Chen, J.C.; Castillo, C.D.; Chellappa, R.; White, D.; O'Toole, A.J. Face recognition accuracy of forensic examiners, superrecognizers, and face recognition algorithms. *Proceedings of the National Academy of Sciences* **2018**, 115(24), 6171-6176.
11. Kortli, Y.; Jridi, M.; Al Falou, A.; Atri, M. Face Recognition Systems: A Survey. *Sensors* **2020**, 20, 342.
12. Parkhi, O.M.; Vedaldi, A.; Zisserman, A. Deep Face Recognition. In Proceedings of the British Machine Vision Conference (BMVC), Swansea, UK, 2015, pp. 1-12.
13. Rahman, J.U.; Chen, Q.; Yang, Z. Additive Parameter for Deep Face Recognition. *Communications in Mathematics and Statistics* **2019**, 8, 203-217.
14. Benzaoui, A.; Boukrouche, A. Ear recognition using local color texture descriptors from one sample image per person. In Proceedings of the 4th International Conference on Control, Decision and Information Technologies (CoDIT), Barcelona, Spain, 2017, pp. 0827-0832.
15. Vapnik, V.N.; Chervonenkis, A. Learning theory and its applications. *IEEE Transactions on Neural Networks* **1999**, 10(5): 985-987.

16. Kannala, J.; Rahtu, E. BSIF: binarized statistical image features. In Proceedings of the 21th International Conference on Pattern Recognition (ICPR), Tsukuba, Japan, 2012, pp. 1363-1366.
17. Djeddi, M.; Ouahabi, A.; Batatia, H.; Basarab, A. ; Kouamé, D. Discrete wavelet for multifractal texture classification: application to medical ultrasound imaging," *2010 IEEE International Conference on Image Processing*, Hong Kong **2010**, 637-640.
18. Ouahabi, A. Multifractal analysis for texture characterization: A new approach based on DWT. 10th International Conference on Information Science, Signal Processing and their Applications (ISSPA 2010), Kuala Lumpur, **2010**, 698-703.
19. Ouahabi, A. Signal and Image Multiresolution Analysis. 1st ed.; ISTE-Wiley, London, 2012.
20. Ouahabi, A. A review of wavelet denoising in medical imaging. In Proceedings of the 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA), Tipaza, Algeria, 2013, pp. 19-26.
21. Sidahmed, S.; Messali, Z.; Ouahabi, A.; Trépout, S.; Messaoudi, C.; Marco, S. Nonparametric denoising methods based on contourlet transform with sharp frequency localization: Application to low exposure time electron microscopy images. *Entropy* **2015**, 17(5), 3461-3478.
22. Ding, C.; Bao, T.; Karmoshi, S.; Zhu, M. Single sample per person face recognition with KPCANet and a weighted voting scheme. *Signal, Image and Video Processing* **2017**, 11, 1213-1220.
23. Vetter, T. Synthesis of novel views from a single face image. *International Journal of Computer Vision* **1998**, 28(2), 103-116.
24. Zhang, D.; Chen, S.; Zhou, Z.H. A new face recognition method based on SVD perturbation for single example image per person. *Applied Mathematics and Computation* **2005**, 163(2), 895-907.
25. Gao, Q.X.; Zhang, L.; Zhang, D. Face recognition using FLDA with single training image per person. *Applied Mathematics and Computation* **2008**, 205(2), 726-734.
26. Hu, C.; Ye, M.; Ji, S.; Zeng, W.; Lu, X.; A new face recognition method based on image decomposition for single sample per person problem. *Neurocomputing* **2015**, 160, 287–299.
27. Dong, X.; Wu, F.; Jing, X.Y. Generic Training Set based Multimanifold Discriminant Learning for Single Sample Face Recognition. *KSII Transactions on Internet and Information Systems* **2018**, 12(1), 368-391.
28. Deng, W.; Hu, J.; Guo, J. Extended SRC: undersampled face recognition via intraclass variant dictionary. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2012**, 34(9), 1864-1870.
29. Yang, M.; Van, L.V.; Zhang, L. Sparse variation dictionary learning for face recognition with a single training sample per person. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Sydney, Australia, 2013, pp. 689–696.
30. Zhu, P.; Yang, M.; Zhang, L.; Lee, L. Local Generic Representation for Face Recognition with Single Sample per Person. In Proceedings of the Asian Conference on Computer Vision (ACCV), Singapore, Singapore, 2014, pp. 34-50.
28. Zhu, P.; Zhang, L.; Hu, Q.; Shiu, S.C.K. Multi-scale patch based collaborative representation for face recognition with margin distribution optimization. In Proceedings of the European Conference on Computer Vision (ECCV), Florence, Italy, 2012, pp. 822-835.
29. Zhang, L.; Yang, M.; Feng, X. Sparse representation or collaborative representation: which helps face recognition? In Proceedings of the International Conference on Computer Vision (ICCV), Barcelona, Spain, 2011, pp. 471-478.
30. Lu, J.; Tan, Y.P.; Wang, G. Discriminative multimanifold analysis for face recognition from a single training sample per person. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2012**, 35(1), 39–51.
31. Zhang, W.; Xu, Z.; Wang, Y.; Lu, Z.; Li, W.; Liao, Q. Binarized features with discriminant manifold filters for robust single-sample face recognition. *Signal Processing: Image Communication* **2018**, 65, 1-10.
32. Gu, J.; Hu, H.; Li, H. Local robust sparse representation for face recognition with single sample per person. *IEEE/CAA Journal of Automatica Sinica* **2018**, 5(2), 547-554.
33. Zhang, Z.; Zhang, L.; Zhang, M. Dissimilarity-based nearest neighbor classifier for single-sample face recognition. *The Visual Computer* **2020**, https://doi.org/10.1007/s00371-020-01827-3.
34. Zeng, J.; Zhao, X.; Qin, C.; Lin, Z. Single sample per person face recognition based on deep convolutional neural network. In Proceedings of the 3rd IEEE International Conference on Computer and Communications (ICCC), Chengdu, China, 2017, 1647-1651.
35. Ding, C.; Bao, T.; Karmoshi, S.; Zhu, M. Single sample per person face recognition with KPCANet and a weighted voting scheme. *Signal, Image and Video Processing* **2017**, 11, 1213-1220.

36. Zhang, Y.; Peng, H. Sample reconstruction with deep autoencoder for one sample per person face recognition. *IET Computer Vision* **2018**, 11(6), 471-478.
37. Mimouna, A.; Alouani, I.; Ben Khalifa, A.; El Hillali, Y.; Taleb-Ahmed, A.; Menhaj, A.; Ouahabi, A.; Ben Amara, N.E. OLIMP: A Heterogeneous Multimodal Dataset for Advanced Environment Perception. *Electronics* **2020**, *9*, 560
38. Du, Q.; Da, F. Block dictionary learning-driven convolutional neural networks for few-shot face recognition. *The visual Computer* **2020**, https://doi.org/10.1007/s00371-020-01802-y.
39. Ataman, E.; Aatre, V.; Wong, K. A fast method for real-time median filtering. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **1980**, 28(4), 415-421.
40. Benzaoui, A.; Hadid, A.; Boukrouche, A. Ear biometric recognition using local texture descriptors. *Journal of Electronic Imaging* **2014**, 23(5), 053008.
41. Stone, J.V. Independent component analysis: An introduction. *Trends in Cognitive Sciences* **2002**, 6(2), 59-64.
42. Zehani, S.; Ouahabi, A.; Oussalah, M. ; Mimi, M. ; Taleb-Ahmed, A. Bone microarchitecture characterization based on fractal analysis in spatial frequency domain imaging. *Int J Imaging Syst Technol*. **2020**; 1– 19.
43. Ojala, T.; Pietikainen, M.; Maenpaa, T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2002**, 24(7): 971-987.
44. Ojansivu, V.; Heikkil, J. Blur insensitive texture classification using local phase quantization. In Proceedings of the 3rd International Conference on Image and Signal Processing (ICSIP), Paris, France, 2012, pp. 236-243.
45. Martinez, A.M.; Benavente, R. The AR face database. *CVC Technical Report* **1998**, 24, 1-10.
46. Huang, G.B.; Mattar, M.; Berg, T.; Learned-Miller, E. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments; Technical Report; University of Massachusetts: Amherst, MA, USA, 2007; pp. 7–49.
47. Mehrasa, N.; Ali, A.; Homayun, M. A supervised multimanifold method with locality preserving for face recognition using single sample per person. *Journal of Center South University* **2017**, 24, 2853-2861.
48. Ji, H.K.; Sun, Q.S.; Ji, Z.X.; Yuan, Y.H.; Zhang, G.Q. Collaborative probabilistic labels for face recognition from single sample per person. *Pattern Recognition* **2017**, 62, 125–134.
49. Turk, M.; Pentland, A. Eigenfaces for recognition. *Journal of Cognitive Neuroscience* **1991**, 3(1), 71-86.
48. Wu, J.; Zhou, Z.H. Face recognition with one training image per person. *Pattern Recognition Letters* **2002**, 23(14), 1711-1719.
49. Chen, S.; Zhang, D.; Zhou, Z.H. Enhanced (PC)2A for face recognition with one training image per person. *Pattern Recognition Letters* **2004**, 25(10), 1173-1181.
50. Yang, J.; Zhang, D.; Frangi, A.F.; Yang, J.Y. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2004**, 26(1), 131–137.
51. Gottumukkal, R.; Asari, V.K. An improved face recognition technique based on modular PCA approach. *Pattern Recognition Letters* **2004**, 25(4), 429–436.
52. Chen, S.; Liu, J.; Zhou, Z.H. Making FLDA applicable to face recognition with one sample per person. *Pattern Recognition* **2004**, 37(7) 1553–1555.
53. Zhang, D.; Zhou, Z.H. (2D)2PCA: two-directional two-dimensional PCA for efficient face representation and recognition. *Neurocomputing* **2005**, 69(1–3), 224–231.
54. Tan, X.; Chen, S.; Zhou, Z.H.; Zhang, F. Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft K-NN ensemble. *IEEE Transactions on Neural Networks* **2005**, 16(4), 875–886.
55. He, X.; Yan, S.; Hu, Y.; Niyogi, P.; Zhang, H.J. Face recognition using Laplacian faces. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2005**, 27(3), 328-340.
56. Deng, W.; Hu, J.; Guo, J.; Cai, W.; Fenf, D. Robust, accurate and efficient face recognition from a single training image: a uniform pursuit approach. Pattern Recognition **2010**, 43(5), 1748-1762.
57. Chu, Y.; Zhao, L.; Ahmad, T. Multiple feature subspaces analysis for single s ample per person face recognition. *The Visual Computer* **2019**, 35, 239-256.
58. Seetafaceengine, 2016, https:// github.com/ seetaface/ SeetaFaceEngine.
59. Cuculo, V.; D'Amelio, A.; Grossi, G.; Lanzarotti, R.; Lin, J. Robust Single-Sample Face Recognition by Sparsity-Driven Sub-Dictionary Learning Using Deep Features. *Sensors* **2019**, 19, 146.

 doi:10.20944/preprints202012.0237.v1

60. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust Face Recognition via Sparse Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2009**, 31(2), 210-227.
61. Su, Y.; Shan, S.; Chen, X.; Gao, W. Adaptive generic learning for face recognition from a single sample per person. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, 2010, pp. 2699-2706.
62. Zeng, J.; Zhao, X.; Gan, J.; Mai, C.; Zhai, Y.; Wang, F. Deep Convolutional Neural Network Used in Single Sample per Person Face Recognition. *Computational Intelligence and Neuroscience* **2018**, 2018, 3803627.
63. Adjabi, I.; Ouahabi, A.; Benzaoui, A.; Taleb-Ahmed, A. Past, Present, and Future of Face Recognition: A Review. *Electronics* **2020**, 9, 1188.