Journal of AI and Data Mining
Vol 7, No 2, 2019, 263-278
DOI: 10.22044/JADM.2019.7215.1857

A Novel Face Detection Method Based on Over-complete Incoherent Dictionary Learning

S. Mavaddati*

Electronic Department, Faculty of Engineering and Technology, University of Mazandaran, Babolsar, Iran

Received 08 July 2018; Revised 01 January 2019; Accepted 23 January 2019
**Corresponding author: s.mavaddati@umz.ac.ir(S. Mavaddati).*

Abstract

In this paper, the face detection problem is considered using the concepts of compressive sensing technique. This technique includes the dictionary learning procedure and the sparse coding method to represent the structural content of input images. In the proposed method, dictionaries are learned in such a way that the trained models have the least degree of coherence to each other. The novelty of the developed method involves the learning of comprehensive models with atoms that have the highest atom/data coherence with the training data and the lowest within-class and between-class coherence parameters. Each one of these goals can be achieved through the proposed procedure. In order to achieve the desired results, a variety of features are extracted from the images and used to learn the characteristics of the face and non-face images. Also the results of the proposed classifier based on the incoherent dictionary learning technique are compared with the results obtained from the other common classifiers such as neural network and support vector machine. The simulation results along with a significance statistical test show that the proposed method based on the incoherent models learned by the combinational features is able to detect the face regions with a high accuracy rate.

Keywords: *Face Detection, Compressive Sensing, Incoherence Dictionary Learning, Neural Network, Statistical Test.*

1. Introduction

Face detection, as one of the major areas in image processing, has attracted many researchers in the recent years. Face detection means determining the presence or absence of a face in a specific image as well as considering the position of the face or faces. A face detection system can be used in a variety of applications including identification on military, security, law, banking, face recognition, etc. [1-4]. Also this field can be considered as the first step in all the face recognition systems [5-6]. Different algorithms have been presented to solve the face detection problem so far. The two general categories of these proposed methods are based on the processing of colored and gray-scale images. In the first category, different color models are used. The most important feature in these methods that are based upon the color index is skin color as a basic factor [7-9]. In the second category, different features in various transformation areas are applied to distinguish between the face and non-face sections of the image [10]. These algorithms are discussed in the following. In [7], face detection has been investigated in color images using skin color properties, and the detection method is based upon an integrated method of template matching, the color space transformation, and Gaussian color model filter. In [8], a face detection algorithm has been investigated in color images using neural networks. In this method, the skin color feature used to explore the face and the neural network is employed to separate the skin areas. In [9], a fast algorithm for face detection based on skin color and the edge-to-edge information has been provided for use in real-time applications. The features used in this algorithm include the skin color histogram. The appropriate feature applied in [10] has led to good results for a face detection problem based on the support vector machine

(SVM) classifier. In [11], a face detection method based on the extracted features from different parts of the input image has been presented, and the facial sections are considered and geometric relationships between these components are learned. In the test step, the trained models are adapted to the observed image, and the existing faces in the image are identified. A face detection method based on skin color and facial features has been introduced in [12]. Skin areas are captured firstly using a skin color model, and non-face areas are ignored with this model. In [13], different methods presented in the field of face detection based on various types of neural networks and different extractive features of the image have been investigated. Then the efficiency of these classifiers is considered. In [14], a face detection method is proposed using the Gabor features and a feed-forward neural network. Also in [15], the face detection routine has been performed by the properties of the Gabor wavelet transform and the neural network. In [16], a face detector algorithm is proposed based on the coefficients of the wavelet transform, in which the dimensional reduction of the characteristics is performed using principal component analysis algorithm. Also the classification process is carried out based on the radial basis function neural network. In [17], different specified features extracted from the meaningful pieces of the image have been calculated and the face has been detected using a Bayesian classifier. In this method, the probability of occurrence for small pieces in the face is calculated and used to train the facial structure. The detection procedure in [18] is based upon the local features extracted from image patches including eyes, mouth, and nose. Face detection is carried out by comparing these extracted features with the features obtained from the input image. In [19], different face detection methods have been analyzed, and the results have been compared with each other. Also the face detection methods based on deep learning have been provided in [20-22].

In [23], a general object recognition algorithm has been introduced based on a sparse representation calculated by l_1-norm minimization. Also a new sparsity-based feature vector has been presented to yield a high-performance classification procedure for high-dimensional input data. A supervised dictionary learning approach has been presented in [24], named by label consistent K-SVD to sparsely code the input data. In the learning process, the labeled information is associated with each dictionary atom to enforce discriminability in the sparse code procedure. In [25], a projective dictionary pair learning method has been proposed for pattern recognition. In this method, a pair of dictionaries, synthesis, and analysis dictionaries are considered to perform the representation and classification procedures simultaneously. This learning procedure can reduce the computational time than conventional dictionary learning methods. A generalization of the K-SVD learning algorithm as a sparse-coding-based classification procedure has been introduced in [26]. This learning process avoids the local minima that may occur in the learning stage of other algorithms by adding a discriminative term into the objective function of the original K-SVD algorithm. An incoherent dictionary learning algorithm has been presented in [27]. In this method, a fixed coherence dictionary is designed close to a given dictionary with a decorrelation procedure as an intermediate step in the K-SVD learning algorithm. The approximation quality is preserved in this decorrelation process.

In this paper, a new method is presented for face detection based on the concepts of the compressive sensing technique. This technique involves dictionary learning and sparse representation to achieve the over-complete models of face and non-face in a digital image. The dictionary learning process is carried out in such a way that the learned models for face and non-face sections are incoherent as much as possible. Using the energy of the sparse coefficients obtained, an effective classifier can be designed to identify the face and non-face sections of an input image without having a common classifier such as neural network or support vector machine.

In Section 2, the dictionary learning and sparse representation procedures are introduced, and then different feature extraction methods are investigated in Section 3. In Section 4, the proposed face detection algorithm is proposed. In Section 5, the results of the developed face detection method are reported and compared with other common classifiers. In the last section, the paper is concluded.

2. Dictionary learning and sparse representation

The input image I can be modeled using the dictionary learning technique as:

$$\mathrm{I_m} = \mathrm{DX} \quad (1)$$

where, $\mathrm{I_m} \in \mathbb{R}^{\mathrm{P}\times\mathrm{N}}, \mathrm{N} \gg \mathrm{P}$ is a data matrix including different patches of the input image I. The input image is divided into N different patches $\mathrm{I_{m\in M}}$, where M involves the dimension of these patches. Usually the dimension of each patch is 8×8. Each column of I_m is a vectorized form of a sample patch, so $P = 64$. The data matrix I_m in the sparse representation can be coded by a linear combination of defined atoms in an over-complete dictionary $D \in \mathbb{R}^{P \times L}, L > P$, as shown in (1). This dictionary includes L columns or atoms $\{d_l\}_{l=1}^{L}$ with unit norm $\|d_{(:,l)}\|_2 = 1, \forall l = 1, \ldots, L$. The over-complete dictionary means that the number of columns or atoms of a dictionary is greater than the number of rows or dimension of the feature space. The ratio of the number of dictionary columns to the dictionary rows is called the redundancy rate. Also the coding matrix X with K cardinality parameter and $L \gg K$ consists of the sparse coefficients of I_m [28-30]. The value of the cardinality parameter K determines how many columns of D matrix can participate in the representation of each input data. Each column of X includes only K non-zero elements. The sparse representation problem with the approximation error and sparsity constraint terms can be formulated as [28-30]:

$$X^* = \arg\min_X \|I - DX\|_2^2 \ \text{s.t.} \ \|X\|_0 \leq K \tag{2}$$

where, $\|X\|_0$ denotes the number of non-zero coefficients in each row of X. The sparse representation technique makes it possible to show the major information of the input data based on a smaller dimension of the original spatial bases. The over-complete dictionary learning was first presented for image denoising using a definition of the K-SVD technique with proper results in this area [31]. The over-complete dictionary learning consists of two stages: sparse representation and updating of atoms. Due to the flexibility of these defined steps, each one of these steps can be carried out with any arbitrary procedure. The methods used for sparse coding are different according to how their coding parameters are set. These tuning procedures should be implemented carefully. In addition to K-SVD, other dictionary learning methods exist such as maximum likelihood (ML), method of optimal direction (MOD), and maximum posteriori (MAP). In all of these algorithms, the convergence rate decreases with increase in the training data [32-33].

3. Feature extraction

The feature extraction step is the first phase in the field of face detection similar to all common classification problems. Two common types of features are utilized in this processing filed including texture-based feature extraction and statistical-based feature extraction. In the following, different methods for extracting these properties are investigated.

3.1. Gabor Filter

One of the important features in the texture-based image analysis is to use the magnitude and phase coefficients generated by applying the Gabor filter to the image. Using this linear filter, frequency components in different directions of an image are calculated, which is very important in order to distinguish between different regions of the image. By applying any Gabor filter in the specified direction of an image or a region of it, a filtered image is obtained. The Gabor filter or wavelet is yielded by the expansion and rotation of a Gabor function [34-36]. The 2D Gabor function $g(x,y)$ is expressed as follows:

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + j\omega(x\mathrm{Cos}\theta + y\mathrm{Sin}\theta)\right] \tag{3}$$

where, σ_x and σ_y are the standard deviations along the x and y directions. Also ω and θ are the frequency and desired direction, respectively. In order to eliminate the intensity effect in the image brightness, the value of DC coefficient for each output of the Gabor function is ignored. The output of each Gabor function at the specified direction is two magnitude and phase matrices with the same dimension as the original image. The two final magnitude and phase matrices for the Gabor features are obtained based on averaging in different specified directions [34-36].

3.2. Principal component analysis

The principal component analysis is a statistical method used for data analysis and feature extraction, which is widely used in various image processing areas. The analysis routine of this algorithm is that the input data is converted into a new uncorrelated dataset with orthogonal components that consist of all characteristics of the main components. Also only a small number of these orthogonal components involve the maximum variation of the input data. To apply this transformation, the initial data is first normalized and then the covariance matrix is calculated. The input data is projected to these special vectors, which are orthogonal to each other. Then the variance of the projected data is maximized along with these new vectors or the orthogonal special bases [37].

3.3. Discrete cosine transform (DCT)

Using this transform, the image is decomposed into various frequency bands to which the human eye system is high sensitivity. These bands include low-, middle-, and high-frequency bands. The coefficients in the low bandwidth contain a large portion of the image energy, and the coefficients in high frequency bands include the minimum amount of energy. The coefficients of the 2D cosine transform $G(k, l)$ for an input image I with dimension $N \times M$ can be calculated as follows [34-36]:

$$G(x,y) = \frac{2\alpha(x)\alpha(y)}{\sqrt{M \times N}} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I(i,j) \times \mathrm{Cos}\big((2i+1)x\pi/2M\big) \times \mathrm{Cos}\big((2j+1)y\pi/2N\big) \quad (4)$$

$$\alpha(x), \alpha(y) = \begin{cases} 1/\sqrt{2} & \text{if } x, y = 0 \\ 1 & \text{Otherwise} \end{cases}$$

This transform results in an image in the frequency domain with the same dimension as the initial image.

3.4. Discrete Fourier transform (DFT)

By applying this transform, the input image is split into sinusoidal and cosine components and transfer into the frequency domain. Therefore, the original gray level scale image is approximated by an unlimited set of these functions. The Fourier transformation function for an input image I with dimension $N \times M$ can be expressed as [34-36]:

$$F(x,y) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} I(i,j) e^{-i(x_{i/M} + y_{j/M})} \quad (5)$$

$$e^{ix} = \mathrm{Cos}x + i\mathrm{Sin}x$$

In this regard, F represents the input image in the frequency domain with a set of complex numbers and the same dimension as the original image. Therefore, the representation of these Fourier coefficients can be in the form of magnitude and phase of the image. Usually the magnitudes of the Fourier coefficients are used in the image processing algorithms because these coefficients include sufficient information about the content of the input image [34-36].

3.5. Local binary pattern (LBP)

The LBP algorithm is one of the most robust feature extraction procedures that is widely used in research fields related to object detection and image retrieval [33]. This algorithm is a rotation-invariant descriptor used to analyze grayscale images in order to extract the properties of adjacent textures in an image. This feature makes it that the change in the position, light or rotation of a person has a less adverse effect on a face detection algorithm. There are several procedures to calculate the local binary pattern coefficients that depend on the choice of neighborhood type. These neighborhoods can be considered as diagonal or circular with different directions [38].

3.6. Histogram of oriented gradients

One of the feature extraction algorithms in machine learning, which is very effective in detecting purposes, is histogram of oriented gradient (HOG) method [39-40]. In this method, the number of gradient occurrence in different directions in the local sections of the image is calculated. This counting is done on different cells that are considered in overlapping blocks on the image. In this case, the image is first divided into blocks with 50% overlap, and then each block is divided into four cells. In the following, the magnitude and angle of gradient in each pixel of image are calculated. Then a histogram with 9 directions consisting of different gradient angles in this cell is obtained. The values of each column are the sum of gradient magnitudes that have the same angle with the value of this column. HOG uses gradient directions in the range of 0-180° to create a histogram. It should be noted that this descriptor varies with respect to the rotation but the difference in the brightness intensity will have little effect on the extracted properties [39-40]. In the experiments carried out in this work, a histogram descriptor with the size of cells 8×8 and blocks 2×2 has been used to extract the HOG features.

3.7. Gray level co-occurrence matrix

Another feature vector that determines the texture properties of an image is the extracted parameters of gray level co-occurrence matrix (GLCM) [41]. The textural information of an image can be expressed by means of a matrix with relative abundances $P(i, j)$, where the value of each row with the number of row i and the column number j represent the number of occurrence neighborhoods with the same gray level value in different directions (for example, 0°, 45°, 90°, 135°, and 180°). The coefficients of this matrix are not used directly as a feature for classification, and the statistical parameters calculated from these coefficients determine the content of the image. These parameters include mean, variance, energy, range of variations in relative abundance, contrast, homogeneity, uncertainty, maximum relative frequency, correlation, and entropy. Selection of all or part of these second-order statistical properties will be effective in texture analysis [41].

3.8. Moments

Another statistical feature that is very important due to its rotation-invariant property is the extracted moments of the image [36, 42]. The values of this feature set that consists of seven coefficients of the first to seventh moments do not change with the image rotation in each direction. Therefore, using these feature sets is very effective on the face detection algorithms that are encountered with some fundamental challenges such as rotation problem.

4. Face detection in proposed method

In order to solve a face detection problem in the proposed algorithm, at first, an over-complete model is learned for each dataset including the face and non-face images. Some samples of this dataset are shown in figure 1. The face images are recorded in various situations including different facial angles, facial expressions (laugh, ordinary, upset, angry), different ages (child, young, middle-aged, elderly), different gender (woman and man), face with additional essentials (hat, glass, etc.), and different skin colors.

The block diagram of the proposed face detection method, which includes all the steps in the training and test phases, is shown in figure 2. All sections of this block diagram including sparse coding, dictionary learning, projection constraints algorithm, and correction step are fully described in the following sub-sections.

4.1. Sparse representation in proposed method

In order to categorize the face/non-face images properly, a structured set of data related to each class can be used as a comprehensive model using a specific procedure called the dictionary learning technique. The first step in the dictionary learning algorithm is the sparse coding of training data. Therefore, each patch of image is represented by several atoms. The concept of sparsity in this representation means that each data patch will only be represented by a linear combination of different atoms determined based on the cardinality rate.

The next step in the dictionary learning algorithms is dictionary update according to the input data patches. Since the dictionary is over-complete and the dimensions of the problem in Eq. 1 is high, this problem leads to the under-determined solving procedure where the number of linear equations is much lower than the parameters. Therefore, the atom learning procedure should be performed in the mentioned steps including sparse representation and update of the dictionary atoms. In the first step, an initial dictionary is selected randomly from the training data. Then using the sparse coding procedure, the sparsity coefficients are obtained. In the second step, the dictionary atoms are updated according to this coefficient matrix. The fundamental difference between different dictionary learning algorithms is in the coding procedure or in the employed dictionary learning approach.

The sparse coding method employed in this paper is introduced in this section, and the employed dictionary learning algorithm is expressed in the next section. In the proposed face detection algorithm, the least angle regression with coherence criterion (LARC) sparse representation method is used, which is an extension of the LARS algorithm [43]. In the LARC coding, a stop condition based on the degree of coherence between the atoms and training data is utilized [43-44]. Using this algorithm, only the atoms whose coherence with data patches remains more than a certain value, named as residual coherence, are included in the dictionary. In this case, the atoms will be considered coherent with the data patches in order to better represent the content of the training set. Another characteristic in this algorithm is that the LARC coding technique uses the variable cardinality parameter instead of the constant rate used in most representation methods. In this algorithm, the upper bound of the cardinality parameter K is determined. It means that each data patch can be represented at most with K atoms. The LARC technique was first introduced in order to represent the speech signal [43] and this paper is applied in a face detection problem. The sparse representation based on this algorithm can be expressed as follows:

$$\mathrm{X}^* = \mathrm{LRAC}(\mathrm{D}, \mathrm{X}, \mathrm{K}, \mathrm{Coh}) \tag{6}$$

Figure 1. Training data used in face detection problem. First row: face images. Second row: non-face images.

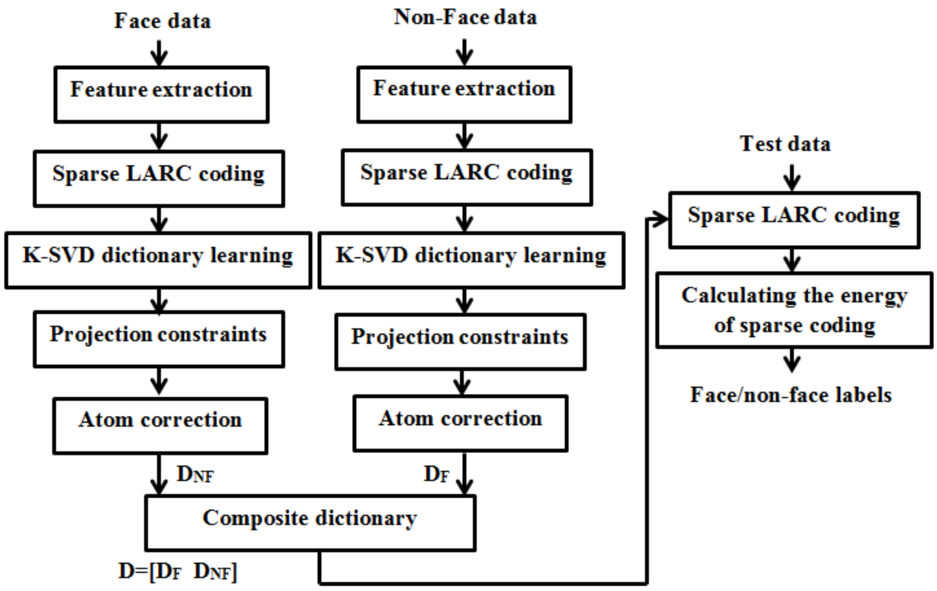


Figure 2. A block diagram of the proposed face detection method based on the learning incoherent models.

where, K is the variable cardinality parameter and Coh indicates the value of the residual coherence. The parameter setting procedure for cardinality value should be carried out carefully to the source confusion or source distortion does not occur. The source confusion occurs due to the selection of a large value for the K parameter, which results in a very dense sparse coding. On the other hand, source distortion occurs when the sparse coding is performed by a low cardinality value or too sparse coding. It means that the linear combination of these atoms is not sufficient to represent the training data so the approximation error will increase. Also the Coh parameter that expresses the coherence between the atom and the training data should be set precisely. If the value for this parameter is high, there will be a few atoms that have a coherence value higher than the Coh parameter so a problem may occur in this learning procedure such as a high approximation error. Also if the value of this parameter is low, this coherence parameter will not affect the procedure of sparse coding.

4.2. Dictionary learning procedure in proposed method

As expressed, the K-SVD dictionary learning algorithm is an appropriate procedure for atom training based on a set of input data [31]. In this algorithm, each patch of the input image is coded by a linear combination of the learned atoms. The consideration of coherence parameter in the learning procedure is important. The atom/data coherence parameter introduced in the previous section should have a high value. Also it is desirable that the coherence parameter between the atoms has a low value to a lower approximation error according to (1). In the case of the coherence between the dictionary atoms, it should be mentioned that there is a least coherence value between the dictionary atoms when the spatial bases are independent from each other, and the best representation for the content of the data patches is achieved. These parameters should be considered with more precision when the training data belonging to different classes is structurally similar. The higher coherence value between data/atoms and also lower mutual coherence between atoms (within-class coherence) result in the minimum reconstruction error. Therefore, the learned atoms can better represent the structural characteristics of the training data. The mutual coherence between the dictionary atoms is obtained based on the maximum value of the correlation between different dictionary atoms, as follows:

$$\mu(D) = \max_{1\le i,j\le L,\, i\neq j} |d_i . d_j| \tag{7}$$

The maximum absolute value obtained from the correlation between the atoms should be as small as possible to result in a trained model with independent spatial bases [43].

It is usually difficult to find a dictionary with this characteristic when the dictionary dimension is high. Therefore, the approximate methods are used to access the dictionary with the incoherent atoms. The approximation procedure used to solve this problem leads to the definition of Gram matrix $G = D^T D$. It should be noted that if the Gram matrix of a dictionary is in the form of the unitary matrix, then the atoms of this dictionary will be independent [44-45]. In order to achieve this Gram matrix for any desired dimension of the dictionary, different approximate solutions are considered. One of these solutions is the post-processing algorithms in the dictionary learning process. In [46], an iterative projection and rotation (IPR) technique has been introduced to obtain this desired Gram matrix. In the first step, the non-diagonal coefficients of this matrix are bounded by a constraint set called the structural constraints, and then the number of Eigenvalues are limited. In this thresholding procedure, the non-diagonal coefficients that must be zero in the ideal form of Gram matrix are set to a pre-defined small coherent value μ_0 to reduce Frobenius norm between the Gramm matrix and unitary matrix as $\|G - I\|_F$. In the following, the Eigenvalues of the Gramm matrix are limited. This procedure is performed by maintaining only the N largest Eigenvalues. In the second step, the obtained atoms are rotated with an orthogonal matrix W since the approximation error may increase with respect to the thresholding procedure in the first step. The rotation step tries to decrease the approximation error $\|Y - WDX\|_F$. This technique was first used to improve the reconstruction error for the music signal. This technique is used in this paper to model the training data in the face detection problem, and results in a better classification rate [5].

4.2.1. Correction of dictionary atoms

In the learning-based face detector algorithms, it is important that the designed dictionaries related to each data class do not have any similarity with each other to make a prominent distinction between different categories. Therefore, the learned dictionaries must have the least coherence value with the atoms of other dictionary classes. In this section, this issue is considered whether the atoms with the same structure exist in the representation of data in the dictionary related to face/non-face classes. If this problem is confirmed, then a routine is designed to reduce this similarity. In the proposed correction step, a composite dictionary $D = [D_F \; D_{NF}]$ including the dictionaries related to the face data D_F, and the non-faces data D_{NF} is regarded. Then the face data is coded over this composite dictionary:

$$X_F^*, X_{NF}^* = \mathrm{LARC}(Y, [D_F \; D_{NF}], \mathrm{Coh}) \rightarrow \underset{X_F, X_{NF}}{\arg\min} \left\| Y - [D_F \; D_{NF}] \begin{bmatrix} X_F \\ X_{NF} \end{bmatrix} \right\|_F^2 \quad (8)$$

In the following, the energy of the sparse coefficients is calculated for the face and non-face data:

$$E_F = {}^1\!/_L \sum_{l=1}^{L} X_{F,l}^{*2} \;\; , \;\; E_{NF} = {}^1\!/_L \sum_{l=1}^{L} X_{NF,l}^{*2} \quad (9)$$

Since the face data should not be represented on the non-face dictionary D_{NF}, the atoms of the D_{NF} dictionary that have the largest energy in this representation are replaced by atoms of the D_{NF} dictionary that have the least energy in this sparse coding. Also the sparse coding of non-face data is performed on this composite dictionary, and the energy of this representation related to face data is calculated. Since non-face data should not be represented on the D_F dictionary, the atoms of the D_F dictionary that have the most energy in this representation are chosen and replaced by atoms of the D_F dictionary that have the least energy. In this case, the possible error in the dictionary learning procedure for different data class is resolved.

4.3. Proposed procedure for face detection

It is important in the dictionary learning procedure to find the atoms that have the highest coherence value with the training data and the lowest coherence value with each other and also with atoms of other data classes. In this section, the proposed procedure for face detection based on the trained incoherent models is introduced. In this paper, in order to categorize the label of input data as face or non-face, conventional classifiers such as different types of neural networks and support vector machine are not used but it is suggested that a dictionary-based classifier is designed using the extracted properties in the sparse representation algorithm described in Section 4.1. In the first step of the proposed detection procedure, feature extraction will be performed, and then the LARC sparse representation over the composite dictionary introduced in the previous section is carried out. In this step, the cardinality parameter is set to the same value as defined in the training phase. Then the energy of sparse coefficients for each dictionary is calculated. It is clear that if the input image includes face content, the energy of this representation is more over face dictionary than non-face dictionary, and if the input image does not include the face content, the representation energy will be higher on the non-face dictionary. Therefore, the energy measure for the calculated sparse coefficients is used to classify the input data into the desired classes. In this case, there will be no need to use other classifiers, and estimation of the input data label will only be possible using the sparse representation technique. Also in order to categorize different parts of an image, the detection procedure is performed for all patches of the input image using the calculated energy of sparse representation over the composite dictionary. If the energy of these coefficients over the face dictionary is greater, then the face label is assigned to this patch. The results of the proposed algorithm that is very successful are presented in the next section.

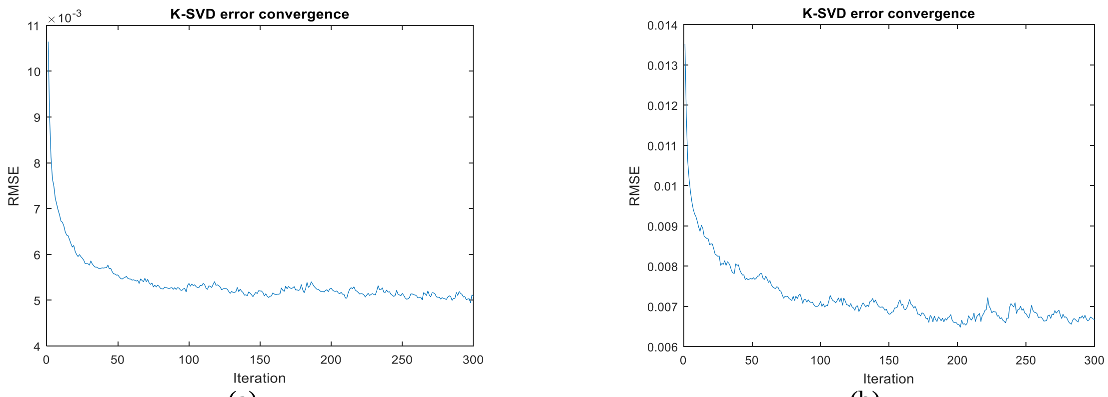
5. Simulation results

In order to simulate the results of the proposed method, face and non-face images of the MIT database are used [47]. This collection consists of face and non-face data with the 30 × 30 dimension. As expressed in Section 4, the image faces of this database are recorded in different angles, states, gender, ages, skin color, and various additional essentials. Also non-face data is captured in different spaces. In the proposed method, the K-SVD technique is used to detect the face in the input image. In the simulations, the various features introduced in Section 3 are used as the extracted training data from the images. The cardinality value of the LARC sparse coding depends on the dimension of the training data and is different for each feature extraction algorithm. The Coh parameter expressed in (6) and (8) is set to 0.2 for face and non-face data according to the experimental results. This means that the dictionary only contains the learned atoms with a coherence value more than 0.2 with the training data patch. The redundancy rate for face and non-face dictionaries is set to 4 for all features. This means that the learned dictionaries are over-complete with a redundancy rate of 4. In order to learn the dictionary in the training step, 400 face and 400 non-face images are used. Also 100 face and 100 non-face images are selected to test the performance of the comparison algorithms. The performance evaluation of different methods is determined using the classification accuracy rate calculated by the percentage of true categorized data divided by the total number of test data.

5.1. Feature extraction based on HOG coefficients

The extracted feature in this experiment is based on the HOG coefficients with 8 × 8 cell size and 50% overlap and also 2 × 2 block size. The dimension of HOG-based features for an initial image of 30 × 30 including 16 blocks with 8 × 8 cell size, 9 bins, and 4 directions is 144. The convergence diagram of the proposed dictionary learning for the extracted HOG-based features from the face and non-face images is shown in figure 3. As it can be seen, a suitable convergence with a low approximation error is obtained based on the calculated root mean square error (RMSE). Therefore, the convergence plot to learn over-complete dictionaries with a redundancy rate of 4 is achieved to train a data matrix with 144 × 400 dimension for each face or non-face dataset. The cardinality rate in the training and test steps of this simulation is set to 15, which is achieved due to the smallest calculated approximation error according to the experimental results.

The results of classification accuracy rate using the HOG feature for different classifiers including a feed-forward neural network with 25 hidden layers, support vector machine, and incoherent dictionary learning technique are reported in table 1. As shown, the classification accuracy for the proposed method is greater than the results of the other two categories. Also the results of calculating the relative confusion matrix using the HOG coefficients are given in table 2, which is the result of table 1. These results indicate that the learned dictionaries using HOG coefficients as a texture-based feature can properly model the content of face and non-face training data, and lead to a face detector algorithm with an appropriate accuracy. The HOG coefficients represent the details of each data class using the block-based analysis over different directions of the input image gradient.



(a) (b)

Figure 3. The convergence plot of the proposed dictionary learning algorithm using the HOG features: a) for face data, b) for non-face data.

5.2. Feature extraction based on Gabor coefficients

According to Section 3.1, the magnitude and phase of Gabor coefficients are the important features in the texture analysis of image using the extracted frequency components in different directions of image. Using the Gabor filter, the magnitude and phase matrices of Gabor coefficients are obtained as shown in Fig 4 having the same dimensions as the initial image. The Gabor coefficients in this simulation are calculated in four directions: 0°, 45°, 90°, and 135°. The magnitude matrix of Gabor coefficients in the specified directions for the face image is shown in Fig. 5, which shows the content of the frequency components in different directions. Then the sum of the row and column elements of the magnitude Gabor matrix is considered as the final features for dictionary learning. As a result, the 240 × 400 training data matrix is used for each face and non-face image. The cardinality rate using these features in the training and test steps is set to 20 with respect to the data dimension of 240, which is estimated from the experimental tests to achieve the lowest approximation error.

The results of classification accuracy rate using the defined Gabor features calculated for different classifiers including a feed-forward neural network with 25 hidden layers support vector machine, and an incoherent dictionary learning algorithm are presented in table 3.

The obtained results obtained show that the proposed classifier achieves the desired results using the magnitude feature of the Gabor coefficients. Therefore, this category of features can provide an appropriate distinction between different data classes. In addition, along with the magnitude of the Gabor coefficients, the phase of these features is also used, and the final feature vector is set based on the column mean of the magnitude and phase matrices in order to learn the dictionaries. Therefore, for each face and non-face image data, a 60 × 400 training data is used. The cardinality rate for these features in the training and test steps is adjusted to 12 according to the data dimension. It means that each training data is maximally represented with 12 atoms. The results of the classification accuracy rate using this category of features for the mentioned classifiers are reported in table 4. It can be observed that the results are lower than the results obtained with other defined features introduced so far. Therefore, a combination of the magnitude and phase of the Gabor coefficients cannot adequately distinguish the face and non-face data since the phase of the Gabor coefficients does not contain distinctive information between the structures of different input classes. The simulation shows that the effect of the Gabor coefficient magnitude is more prominent than the phase information in the face detection procedure.

Table 1. Classification accuracy rate for face and non-face images based on HOG feature for neural network, support vector machine, and dictionary-based classifiers.

	Support vector machine	Feed-forward neural network	Dictionary-based classifiers (proposed)
Face image	95	94	**97**
Non-face image	93.5	95	**96.5**

Table 2. Results of relative confusion matrix based on HOG feature for the proposed dictionary-based classifiers.

	Face image	Non-face image
Face image	97	3
Non-face image	3.5	96.5

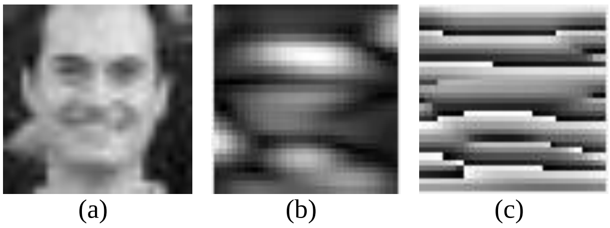


Figure 4. a) Face image b) A representation of the magnitude for the Gabor coefficient matrix c) A representation of the phase for the Gabor coefficient matrix.

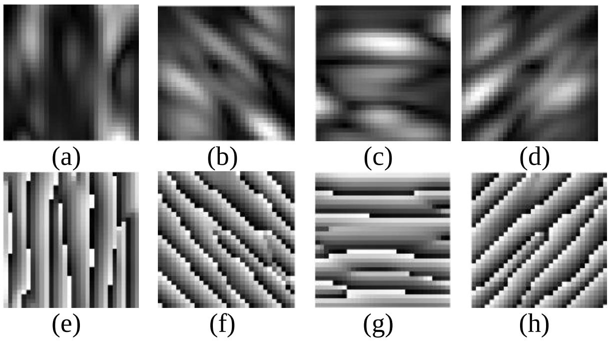


Figure 5. A representation of the magnitude for the Gabor coefficient matrix in the directions: a) 0°. b) 45°. c) 90°. d) 135°, and the phase of the Gabor coefficient matrix in the directions: e) 0°. f) 45°. g) 90°. h) 135°.

Table 3. Accuracy percentages of face/non-face image classifier based on the magnitudes of Gabor coefficients for the neural network, support vector machine, and the proposed method based on learning incoherent dictionary.

	Support vector machine	Feed-forward neural network	Dictionary-based classifiers (proposed)
Face image	91.5	94	**95**
Non-face image	96	95	**97.5**

Table 4. Accuracy percentages of face/non-face image classifier based on the magnitude and phase of the Gabor coefficients for different classifiers and the proposed approach.

	Support vector machine	Feed-forward neural network	Dictionary-based classifiers (proposed)
Face image	74	82.5	**86**
Non-face image	65	78.5	**88.5**

Table 5. Accuracy percentages of face/non-face image classifier based on the combination of magnitude and phase of Gabor coefficients with PCA for neural network, support vector machine, and the proposed dictionary-based classifier.

	Support vector machine	Feed-forward neural network	Dictionary-based classifiers (proposed)
Face image	89.5	92	**94**
Non-face image	86	84	**96.5**

Table 6. Accuracy percentages of face/non-face image classifier based on PCA coefficients for neural network classifications, support vector machine, and the proposed dictionary-based classifier.

	Support vector machine	Feed-forward neural network	Dictionary-based classifiers (proposed)
Face image	60	82.5	**64**
Non-face image	65.5	78.5	**69.5**

For more consideration, the PCA algorithm is used to achieve a more appropriate set of features to solve the face detection problem. Firstly, the Gabor magnitude and phase in different directions is extracted (eight images for the magnitude and phase of coefficients). After averaging these feature matrices and achieving the final matrix of magnitude and phase, the PCA algorithm is applied, and the mean of the columns are calculated. Finally, a 60×400 training data matrix is obtained. The cardinality rate is set to 12 in the training and test steps with respect to the dimension of the Gabor/PCA combinational feature. The classification results based on this defined feature are expressed in table 5, which shows that the effect of the phase properties of the Gabor coefficients to distinguish between classes has been increased by applying PCA.

The PCA techniques help to achieve better results by deleting additional information and applying distinctive features. Also the results of the extracted feature using PCA algorithm are shown in table 6, which shows that this feature alone cannot achieve the appropriate results to solve the face detection problem. In fact, the gray levels alone cannot distinguish between different data classes so the statistical or texture-based features should be utilized.

5.3. Feature extraction based on Fourier transform and discrete cosine transform

The purpose of this paper is to consider the main extracted features from the image and determine their efficiency to solve the face detection problem. Therefore, it was tried to analyze the results of these features, and, finally, choose a desirable feature category with the ability to make more distinctions between the face/non-face classes. In this regard, the results of the face/non-face classification using the extracted feature of DFT and DCT are shown in table 7. Also the results of these characteristic coefficients after applying the PCA algorithm are reported. The results show that the combination of these features has not succeeded to solve this problem since this feature vector cannot represent the structural information of texture in face/non-face images. These results show that the frequency content of the input data does not have a significant effect to create a distinction between the input data in the face detection problem. The dimension of each learning matrix is 60×400 and the cardinality rate is set to 10 in the training and test steps.

5.4. Feature extraction based on hybrid feature coefficients

In this section, the feature extraction procedure is considered using the GLCM coefficients and its important parameters. These selected parameters from the GLCM feature are nine characteristics including mean, variance, energy, range of relative abundances, contrast, homogeneity, maximum relative frequency, correlation, and entropy. The GLCM coefficients are calculated in the four directions of 0°, 45°, 90°, and 135°. Then the mentioned parameters in these directions are calculated. As a result, a feature vector with 36 coefficients is obtained for the input image. Since the simulations concluded that this dimension was not sufficient to categorize this issue and did not produce satisfactory results, the extracted feature vector was used in combination with other features. Among these features, the coefficients derived from HOG with a dimension of 144 for each image data, a local binary pattern LBP with a dimension of 10 for each image data and moments with a dimension of 7 for each image data introduced in Section 3 is applied. The results of combining these features to solve the face detection problem are presented in table 8. The dimension of the training matrix for using the GLCM/HOG, GLCM/HOG/LBP, and GLCM/HOG/LBP/Moment composite properties for each face/non-face class are 400×144, 400×190, and 400×197, respectively. The cardinality rate of these features in the training and test steps is set to 20 and 25, respectively. For more evaluations, the results of the proposed detection algorithm using Bao Face Database are also reported [48]. In the simulation with this database, 200 face images and 200 non-face images are randomly selected. Also 50 face images and 50 non-face images are considered to test the performance of the proposed algorithms. Also the feature extraction procedure is carried out over the gray levels of the color images of this database. The learned dictionaries are over-complete with 2 redundancy rates, and the cardinality rate is set to 20 and 25 in the training and test steps, respectively. The results show that the use of statistical characteristics such as the extracted parameters from the GLCM coefficients along with other rotation-invariant features such as LBP and moments leads to the desired results in the face/non-face classification. These desired results are achieved based on a combination of the statistical and texture-based features to learn the face and non-face models. These incoherence models are learned in such a way to have a maximum coherence to the training data and can sparsely represent the details of each data class. Also the role of between-class coherence parameter to achieve these results is very prominent since it leads to a proper distinction between the learned models of different classes. However, the results reported in Section 5.2 show that the magnitude of the Gabor coefficient matrix is effective to solve the face detection problem but the experiments show that these coefficients do not obtain the appropriate results in combination with other features, and therefore, these results are not reported in this section. In order to have more consideration about the performance of the proposed algorithm, the results of face detection are considered over different images. The implementation routine of the proposed algorithm is that, at first, 60 × 60 blocks are extracted from the whole input image with a 25% overlap, and then the dimension of these blocks is reduced by half. Then the sparse representation of these blocks is carried out using the LARC algorithm over the composite dictionary including the face and non-face models. The calculated energy calculated by these representations determines the face/non-face labels for each block. Two examples of these results are shown in figure 6, which illustrates the success of the proposed method. The features used in this detection process include the combinational features derived from GLCM, HOG, LBP, and moments that yield appropriate results to solve this problem. The results from the combination of these features are expressed in the end row of table 8. In this table, it is neglected to express the weakest results with other combinational feature vectors defined in the paper. Also the results of the proposed dictionary-based detection algorithm (using MIT database) based on the GLCM/HOG/LBP/Moment feature vector for the different cardinality rates 10, 15, 20, 25, and 30 are reported in table 9. As it can be seen, the best results are obtained when this parameter in the LARC sparse representation step of dictionary learning algorithm is set to 20, which is consistent with the previous result.

Table 7. Percentage of classification accuracy rate for face/non-face image based on DFT and DCT coefficients as well as these coefficients in combination with PCA algorithm for neural network classifications, support vector machine, and proposed dictionary-based approach.

		Support vector machine	Feed-forward neural network	Proposed dictionary-based classifiers (proposed)
DCT feature	Face image	54	49	**59**
	Non-face image	56.5	55.5	**63**
DCT/PCA feature	Face image	59	63	**66.5**
	Non-face image	61.5	64	**68.5**
DFT feature	Face image	58	52.5	**60.5**
	Non-face image	56	58.5	**63.5**
DFT/PCA feature	Face image	61.5	63	**67**
	Non-face image	63	66.5	**68.5**

Table 8. Percentage of classification accuracy rate for face/non-face image based on combinational feature coefficients for neural network classifications, support vector machine, and proposed dictionary-based classifier.

		Support vector machine	Feed-forward neural network	Proposed classifier (MIT database)	Proposed classifier (BaoFace database)
GLCM/HOG	Face image	88	89	**95**	94
	Non-face image	90.5	92.5	**95.5**	94.5
GLCM/HOG/LBP	Face image	94.5	93	**98**	96
	Non-face image	93.5	95	**97.5**	96.5
GLCM/HOG/LBP/Moment	Face image	95	96.5	**100**	98.5
	Non-face image	94.5	96	**98.5**	97.5

Table 9. Percentage of classification accuracy rate for face/non-face image based on the proposed dictionary-based classifier (MIT database) with GLCM/HOG/LBP/Moment feature vector for different cardinality values.

Cardinality rate ↓	Face image	Non-face image
10	91.5	90
15	97	96.5
20	**100**	**98.5**
25	96	94.5
30	92.5	91.5

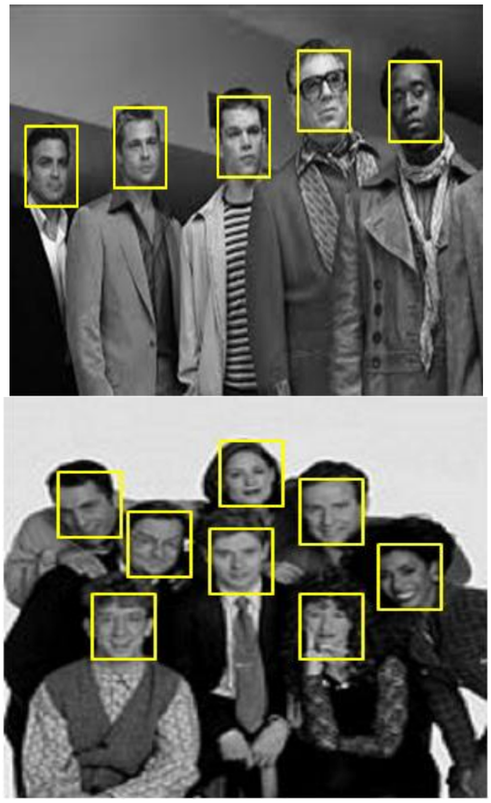


Figure 6. Face detection results by the proposed detection algorithm based on the extracted combinational feature GLCM/HOG/LBP/Moment.

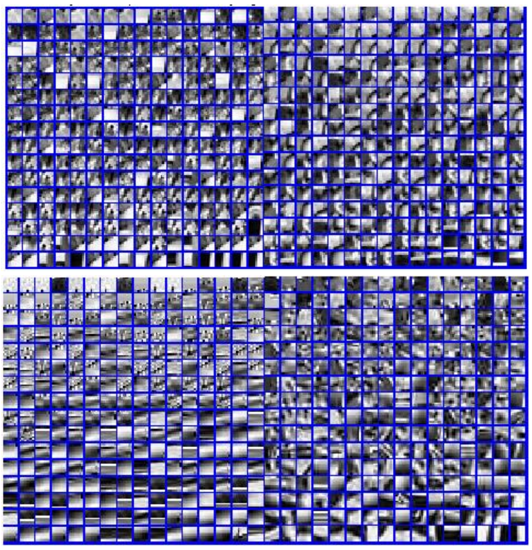


Figure 7. The learned dictionary by the proposed detection algorithm (MIT database) using GLCM/HOG/LBP/Moment feature vector: First row for face, and second row for non-face training data.

Figure 7 shows a section of the learned dictionaries for the face and non-face data by the proposed learning algorithm using the GLCM/HOG/LBP/Moment feature vector.

The simulations are carried out using MATLAB software on a Windows 64-bit based computer with Core i5 3.2GHz CPU. The mean detection procedure time in s for the proposed classifier and other comparison algorithms are reported in table 10. These results are the average detection time of 50 face and non-face images for each mentioned algorithm based on the GLCM/HOG/LBP/Moment feature vector. As it can be seen, the computation time of the proposed method is slightly more than the other algorithms since the test step in the proposed method contains LARC sparse representation with coherence constraint, which has a more computation time.

As mentioned in different parts of Section 4, three coherence parameters are considered in this paper: 1) the atom/data coherence measure that considers the coherence value between the input training data and the dictionary atoms. This coherence value should be high. This means that the dictionary atoms can correctly represent the structure of the training data. The LARC sparse representation algorithm is the first coding algorithm that works based on the atom/data coherence measure. This measure is evaluated using the Coh or residual coherence parameter shown in (6), defined in Section 4.1. 2) The mutual coherence between different atoms of a fixed dictionary or within-class coherence parameter: this parameter should have a low value. This means that the dictionary atoms are incoherent or independent from each other as much as possible. This goal is yielded with an IPR technique based on thresholding and rotation of the Gram matrix of a fixed dictionary, as introduced in Section 4.2. 3) The mutual coherence between atoms of different dictionaries (such as face and non-face dictionaries) or between-class coherence parameter: this coherence parameter should have a low value. This means that the atoms of different dictionaries are incoherent or independent from each other as much as possible. This goal is satisfied with the correction step proposed in Section 4.2.1. These coherence values for different dictionary learning methods are reported in table 11. The first learning method includes orthogonal matching pursuit (OMP) as a basis sparse representation method followed by the K-SVD algorithm [49]. The worst coherence values are obtained using the OMP/K-SVD learning method because this algorithm does not use the coherence constraint.

The atom/data coherence parameter in the LARC spars coding causes to increase this measure, and the atoms can represent the structure of the training data with more precision. Using the LARC/K-SVD/IPR learning method, more incoherent atoms are obtained, and the value of the within-class coherence parameter reduces. After the correction step in the proposed learning procedure, the coherence atoms in two face and non-face dictionaries are corrected, and the value of between class coherence parameter reduces. The results obtained (Table 11) indicate that the learned dictionaries based on the proposed method have the highest coherence with the training data and the least coherence with each other.

In order to make a proper decision on the efficiency of the different methods in the mentioned different conditions, a statistical significance test has been used. In this work, the Friedman test with the Holms post hoc test is used to compare the results of more than two algorithms [50-51]. This test is applied to the results of all four classifiers (neural network, support machine, and learning-based techniques (using the MIT and Bao Face databases), in two states of the face and non-face detection, as well as eight extracted features specified in tables 1, 3-4, and 7-8. Therefore, in this test, the numbers of different methods and conditions are $J = 4$ and $I = 8 \times 2$, respectively. It should be noted that the better results will be obtained from this test if the number of conditions compared to the number of examined methods is higher. The non-parametric Friedman test is one of the best methods available to compare several methods in different situations without the need for initial assumptions. The average performance rating of R_j is initially calculated for the j-th method from the J method in I different conditions as follows:

Table 10. Mean computation time in s to detect face/non-face images based on different methods with GLCM/HOG/LBP/Moment feature vector.

Support vector machine	8
Feed-forward neural network	6
Proposed dictionary-based classifier	10

Table 11. The coherence parameter values for different dictionary learning methods.

Dictionary learning method ↓	Atom/data coherence	Mutual coherence within-class	Mutual coherence between-class
OMP/K-SVD	0.103	0.658	0.756
LARC/K-SVD	0.627	0.693	0.742
LARC/K-SVD/IPR	0.636	0.362	0.729
LARC/K-SVD/IPR/ Correction(Proposed)	**0.654**	**0.325**	**0.426**

Table 12. The results of statistical test to compare the performance of the neural network, support vector machine, and dictionary-based classifier in order to detect face regions of an input image based on different features.

	R_j	$\mathcal{Z}$	ρ -value	Holm (α/(J-i))
Proposed dictionary-based classifier (MIT database)	1.03	4.949	0	0.0167
Proposed dictionary-based classifier (BaoFace database)	1.65	3.561	0.0004	0.025
Feed-forward neural network	1.88	2.545	0.0109	0.050
Support vector machine	2.78	0	-	-

$$R_j = \frac{1}{I}\sum_{i=1}^{I} r_{ij} \tag{10}$$

where r_{ij} is the performance rank of the j-th method in the i-th test state. It should be noted that the method with the best performance in this statistical test will earn the lowest rate value.

This significant test starts with a null-hypothesis the algorithms of which have the same performance, and then it should be proved that this assumption is wrong and then the rank of different methods is calculated according to their efficiency [52]. This test begins with the introduction of the critical value as follows:

$$\chi_F^2 = \frac{12I}{J(J+1)}\left[\sum_{j=1}^{J} R_j^2 - J(J+1)^2/4\right] \tag{11}$$

Moreover, the modified statistical value of the Friedman test, which is based upon the F distribution, is defined by $(J - 1)$ and $(I - 1) \times (J - 1)$ degrees of freedom:

$$F_F = (I-1)\chi_F^2/\left(I(J-1)-\chi_F^2\right) \tag{12}$$

If the F_F value in this statistical test is greater than the critical value χ_F^2, the null-hypothesis is rejected. It means that the mentioned algorithms do not have the same performance. The next step is to determine which algorithm will perform better. This decision is made using the post-hoc test. In this case, at first, the $\mathcal{Z}_j$ parameter for each evaluation method in this test is calculated as follows [53]:

$$\mathcal{Z}_j = \left(R_0 - R_j\right)/\sqrt{J(J+1)/6I} \tag{13}$$

In this regard, R_0 is the method that achieves the lowest performance level, and the highest average rate obtained is 10. The $\mathcal{Z}$ value to achieve the ρ parameter is calculated at a statistical significance level of $\alpha = 0.05$. The value of ρ in the probability density of the standard normal distribution, corresponding to the underlying surface of this distribution, will be outside the range (-$\mathcal{Z}$, $\mathcal{Z}$) [54]:

$$\rho = 1 - \int_{-\mathcal{Z}}^{\mathcal{Z}} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \qquad (14)$$

The simulation results of this statistical test based on the Holm assessment measure for the classification methods mentioned in this paper are reported in table 12 [52]. These results are calculated to determine the averaging rank for different classifiers based on various features. The value of the parameter χ_F^2 for the results of this table according to (11) is 43.71. Also the value of the F_F parameter according to (12) will be 27.92. The critical F value with the degrees of freedom (4-1) and (4 - 1) × (16 - 1) is 2.8115 [53].
It can be seen from the reported results that all the F_F values obtained in this statistical test are more than the critical values of F, and therefore, the initial null hypothesis based on the similarity of the performance for all mentioned algorithms is rejected. Therefore, the post-hoc testing can be used to determine the performance rank of different methods. As expressed, the rank R_0 belongs to a method with the lowest level of performance and the highest average rank among the mentioned methods. This rank in the statistical test is related to the support vector machine. In order to calculate the $\mathcal{Z}$ parameter to calculate the ρ-value at a statistical significance level $\alpha = 0.05$, (13) is used. Also the value of i in $(\alpha/(J - i))$ to calculate the Holm measure corresponds to the row number in table 12 and starts from 1, which includes a method with a better performance and the lowest average rank, up to $J-1$, which means a method with the lowest performance and higher average rank value. Also the results for the last row of table 12 are not achieved since the comparisons are based on the method with a lower performance according to (13). In this table, the ρ-value is lower than the corresponding Holm critical value for all methods. Therefore, it can be concluded that the methods are listed in table 12 from the best performance to the method with the worst performance in the face detection problem. It can be concluded that the results of the statistical test for the proposed method have a ranked value that is close to 1; this method has a better performance than the other ones. Also, in this table, the detection algorithm based on the support vector machine is introduced as a method with an average rating of R_0, and the value of the $\mathcal{Z}$ parameter for it is zero, so the parameter value of ρ for this method cannot be calculated.

6. Conclusion

Face detection is one of the most widely used fields in image processing, and due to its importance, in this paper, we presented a solution to solve this problem. In the proposed method, the sparse coding and dictionary learning techniques are applied to learn the over-complete models, and the extracted features derived from these concepts are also employed to design a classifier. The proposed procedure for sparse coding and updating the atoms leads to a dictionary learning procedure that includes three important coherent measures. These parameters are the high coherence value between data/atoms, the least coherence value between the atoms of a dictionary, and also between the atoms of different dictionaries. Different feature vectors were applied in this work based on different statistical-based and texture-based properties, and the performance of the proposed incoherent dictionary-based classifier was compared with other classifiers such as neural network and support vector machine. These comparisons were performed using the classification accuracy rate and statistical significance test. The simulation results showed that the proposed algorithm based on the designed learning procedure correctly performed face detection in comparison with other classifiers using the combinational feature vectors.

Reference

[1] Danish, A., Rehman, A., Ahmad, F. & Ahmad, A. (2013). Smart security system for sensitive area using face recognition. IEEE Conference on Sustainable Utilization and Development in Engineering and Technology (CSUDET), pp.11-14.

[2] Yinghong, D., Lele, W. & Jing, B. (2012). The research based on face detection and location system. In: Zhang W. (eds) Software Engineering and Knowledge Engineering: Theory and Practice. Advances in Intelligent and Soft Computing, Springer, Berlin, Heidelberg, vol. 162, pp. 337-344.

[3] Pirolaga, M. D. (2016). The implementation of biometric technologies for military recognition systems. National Strategies Observer, vol. 2, no. 2, pp. 233-247.

[4] Nikpour, M., Karami, R. & Ghaderi, R. (2017). Face recognition using an affine sparse coding approach. Journal of AI and Data Mining, vol. 2, no. 2, pp. 223-234.

[5] Yang, M. H., Kriegman, D. J. & Ahuja, N. (2002) Detecting faces in images: a survey. IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 24, no. 1, pp. 1-25.

[6] Rowley, H. A., Baluja, S. & Kanade, T. (1998). Neural network-based face detection. IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 20, no. 1, pp. 39-51.

[7] Nasrabadi, A., Azadi, S. & Hadadnia, j. (2009). Manifestation of human face in color images based on Gaussian filter. Journal of Modeling in Engineering, vol. 3, no. 17.

[8] Malhouji, M. & Menhaj, M. (2006). Face detection with neural networks in color images. The 9th Student Conference on Electrical Engineering, Tehran, University of Tehran, pp. 1-6.

[9] Sandeep, K. & Rajagopalan, A. N. (2002). Human face detection in cluttered color images using skin color and edge information. Indian Institute of Technology, ICVGIP Proceeding.

[10] Ziaee, Z., Bagherinia, A. & Rad, F. (2015). Face detection using intelligent algorithms. The 7th National Conference on Electrical and Electronic Engineering, Gonabad, Islamic Azad University, Gonabad, Iran.

[11] Hosseini, F., Shahebrahimi, A., Hajarian, M. & Khalilidizji, N. (2016). Detection of a non-invariant face using a section-based model. Computing science Journal, vol. 1, no. 1.

[12] Hu, W. C., Yang, C. Y., Huang, D. Y. & Huang, C. H. (2011). Feature-based face detection against skin-color like backgrounds with varying illumination. J. Inf. Hiding Multimed. Signal Process. vol. 2, no. 2.

[13] AL-Allaf, O. N., (2014). Review of face detection systems based artificial neural networks algorithms. The International journal of Multimedia & Its Applications, vol. 6, no. 1, pp. 1-16.

[14] Gupta, S., Gupta, A. & Tiwari, K., (2010). Face detection using Gabor feature extraction and artificial neural network, Proceedings from ISCET, pp. 18-23.

[15] Kaushal, A. & Raina, J. P. S. (2010). Face detection using neural network & Gabor wavelet transform. Int. J. Comput. Sci. Technol, vol. 1, pp. 58-63.

[16] Kamruzzaman, S. M., Ahmed, F., Saiful, M. D., Emdadul, M. D. & Shamsul, M. (2010). Rotation invariant face detection using wavelet, PCA and Radial Basis Function Networks. 12th International Conference on Human Computer Interaction, Beijing, China, Vol. 18, pp.1-5.

[17] Colmenarez, A. J. & Huang, T. S. (1997). Face detection with information based maximum discrimination. In IEEE Conference on Computer Vision and Pattern Recogntion, pp. 782-787.

[18] Zhu, X. & Deva, R. (2012). Face detection, pose estimation, and landmark localization in the wild. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2879-2886.

[19] Zafeiriou, S., Zhang, C. & Zhang, Z. (2015). A survey on face detection in the wild: past. present and future. Computer Vision and Image Understanding, vol. 138, pp. 1-24.

[20] Sun, X., Wu, P. & Hoi, S. C. (2017). Face detection using deep learning: An improved faster RCNN approach. Neurocomputing, vol. 299 pp. 42-50.

[21] Triantafyllidou, D., Paraskevi, N. & Anastasios, T. (2018). Fast Deep Convolutional Face Detection in the Wild Exploiting Hard Sample Mining. Big Data Research, vol. 11: pp. 65-76.

[22] Ye, X., Chen, X., Chen, H., Gu, Y. & Lv, Q. (2015). Deep learning network for face detection. IEEE 16th International Conference on Communication Technology, Hangzhou, China, pp. 504-509.

[23] Wright, J., Yang, A., Ganesh, A., Sastry, S. & Ma, Y., (2009). Robust face recognition via sparse representation. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210- 227.

[24] Jiang, Z., Lin, Z. & Davis, L. S. (2013). Label consistent K-SVD: learning a discriminative dictionary for recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 11, pp. 2651-2664.

[25] Gu, S. H., Zhang, L., Zuo, W. M. & Feng, X. C. (2014). Projective dictionary pair learning for pattern classification. Advances in Neural Information Processing Systems, pp. 793-801.

[26] Zhang, Q. & Li, B. X. (2010). Discriminative K-SVD for dictionary learning in face recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2691-2698.

[27] Mailhé, B., Barchiesi, D. & Plumbley, M. D. (2012). INK-SVD: Learning incoherent dictionaries for sparse representations. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '12), pp. 3573-3576.

[28] Davis, G., Mallat, S. & Zhang, Z., (1994). Adaptive time-frequency decompositions. Optical-Engineering, vol. 33, pp. 218-391.

[29] Agarwal, A., Anandkumar, A., Jain, P., Netrapalli, P. & Tandon, R. (2014). JMLR: Workshop and Conference Proceedings, vol. 35, pp. 1-15.

[30] Lee, H., Battle, A., Raina, R. & Ng, A. Y. (2006). Efficient sparse coding algorithms. Advances in Neural Information Processing Systems, pp. 801-808.

[31] Aharon, M., Elad, M. & Bruckstein, A., (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Trans. Signal Process, vol. 54, no. 11, pp. 4311-4322.

[32] Engan, K., Aase, S. O. & Hakon-Husoy, J. H. (1999). Method of optimal directions for frame design.

IEEE International Conference on Acoustics, Speech, and Signal Processing., vol. 5, pp. 2443-2446.

[33] Delgado, K., Murray, J. F., Rao, B. D., Engan, K., Lee, T. & Sejnowski, T. J. (2003). Dictionary learning algorithms for sparse representation", Neural Computation., vol. 15, no. 2, pp. 349-396.

[34] Nixon, M. & Aguado, A. (2012). Feature extraction & image processing for computer vision. Third Edition, Academic Press.

[35] Mathur, N. P. & Gupta, S. K. (2017). Performance analysis of feature extraction techniques for facial expression recognition. International Journal of Computer Applications, vol. 166, no. 1, pp. 1-3.

[36] Nixon, M. & Aguado, A., (2008). Feature Extraction & Image Processing. 2nd Edition, Academic Press, Cambridge.

[37] Guo, H., Marfurt, K. J., Liu, J. & Dou, Q. (2006). Principal components analysis of spectral components. 76th Annual International Meeting, SEG, Expanded Abstracts, pp. 988-992.

[38] Ojala, T. & Pietikäinen, M. (1999). Unsupervised texture segmentation using feature distributions. Pattern Recognition, vol. 32, no. 3, pp. 477-486.

[39] Freeman, W. T. & Roth, M. (1994). Orientation histograms for hand gesture recognition. In international workshop on automatic face and gesture recognition, vol. 12, pp. 296-301.

[40] Dalal, N. & Triggs, B. (2005). Histograms of oriented gradients for human detection. In Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 1, pp. 886-893.

[41] Majumdar, S. & Jayas, D. S. (2000). Classification of cereal grains using machine vision: I. Morphology models. Transactions of the ASAE, vol. 43, no. 6, pp. 1669-1675.

[42] Flusser, J., Suk, T. & Zitová, B. (2009). Moments and moment invariants in pattern recognition. Chichester, West Sussex, U.K: J. Wiley.

[43] Sigg, C. D., Dikk, T. & Buhmann, J. M. (2012). Speech enhancement using generative dictionary learning.IEEE Transactions on Audio, Speech and Language Processing, vol. 20, no. 6, pp.1698-1712.

[44] Efron, B., Hastie, T., Johnstone, I. & Tibshirani, R. (2004). Least angle regression. Ann. Stat., vol. 32, pp. 407-499.

[45] Tropp, J., Dhillon, I., Heath, R. J. & Strohmer, T. (2005), Designing structural tight frames via an alternating projection method. IEEE Trans. on Information Theory, vol. 51, no.1, pp.188-209.

[46] Barchiesi, D. & Plumbley, M. D. (2013). Learning incoherent dictionaries for sparse approximation using iterative projections and rotations. IEEE Transactions on Signal Processing, vol. 61, no. 8, pp. 2055-2065.

[47] http://cbcl.mit.edu/projects/cbcl/software-datasets/FaceData2.html.

[48]https://facedetection.com/downloads/BaoDataBase.zip.

[49] Rubinstein, R., Zibulevsky, M. & Elad, M. (2008). Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit. Technical report, Technion, Haifa, vol. 40, no. 8, pp.1-15.

[50] Benesty, J. (2008). Springer handbook of speech processing, Springer's publication, pp. 843-871.

[51] Demsar, J. (2006). Statistical comparisons of classifiers over multiple data set. The Journal of Machine Learning Research, vol. 7, pp. 1-30.

[52] Sheskin, D. J. (2000). Handbook of Parametric and Nonparametric Statistical Procedures, 4nd ed. Boca Raton, FL: Chapman & Hall/CRC.

[53]https://www.danielsoper.com/statcalc/calculator.aspx?id=4.

[54]http://onlinestatbook.com/2/calculators/normal_dist.html.

نشریه هوش مصنوعی و داده کاوی

ارائه یک روش جدید آشکارسازی چهره مبتنی‌بر یادگیری مدل‌های فراکامل ناهمدوس

سمیرا مودّتی

دانشکده فنی و مهندسی، دانشگاه مازندران، بابلسر، ایران.

ارسال ۲۰۱۸/۰۷/۰۸؛ بازنگری ۲۰۱۹/۰۱/۰۱؛ پذیرش ۲۰۱۹/۰۱/۲۳

چکیده:

آشکارسازی چهره در تصویر، همواره به عنوان یک گام پردازشی اساسی در کاربردهای مختلف حائز اهمیت بوده است. در این مقاله، آشکارسازی چهره به کمک مفاهیم حسگری فشرده مورد بررسی قرار می‌گیرد. این مفاهیم شامل بکارگیری الگوریتم‌های آموزش واژه‌نامه و بازنمایی تنک به منظور دستیابی به مدل‌های فراکامل که نمایانگر محتوای ساختاری تصاویر می‌باشند، خواهد بود. همچنین واژه‌نامه‌ها به صورتی یادگیری می‌شوند که مدل‌های آموزش‌دیده حداقل میزان همدوسی را با یکدیگر داشته باشند. به منظور دستیابی به نتایج مطلوب، دسته ویژگی‌های مختلفی از تصاویر استخراج شده و برای یادگیری مدل‌های بازنمایی‌کننده مشخصات چهره مورد استفاده قرار می‌گیرند. همچنین نتایج حاصل از طبقه‌بند مبتنی‌بر یادگیری مدل پیشنهادی با نتایج بدست آمده از طبقه‌بندهای شبکه عصبی و ماشین بردار پشتیبان مورد مقایسه قرار می‌گیرد. نتایج شبیه‌سازی‌های به همراه تست معناداری آماری نشان می‌دهد که روش پیشنهادی مبتنی‌بر واژه‌نامه‌های حاصل از ویژگی‌های ترکیبی معرفی‌شده قادر به آشکارسازی چهره در تصاویر مورد ارزیابی با دقت بالایی هستند.

کلمات کلیدی: آشکارسازی چهره، یادگیری مدل، واژه‌نامه فراکامل، ناهمدوسی، شبکه عصبی، ماشین بردار پشتیبان.