



A novel partition selection method for modular face recognition approaches on occlusion problem

Mehmet Koc¹

Received: 28 June 2020 / Revised: 28 June 2020 / Accepted: 6 November 2020 / Published online: 7 January 2021
© Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Recognizing the face with partial occlusion is an important problem for many face recognition applications. Since the occluded parts have no contribution to recognize the face, these parts should be excluded when performing the classification. In this paper, we propose a new method to detect and to use the non-occluded parts of face image for modular face recognition approaches. The occlusion of a partition is decided using the combination of three coefficients which can be easily derived: (i) image entropy, (ii) image correlation, (iii) root-mean-square error. The performance of the proposed partition selection method is tested using the modular extensions of three subspace-based approaches, namely linear regression classification (LRC), common vector approach (CVA), and discriminative common vector approach (DCVA). Modular DCVA is also proposed for the first time in this paper. After the selection of the non-occluded partitions of the face image, LRC, CVA, and DCVA are applied to each of the partitions independently. Then the classifier supports acquired from each of the partitions are combined using three well-known (product, sum, and Borda count) methods to get the final decision. The experiments implemented on the AR and the Extended Yale B face databases show that selection of the face partitions using the proposed strategy improves the recognition accuracy and outperforms state-of-the-art methods.

Keywords Decision fusion · Modular face recognition · Partial occlusion · Subspace methods

1 Introduction

Face recognition is an active and popular visual pattern recognition problem in machine learning. A lot of methods proposed in automatic face recognition area [9,31]. These methods generally work well under controlled conditions, but their performances decrease in real-world problems, i.e. uncontrolled environments. There are several factors such as lighting, partial occlusion, exposure and expression differences which affect negatively to a face recognition system's performance. Lighting, exposure, and expression factors are extensively investigated in the literature [2,4,8,21,25,28]. Face occlusion with other objects is quite common in uncontrolled real-world problems. There are also several works on the occluded faces to improve the performance of facial recognition system [3,22,26,35,36].

Linear regression classification (LRC) which is a subspace-based classification method uses linear regression idea and is successfully applied to face recognition problem in [22]. Several studies are done to improve the classification performance of LRC [17,23]. Authors also define an extension of LRC which is called modular LRC to overcome the scarf occlusion problem [22]. In modular LRC, face image is divided into different (4,6, and 8) nonoverlapping rectangular subregions. Then LRC is applied to each of the regions separately. Using *distance-based evidence fusion method*, the scores of each region are combined to get the final result for decision. Common vector approach (CVA) is another subspace-based classification method. The first work about CVA in pattern recognition area was published in 1999 [6]. Similar with LRC, Koc and Barkana propose a modular extension of CVA to overcome the occlusion problem [12]. The effect of different classifier combination methods on classification performance of modular CVA is investigated in [14]. Also in [15], author used Local Binary Patterns (LBP) as facial features to analyze the effect of several classifier combination methods on the classification performance of modular CVA. Discriminative common vector approach

✉ Mehmet Koc
mehmet.koc@bilecik.edu.tr

¹ Department of Electrical and Electronics Engineering, Bilecik Seyh Edebali University, Bilecik, Turkey

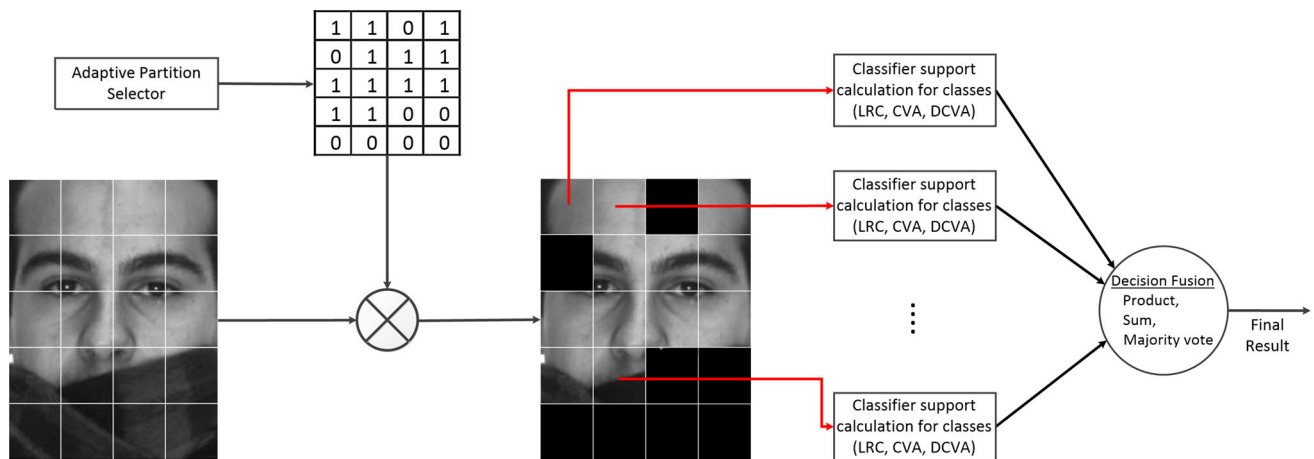


Fig. 1 The overview of the classification procedure. The partitions of the face image are selected using adaptive partition selector which generates one and zero for selected and unselected partitions respectively.

At each partition, classifier generates supports for each class. Finally, the supports of a class are combined using product, sum, or majority vote rule

(DCVA) was proposed by Cevikalp et al. in 2005 [1]. The method uses common vectors of subject classes to find a discriminative single subspace. The modular extension of DCVA has never been proposed before. In this work, we also define modular DCVA for the first time and use in the experiments. The modular extensions of these three subspace-based classification methods use distance-based evidence fusion method to combine the scores evaluated at the modules.

In the literature, several methods were proposed which divide face image into subregions and apply the method to each of the subregions. These methods reach the final result by combining the scores of the subregions after weighting the scores [21,24,29,32,34,37]. Wriht et.al use l_1 minimization to recognize human faces with varying expression and illumination from frontal viewpoints as well as occlusion [35]. The method known as sparse representation-based classification (SRC) uses the entire set of training samples as the dictionary for sparse discriminative code. In the method, nonzero sparse coefficients are assumed to lie heavily on the training samples, with the test probe having the same class label. Mi et.al propose a new matrix regression-based method (MRC-block) which uses matrix form of the images instead of vectors to get more discriminatory data than other regression-based approaches [20]. Authors experimentally show that the classification performance of MRC-block is better than many state-of-the-art methods in occluded face recognition problem. Yan et.al propose a method (MSR-MSF-VQ) based on multiple image subregions. The method captures the structure of the spatial information to overcome the vector quantization (VQ) limitations by using the Markov stationary features (MRF) [37]. In [24], authors combine the scores of each partition (included occluded ones) using some weights which are evaluated using entropy of partition.

In this work, we firstly find the occluded/nonoccluded partitions using our proposed occlusion detection method. We use three coefficients to detect the occluded partitions: (i) entropy, (ii) correlation, and (iii) root-mean-square error of the partition. Then, we combine the scores (also called distance or support) given by classifier for a class at each of the non occluded regions using sum, product, and majority vote rules to get the final result. Fig. 1 illustrates the classification procedure of the proposed partition selection method. We can summarize the contributions of this work as follows:

- A novel partition selection method is proposed. In the method, partitions are selected according to three coefficients which can be easily obtained from image. Thus, the partitions which may negatively affect the classifier performance are eliminated.
- The modular extension of well-known subspace-based classifier DCVA is proposed.

The rest of this paper is organized as follows. In Sect. 2 we briefly give the subspace methods used in this work. Sect. 3 introduces the generalization of the modular approaches, in Sect. 4 we explain our proposed partition selection approach. Decision fusion methods are given in Sect. 5 and the experimental work is given in Sect. 6 and also we discuss the results in the same section. Finally, the conclusion is given in Sect. 7.

2 Subspace methods

In this section LRC, CVA, and DCVA methods are briefly explained. In the following methods, N is the number of

feature vectors in each class, C is the number of classes, and $\mathbf{x}_i^k \in \mathbb{R}^{n \times 1}$ is the k th feature vector of the i th class.

2.1 Linear regression classification

In this method, each class is represented by an $n \times N$ dimensional transformation matrix named *predictor* or *regressor*. The predictor of the i th class is given as follow:

$$\mathbf{W}_i = [\mathbf{x}_i^1 \ \mathbf{x}_i^2 \ \dots \ \mathbf{x}_i^N]. \tag{1}$$

Assume that \mathbf{y} be an unknown vector to be assigned to the one of the $i = 1, 2, \dots, C$ classes. In the LRC approach, if \mathbf{y} is a member of the i th class, it is assumed that \mathbf{y} lies on the subspace spanned by the training feature vectors of the i th class, thus it can be represented as a linear combination of that vectors, i.e.,

$$\mathbf{y} = \mathbf{W}_i \beta_i, \quad i = 1, \dots, N, \tag{2}$$

where β_i is the $N \times 1$ dimensional coefficient vector. After the least squares minimization, the coefficient vector estimation becomes

$$\hat{\beta}_i = (\mathbf{W}_i^T \mathbf{W}_i)^{-1} \mathbf{W}_i^T \mathbf{y}, \quad i = 1, \dots, N. \tag{3}$$

We can find predictions of \mathbf{y} for each class using the estimated coefficient vectors in Eq.2 as follows:

$$\hat{\mathbf{y}}_i = \mathbf{W}_i \hat{\beta}_i = \mathbf{W}_i (\mathbf{W}_i^T \mathbf{W}_i)^{-1} \mathbf{W}_i^T \mathbf{y}, \quad i = 1, \dots, N. \tag{4}$$

As it is seen from Eq. 4 that, $\hat{\mathbf{y}}_i$ is found by projecting \mathbf{y} onto the range space of \mathbf{W}_i , i.e., subspace spanned by the training vectors of the i th class. The best estimate of test probe \mathbf{y} is expected to be the closest to \mathbf{y} . Then \mathbf{y} is classified according to the following decision criterion.

$$C^* = \arg \min_i \{ \|\mathbf{y} - \hat{\mathbf{y}}_i\| \}, \quad i = 1, \dots, C. \tag{5}$$

2.2 Common vector approach

The assumption in CVA is that each class can be represented with a specific vector which is unique for a class [7]. This vector represents the common characteristics of the class members. After subtracting the common vector from the members of a class, the variations from the common characteristics remain. Common vector of a class can be determined using three approaches: (i) Using Gram–Schmidt orthogonalization procedure, (ii) Within-class covariance-based approach, and (iii) linear regression-based approach [13]. In the following, we summarize the second approach:

Assume that within-class covariance matrix of the i th class is

$$\mathbf{S}_i = \mathbf{A}_i \mathbf{A}_i^T \tag{6}$$

where \mathbf{A}_i is a matrix of the form

$$\mathbf{A}_i = [\mathbf{x}_i^1 - \boldsymbol{\mu}_i \ \mathbf{x}_i^2 - \boldsymbol{\mu}_i \ \dots \ \mathbf{x}_i^N - \boldsymbol{\mu}_i]. \tag{7}$$

Here $\boldsymbol{\mu}_i$ is the average vector of the i th class. If the dimension of the feature vectors is larger than or equal to the number of the feature vectors ($n \geq N$), there exist $n - N + 1$ zero eigenvalues. The common vector of a class lies in the null space of the within-class covariance matrix [7]. Thus, common vector is a linear combination of the eigenvectors corresponding to the zero eigenvalues of within-class scatter. Let $(\lambda_k, \mathbf{u}^k), k = 1, \dots, (n - N + 1)$ be the eigenvalue-eigenvector pairs of the null space of within-class covariance matrix. Then the projection matrix of the null space's within-class covariance matrix for the i th class is

$$\mathbf{P}_i^\perp = \mathbf{U}_i \mathbf{U}_i^T. \tag{8}$$

\mathbf{U}_i is generated by stacking the within-class scatter's eigenvectors that correspond to the zero eigenvalues i.e.,

$$\mathbf{U}_i = [\mathbf{u}_i^1 \ \mathbf{u}_i^2 \ \dots \ \mathbf{u}_i^{n-N+1}]. \tag{9}$$

The common vector of the i th class is calculated by projecting any feature vector used in the training onto the null space (also called *indifference subspace*) of the within-class covariance matrix as follows:

$$\mathbf{a}_{com,i} = \mathbf{P}_i^\perp \mathbf{x}_i^j, \quad j = 1, \dots, N \tag{10}$$

As seen from Eq.10, common vector is unique for each class [7]. The unknown query \mathbf{y} should have the common properties of a class where it belongs to. Then, it is classified according to the following decision criterion:

$$C^* = \arg \min_i \left\{ \left\| \mathbf{a}_{com,i} - \mathbf{P}_i^\perp \mathbf{y} \right\| \right\}, \quad i = 1, \dots, C \tag{11}$$

2.3 Discriminative common vector approach

Discriminative common vector approach (DCVA) is a popular subspace-based face recognition method. In DCVA, the first step is the calculation of the common vectors. But when calculating the common vectors, instead of using the classes' within-class covariance matrices as in CVA, the sum of the within-class covariance matrices is used. In this method, the common vectors are unique for each class [1]. We can summarize DCVA method as follows.

The within-class covariance matrix of the i th class is given in Eq.6. DCVA uses the total within-class covariance matrix (the summation of all within-class covariance matrices of classes), i.e.,

$$\mathbf{S}_T = \mathbf{S}_1 + \mathbf{S}_2 + \dots + \mathbf{S}_C. \tag{12}$$

\mathbf{S}_T can be rewritten by using the matrices \mathbf{A}_i in Eq.7 as follows:

$$\mathbf{S}_T = \mathbf{A}\mathbf{A}^T, \tag{13}$$

where \mathbf{A} is a matrix of the form

$$\mathbf{A} = [\mathbf{A}_1 \ \mathbf{A}_2 \ \dots \ \mathbf{A}_C]. \tag{14}$$

Let \mathbf{Q} and $\tilde{\mathbf{Q}}$ are the projection matrices whose columns are the eigenvectors which span the range space and the null space of \mathbf{S}_T respectively. Then the common vectors of the classes are calculated by using either \mathbf{Q} or $\tilde{\mathbf{Q}}$ projection matrices as follows:

$$\begin{aligned} \mathbf{a}_{com,i} &= \mathbf{x}_i^j - \mathbf{Q}\mathbf{Q}^T\mathbf{x}_i^j = \tilde{\mathbf{Q}}\tilde{\mathbf{Q}}^T\mathbf{x}_i^j \\ i &= 1, \dots, C, \quad j = 1, \dots, N. \end{aligned} \tag{15}$$

The common vectors are independent from the selection of the training vectors, i.e., they are independent from index j in Eq.15 [1]. The difference subspace of the common vectors is obtained by using the scatter matrix of the common vectors \mathbf{S}_{com} as follows:

$$\mathbf{S}_{com} = \mathbf{A}_{com}\mathbf{A}_{com}^T \tag{16}$$

where \mathbf{A}_{com} is a matrix of the form

$$\mathbf{A}_{com} = [\mathbf{a}_{com}^1 - \boldsymbol{\mu}_{com} \ \mathbf{a}_{com}^2 - \boldsymbol{\mu}_{com} \ \dots \ \mathbf{a}_{com}^N - \boldsymbol{\mu}_{com}]. \tag{17}$$

Here, $\boldsymbol{\mu}_{com} = \frac{1}{C} \sum_{i=1}^C \mathbf{a}_{com,i}$ is the mean of the common vectors. The discriminative common vector of each class is obtained by projecting any feature vector from the training set of that class onto the range space of \mathbf{S}_{com} . That is,

$$\boldsymbol{\Omega}_i = \mathbf{W}^T \mathbf{x}_i^j, \quad i = 1, \dots, C, \quad j = 1, \dots, N \tag{18}$$

Here, \mathbf{W} is the range space projection matrix of \mathbf{S}_{com} . \mathbf{W} is formed by stacking the eigenvectors of \mathbf{S}_{com} corresponding to the nonzero eigenvalues. The test query \mathbf{y} is classified according to the following criterion:

$$C^* = \arg \min_i \left\{ \left\| \boldsymbol{\Omega}_i - \mathbf{W}^T \mathbf{y} \right\| \right\}, \quad i = 1, \dots, C \tag{19}$$

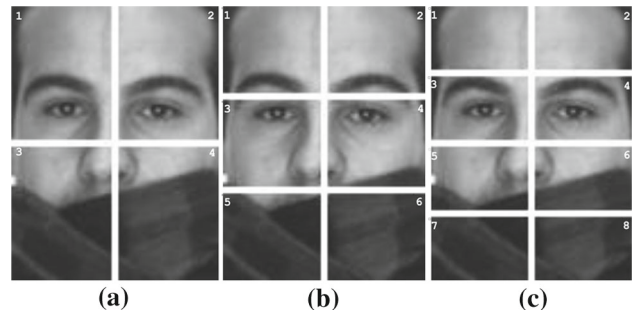


Fig. 2 A sample occluded face image divided into 4, 6, and 8 partitions [12]

3 Modular extensions of subspace methods

In modular extension of a conventional face recognition method, generally the face image is divided into rectangular subregions. Then the classification method is applied to these subregions independently. The final decision is obtained by combining the independent outputs using several ensemble approaches. However, the sub-block size is an important issue. Large sized sub-blocks increase the computational cost of the algorithm [25], whereas small-sized blocks do not meet the expectations from the modular approaches [24]. If $P \times Q$ dimensional face image is divided into $p \times q$ dimensional sub-blocks, there exist $K = \frac{P}{p} \times \frac{Q}{q}$ partitions.

The main aim of modular approaches is to separate the occluded and nonoccluded regions. Majority voting is popular method for decision fusion [27]. However it gives equal weight to all of the noisy and clean subregions.

In [22], authors propose simple but efficient decision rule for modular LRC approach, namely *distance-based evidence fusion*. In this method, firstly for the k th partition the distance between the test probe \mathbf{y} and its predictions $\hat{\mathbf{y}}_i$ in Eq.4 are computed. Then the minimum of these distances is chosen as a winner for this partition, i.e.,

$$d_i^k = \min_i \left\| \mathbf{y}^k - \hat{\mathbf{y}}_i^k \right\|, \quad i = 1, \dots, C \tag{20}$$

where \mathbf{y}^k and $\hat{\mathbf{y}}_i^k$ are the k th partitions of test probe and its prediction generated using the i th class predictor. Then, we get K decisions from K partitions. The final decision is decided in the favor of minimum d_i^k , that is,

$$C^* = \arg \min_i \left\{ d_i^k \right\}, \quad k = 1, \dots, K \tag{21}$$

Distance-based evidence fusion method is successfully applied to CVA in [12]. Similar with modular LRC approach, CVA is applied and a winner distance is chosen for each of the partition. Finally, test probe is classified in the favor of minimum of these distances. By following the similar idea, the modular extension of DCVA can be defined.

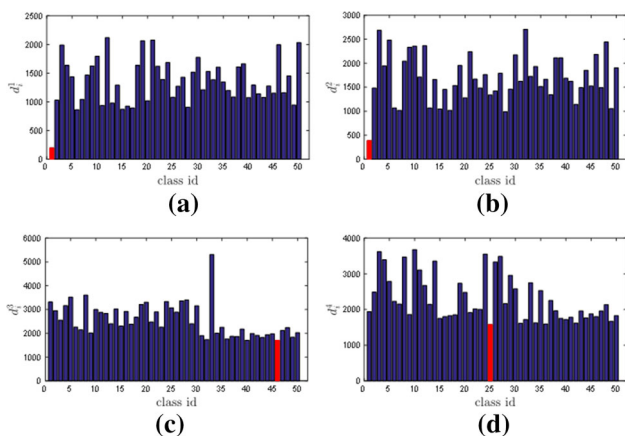


Fig. 3 The d_i^k distances in Eq.20 for CVA. d_i^k calculates the distance between the common vector of the i th class and the projection of the test probe y in the null space of the i th class

Sample partitions of an occluded image from AR face database [19] is given in Fig. 2. The 5th and the 6th partitions of the second image in this figure are completely occluded with scarf. It is obvious that these parts have no contribution to classify the face image. On the contrary, these parts may cause the image to be classified to the wrong class. Distance-based evidence fusion method successfully eliminates the occluded parts since the distances in Eq.20 evaluated from the occluded partitions are generally higher than the nonoccluded partitions.

As a case study, assume that we want to classify the image in Fig. 2 with CVA using four partition in a 50-class classification problem. It should be noted that the test image is selected from the first class. In Fig. 3, it is seen that by using the first and the second partition the test image is classified to the first class, whereas it is classified to the 46th and 25th classes by using the 3rd and the 4th partitions respectively. The 3rd and the 4th partitions correspond to occluded parts of the image and using them for classification causes wrong decision. Also, distance-based evidence fusion method finds the minimum distance in Fig. 3 and classifies the test image to the class i where the minimum distance appears.

4 Partition selection strategy

The main advantage of modular face recognition approaches is to eliminate the occluded parts in the image since they have no contribution to classification. If one combines the classifier outputs of the first and the second partitions in Fig. 3, the performance of the classifier may increase. However, before the combination of the classifier outputs, the occluded parts in the image must be determined and then eliminated, because it is obvious that classification using the occluded parts decreases the classifier performance.



Fig. 4 A sample scarf-occluded face image divided into 20 partitions

In this part, a new partition selection method is proposed by using three information acquired from the test image: (i) Entropy of partition, (ii) a correlation information acquired from partition, and (iii) root-mean-square error value computed using the partitions of the test and mean images of the training set. These three information can be easily derived from the test image. After partition selection, the classifier supports of the selected partitions are combined using several combination rules.

4.1 Image entropy

The entropy of an image can be defined as a measure of the amount of randomness in the image. Entropy is used to determine the weights of the partitions in several works[21, 24]. The entropy for partition X_k can be given as follows

$$H_k = - \sum_{i=1}^{N_i} p_i \log(p_i) \tag{22}$$

where N_i is the number of gray levels and p_i is the probability associated with gray level i . If the entropy of a partition is low, then the information contained by the partition is low. Thus partition may have little or no contribution to the classification.

Figure 4 shows a sample occluded face image from the AR face database divided into $5 \times 4 = 20$ partitions. The entropy values for the partitions are given in Fig. 5a. The

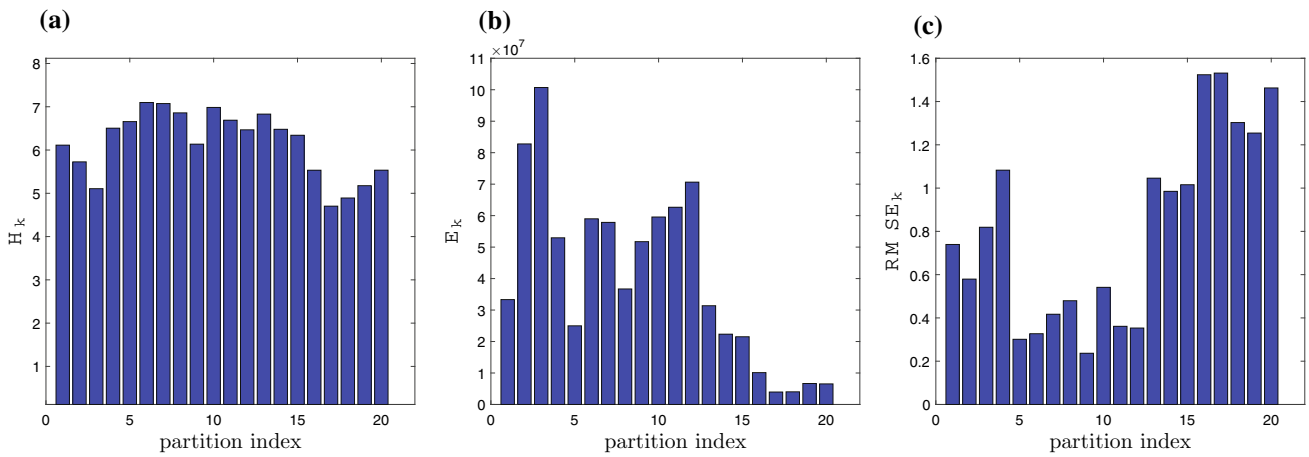


Fig. 5 For an image from AR database **a** entropy values, **b** eigenvalue sum of the correlation matrices, **c** RMSE values of 20 partitions

partitions 3, 16, 17, 18, 19, and 20 have low entropy values when compared with the other partitions, i.e. they contain little information which may lead wrong classification.

4.2 Image correlation

In this part we define a new criterion to detect the occlusion in an image. Let \mathbf{X}_k be the k th partition of the test image. Then we define a correlation matrix for the k th partition as follows:

$$\mathbf{R}_k = \mathbf{X}_k \mathbf{X}_k^T \tag{23}$$

Let E_k denotes the sum of the eigenvalues of \mathbf{R}_k . If the k th partition is occluded, it is expected that the relations between pixels are low. Therefore E_k will be lower than the eigenvalue sum of other partitions' correlation matrices. In Fig. 5b the eigenvalue sums of the correlation matrices for the partitions of the face image in Fig. 4 are given. The lowest five eigenvalue sums are evaluated for the partitions 16, 17, 18, 19, and 20 ($E_{16} - E_{20}$), which are completely occluded by scarf. Also, partitions 1, 5, 13, 14, and 15 have small eigenvalue sums when compared to the other partitions. Therefore, image correlation can be used for detecting the occluded parts of the face image.

4.3 Root-mean-square error

Before giving the root-mean-square error (RMSE) coefficient for the partition \mathbf{X}_k , we need to define the mean image of the database. The mean image of the database is evaluated by averaging the images in the training set taken under good illumination conditions with no facial expression. If \mathbf{X}_k is not occluded, it is expected that the difference between mean image and \mathbf{X}_k is small. Let \mathbf{I}_k represent the k th partition of the mean image of the database. Then the RMSE of \mathbf{X}_k is

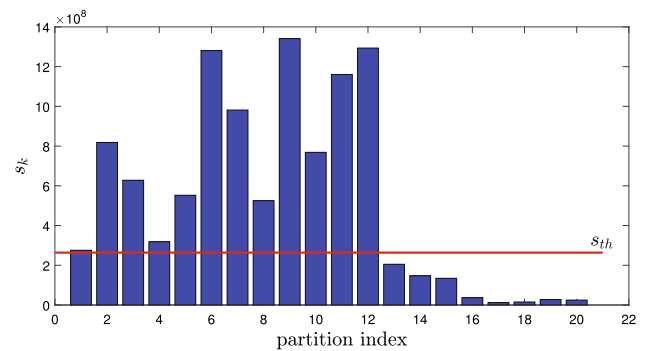


Fig. 6 Adaptive distance measures of an image with 20 partitions

calculated by using the following formula:

$$RMSE_k = \sqrt{\frac{\sum_{p=1}^P \sum_{q=1}^Q (\mathbf{X}_k(p, q) - \mathbf{I}_k(p, q))^2}{PQ}} \tag{24}$$

Here p and q represent the pixel coordinates, $\mathbf{X}_k(p, q)$ and $\mathbf{I}_k(p, q)$ represent the corresponding pixel gray level intensities, P and Q represent the dimensions of the partition.

If the partition is not occluded, then it is expected that the RMSE value for the partition is small with respect to the occluded partitions' RMSE values. In Fig. 5c, RMSE values for the partitions of the image in Fig. 4, are given. RMSE values for the partitions 4, 13-20 are large which are occluded by scarf except the 4th partition. Then, RMSE values can be used for occluded partition selection.

4.4 Adaptive partition selector

The disuse of occluded partitions in classification may increase the system performance. Careful selection of nonoccluded image partition becomes more important. By using the image entropy, image correlation, and root-mean-square

error measures, we define adaptive distance measure for the k th partition as,

$$s_k = \frac{H_k \times E_k}{RMSE_k} \quad (25)$$

E_k and H_k are expected to be higher and $RMSE_k$ is expected to be lower for nonoccluded image partitions. Then s_k may have higher values for nonoccluded partitions. We define a threshold value s_{th} as half of mean of the adaptive distance measures for the partitions of a face image. Then the k th partition is assigned as nonoccluded if its adaptive distance measure s_k is larger than s_{th} .

Adaptive distance measures, s_k , and threshold s_{th} are given in Fig. 6. Since the s_k values for partitions 1–12 are larger than the threshold value s_{th} , they are assigned as nonoccluded partitions. The partitions 1–12 are completely nonoccluded and contain the eyes, their surroundings, forehead, some parts of the cheeks. Our algorithm eliminates the rest of the partitions (13–20) since they are mostly or fully covered with scarf.

5 Decision fusion

After the selection of the nonoccluded partitions, distance-based evidence fusion method can be applied to classify the test sample. But the result will be the same when we apply the method without any prior partition selection. Instead, we apply several decision fusion rules to combine the classifier distance values of selected partitions. In the following, we briefly explain three very well-known distance combination methods, namely product, sum, and majority voting rules [11,30].

Assume that image is partitioned into K subimages and $D_i^k, i = 1, \dots, C$ are the distances evaluated for the k th partition using an individual classifier. When classifying the k th partition of a test sample, classifier compares C distances, in the other words, it produces a distance for each class.

5.1 Product rule

In the product rule, the response of the i th class is computed by multiplying the distances evaluated for K partitions. The method is very sensitive to outliers, since very high or very low scores have great effect on the final score. Distance values for product rule are computed as,

$$Dist_i = \prod_{k=1}^K D_i^k, \quad i = 1, \dots, C \quad (26)$$

Final decision is given according to the criterion given below:

$$C^* = \arg \min_i \{Dist_i\}, \quad k = 1, \dots, K. \quad (27)$$

5.2 Sum rule

The fused distance value for the i th class is computed by summing the responses of the class for K partitions as follows

$$Dist_i = \sum_{k=1}^K D_i^k, \quad i = 1, \dots, C \quad (28)$$

The decision criterion for sum rule is the same as product rule. Since the classification methods LRC, CVA, and DCVA methods decide according to the nearest neighbor rule, the decision criterion given in Eq. 27 is ruled in favor of the minimum distance.

5.3 Majority voting rule

It is the way of making decisions via voting. Majority voting has three variations: (i) most of the classifiers have the same decision, (ii) more than 50% of the classifiers have the same decision, (iii) all classifiers have the same decision[30]. We used the first version in the experimental work. The outputs of majority voting rule classifiers are binary values, i.e., $D_i^k \in \{0, 1\}$. If a test query satisfy the following equation, then it is assigned to the class I:

$$\sum_{k=1}^K D_i^k = \max_{j=1}^C \sum_{k=1}^K D_i^k. \quad (29)$$

The summation on the right side of Eq. 29 simply counts the votes from classifiers for each class and finds the winner one.

6 Experiments

We test the performance of the proposed partition selection method on AR [19] and Extended Yale B [5] face databases using LRC, CVA, and DCVA.

6.1 The AR face database Experiments

The database contains 26 images from each of 126 people. Images are captured in two sessions, under several illumination conditions, occlusions (sun glasses, scarf), and facial expressions (smile, neutral, anger, and scream). Original images are of 768×576 pixels. After preprocessing operations (converting grayscale, cropping, and scaling) the final size of the images is 299×222 . In the experiments we used

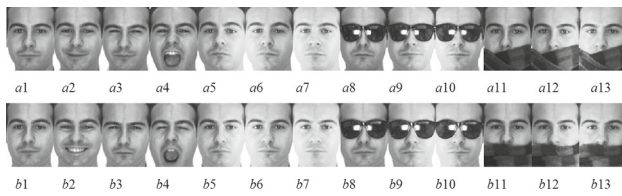


Fig. 7 Images of a subject from the AR face database after the preprocessing operations

the same subset of AR (30 male, 20 female) with Ref. [1]. Figure 7 shows a subject's images from the database after application of the preprocessing steps. The same poses are taken from the subjects on different days; thus, database contains two same samples from each illumination conditions, facial expressions, and occlusions.

In the first part of the experiments, we used all of the 14 nonoccluded face images (a1–a7, b1–b7 in Fig. 7) for training and six occluded images with scarf (a11, a12, a13, b11, b12, and b13 in Fig. 7) for testing purposes. Thus, we used $50 \times 14 = 700$ and $50 \times 6 = 300$ images for training and testing purposes. The scarf images also have different illumination conditions. All images are divided into $5 \times 4 = 20$ modules with the size 59×55 .

The results are given in Table 1. We use Prod_S, Sum_S, and MV_S to represent the product, sum and majority vote results after applying the proposed partition selection algorithm. Conventional LRC, CVA, and DCVA methods give the worst results. Among the modular methods, CVA has the highest classification accuracy. But the recognition results are not satisfactory level at all. The fusion of modules' classification results using product, sum, or majority vote rules increase the recognition results in all of the three subspace methods. Combining the classifier outputs using majority vote gives the best classification results. When we select the partitions of the face images using the proposed method and apply product, sum, and majority vote rule to the classifier supports, the subspace methods reach their best recognition performances. LRC method correctly classifies all of the occluded images when we combine the classifier outputs of the partitions using the product or majority vote rule after applying the proposed partition selection algorithm. CVA method misclassifies just one sample from 300 occluded images when we combine the classifier outputs using product rule. As a result, classification performances of the subspace methods clearly increase when the partitions are selected by our proposed algorithm.

In the second part of the experiments, we used the same training set with the first experiments and the six occluded images with sun glasses (a8, a9, a10, b8, b9, and b10 in Fig. 7) for testing. Images of the test set also have different illumination conditions. 700 images are used in training stage and 300 images are used for testing the performances of the sub-

Table 1 Comparison of the methods at scarf-occlusion problem in terms of recognition rate (%)

| Methods | LRC | CVA | DCVA |
|---------------|-------|-------|-------|
| Conventional | 12.67 | 19.00 | 72.33 |
| Modular | 76.33 | 97.67 | 81.00 |
| Product | 99.33 | 97.00 | 89.67 |
| Prod_S | 100 | 99.67 | 96.33 |
| Sum | 98.33 | 96.67 | 87.67 |
| Sum_S | 98.67 | 99.00 | 95.00 |
| Majority Vote | 99.67 | 98.67 | 97.00 |
| MV_S | 100 | 99.00 | 98.33 |

Table 2 Comparison of the methods at glass-occlusion problem in terms of recognition rate (%)

| Methods | LRC | CVA | DCVA |
|---------------|-------|-------|-------|
| Conventional | 34.33 | 55.00 | 17.67 |
| Modular | 64.33 | 68.67 | 34.33 |
| Product | 87.67 | 86.33 | 39.33 |
| Prod_S | 86.67 | 85.67 | 41.33 |
| Sum | 69.67 | 67.00 | 29.33 |
| Sum_S | 69.67 | 68.67 | 32.00 |
| Majority Vote | 73.00 | 73.33 | 40.00 |
| MV_S | 76.00 | 75.67 | 38.00 |

space methods. Similar with the first experiment, the images are divided into 59×55 dimensional $5 \times 4 = 20$ modules.

Table 2 shows the classification results of the second experiment. As in the first experiments, the conventional subspace methods have the worst classification performances. Modular type subspace methods are better than their conventional versions. Among the modular methods, CVA takes the lead but the results are not at satisfactory level. When we combine the classifier outputs using product, sum, and majority vote rules, the performances of the subspace methods increase at a certain level. Also its obvious from the results in Table 2 that selecting the partitions using our proposed algorithm mostly increase the performance of LRC, CVA, and DCVA. Among the subspace methods, LRC is the better than CVA and DCVA. DCVA is the worst classifier since it generates single subspace for all classes, whereas LRC and CVA generate different subspace for each of the classes. Also classification performance of DCVA strongly depends on the size of feature vector [1].

We also compare our proposed method, with different methods on occlusion problem in the AR face database. Different experiment scenarios are realized for the works given in the references [17,20,35,37]. We implement our method using the same experiment scenarios for fair comparison. We use 100 subjects (50 males, 50 females) from the AR

Table 3 Comparison of the methods at scarf and glass-occlusion problems in terms of recognition rate (%)

| | MRC-Block [20] | | | SRC [35] | | MSR-MSF-VQ [37] | Li et. al [17] |
|--------------|----------------|--------------|--------------|--------------|--------------|-----------------|----------------|
| | Scarf | Glass | Scarf+Glass | Scarf | Glass | Scarf | Scarf |
| | 60.56 | 64.56 | 60.28 | 93.50 | 97.50 | 98.00 | 95.70 |
| MLRC+Prod_S | 95.42 | 94.25 | 94.64 | 97.00 | 95.50 | 97.33 | 96.33 |
| MLRC+Sum_S | 92.25 | 85.92 | 88.53 | 91.50 | 80.50 | 95.67 | 95.33 |
| MLRC+MV_S | 95.17 | 94.67 | 94.64 | 97.00 | 91.00 | 97.33 | 93.33 |
| MCVA+Prod_S | 93.50 | 92.33 | 93.06 | 84.00 | 84.50 | 97.00 | 95.67 |
| MCVA+Sum_S | 90.42 | 85.33 | 87.77 | 60.00 | 62.50 | 94.67 | 91.67 |
| MCVA+MV_S | 93.08 | 93.25 | 92.88 | 90.00 | 80.00 | 97.33 | 94.33 |
| MDCVA+Prod_S | 61.33 | 63.08 | 59.00 | 69.50 | 67.00 | 73.67 | 76.33 |
| MDCVA+Sum_S | 57.67 | 57.17 | 54.00 | 64.00 | 52.50 | 66.33 | 69.00 |
| MDCVA+MV_S | 68.00 | 66.92 | 62.77 | 79.00 | 62.50 | 80.67 | 83.00 |

Bold represents the best recognition performances



Fig. 8 Sample test face with random block occlusions changing from 0 to 50%

face database. The results are summarized in Table 3. The value at the topmost of a column is the performance of the method which is compared with our proposed approach. Each of the column represents an independent experiment whose scenario is the same with the one given in the compared reference. It is clearly seen from the table that our proposed method outperforms the methods given in Table 3 except SRC with glass occlusion problem and MSR-MSF-VQ with scarf occlusion problem. Modular LRC is better subspace-based classifier than Modular CVA and Modular DCVA. When combining the classifier supports, product rule generally has the best performance. Thus, choosing LRC for classification method and product rule for classifier support combination will give the best result.

6.2 The YALE face database experiments

The extended Yale Face Database B consists of 38 subject classes. 64 images are taken under different illumination settings for each subject [5]. The images are cropped, resized, and normalized to 192 × 168 pixels [16]. The database is divided into 5 subsets based on the angle of the light source. The angles covered by the first two subsets are between 0° and 25°, the third subset covers from 25° to 50°, the fourth subset covers 50° to 77°, and the fifth subset covers above 78° [16]. The number of images in the subsets for each subject is 7, 12, 12, 14, and 19 respectively.

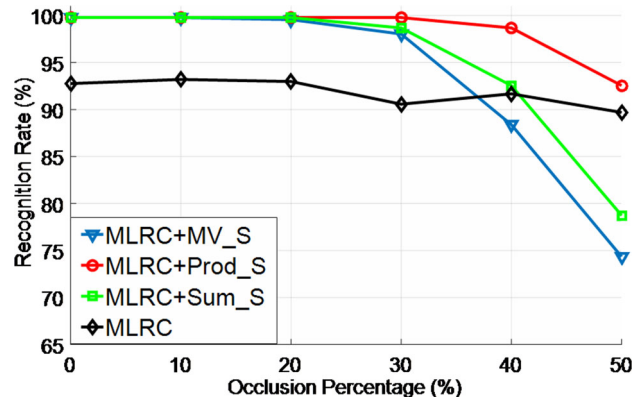


Fig. 9 Recognition rate versus occlusion percentage changing from 0 to 50%

Table 4 Comparison of SRC and our proposed method for various level of occlusion

| | 0% | 10% | 20% | 30% | 40% | 50% |
|-------------|------------|------------|-------------|-------------|-------------|-------------|
| SRC | 100 | 100 | 99.8 | 98.5 | 90.3 | 65.3 |
| MLRC+Prod_S | 99.8 | 99.8 | 99.8 | 99.8 | 98.7 | 92.6 |

Bold represents the best recognition performances

In the experiments, we choose the first and the second subsets for training (722 images, normal to moderate lighting conditions) and the third subset for testing (456 images, more extreme lighting conditions). We randomly add square block unrelated image to the testing probes to investigate the contiguous occlusion. The size of the block is varied from 0 to 50% of testing image as shown in Fig. 8. Since the results in Sect. 6.1 show that LRC outperforms CVA and DCVA, we use LRC to compare the recognition rates. Fig. 9 shows the classification results versus various level of occlusion. It is clearly seen that our proposed method outperforms modular LRC method on the occlusion problem. Also, proposed

method behaves nearly perfect up to 40% occlusion, i.e. it misclassifies just single probe out of 456 face images when the occlusion level is 30%.

In the work [35], authors compare the SRC recognition rate with several popular methods under several synthetic block occlusion levels on the Extended Yale B database. SRC has superior classification results over Independent Component Analysis [10], Principal Component Analysis (PCA)[33], Nearest Subspace (NS) [16], and Localized Non-negative Matrix Factorization (LNMF) [18]. We implement the same experiment with our proposed method. As it is seen from Table 4, even if SRC reaches 100% recognition rate when the occlusion is 10% (also our method has 99.78% recognition rate), our method is more robust to occlusion than SRC.

7 Conclusion

The modular subspace-based classification methods obviously surpass their conventional variations. But in occluded face recognition problem, their performances are not at the desired level.

In this paper, the modular version of DCVA is proposed firstly and used. We proposed a new partition selection strategy for modular face recognition methods. In the propose method, the partitions are selected using the combination of three coefficients which can be easily generated from face image. The supports of the selected regions are combined using product, sum, and majority vote rules. The performance of the proposed partition selection strategy is tested on scarf, sun glasses, and random block occlusion problems. In the experiments, it is seen that combining the classifier outputs of all partitions using several decision fusion strategies increases the classifier performances. The subspace-based methods reach their best recognition performance after selecting the partitions using the proposed approach.

References

1. Cevikalp, H., Neamtu, M., Wilkes, M., Barkana, A.: Discriminative common vectors for face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(1), 4–13 (2005)
2. Comenau, C.A., Simón, M.O., Cohn, J.F., Escalera, S.: Survey on RGB, 3D, thermal, and multimodal approaches for facial expression recognition: history, trends, and affect-related applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(8), 1548–1568 (2006)
3. Dagnes, N., Vezzetti, E., Marcolin, F., Tornincasa, S.: Occlusion detection and restoration techniques for 3D face recognition: a literature review. *Mach. Vis. Appl.* **29**, 789–813 (2018)
4. Ding, C., Tao, D.: A comprehensive survey on pose-invariant face recognition. *ACM Trans. Intell. Syst. Technol.* **7**(37), 1–42 (2016)
5. Georgiades, A.S., Belhumeur, P.N., Kriegman, D.J.: From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(6), 643–660 (2001)
6. Gulmezoglu, M.B., Dzhafarov, V., Keskin, M., Barkana, A.: A novel approach to isolated word recognition. *IEEE Trans. Speech Audio Process.* **7**(6), 620–628 (1999)
7. Gulmezoglu, M.B., Dzhafarov, V., Barkana, A.: The common vector approach and its relation to principal component analysis. *IEEE Trans. Speech Audio Process.* **9**(6), 655–662 (2001)
8. Han, H., Shan, S., Chen, X., Gao, W.: A comparative study on illumination preprocessing in face recognition. *Pattern Recognit.* **43**, 1691–1699 (2013)
9. Jafri, R., Arabnia, H.R.: A survey of face recognition techniques. *Int. J. Inf. Process. Syst.* **5**(2), 41–68 (2009)
10. Kim, J., Choi, J., Yi, J., Turk, M.: Effective representation using ICA for face recognition robust to local distortion and partial occlusion. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(12), 1977–1981 (2005)
11. Kittler, J., Hatef, A., Duin, R.P.W., Matas, J.: On combining classifiers. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(3), 226–239 (1998)
12. Koc, M., Barkana, A.: Modular common vector approach. In: *Proceedings of the IEEE 22nd Signal Processing and Communications Applications Conference*, pp. 533–535. SIU (2014)
13. Koc, M., Barkana, A.: A novel implementation algorithm for calculation of common vectors. *Anadolu Univ. J. Sci. Technol. A Appl. Sci. Eng.* **7**(2), 251–262 (2016)
14. Koc, M.: Different decision fusion methods for modular common vector approach. In: *Proceedings of the XIVth International Scientific and Technical Conference on Computer Sciences and Information Technologies*, pp. 67–70. CSIT (2019)
15. Koc, M.: Decision fusion for partially occluded face recognition using common vector approach. Chapter In: *Advances in Intelligent Systems and Computing IV*. **1080**, 351–360, Springer (2020)
16. Lee, K., Ho, J., Kriegman, D.: Acquiring linear subspaces for face recognition under variable lighting. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**(5), 684–698 (2005)
17. Li, J., Leng, J., Zhang, W., Guo, M.: Probabilistic robust regression with adaptive weights—a case study on face recognition. *Front. Comput. Sci.* **14**(5), 1–2 (2020)
18. S. Li, X. Hou, H. Zhang, and Q. Cheng. Learning spatially localized, parts-based representation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pp 1–6, 2001
19. Martinez, A.M., Benavente, B.: The AR face database. CVC Technical Report No: 24, (1998)
20. Mi, J.-X., Zhu, Q., Luo, Z.: Matrix regression-based classification with block-norm. *Pattern Recognit. Lett.* **125**, 654–660 (2019)
21. Nabatchian, A., Abdel-Raheem, E., Ahmadi, M.: Illumination invariant feature extraction and mutual-information-based local matching for face recognition under illumination variation and occlusion. *Pattern Recognit.* **44**(10–11), 2576–2587 (2011)
22. Naseem, I., Togneri, R., Bennamoun, M.: Linear regression for face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**(11), 2106–2112 (2010)
23. Naseem, I., Togneri, R., Bennamoun, M.: Robust regression for face recognition. *Pattern Recognit.* **45**(1), 104–118 (2012)
24. Nikan, S., Ahmadi, M.: Human face recognition under occlusion using LBP and entropy weighted voting. In: *Proceedings of the 21st International Conference on Pattern Recognition*, pp 1699–1702. ICPR (2012)
25. Nikan, S., Ahmadi, M.: Local gradient-based illumination invariant face recognition using local phase quantisation and multi-resolution local binary pattern fusion. *IET Image Process* **9**(1), 12–21 (2015)

26. Ou, W., Luan, X., Gou, J., Zhou, Q., Xiao, W., Xiong, X., Zeng, W.U.: Robust discriminative nonnegative dictionary learning for occluded face recognition. *Pattern Recognit. Lett.* **107**, 41–49 (2018)
27. Pentland, A., Moghaddam, B., Starner, T.: View-based and modular eigenspaces for face recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp 84–91. CVPR (1994)
28. Peng, Y., Li, L., Liu, S., Li, J., Wang, X.: Extended sparse representation-based classification method for face recognition. *Mach. Vis. Appl.* **29**, 991–1007 (2018)
29. Pereira, J.F., Barreto, R.M., Cavalcanti, G.D.C., Tsang, I.R.: A robust feature extraction algorithm based on class-Modular Image Principal Component Analysis for face verification. In: *Proceedings of the IEEE International Conference Acoustic Speech Signal Processing*, pp. 1469–1472. ICASSP (2011)
30. Polikar, R.: Ensemble based systems in decision making. *IEEE Circuits Syst. Mag.* **6**(3), 21–45 (2006)
31. Sepas-Moghaddam, A., Pereira, F., Correia, P.L.: Face recognition: a novel multi-level taxonomy based survey (2019). [arXiv:1901.00713](https://arxiv.org/abs/1901.00713)
32. Tan, K., Chen, S.: Adaptively weighted sub-pattern PCA for face recognition. *Neurocomputing* **64**, 505–511 (2005)
33. Turk, M., Pentland, A.: Eigenfaces for recognition. *J. Cognit. Neurosci.* **3**(1), 71–86 (1991)
34. Wang, J., Zhang, B., Wang, S., Qi, M., Kong, J.: An adaptively weighted sub-pattern locality preserving projection for face recognition. *J. Network Comput. Appl.* **33**, 323–332 (2010)
35. Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(2), 210–227 (2009)
36. Wu, C.Y., Ding, J.J.: Occluded face recognition using low-rank regression with generalized gradient direction. *Pattern Recognit.* **80**, 256–268 (2018)
37. Yan, Y., Lee, F., Chen, Q.: Face recognition algorithm using extended vector quantization histogram features. *PLoS ONE* **13**(1), 1–24 (2018)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Mehmet Koc received his two BSc degrees in Mathematics (2003), and in Electrical and Electronics Engineering (2004). He received MSc and PhD degrees in Electrical and Electronics Engineering from Eskisehir Osmangazi University (2006) and Anadolu University (2012), respectively. Currently, he is an Assistant Professor in Department of Electrical and Electronics Engineering at Bilecik Seyh Edebali University. His research interests include pattern recognition and machine learning.