International Journal of Innovative Computing, Information and Control
Volume **13**, Number **5**, October **2017**

ICIC International ©2017 ISSN 1349-4198
pp. 1721–1733

MULTI-CLASSIFIERS FACE RECOGNITION SYSTEM USING LBPP FACE REPRESENTATION

Abdellatif Dahmouni*, Nabil Aharrane, Karim El Moutaouakil and Khalid Satori

Laboratory of Computer-Imaging and Digital-Analysis
Department of Mathematics and Computer Science
Dhar-Mahraz Sciences Faculty
Sidi Mohamed Ben Abdellah University
B.P 1796 Atlas, Fez 30000, Morocco
*Corresponding author: Abdellatif.dahmouni@usmba.ac.ma
Aharranenabil@gmail.com; { Karimmoutaouakil; Khalidsatorim3i }@yahoo.fr

Received March 2017; revised June 2017

Abstract. *Local Binary Probabilistic Pattern (LBPP) is a local descriptor able to improve the recognition capabilities of a typical pattern recognition system. It is a new alternative of the famous Local Binary Pattern (LBP) descriptor based on confidence interval concept. To achieve an enhanced representation for face's principal components, LBPP evaluates each current pixel using a probabilistic confidence interval related to its neighborhood. In this paper, to improve face recognition performance we propose a new methodology based on the combinative use of LBPP descriptor, Two Dimensional Discrete Cosine Transform (2DDCT) frequency subspace, and some machine learning algorithms. The main idea behind this methodology is to elevate the weak points of each one of them, while making use of their major advantages. Hence, after the LBPP processing phase, 2DDCT method decomposes obtained image into set of local features vectors. Each local vector is formed by the k-first zigzag coefficients for each sub-image. Then, we carefully concatenate all local vectors into a single features vector. In addition, obtained features dataset will be classified using relevant machine learning classifiers. To access our solution, we applied it on ORL, Yale and AR face databases. Obtained results clearly show the effectiveness of the proposed approach compared to the existing state of the art techniques. Indeed, the LBPP capacity to discriminate face components, the small size of 2DDCT features vector, and the efficiency of used classifiers, allow justifying the proposed approach's good performance.*
Keywords: LBPP, Confidence interval, 2DDCT, KNN, MLP, LIBSVM, SMO, Vote rules

1. **Introduction.** Face recognition is a natural and promising modality for person identification. It is largely used in different fields where the human-computer interaction is a decisive stage, such as information security, smart cards, entertainment, law enforcement, and video surveillance. All face recognition methods can be regrouped into two categories according to their properties: global and local methods. The global methods use the entire facial surface as features vector and reduce the representation space by linear transformations. Among the most known global methods we find Principal Component Analysis (PCA) [1,2], Linear Discrimination Analysis (LDA) [3,4], Locality Preserving Projection (LPP) [5], Unsupervised Discriminant Projections (UDP) [6], Discrete Wavelet Transforms (DWT) [7], and Discrete Cosine Transforms (DCT) [8,9]. The local methods isolate the critical face areas and build features vector as a set of relations between principal components of each area. Among the most known local methods we

find classical methods such as Speeded-Up Robust Features (SURF) [10], Weber Local Descriptor (WLD) [11], Histogram Oriented Gradient (HOG) [12], and some LBP varieties [13,14].

LBP is a most popular descriptor that describes the local structure of an image while evaluating each current pixel using its neighborhood. Since it was introduced by Ojala et al. [13], LBP is extended into many variants which covert different pattern recognition fields. For example, Tan and Triggs [15] proposed a more discriminant variety called Local Ternary Pattern (LTP), which evaluates current pixel using a three bits code. Houam et al. [16] proposed one-dimensional LBP, which uses one-dimensional neighborhood to adapt LBP for one-dimensional signals. Jabid et al. [17] proposed Local Directional Pattern (LDP), which detects the edge of different orientations using eight Kirsch kernels. To reduce the complexity of Local Directional Pattern (LDP), Srinivasa and Chandra Mouli [18] introduced the Dimensionality Reduced LDP (DR-LDP), which computes a single code by X-ORing the eight LDP codes. Liu et al. [19] proposed Weber LBP (WLBP), which combines the advantages of WLD and LBP to splendidly describe local features. Guo et al. [20] proposed Completed LBP (CLBP), which combines sign and magnitude information to achieve robust local description. Jun and Kim [21] proposed Local Gradient Pattern (LGP), which uses the local gradient neighborhood to reach local and global invariance. Nguyen and Caplier [22] proposed Elliptical LBP (ELBP), which applies LBP to Local Neighborhood of elliptical structure. In a previous study, we have proposed a new Local Binary Probabilistic Pattern (LBPP) face representation, which uses the confidence interval concept to evaluate the current pixel [24,25]. The confidence interval will be calculated according to a probabilistic law in the current neighborhood. In this paper, a new improved face recognition system is proposed. It consists to improve recognition capacities by combining the major advantages of LBPP descriptor, 2DDCT subspace, and some machine learning classifiers. Indeed, to achieve the suggested objectives we use, firstly, the local behavior of LBPP descriptor. In order, LBPP allows separating between the almost homogeneous areas and peak areas. Secondly, during 2DDCT decomposition, one must treat three types of blocks. The first type is made of pixels having grayscale values near to ‘255’; the second is made of pixels having grayscale values near to ‘0’; the third is made of pixels of various grayscale values. For the blocks of the first and second type, which cover the majority of the LBPP image, 2DDCT stores the majority of information in just first zigzag coefficient. That interprets the reduced features vectors length. The choice of a powerful classifier that discriminates between features vectors remains a decisive phase for each recognition system; in this work a well-known classifiers comparative study is completed to determine the most selective between them.

The remainder of this paper is organized as follows. We start by describing our LBPP descriptor in Section 2. In Section 3, we present the 2DDCT frequency subspace. We describe the used machine learning algorithms in Section 4. After that, we present the experimental results in Section 5. Finally, we conclude the paper.

2. Local Binary Probabilistic Pattern (LBPP).

The main idea of our LBPP descriptor is basing to following hypothesis: “as many natural phenomena, the distribution of pixels in the almost homogeneous areas on the face surface follows approximately sum of normal laws”. This propriety is used to evaluate the confidence interval that characterizes current neighborhood. Therefore, we pass from a fixed thresholding mode for several LBP varieties, to a probabilistic thresholding mode.

2.1. Confidence interval evaluation.

Knowing that each normal law $N(\mu, \sigma)$ is perfectly defined using its probability density and its own statistical parameters, such as, the average: $\mu = \sum_i x_i$, the standard derivation: $\sigma = \sum_i (x_i - \mu)^2$, the coefficient of variation: $\delta = \sigma/\mu$, and the asymmetry coefficient: $S = \sum_i (x_i - \mu)^3/\sigma^3$. In order to evaluate the current pixel, we generate a confidence interval $[\alpha_1, \alpha_2]$ related to the empirical distribution, which governs current neighborhood of size 5×5 pixels. It is therefore necessary to estimate data dispersion using the deviation of this distribution to a normal low defined by same values of μ and σ. In this study, we evaluate data dispersion using δ and S coefficients. Moreover, as it is illustrated in Figure 1, the dispersion of data is mainly influenced by the shape of the probability density curve. Indeed, the more this curve is similar to a bell; the confidence interval length will be equal to 8σ. The more it is flattened or dissymmetrical; the more the confidence interval length tends towards 2σ. For more dispersed data, this length migrated towards a limiting value equal to 0.4μ.

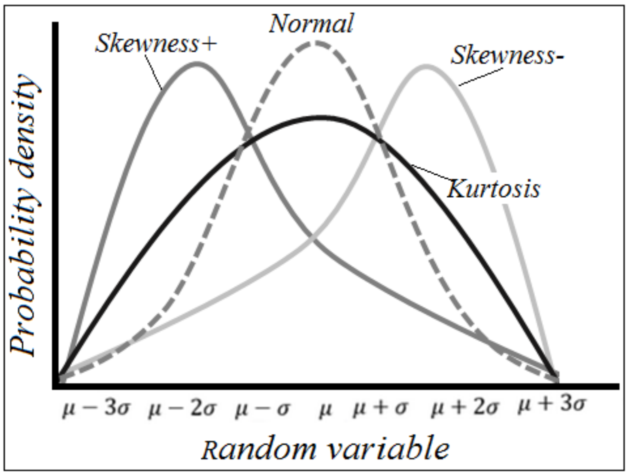

Figure 1. Normal low probability density curve

In this context, we associate to the gray level value i_n a random variable X. Let us suppose that any variable X fulfilling condition of Equation (1) follows the normal law $N(\mu, \sigma)$, it is a homogeneous neighborhood in which the confidence interval is defined by (2).

$$\text{if } S = 0 \text{ or } \delta < \beta \text{ where } 0.1 < \beta < 0.2 \text{ (β is experimentally determined)} \tag{1}$$

$$[\alpha_1, \alpha_2] = [\boldsymbol{\mu} - k\boldsymbol{\sigma}, \boldsymbol{\mu} + k\boldsymbol{\sigma}];\ \text{where: } k = 1, 2, 3, 4 \tag{2}$$

For any variable X not fulfilling (1) one takes:

$$[\alpha_1, \alpha_2] = [\boldsymbol{\mu} - K, \boldsymbol{\mu} + K];\ \text{where: } K = \begin{cases} \boldsymbol{\sigma} & \text{if } \beta < \delta < 0.2 \\ 0.2\boldsymbol{\mu} & \text{others} \end{cases} \tag{3}$$

2.2. **LBPP formulation.** According to the three previously equations, that define the confidence interval, LBPP descriptor computes the new grayscale value of current pixel using following process. Neighborhood's pixels having grayscale value located between α_1 and α_2 take a value equal to '1'. While the others take a value equal to '0'. Then, the obtained byte is converted into gray level value; see Figure 2. Finally, basing the confidence interval defined in (1), (2) and (3), LBPP descriptor is given by:

$$LBPP_{P,R,P',R'} = \sum_{n=0}^{P-1} S(i_n)2^n \text{ where } S(x) = \begin{cases} 1, & \text{if } \alpha_1(\mu, \sigma, k) \le x \le \alpha_2(\mu, \sigma, k) \\ 0, & \text{others } (k \in [-4, 4]) \end{cases} \tag{4}$$

The parameters R and P are respectively the neighborhood radius and number of neighboring pixels that contribute to binary code calculation.

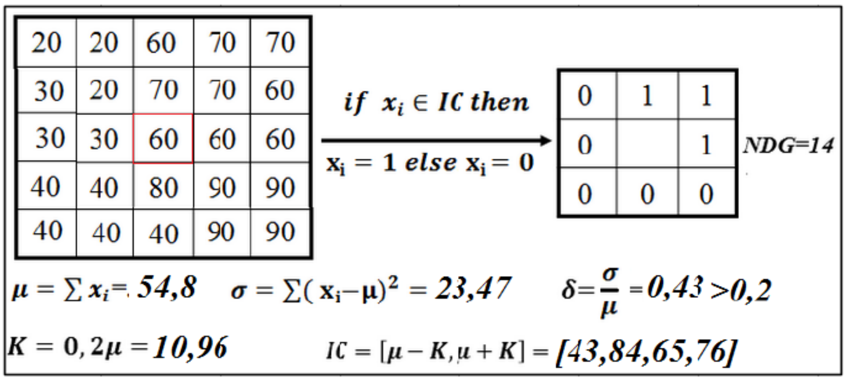

FIGURE 2. Step of LBPP grayscale value calculation

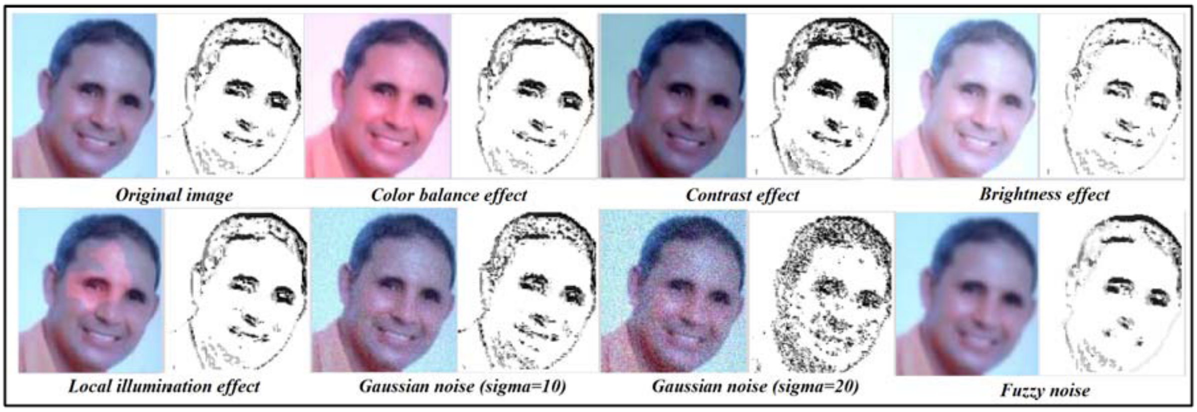

FIGURE 3. Illustration of LBPP behavior through different changes

The parameters R' and P' are respectively the neighborhood radius and number of neighboring pixels that contribute to the statistical moments (μ, σ, δ and S) calculation.

Contrary to some LBP varieties, which are deterministic models, our LBPP considers the face areas as distributions of pixels that follow a probability law. The confidence interval concept allows to perfectly locating different areas having strong gradient such as the nose, the eyes and the mouth. The others areas considered as almost homogeneous areas are coded with values near the maximum gray level value. Similarly, LBPP descriptor is robust to different illumination changing and noises adding. As it is shown in Figure 3, for different changes that affect the original image (to the left), the associated LBPP images (to the right) are visibly comparable.

To optimize the processing complexity and the overall recognition performance generally affected by the great dimension of original face, we construct a new face representation that allows managing and processing less coefficients. Reduced face features vectors must preserve all pertinent information. In order to improve processing time while maintaining recognition performance, 2DDCT frequency subspace method will be used.

3. **2DDCT Features Extraction.** 2DDCT is a linear orthogonal and reversible transformation that converts a signal from spatial or temporal domain towards frequency domain. It permits to reduce the redundancy of the signal while preserving the maximum of information into small lower frequency components. Therefore, 2DDCT is commonly used in various applications in which the data compression is a decisive stage; see Figure 4. As a subspace method in frequency domain, 2DDCT becomes an effective tool in pattern recognition field. It consists in decomposing an image $I_{N,M}$ in a local cosine basis

according to the following equation:

$$F(u,v) = \alpha(x)\alpha(y)\sum_{x=0}^{N-1}\sum_{y=0}^{M-1} I(x,y)\beta(x,u)\beta(y,v) \tag{5}$$

where

$$\alpha(k) = \begin{cases} \sqrt{2/N} & \text{for } k = 0 \\ \sqrt{1/N} & \text{for } k > 0 \end{cases} \text{ and } \beta(x,u) = \cos\left(\frac{(2x+1)u\pi}{2N}\right) \tag{6}$$

The success of a face recognition algorithm mainly depends on the compromise between the information quantity to be treated and the processing time. To reply this compromise we tend to minimize the features vector length without significant loss in terms of recognition rate. Hence, each face image is subdivided into N non-overlapping blocks of n^2 pixels, especially, $n = 8$, for the JPEG standard. Then, 2DDCT is applied to each block to generate n^2 coefficients. The first coefficient named Direct Current (DC) represents the block intensity average and stores almost 95% block energy. The remaining coefficient is Alternative Current (AC) that represents the smallest amount of information. This property is employed to reduce the face features vector length. To illustrate the local characteristics of the image, it is useful to traverse each block using zigzag technique. This particular path makes to collect the K-First Neighboring Coefficients of Direct Current (KFNC-DC $= b$) related to the low frequencies. The KFNC-DC represents the local features vector denoted $F_{lv} = (z_{11}, \ldots, z_{1N})$. After that, all local features vectors are concatenated into a single global one representing face features, $F_{gv} = (z_{11}, \ldots, z_{1N}; \ldots; z_{bN}, \ldots, z_{bN})$; see Figure 5. Then, the set of all global vectors will be the inputs of classification phase in which the most known classifiers are applied to evaluating the proposed method. As it

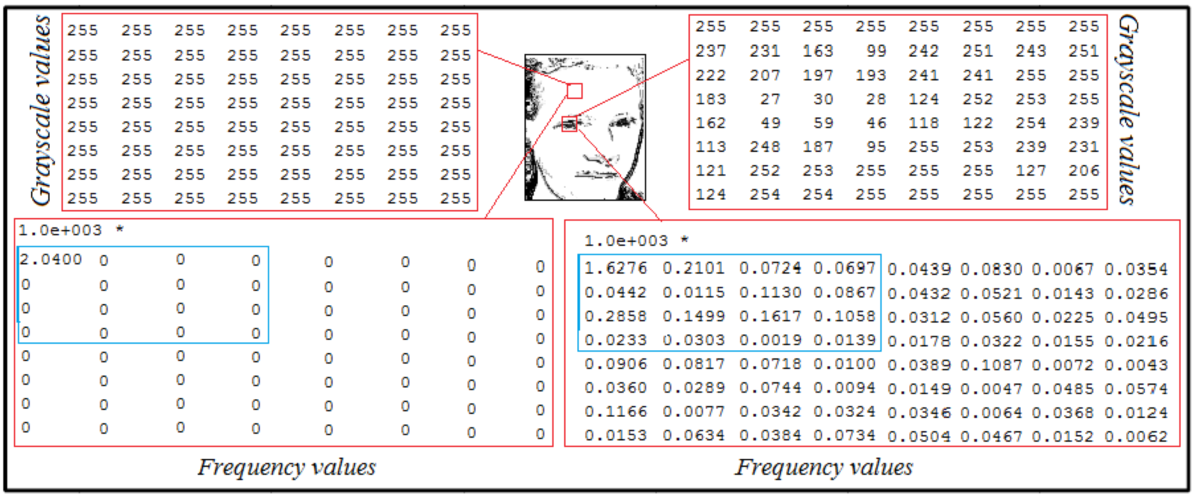

FIGURE 4. Example of 2DDCT blocks frequency transform

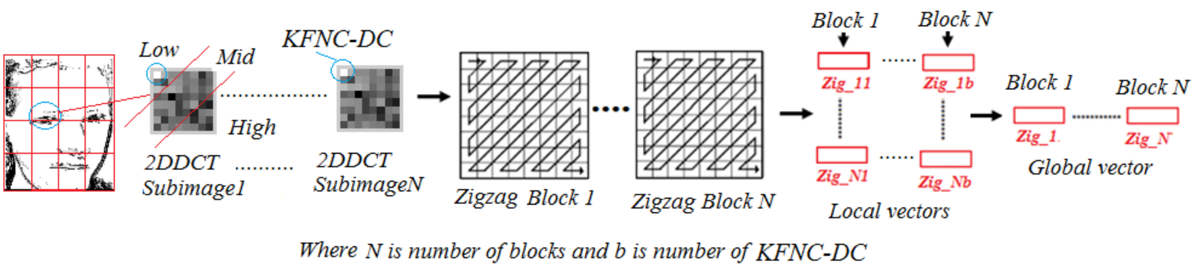

FIGURE 5. Steps of 2DDCT face features extraction

is shown in Figure 4, the LBPP blocks associated to the almost homogenous areas are transformed using 2DDCT into single DC value equal to 2.04×10^3Hz. The others are transformed into b significant values ($b < 10$). This propriety reflects the fact that effective information of an image is concentrated in the low frequencies, where the eye has a strong acuity.

4. **Machine Learning Classifiers.** In this section, we present a brief overview of machine learning algorithms will be used in classification phase. Mainly, Support Vector Machine (SVM), Multilayer Perceptron (MLP), K-Nearest Neighbors (KNN), and their combination in terms of majority voting. Moreover, it should be noted that several versions of each method are proposed in the literature. In our case, we focus on the most known ones.

- SVM Classifier: since it was proposed by Vapnik, SVM became a powerful tool for supervised learning. SVM is based to find the optimal hyperplanes presenting the maximum margin using a set of training instances. In this setting, binary classification problems can be reduced into the minimum margin using optimization criterion [26]. To extend SVM to multi-class problems, such as faces classification, two strategies are used: one-against-all and one against-one. In the one-against-all method, SVM separates a single class from all remaining classes. In the one-against-one method, SVM arranges the pairwise classifiers in binary trees. It is noted that, the first strategy is the most replied in the literature. In our experiments, we use two recent SVM learning algorithms: Library for Support Vector Machines (LIBSVM) [27,28] and Sequential Minimal Optimization (SMO) [29,30]. Moreover, LIBSVM is an integrated software package that allows users to explore the potentials of SVM algorithms in their purpose. It can be easily added to the WEKA system. Whereas, SMO is an SVM improvement that solves SVM quadratic programming problem (QPP) as a series of smallest QP sub-problems using two components; the first is an optimization of two Lagrange multipliers, and the second is a heuristic that qualifies which multipliers to optimize.
- MLP Classifier: as a feed-forward artificial neural network, MLP is widely used in classification problems. To adjust the weights of its neurons, MLP uses many supervised training procedures mainly, the backpropagation that learns the network through three phases: forward propagation, backward propagation, and weights adapting [31]. In MLP, the input layer has a number of neurons equal to the number of selected features. The number of neurons in the output layer is always the number of classes. The major challenge of MLP remains to determine the optimal number of hidden layers and the number of neurons in each hidden layer. Viewing its simplicity, the topology based on one hidden layer is used in several classification problems.
- KNN Classifier: it is among the simplest supervised learning algorithms. It was proposed to perform discriminant analysis when it is difficult to determine reliable parametric estimates of the probability densities. To classify each new sample from the training set, KNN finds its k neighbors nearest samples based on some similarity metrics, especially the Euclidean norms [32].

Classifiers Combination: The fusion of multiple classifiers can be interpreted as a decision-making problem by combining their outputs [33,34]. In our case, to make the best decision we use the voting strategy. It assigns each instance to the class that receives the largest votes' number. The use of the output probabilities of each classifier permits to predict the output class of the fused classifier using the five following rules: the average, the product, the maximum, the minimum, and the majority rules. In order to improve the system performance, we combined KNN, MLP, LIBSVM and SMO classifiers using the voting strategy by the five prediction rules mentioned bellow.

5. **Experimentation.** In this section, we evaluate the performance of proposed face recognition system; see Figure 6. Hence, we tested different scenarios using ORL, Yale, and AR standard databases. First LBPP extracts principal face components by separating the almost homogeneous areas (Forehead, Cheek), and the peaks areas (Eyes, Nose, Mouth). Then, 2DDCT builds the reduced features vector based on KFNC-DC of each block. Finally, we use LIBSVM, SMO, MLP, and KNN machine learning classifiers individually applied and then combined to generate best classification results.

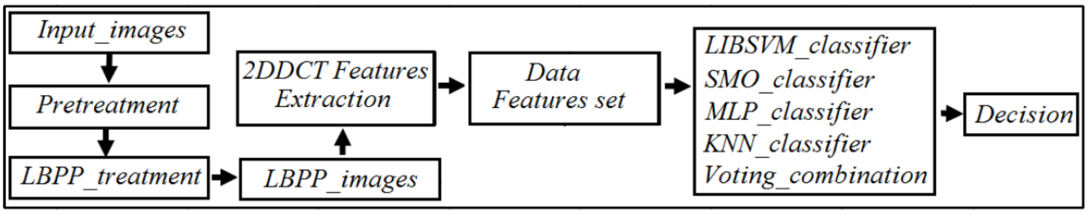

FIGURE 6. Diagram of proposed face recognition system

5.1. **Face images databases.** The ORL database is made of 40 subjects having each one 10 different views [41]. Images are in gray level with the same size 112×92 pixels. Moreover, they were taken at different time and contain variations of lighting, facial expressions, and face's details. The Yale database is made up of 165 images of 15 subjects representing 11 conditions of lighting, pose, and facial expressions [42]. All Yale images are centered and resized into 112×92 pixels. The AR database contains 126 people with 26 images per person [43]. In this study, we selected an AR subset composed by 100 individuals; each individual has 14 images presenting the variations of facial expressions and illumination. All images are resized into 112×92 pixels. Figures 7, shows respectively the obtained images using LBP and different variants of LBPP according to the confidence intervals size defined by: $k = 1, 2, 3$ or 4. We note that the face components belonging to strong variations areas such as the nose, the mouth, and the eyes are clearly distinguished by increasing the confidence interval length, especially for LBPP ($k > 2$). This property can be subsequently used in many pattern recognition applications. In our case, it will be used for selecting the small features vectors using 2DDCT method.

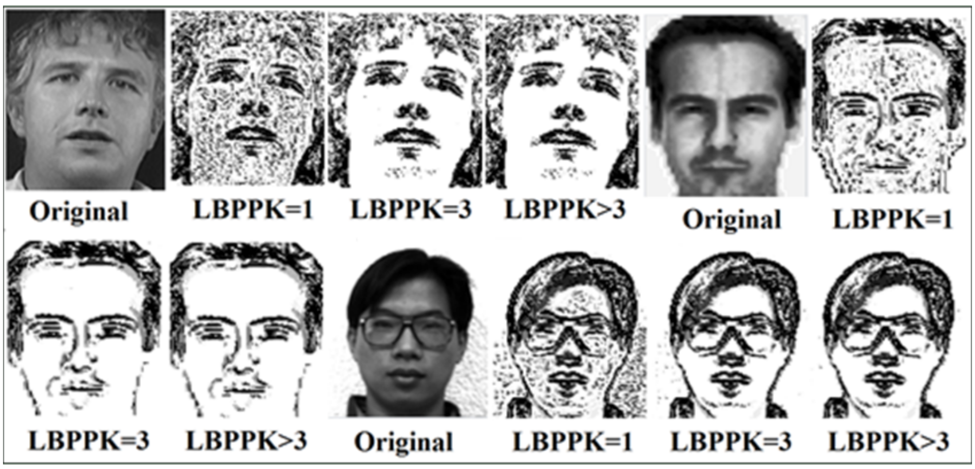

FIGURE 7. Illustration of LBPP behavior through different changes

5.2. **Features vector parameters.** To take advantage of LBPP and 2DDCT performances, we must define the own optimal parameters, namely the optimal Confidence Interval (CI), the optimal block size and the optimal local vector length. Indeed, we repeat hundred times an experiment in which we randomly took 50 percent from each class as training set. The other 50% images are used for tests. Thereafter, we calculate in same conditions the average and the standard deviation for every test. Table 1 shows that the choice of the confidence interval associated to $k = 4$ achieves the best results compared to other cases $k = 1, 2, 3$. Accordingly, we reach a recognition rate of 94.87±1.1% on ORL, 99.23±0.9% on Yale and 96.51±1.2% on AR.

Additionally, Figure 8 shows that the 14×14 pixels block size as well as the three KFNC-DC per block are the optimal values for the 2DDCT parameters. They do achieve

Table 1. Recognition rate using optimal block parameters on ORL, Yale and AR databases

	Confidence Interval Size	LBPP $k = 1$	LBPP $k = 2$	LBPP $k = 3$	LBPP $k > 4$
Rate %	ORL (5 train/5 test)	94.72±1.13	94.81±1.11	94.85±1.12	94.87±1.10
	Yale (5 train/6 test)	98.83±1.09	99.13±1.08	99.23±1.01	99.23±0.90
	AR (7 train/7 test)	96.50±1.14	96.45±1.16	96.50±1.17	96.51±1.20

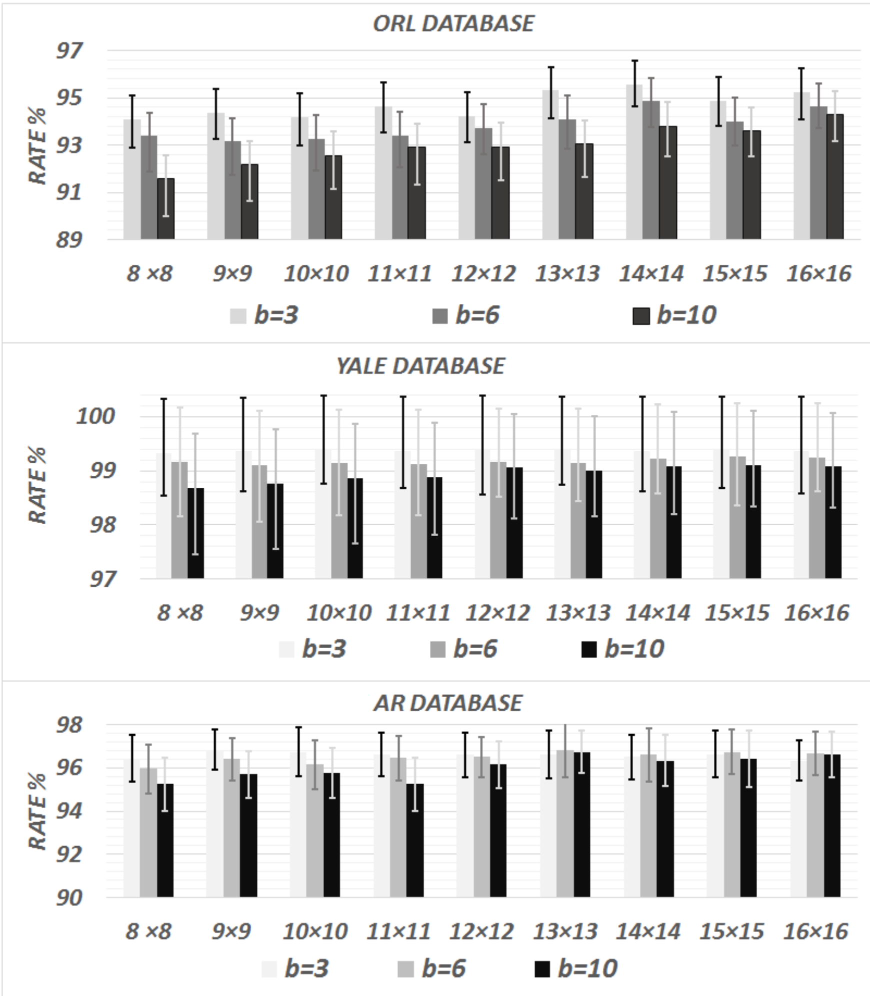

Figure 8. LBPP recognition rate using different parameters and databases

better results than the other tested values. Consequently, we reach a recognition rate of 95.12±0.95% on ORL, 99.37±0.74% on Yale, and 96.6±1.04% on AR. After the choice of the optimal values of used parameters, we start the classification phase. Hence, we represent each face using a feature vector of length $n_v = N \times b$, where, N is number of blocks of size 14×14 pixels, and b is number of KFNC-DC; in our case, $n_v = 48 \times 3 = 144$. Then, the obtained vectors are employed to prepare the dataset in an ".arff" format that is the basis file to start the classification process using Weka platform.

5.3. **Classifiers evaluation.** To achieve the creation and the study of proposed classifiers, we use the Weka tool version 3.6.0 [44]. With its greatest GUI, it offers the possibility to explore easily many classifiers families.

TABLE 2. Machine learning parameters

	Observed: Positive	Observed: Negative	Recall: R = TP/(TP+FN). Precision: P = TP/(TP+FP). F-measure = 2PR/(P+R). Accuracy = (TP+TN)/N_{total}.
Predicted: Positive	TP	FP	
Predicted: Negative	FN	TN	

In order to analyze the performance of each proposed classifier, we conducted separately its evaluation using the results average for $k = \{4, 5, 6, 7, 8, 9, 10\}$ folds cross validation. Moreover, we had to define some machine learning parameters: True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), Precision (P), Recall (R), F-Measure and Accuracy [35,36]. While assessing our method, we used the specific configuration for each classifier. Indeed, we choose Euclidean distance for KNN algorithm, polynomial kernel function for SVM algorithms, and optimal auto-Building of hidden layers for MLP algorithm.

Hence, we obtained a confusion matrix for each classifier. These matrices are given by their diagonal element's representing the correctly classified instances all others are incorrectly classified. As obtained results, almost all diagonal elements of the matrices are equal to the person's number of each class. According to classification measures, Precision, Recall, F-measure and Accuracy, are shown in Tables 3, 4 and 5, and it is clear that SVM classifiers give the best results compared to KNN and MLP. Particularly, for ORL database, we reach an F-measure of 98.58±0.51% for LIBSVM and 98.23±0.72% for SMO. For Yale database, we reach an F-measure of 99.4±0.23% for LIBSVM and

TABLE 3. Comparison of different used classifiers on ORL database

Classifiers	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)
LIBSVM	98.67±0.47	98.64±0.51	98.58±0.51	98.61±0.52
SMO	98.34±0.65	98.27±0.75	98.23±0.72	98.25±0.74
MLP	97.59±0.49	97.44±0.55	97.37±0.55	97.39±0.56
KNN	97.00±0.49	96.77±0.63	96.66±0.61	96.72±0.59

TABLE 4. Comparison of different used classifiers on Yale database

Classifiers	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)
LIBSVM	99.4±0.23	99.4±0.23	99.4±0.23	99.39±0.22
SMO	99.4±0.21	99.4±0.21	99.4±0.21	99.39±0.18
MLP	99.23±0.27	99.23±0.27	99.23±0.27	99.23±0.24
KNN	99.08±0.54	99.08±0.54	99.08±0.54	99.08±0.53

Table 5. Comparison of different used classifiers on AR database

Classifiers	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)
LIBSVM	98.58±0.12	98.5±0.13	98.48±0.14	98.58±0.12
SMO	98.35±0.25	98.23±0.29	98.23±0.29	98.35±0.25
MLP	98.38±0.13	98.3±0.13	98.3±0.13	98.28±0.13
KNN	97.68±0.32	97.51±0.34	97.51±0.34	97.68+0.32

Table 6. Comparison of different used classifiers on ORL database

Classifiers	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)
V_{maj}	98.46±0.60	98.54±0.53	98.48±0.58	98.44±0.61
$V_{\max}$	98.60±0.51	98.67±0.46	98.63±0.52	98.58±0.51
$V_{\min}$	98.57±0.55	98.65±0.48	98.58±0.56	98.55±0.54
V_{prod}	98.60±0.52	98.67±0.47	98.62±0.52	98.58±0.51
V_{avg}	98.71±0.66	98.78±0.61	98.74±0.67	98.70±0.68

Table 7. Comparison of different used classifiers on Yale database

Classifiers	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)
V_{maj}	99.74±0.23	99.74±0.23	99.74±0.23	99.74±0.23
$V_{\max}$	99.10±0.29	99.10±0.29	99.10±0.29	99.10±0.29
$V_{\min}$	99.12±0.27	99.12±0.27	99.12±0.27	99.12±0.27
V_{prod}	99.23±0.21	99.23±0.21	99.23±0.21	99.23±0.21
V_{avg}	99.48±0.32	99.48±0.32	99.48±0.32	99.48±0.32

Table 8. Comparison of different used classifiers on AR database

Classifiers	Precision (%)	Recall (%)	F-measure (%)	Accuracy (%)
V_{maj}	98.80±0.18	98.68±0.18	98.68±0.19	98.70±0.12
$V_{\max}$	98.70±0.21	98.60±0.28	98.60±0.28	98.68±0.18
$V_{\min}$	98.68±0.19	98.60±0.26	98.60±0.26	98.60±0.23
V_{prod}	98.68±0.19	98.60±0.25	98.60±0.25	98.60±0.24
V_{avg}	98.74±0.19	98.64±0.19	98.64±0.19	98.66±0.21

99.4±0.21% for SMO. Finally, for AR database we reach an F-measure of 98.48±0.14% for LIBSVM and 98.23±0.29% for SMO. Hence, the SVM classifier is the most qualified one to be compared with the classifiers voting combination. To enhance the performances of the proposed recognition system, we recommend combining the four already mentioned classifiers using the meta voting strategy based on the five following voting rules: majority voting (V_{maj}), maximum of probabilities ($V_{\max}$), minimum of probabilities ($V_{\min}$), product of probabilities (V_{prod}), and average of probabilities (V_{avg}). Experimental results shown in Tables 6, 7 and 8 are encouraging. They illustrate that the classifiers combination leads to more accurate face recognition than individual classifier. We noted that, the majority-voting rule reaches the highest results for all proposed databases; we are getting an F-measure of 98.48±0.58% for ORL, 99.74±0.23% for Yale, 98.68±0.19% for AR. However, obtained confusion matrices show that the number of incorrectly classified elements using separate classifiers decreases while using meta voting classifiers.

To further confirm the effectiveness of proposed method, we compare it with other most known state-of-the-art approaches, including 2DPCA, PCA+SVM, 2DPCA+SVM,

Table 9. Comparison of proposed approach and others face recognition methods

Rate (%)	Methods	HDCT [8]	GWT+ DCT [37]	PCA+ DCT [37]	2DPCA [38]	PCA+ SVM [39]	LBP	ULBP/ LPQ	Our approach
	ORL	88.43	96.5	95.75	96	97	91 [23]	91.5 [23]	98.54
	Yale	68	96	95.5	84.24	99.39	96 [40]	98 [40]	99.74
	AR	****	****	****	96.1	92.67	8 [40]	92.7 [40]	98.68

ICA+SVM, HMM+LBP, HDCT, LBP, ULBP+LPQ, GWT+DCT, and DCT+PCA. The illustrated results in Table 9 show that the proposed approach gives a best success rate with ORL, Yale and AR databases.

6. Conclusions.

In this paper, a hybrid approach based on combining the Local Binary Probabilistic Pattern (LBPP) face representation, the 2DDCT frequency subspace, and the most known machine learning classifiers is used to construct an enhanced face recognition system. For this recognition system, we discussed the effect of using face recognition methods and machine learning classifiers. Indeed, LBPP descriptor is employed to enhance the extraction of relevant face information. Therefore, it allows generating a new face description through the dispersion measurement of its neighboring pixels. These favorable LBPP properties are employed to build small face features vector using 2DDCT frequency subspace. To build face features vector, we have concatenated all local features vectors, associated of each block, into a single global vector. Each local vector is formed by the low frequency components obtained with the zigzag construction. To build face features vector, we have divided it into a set of non-overlapping blocks and concatenated all local features vectors into a single global features vector. Each local vector is formed by the K-First Neighboring Coefficients of Direct Current using zigzag construction. Finally, we have applied LIBSVM, SMO, KNN and MLP machine learning classifiers and their combination using the majority vote strategy to make the final decision using Weka software. According to the experimental results, the proposed system gives a very satisfactory recognition rate while it is combined with different supervised classifiers.

REFERENCES

[1] X. Luan, B. Fang, L. Liu, W. Yang and J. Qian, Extracting sparse error of robust PCA for face recognition in the presence of varying illumination and occlusion, *Pattern Recognition*, vol.47, no.2, pp.495-508, 2014.

[2] Q. Zhu and Y. Xu, Multi-directional two-dimensional PCA with matching score level fusion for face recognition, *Neural Computing and Applications*, vol.23, no.1, pp.169-174, 2013.

[3] S. Huang, D. Yang, J. Zhou and X. Zhang, Graph regularized linear discriminant analysis and its generalization, *Pattern Analysis and Applications*, vol.18, no.3, pp.639-650, 2015.

[4] P. Forczmański and P. Łabędź, Improving the recognition of occluded faces by means of two-dimensional orthogonal projection into local subspaces, *Image Analysis and Recognition*, pp.229-238, 2015.

[5] D. Hu, G. Feng and Z. Zhou, Two-dimensional locality preserving projections (2DLPP) with its application to palm-print recognition, *Pattern Recognition*, vol.40, no.1, pp.339-342, 2007.

[6] J. Yang, D. Zhang, J. Yang and B. Niu, Globally maximizing, locally minimizing: Unsupervised discriminant projection with applications to face and palm biometrics, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.29, no.4, pp.650-664, 2007.

[7] G. Anbarjafari, S. Izadpanahi and H. Demirel, Video resolution enhancement by using discrete and stationary wavelet transforms with illumination compensation, *Signal, Image and Video Processing*, vol.9, no.1, pp.87-92, 2015.

[8] I. Dagher, Highly compacted DCT coefficients, *Signal, Image and Video Processing*, vol.4, no.3, pp.303-307, 2010.

[9] F. Gmira, S. Hraoui, A. Saaidi, A. J. Oulidi and K. Satori, Securing the architecture of the JPEG compression by an dynamic encryption, *Intelligent Systems and Computer Vision (ISCV)*, pp.1-6, 2015.

[10] J. M. Bay, A. Ess, T. Tuytelaars and L. V. Gool, SURF: Speeded up robust features, *Computer Vision and Image Understanding*, vol.110, no.3, pp.346-359, 2008.

[11] A. Pal, N. Das, S. Sarkar, D. Gangopadhyay and M. Nasipuri, A new rotation invariant weber local descriptor for recognition of skin diseases, *Pattern Recognition and Machine Intelligence*, pp.355-360, 2013.

[12] A. D. Calvillo, R. A. Vazquez, J. Ambrosio and A. Waltier, Face recognition using histogram oriented gradients, *International Symposium on Intelligent Computing Systems*, pp.125-133, 2016.

[13] T. Ojala, M. Pietikäinen and T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.24, no.7, pp.971-987, 2002.

[14] T. Ahonen, A. Hadid and M. Pietikäinen, Face description with local binary patterns: Application to face recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.28, no.12, pp.2037-2041, 2006.

[15] X. Tan and B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, *IEEE Trans. Image Process.*, vol.19, no.6, pp.1635-1650, 2010.

[16] L. Houam, A. Hafiane, A. Boukrouche, E. Lespessailles and R. Jennane, One dimensional local binary pattern for bone texture characterization, *Pattern Analysis and Applications*, vol.17, no.1, pp.179-193, 2014.

[17] T. Jabid, M. H. Kabir and O. Chae, Local directional pattern (LDP) for face recognition, *Digest of Technical Papers International Conference on Consumer Electronics*, pp.329-330, 2010.

[18] P. R. Srinivasa and P. V. S. S. R. Chandra Mouli, Dimensionality reduced local directional pattern (DR-LDP) for face recognition, *Expert Systems with Applications*, vol.63, pp.66-73, 2016.

[19] F. Liu, Z. Tang and J. Tang, WLBP: Weber local binary pattern for local image description, *Neurocomputing*, vol.120, pp.325-335, 2013.

[20] Z. Guo, L. Zhang and D. Zhang, A completed modeling of local binary pattern operator for texture classification, *IEEE Trans. Image Process.*, vol.19, no.6, pp.1657-1663, 2010.

[21] B. Jun and D. Kim, Robust face detection using local gradient patterns and evidence accumulation, *Pattern Recognition*, vol.45, no.9, pp.3304-3316, 2012.

[22] H. T. Nguyen and A. Caplier, Elliptical local binary patterns for face recognition, *Asian Conference on Computer Vision*, pp.85-96, 2012.

[23] W. Yu, L. Gan, S. Yang, Y. Ding, P. Jiang, J. Wang and S. Li, An improved LBP algorithm for texture and face classification, *Signal Image and Video Processing*, vol.8, no.1, pp.155-161, 2014.

[24] A. Dahmouni, K. Elmoutaouakil and K. Satori, Robust face recognition using local gradient probabilistic pattern (LGPP), *Proc. of the Mediterranean Conference on Information & Communication Technologies 2015*, vol.380, no.29, pp.277-286, 2016.

[25] A. Dahmouni, N. Aharrane, K. E. Moutaouakil and K. Satori, Face recognition using local binary probabilistic pattern (LBPP) and 2DDCT frequency decomposition, *The 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV)*, pp.73-77, 2016.

[26] J. C. burges, A tutorial on support vector machines for pattern recognition, *Data Mining and Knowledge Discovery*, vol.2, no.2, pp.121-167, 1998.

[27] C. C. Chang and C. J. Lin, LIBSVM: A library for support vector machines, *ACM Trans. Intelligent Systems and Technology (TIST)*, vol.2, no.3, 2011.

[28] N. Haridas, V. Sowmya and K. P. Soman, GURLS vs LIBSVM: Performance comparison of kernel methods for hyperspectral image classification, *Indian Journal of Science and Technology*, vol.8, no.24, 2015.

[29] J. Platt, Sequential minimal optimization: A fast algorithm for training support vector machines, *Advances in Kernel Methods – Support Vector Learning*, vol.208, no.1, pp.212-223, 1998.

[30] S. Abe, Fusing sequential minimal optimization and Newton's method for support vector training, *International Journal of Machine Learning and Cybernetics*, vol.7, no.3, pp.345-364, 2016.

[31] N. Aharrane, K. E. Moutaouakil and K. Satori, A comparison of supervised classification methods for a statistical set of features: Application: Amazigh OCR, *Intelligent Systems and Computer Vision*, pp.1-8, 2015.

[32] Y. Zhou, Y. Li and S. Xia, An improved KNN text classification algorithm based on clustering, *Journal of Computers*, vol.4, no.3, pp.230-237, 2009.

[33] H. Li and J. Sun, Majority voting combination of multiple case-based reasoning for financial distress prediction, *Expert Systems with Applications*, vol.36, no.3, pp.4363-4373, 2009.

[34] X. Mu, J. Lu, P. Watta and M. H. Hassoun, Weighted voting-based ensemble classifiers with application to human face recognition and voice recognition, *International Joint Conference on Neural Networks*, pp.2168-2171, 2009.

[35] D. M. Powers, Evaluation: From precision, recall and f-measure to ROC, *Informedness, Markedness and Correlation*, 2011.

[36] S. Malakar, P. Sharma, P. K. Singh, M. Das, R. Sarkar and M. Nasipuri, A holistic approach for handwritten hindi word recognition, *International Journal of Computer Vision and Image Processing (IJCVIP)*, vol.7, no.1, pp.59-78, 2017.

[37] S. Ajitha, A. A. Fathima, V. Vaidehi, M. Hemalatha and R. Karthigaiveni, Face recognition system using combined Gabor wavelet and DCT approach, *IEEE International Conference on Recent Trends in Information Technology (ICRTIT)*, pp.1-6, 2014.

[38] Y. Jian, D. Zhang, A. Frangi and J. Y. Yang, Two-dimensional PCA: A new approach to appearance-based face representation and recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, pp.131-137, 2004.

[39] O. Deniz, M. Castrillon and M. Hernandez, Face recognition using independent component analysis and support vector machines, *Pattern Recognit. Lett.*, vol.24, pp.2153-2157, 2003.

[40] H. Ghaforiyan and M. Emadi, Human face recognition under pose variation with fusion geometric methods, *International Academic Journal of Science and Engineering*, vol.3, no.1, pp.1-10, 2016.

[41] *The ORL Face Database at the AT&T*, http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase, 2015.

[42] *The Yale Face Database*, http://vision.ucsd.edu/content/yale-face-database, 2015.

[43] A. Martinez and R. Benavente, The AR face database, *CVC Tech. Report No. 24*, 1998.

[44] *The WEKA*, http://www.cs.waikato.ac.nz/~ml/weka/, 2016.